Barbara Deml · Patricia Stock
Ralph Bruder · Christopher Marc Schlick *Eds.*

Advances in Ergonomic Design of Systems, Products and Processes

Proceedings of the Annual Meeting of GfA 2015

Springer Vieweg

Advances in Ergonomic Design of Systems, Products and Processes

Barbara Deml • Patricia Stock •
Ralph Bruder • Christopher Marc Schlick
Editors

Advances in Ergonomic Design of Systems, Products and Processes

Proceedings of the Annual Meeting of GfA 2015

Springer Vieweg

Editors
Barbara Deml
ifab-Institute of Human and Industrial Engineering
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany

Patricia Stock
REFA-Institut e.V.
Dortmund, Germany

Ralph Bruder
Institut für Arbeitswissenschaft
Technische Universität Darmstadt
Darmstadt, Germany

Christopher Marc Schlick
Institute of Industrial Engineering and Ergonomics
RWTH Aachen University
Aachen, Germany

ISBN 978-3-662-48659-7 ISBN 978-3-662-48661-0 (eBook)
DOI 10.1007/978-3-662-48661-0

Library of Congress Control Number: 2016932311

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer Vieweg is a brand of Springer
Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media (www.springer.com)

Editorial

These proceedings include a selection of papers presented at the 2015 Annual Meeting of the German Gesellschaft für Arbeitswissenschaft, held at Karlsruhe Institute of Technology (KIT) from February 26 to 28. The conference featured more than 160 presentations and 30 posters reflecting the diversity of subject matter in the field of ergonomics, human factors and industrial engineering.

The first part of the book deals with the *design of work systems* against the background of current socio-technical challenges. The contributions take up relevant research topics caused by the demographic change (Jungmann et al.; Wassmann et al.), important issues concerning occupational health (Bockelmann et al.; Hillebrecht et al.; Penzkofer et al.; Rivkin and Schmidt; Rücker and Brombach) and current changes in working time (Stock; Tegtmeier et al.). Furthermore, the effect on outcome variables, such as motivation (Kassirra and Rausch), quality (Sattler et al.) or efficiency/effectiveness (Tackenberg et al.) is regarded.

Within the second part of the book the *design of products* is considered. Here, too, the contributions cover a wide spectrum addressing the assessment of both cognitive (Arenius et al., Schneider and Deml) and physiological user states (Bürkle and Schmauder; Franzke and Walther; König and Jaschinski; Jaschinski et al.), the design of both input (Meyer et al.) and output (Knott et al.; Nelles et al.; Strenge et al.) processes in the field of human–machine interaction as well as new approaches for measuring working environment variables (Spitzhirn et al.).

Finally, in the third part of the book the *design of processes* is taken into account. Again it is current socio-technical developments that are reflected by the research papers. The contributions address topics such as flexible-mobile working (Gisin et al.), interdisciplinary collaboration (Brandtstädter and Sonntag), age-appropriate working processes (Büttner et al.; Kugler et al.), complex project planning (Terstegen et al.) or ecological aspects (Lüderitz et al.).

Considering the wide range of topics covered and the variety of scientific methods applied, it is apparent that advances in ergonomics may only be achieved by a multidisciplinary approach. Thus, these proceedings address human factors and safety specialists, industrial engineers, work and organizational psychologists, specialists in occupational medicine as well as production planners and design engineers.

Karlsruhe, Germany Barbara Deml
Dortmund, Germany Patricia Stock
Darmstadt, Germany Ralf Bruder
Aachen, Germany Christopher M. Schlick

Contents

Part I

Design of Work Systems

Assessment of the Individual Work Organization During a Service Provision

Sven Tackenberg, Sönke Duckwitz, Julia Seibold, and Christopher M. Schlick

Abstract

Employees of knowledge-intensive service companies organize their work individually. Therefore, an inadequate coordination of people may lead to an exceeding of service time and costs. In order to avoid this, a method-based work analysis provides an appropriate, but also time-consuming procedure to evaluate the efficiency and effectiveness of operations at an individual employee level. In this paper, we introduce a new approach for the assessment of an individual work organization during a service provision. To achieve this objective, a performance measurement system and a software tool for tablets and smartphones are presented. The software tool will be introduced, and tested by a verification study in a service company.

Keywords

Work organization • Service management • Self assessment tool

1 Introduction

For the most part, service research-to-date contains a variety of generic models to describe service productivity [2, 11, 23]. All these models have in common that they focus on the sales department, human resource management, the design of work processes, and the application of technologies [11]. But, knowledge-intensive services are based on service provisions that mainly rely on the expertise of individuals. Therefore, the processes contain a high degree of individualization and interaction between persons, as well as high uncertainty with regard to the performance potential and the work results. Hence, those service productivity

S. Tackenberg • S. Duckwitz (✉) • J. Seibold • C.M. Schlick
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany
e-mail: s.duckwitz@iaw.rwth-aachen.de; c.schlick@iaw.rwth-aachen.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_1

models are of limited suitability in practice. According to this understanding, the work processes of a service provision as well as the individual work organization of the involved working persons determine the success of sustainable services.

In this paper, we have interpreted the measurement of service productivity from the perspective of the individual work organization of an employee. Yet, there are only very few studies about individual planning and knowledge-intensive activities [3, 17, 21], which investigated the relationship between the individual work organizations of employees and service productivity. This is surprising, as the studies were able to show that an individual work organization and the working conditions influence the required time, strain, and the work outcome. Based on the results of [10], service productivity can be defined as the result of the invested time, the quality of results, and of the individual work organization. Therefore, we interpret a service as productive, if there is an efficient and effective collaboration of persons with an optimal individual work organization. Accordingly, it raises the question, which performance indicators of an individual work organization exist, and how is it possible to measure and control them.

2 Work Analysis: State of Research

2.1 Conceptual Classification

Work analysis is often characterized as the initial starting point for almost any other effort to measure and improve work. Thereby, work analysis methodology is the foundation of work analysis [28]. For the most part, work analysis focuses on a systematic structuring of tasks and activities, which are processed by an employee of a work system [26]. Such a work system is a regulatory model to systematically describe a workplace using various elements, such as employees, requirements and tasks [27]. To perform a work analysis, various methods, procedures and instruments are available to capture, process, and interpret information about tasks, the organizational-technical working conditions, the required work equipment, as well as their impact on the involved employees [27]. The different methodological approaches can be applied to a theoretical issue or to specific elements of a work system. Therefore, a work analysis proceeds not just analytically, but also for evaluation and design purposes [14]. In particular, a distinction must be drawn between the psychological and the task-centered work analysis [5, 6]. The former is restricted to the effect of working conditions and the requirements imposed on the employee [5], whereas a task-centered work analysis gather data about the work content and the target oriented processing of tasks [6].

In our approach of a performance measurement system we provide a conceptual framework for the assessment of an individual work organization by focusing on processes that exhibit two critical characteristics: knowledge-intensive and weakly-structured:

– Knowledge-intensive: Such processes are characterized by a heavy reliance on employees and significant incidents of problem solving as well as non-standardized production [1]. These employees are the primary sources of information and knowledge which are used to produce intermediary services for the customer.
– Weakly-structured: Such workflows represent task sequences which are not strict and predetermined [27]. Due to missing standards and incomplete information, an employee has to identify and to choose the best strategy to solve and schedule the assigned tasks.

2.2 Background: Work Analysis

In the last 50 years in fields as task scheduling, organizational behavior, and management science, many studies have been conducted in order to understand how employees distribute their working time [10]. Most studies have focused on the role of managers, to identify the characteristic of their tasks and to analyze their simultaneous management of activities [17, 19, 20]. In the field of service research the focus was solely laid on the relations between actors and communication processes [11]. In real service provision, the employees communicate asynchronously, i.e., via email, messenger, and document postings in a cloud. Therefore, the sender of a message may not know exactly how long a response will take. Due to the lack of information the person might switch to another task and reorganize his/her work plan. Rubinstein et al. examined the effects of task switching depending on the familiarity and complexity of tasks [25]. Their results indicate that employees create a delay before they start working effectively on another task. Furthermore, they could show that each task switch causes a time loss. There is even more time loss, when switching from a familiar to an unfamiliar task and from a simple to a complex one.

An external interruption of work is defined as a synchronous interaction, which is unscheduled and not initiated by the employee of the considered work system. Such an interruption results in a discontinuing of the current activity of the recipient [4, 24]. For improving the productivity of an employee it is essential to know whether and when interruptions constitute disruptions and which kinds of interruptions have negative or positive effects on task performance. Therefore, in the field of interruption research, efficiency and effectiveness are measured based on the work conditions [9, 21, 22].

In the late 1960's, Horne and Lupton found that middle managers spend most of their work time in managing various activities, but very little time for reflection and decision-making [13]. To obtain the relevant data, managers recorded their own activities on a specific form, which contained several terms, the "*Managers Activity Record*". If one of the terms was appropriate for describing the work day, the manager had to mark it.

Mintzberg proposed that the best way to describe and evaluate the work of managers is to focus on the managers' activities [17–19]. He conducted a 5-day long observational analysis, during which he recorded the type, the purpose and the duration of each activity of the observed managers. Based on the data of five chief executive officers he developed his role theory. Thereby, Mintzberg [17] concluded that managers perform a set of ten, basic roles, which describe behaviors belonging to an identifiable office or position.

In a quasi-field experiment, Perlow [22] investigated how team members of a software engineering team use their work time, what kind of tasks they choose and if their planning is time-optimal for them or their work groups. Thereby, the group's collective use of time perpetuated each individual as "time famine", a feeling of having too many tasks and not enough time to properly work on them. Due to interrupting each other, the observed engineers had difficulties to get their individual work done. To reduce the quantity of interruptions, Perlow [22] initiated the implementation of a "*quiet time*". Such a "*quiet time*" is an agreement among coworkers to not interrupt a colleague during the defined designated hours. A result of the "*quiet time*" was a higher percentage of task completion due to a longer processing time.

During a fieldwork observation Gonzáles and Mark [10] discovered that employees (analysts, software developers, and manager) have a high level of discontinuity during the execution of their activities. Thereby, the average to work on a task was about 3 min. If the employee had to use an electronic tool or paper document the period before a further task switch was not longer than 2 min. To enlarge the working period, the authors introduced the concept of "*working spheres*" to conceptualize and organize the work [10]. The result was that the employees worked on average in ten different working spheres for a period of about 12 min.

Other interruption and communication characteristics, including the particular time and the length, the demands of the service process and the initiator, as well as the recipient moderate the efficiency and the stress of project managers [3]. In an observational study of project managers Beuscher-Mackay et al. were able to show that the fragmentation of a work day has an impact on the subjectively perceived productivity [3]. In addition, Cutrell et al. [4] found that the timing of such interruptions has an impact on the performance of an employee. The earlier such an interruption occurs during the completion of a task, the more disruptive is the effect [4].

For the most part, research-to-date focuses on identifying the timing or the frequency of task types of managers and on the characteristics of individual interruptions [24]. While studies indicate that various interruptions decline the performance of an individual, there is no concept to measure the individual work organization. Furthermore, we did not identify a software solution, which can automatically identify and support people's continuous switching between tasks and interruptions.

3 Method for Evaluating Individual Work Organization

3.1 Individual Work Organization

The term work organization is often used in the context of planning and executing actions to solve work tasks. Thereby, the content of a work organization relates to various contexts. For instance, Schlick et al. [27] define a work organization as the planning and the design of separate work systems with a specific focus on division of labor and cooperation. Other definitions focus on the individual's work behavior, which differs in terms of decision-making and the strategies to solve a task [12]. Both aspects are prerequisites for an employee to carry out the necessary actions to fulfill the assigned task. Therefore, a definition and demarcation of the term work organization are necessary.

The individual work organization of an employee comprises in particular the identification of options for decision-making in regard to perform certain acts for solving the assigned tasks to maximize the potential benefits. Thereby, existing restrictions and degrees of freedom have to be considered by the employee. Applying this definition to the systematic of labor organization, as defined by Hacker [12], an individual work organization is only the third step after the distribution of functions between employees and work equipment, and the distribution of work tasks between different employees (Fig. 1).

Hence, there are two perspectives from which an individual work organization can be assessed. Here, the work system determines if there are any degrees of freedom and uncertainty for processing work tasks. Also, the input of the work system as well as the characteristic of the elements of the observed work system can cause the occurrence of interruptions. Furthermore, the data quality is important, because an evaluation of an individual work organization can only be achieved if a baseline measurement is available that can be used as a reference. Consequently, the evaluation and improvement of the individual work organization depend on the design of the work system and the decision-making of the employee. Therefore, the evaluation of a given, unchangeable work system differs from a system, which allows the partitioning of functions between the employee and the work equipment, as well as between different persons.

3.2 Assessment of the Individual Work Organization

To provide an adequate database, a workshop with the knowledge-intensive employees of a service company was carried out to identify the central factors influencing the quality of an individual work organization, and to describe the characteristics of a work system. Only one worker can be assigned to such a work system. Therefore, cooperative activities between several workers have to be described by a corresponding number of work systems. To analyze the individual work organization of an employee, as well as the performance of a service provision, the terms "*effectiveness*" and "*efficiency*" are particularly suitable for

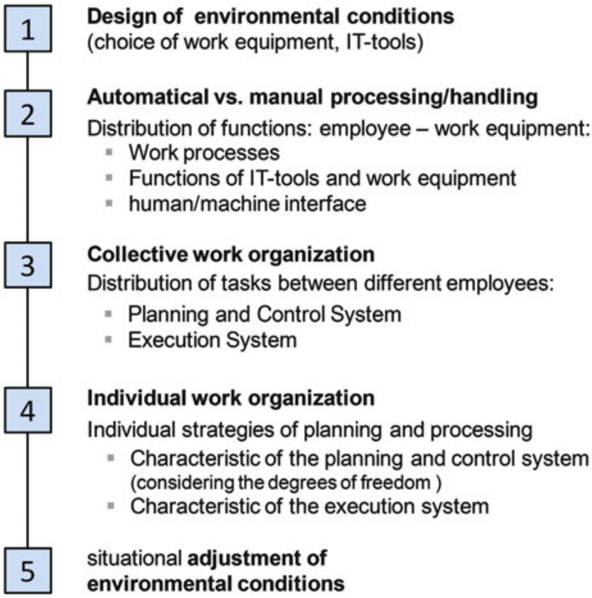


Fig. 1 Classification for designing a work organization

describing the corresponding actions. Thus, identifying and scheduling tasks in connection with the optimal characteristics of a work system can be deliberately focused on an efficient and effective task processing. Moreover, the employee is empowered to understand and analyze the individual productivity as a result of their decisions and actions.

Based on the process of a service provision the essential influential factors were consolidated to latent variables. These variables shall contribute to the endogenous latent construct "*service productivity*". The productivity of a process is related to how effectively input resources of a service process are transformed into economic results for the service provider and value for its customers [11]. In the following sections we refer to the concept of service productivity developed by Lasshof [16] and the results of the empirical study of v. Garrel et al. [8]. They have linked service productivity to the aspect of planning and task execution, measured in terms of efficiency, effectiveness and quality. All of the constructs have the advantage that they can be used for the assessment of the individual planning as well as the action of employees [29].

A work organization is considered effective if the expected output of the work system is achieved [15]. However, a work organization is efficient, if the required quantity or quality of the output is achieved with a minimum number of input factors (e.g. working hours), while the product quantities remain constant [7]. Thereby, the specification of the minimal input factors has to reach a balance between a human-oriented structuring of work, and the productivity of a service provision. Yet, in practice such an equality of economic and human criteria is often not achieved due to considerable private economic interests [29].

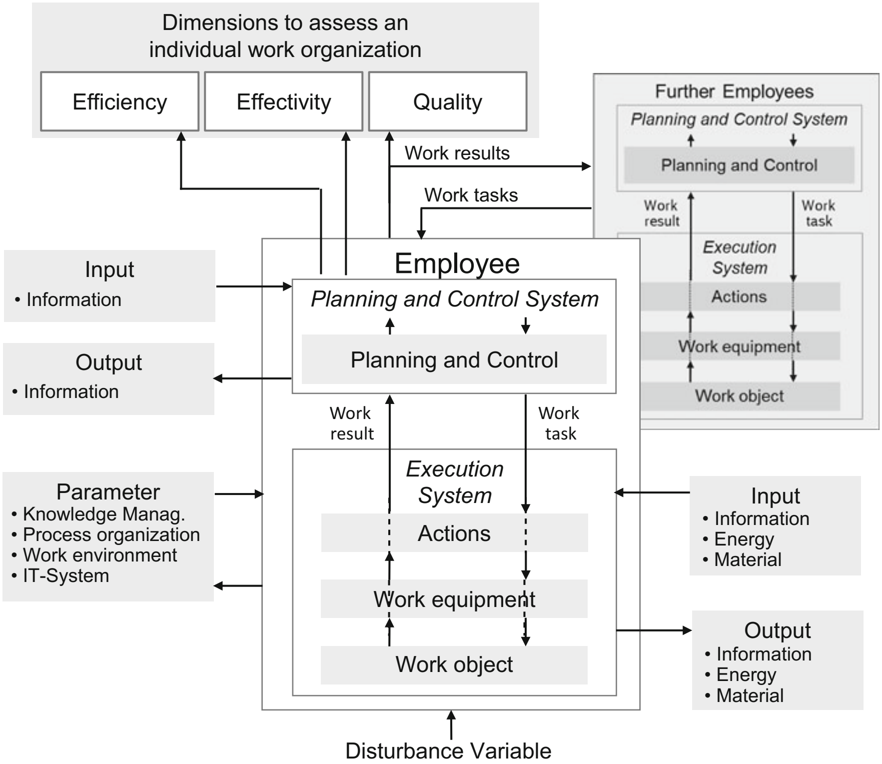


Fig. 2 Work system for assessing the individual work plan

The basic model of a work system [27] was expanded to include three evaluation dimensions for the individual planning of work. Whereas, the efficiency and the effectiveness are a result of the "*planning and control system*" of an employee, the quality of work is determined by the interaction between the "*planning and control system*" and the "*execution system*" of a work system (Fig. 2). Therefore, the key figures of effectiveness and efficiency evaluate the identification and scheduling of tasks. While the indices for the term "*quality*" refer to the work results of the work system.

3.3 Performance Figures for the Individual Work Organization

The following section is based on the work of several previous studies [3, 4, 10, 17–19, 21]. A performance measurement system will be presented, which argues that it is possible to evaluate the individual work organization of employees during a knowledge-intensive service provision. Thereby, we refer to the terms "*effectiveness*" and "*efficiency*" because both are typically used to indicate the desired and the unintended consequences within a work system.

3.3.1 Efficiency of the Individual Work Organization

The result of a knowledge-intensive service provision is based on a strict division of labor and a task-allocation to several decentralized employees. In the field of manufacturing, productivity is a concept related to the efficiency of producing goods [11]. However, the problem about providing an efficient service process is that productivity and perceived service quality are inseparably connected. In contrast to the manufacturing sector there are no methods to calculate standard times for a task execution due to the dependence on the individual problem solving. Therefore, based on the introduced understanding of the terms "*efficiency*" we defined the figures "*work efficiency*", "*perceived work efficiency*" and "*process continuity*". All three figures are indicators for the latent superordinate construct of "*efficiency*" to evaluate the work organization of an individual.

Definition "*expenditure of human labor*"

The measured expenditure of time $ET^m_{i,w}$ of a task i is the sum of all time points t (time: 1 min), a task is processed by person w:

$$ET^m_{i,w} = \sum_{t \in T} a_{i,w,t}, \tag{1}$$

$$\begin{matrix} a_{i,w,t} = 1 & \textit{if} & \textit{task i is processed at time point t by individual w} \\ a_{i,w,t} = 0 & \textit{else} & \end{matrix}. \tag{2}$$

Figure *"work efficiency" (WE)*

The figure "*work efficiency*" (*WE*) is based on the comparison of the expenditure of human labor $ET^m_{i,w}$ needed for a complete processing of a task (current value), and the initial estimated workload $ET^p_{i,w}$ (planned value). To determine the value of "*work efficiency*" for a given period, the mean value WE_{ges} across all completely processed tasks is calculated:

$$WE_{i,w} = \frac{ET^m_{i,w}}{ET^p_{i,w}}, \tag{3}$$

$$WE_{ges} = \frac{1}{\sum_{t=t_{min}}^{t_{max}} y_{i,t}} \sum_{t=t_{min}}^{t_{max}} y_{i,t} WE_{i,w}, \ \forall i \in A, \tag{4}$$

$$\begin{matrix} y_{i,t} = 1 & \textit{if} & \textit{task i is completely processed at time point t} \\ y_{i,t} = 0 & \textit{else} & \end{matrix}. \tag{5}$$

The information provided by the figure "*work efficiency*" has to be questioned in particular when the observed employee w defines the planned value $ET^p_{i,w}$. In that case, the demand of minimizing $WE_{i,w}$'s value, may lead to the definition of time buffers to maximize the robustness of the individual work organization. Yet, such a behavior would be contrary to the objectives of improving the scheduling and processing of tasks.

We consider that an efficient individual work organization exists if $ET^{p}_{i,w} = ET^{m}_{i,w}$. If $ET^{m}_{i,w}$ clearly exceeds $ET^{p}_{i,w}$, the risk of an outrun of the planned duration and costs may exists. However, an overestimated expenditure of human resources indicates an inefficient individual work organization. Because these time buffers cause idle periods, or they are often used to generate a not necessary outcome. Both kinds of deviations can trigger a situation-based adjustment of the individual work organization. But studies have shown that compared to the initial planning, employees are often not able to use interruptions or time buffers, arriving in the short term, efficiently [18]. Therefore, it should be an aim to minimize any deviation: $\Delta ET = ET^{p}_{i,w} - ET^{m}_{i,w}$. Consequently, a learning assistance system can propose a value for $ET^{p}_{i,w}$. Alternatively, based on historical values the input of the user $ET^{p}_{i,w}$ can be adjusted.

Definition *"Processing phase"*
A processing phase of a task *i* is defined as the period of connected points in time at which this specific task is processed. A processing phase ends if an interruption or a task switch occurs.

Figure "*Perceived work efficiency*" (*PE*)
The figure "*perceived work efficiency*" can consider the contribution of the task content as well as of the working conditions to the efficiency of an individual work organization. Both factors are necessary for setting the performance baseline. The "*perceived work efficiency*" is defined as the proportion of the value ($V_{i,n,w}$), perceived by the employee *w* for processing phase *n* of task *i*, to the measured duration of phase *n*. As an input the subjective evaluation of the value $V_{i,n,w}$ has to be measured after each of the phases. In fact, we define that a unit of labor (1 min) represents one unit of a currency. The usage of these time and monetary units should generate a value for the employee of the work system. Later on, the ratio of the perceived benefit of the invested time is compared to the measured time of task processing d_n:

$$PE_{i,n} = \frac{V_{i,n,w}}{d_n}. \tag{6}$$

The key figure value of all finished task shall be calculated as follows:

$$PE_{ges.} = \frac{1}{\sum_{t=t_{min}}^{t_{max}} y_{i,t}} \sum_{t=t_{min}}^{t_{max}} y_{i,t} PE_{i,n,w}, \ \forall i \in A. \tag{7}$$

Figure "*process continuity*" (*PC*)
In real world contexts, processing tasks is often interrupted by rotation of tasks, information delays, interruptions and non-working periods. From the efficiency

perspective, the fragmentation of processing a specific task i has to be evaluated in the context of the measured expenditure of human labor $ET^m_{i,w}$. To determine the respective period, a differentiation between working hours and unavailability can be included or neglected. The basic calculation of the figure $PC_{i,w}$ is based on the quotient of the measured workload and the duration to complete the task i:

$$PC_{i,w} = \frac{ET^m_{i,w}}{t_{i,max} - t_{i,min}}, \tag{8}$$

$$t_{i,\max} = \max\{t\, x_{i,t} : t \in T\}, \tag{9}$$

$$t_{i,\min} = \min\{t\, x_{i,t} : t \in T\}, \tag{10}$$

$$\begin{array}{lll} x_{i,t} = 1 & if & task\ i\ is\ processed\ at\ time\ point\ t \\ x_{i,t} = 0 & else & \end{array}. \tag{11}$$

3.3.2 Effectiveness of the Individual Work Organization

A conceptual differentiation of effectiveness from the term efficiency in the context of individual work organization cannot be presented selectively. Finally, we have defined that the execution of the right tasks to reach the expected outcome is recorded through the figure "*work efficiency*". This is mainly due to the fact that counterproductive operations lead to a higher workload compared to the initial planning. Within a work organization context, the term "*effectiveness*" refers to the rework of an initial completed task. An iteration may be necessary for the following reasons: It is the repetition of initially completed work outcomes due to the availability of new and modified information, or an upstream task need to be repeated when the downstream task discovers some sort of error or incompatibility. Yet, a modification of a task by another person of a different work system can cause iteration, too.

Definition "*Rework*"

Rework of a task is said to occur when the current work results have to be modified that beforehand were regarded as "*finished*", and transferred as an input to the same or a different work system. A possible reason of rework could be a faulty or incomplete output due to missing information or erroneous actions or erroneous transmission of information.

In contrast to a manufacturing process, the quality of content and methodology cannot be compared to a reference value due to the heavy reliance on individuals and on non-standardized production. Nevertheless, in order to enable an assessment, we differentiate the investigated tasks and interruptions by the addressees and the generated results. If a result is the input and the output of the same work system, we imagine that the "*planning and control system*" of the system only considers a result as completed if it has a sufficient quality. In contrast, if the result is handed over to another work system, the external "*planning and control system*" assesses the content and methodology of the result. Afterwards, the previous result is again transferred to the observed work system and the information regarding the results may get revised. Therefore, we focus on the assessment of results, which are handed over to an external work system.

To ensure the consistency of indicators, we are making use of the currency concept. When a result is handed over to another employee, the person of the work system has to estimate the time (earned value) still required to reach an adequate result for the recipient. In case of a response, the required time for rework of the results is an indicator for the genuine quality of the content and the methodology of the previous outcome. In this respect, the relative proportion of time, which is based on a modified or an enlarged task, has to be separated for calculating the figure.

Figure "*effectivity of task processing*" (*EP*)
The figure relates exclusively to tasks for which the output is transferred to another work system and a verification of the results is expected—"*task for resubmission*". As an input, at the date of handover $n = 1$, the expenditure of human labor $ET_{i,w}^{m(1)}$ and the anticipated rework $ET_{i,w}^{n,p}$ is measured. Following a feedback, the required time for rework $ET_{i,w}^{n,m}$ minus the time $ET_{i,w}^{n,a}$ which is caused by a modified or enlarged task, must be identified and are included in the calculation of $EP_{i,w}$:

$$EP_{i,w} = \frac{\sum_{n \in N} ET_{i,w}^{n,m} - ET_{i,w}^{n,a}}{ET_{i,w}^{m(1)} + \sum_{n \in N} ET_{i,w}^{n,p}}. \tag{12}$$

The consolidated key figure value of all finished task shall be calculated as follows:

$$EP_{ges.} = \frac{1}{\sum_{t=t_{min}}^{t_{max}} y_{i,t}} \sum_{t=t_{min}}^{t_{max}} y_{i,t} EP_{i,w}, \ \forall i \in A. \tag{13}$$

Figure "*Relative adherence to schedule*" (*RT*)
An important subsidiary aspect of the quality of an individual work organization is the period between the planned and actually realized date of completion. The former value is imposed by the "*planning and control system*" of the observed or the external work system. Both times describe the robustness of the individual work system. The time units after the scheduled time represent penalty minutes. We consider these time units as an indicator of an ineffective individual work organization, or an unrealistic definition of the planned date of completion. The calculation of the figure $RT_{i,w}$ for the completed task i is based on the quotient of the measured workload $ET_{i,w}^{m}$ and the period between the planned t_i^p and measured date t_i^m of completion:

$$RT_{i,w} = \frac{t_i^p - \sum_{t=t_{min}}^{t_{max}} y_{i,t} t}{ET_{i,w}^m} = \frac{t_i^p - t_i^m}{ET_{i,w}^m}. \tag{14}$$

The consolidated key figure value of all finished task shall be calculated as follows:

$$RT_{ges.} = \frac{1}{\sum_{t=t_{min}}^{t_{max}} y_{i,t}} \sum_{t=t_{min}}^{t_{max}} y_{i,t} \, EQ_{i,w}, \ \forall i \in A. \tag{15}$$

4 Conception and Implementation of the Software

The calculation of the defined key performance indicators (*KPI*) requires a data basis, provided by a measuring tool. Existing tools, *such as ORTIM time, REFA time, Getting Things Done* and *Every Task*, do not allow an assessment of the individual work organization if the work processes are weakly-structured. Furthermore, a second person is required to use the tools. Also, the defined *KPI*s have not been implemented to date. Therefore, a new software framework will be presented and verified in the following sections.

4.1 Software-Framework

The application is structured according to three defined variants. These differ in terms of time, required for data entry and the informative value of the *KPI*s as well as the deducible directives. The specific functions of the software build upon one another and their combination results in three separately usable software versions. By that, the user can select the so-called "*basic version*", "*evaluation version*" or "*expert version*":

– The "*basic version*" offers the possibility of a continuous task analysis. Here, the tasks processed during a work day are recorded with their particular time of operation. In addition to the tasks of the service projects and the day-to-day business, self-initiated and external-initiated interruptions are documented. Therefore, the generated database contains information regarding the dates and the frequency of task-processing and interruptions as well as the characteristic of the observed rotation of tasks. Based on these data, the deducted *KPI* "*work efficiency*" assesses the deviation between the planned and required expenditure of time for each task. Furthermore, the proportion of the defined task categories as well as the fragmentation of a working day is captured.
– The "e*valuation version*" is an enhanced version of the "*basic version*". In particular, it allows the calculation of the *KPI* "*perceived work efficiency*" which can be determined for an individual task as well as for the whole of the completely processed tasks.

– The application of the "*expert version*" leads to the calculation of all *KPI*s, defined in Sect. 3.3. In addition to the evaluation of the planned and measured date of completion, the required time for an iterative task processing is recorded. Furthermore, this version includes an assistance system which proposes the user a probable expenditure of time for a specific task processing. Such a value is calculated based on historical data.

4.2 Software Application

The software-framework was implemented for the Android operating system—a so-called application (*APP*). In order to graphically optimize the user interaction and the dialogs, we focus on tablets and smartphones with a screen size of at least 4 inches. For each type of device, we developed a specific layout, in regard to the orientation—vertical vs. horizontal—and the screen size of the used device.

4.2.1 Dialog Boxes for Interaction

The "*main screen*" of the software comprises four areas: "*task pool*", "*task collection and interruptions*", "*workbench*" and "*menu bar*". Thereby, the "*main screen*" is the starting point of any user interaction. If a tablet is used, these areas are arranged on the right side of the screen while the application- or situation-specific dialog boxes are situated on the left side. An overview of the Human-Machine-Interface is depicted in Fig. 3:

– The area "*task-pool*" includes all tasks which have to be processed by the employee of the work system. Thus, it represents the current status of his or her work basket. Besides the name and the importance of a task, the required time for a complete processing as well as the already consumed time exposure are displayed. To ensure a concise presentation, the current tasks can be filtered regarding their assignment to a project or the day-to-day business.
– The "*workbench*" comprises the currently processed task or interruption. Therefore, its content indicates the status of the employee (user of the *APP*). Besides the task-specific information—name, importance, required finishing time—the consumed expenditure of time is presented up to the respective point in time. This value is depicted in minutes and shown as a bar diagram. The bar has at the beginning a green color and the bar height increases during task processing. If the consumed effort exceeds the planned value, the bar is removed and a new bar with a red color appears. This bar increases also proportionally to the consumed time during the further task processing.
– The "*pool of standard tasks*" contains the buttons for tasks, which are often chosen by the user during a work day. Standard tasks can be defined and sequenced individually in this area.
– The button for indicating the occurrence of an "*interruption*" is arranged next to the "*workbench*".

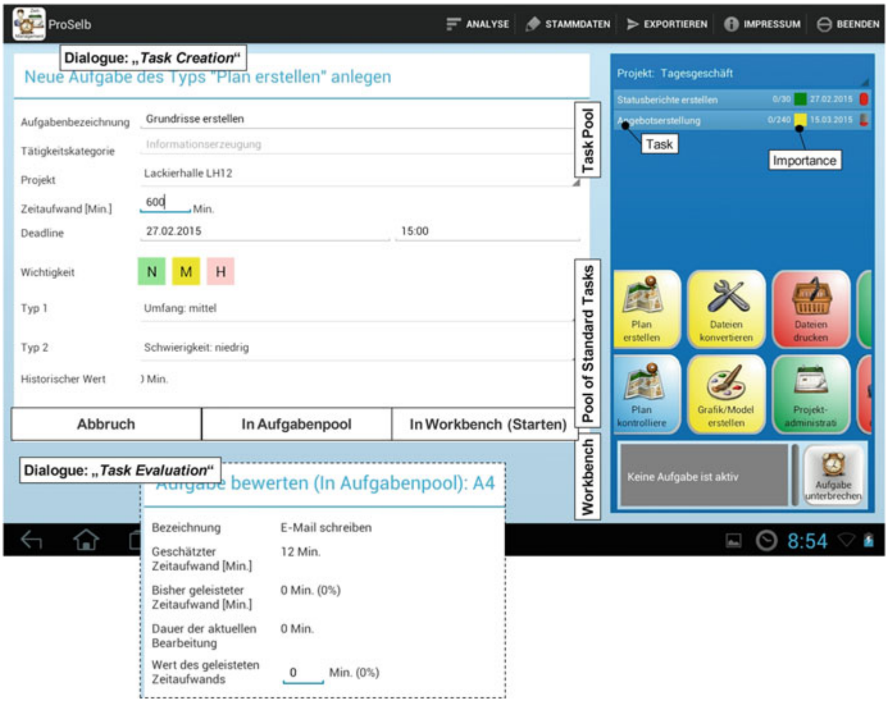


Fig. 3 Structure of the application's user interface

– The "*menu bar*" includes the functions for switching to further dialog boxes. These boxes enable the user to analyze the recorded working day and the calculated *KPIs*. Furthermore, a function allows the termination of a working day.

4.2.2 Software Functions for Data Collection

The software-framework is based on a so called "*project/task – workbench*". This means that in the case of the initial use of the *APP*, the current service projects of the employee as well as the category "*day-to-day business*" have to be generated. For this purpose, the user must select the button "*new project*" in the "*menu bar*". Thereby, it is possible to edit the entries or to create additional projects at a later point in time.

Furthermore, so-called "*standard tasks*" can be defined. Such a "*standard task*" represents a task category or an action, which occurs several times during a regular working day of the user. Due to the individual definition of "*standard tasks*" it takes up much less time to create a new task during the process of data collection. For example, an employee in a service company for factory planning will define a standard task for "*Writing an email*" or "*Submitting a building permission*".

Subsequent to the initial start of the *APP* and the definition of projects and "*standard tasks*", the user has to create his or her current tasks and to assign them to a project or the category "*day-to-day business*". To create such a new task, the user can select one of the buttons representing a "*standard task*" or he/she may use the button "*user-defined task*". In both cases a dialog box appears to enter the relevant information of the task (Fig. 3).

Based on the selected software version (*basic*, *evaluation* or *expert*), a variable number of user inputs is required. After entering the information, the user must decide, whether the task has to be assigned to the "*task pool*" or the "*workbench*". The latter means, that the user will process the new task subsequently to its creation. During the phase of processing, the task will appear in the "*workbench*". Each minute of such a processing leads to a proportional rise of the expenditure of time. If a new task has to be set up, the user is able to select at any time a task from the "*pool of standard tasks*" or can choose the button "*user-defined task*".

If the user wants to interrupt the current task processing to start working on another task, he or she can choose a task from the "*task pool*" or can create a new task. With such an action, the dialog box for the assessment of the past processing phase appears (Fig. 3 – "*expert version*" of the *APP* is shown). The information to be entered in the box provides the basis for the calculation of the *KPI*s (Sect. 3.3).

Besides the evaluation of the current processing phase, the user has to indicate whether: 1) the task is completely processed, 2) the task has to be processed further at a later point in time (reassignment to the *task pool*) or 3) the task has to be indicated for a later resubmission. The latter describes a transfer of the outcome of the task processing to another employee (outside the observed work system), and the expectation of feedback and rework. Task with the status "*resubmission*" are differently colored in the "*task pool*" and can be hidden from the display if necessary.

In case of an externally or self initiated interruption, the user has to select the button "*interruption*". Following this selection, a specification of the type of the current interruption is necessary. Afterwards, the interruption appears in the "*workbench*" and the last processing phase has to be evaluated by the user. Hereby, the previous processed task is automatically moved to the "*task pool*". If the user selects the button "*task restart*", the interruption is terminated and the previous task is reassigned to the "*workbench*". In analogy to a task switch, the dialog box for the assessment of the interruption appears.

The operating philosophy of the *APP* ensures that subsequent to the start of the *APP* and the initial selection of a task, a task or an interruption is always assigned to the "*workbench*". Only if the user selects the function "*termination of the work day*" from the "*menu bar*", the data collection is terminated and the task is removed from the "*workbench*". This guarantees a continuous assignment of all minutes of a working day to a specific task or interruption.

4.2.3 Software Functions for Data Analysis

The screen "*daily overview*" (Fig. 4) presents the tasks and the interruptions which are processed by the user on a working day. If the user selects the menu item "*analysis*" and subsequently the section "*daily overview/schedule*", a daily overview of a single working day occurs on the screen. Thereby, the tasks and

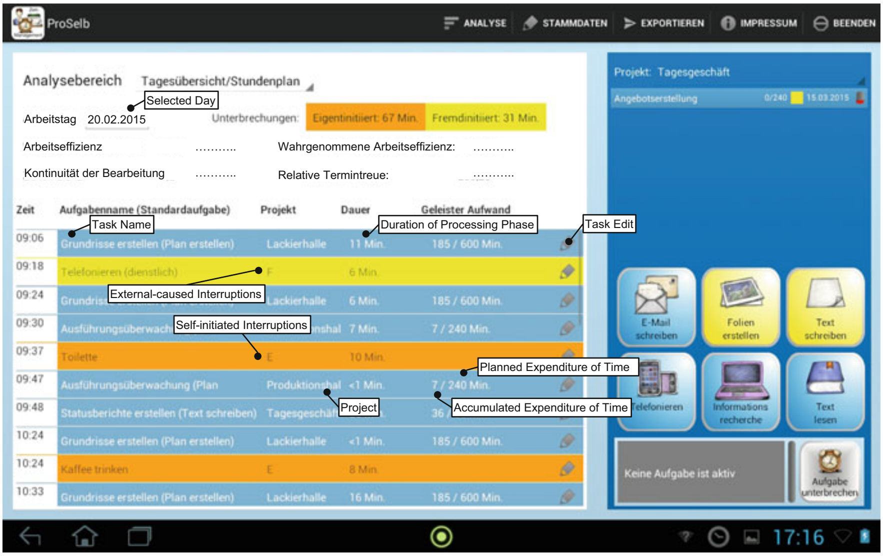


Fig. 4 Depiction of the daily overview and the central KPI

interruptions are chronologically assigned to the presented timeline. Therefore, a task can be listed several times, due to task switches and interruptions. For each processing phase of a task, both the planned as well as the cumulated expenditure of time are presented.

The button "*pencil*" (edit) enables the user to edit a recorded processing phase of a task or an interruption. Besides the adjustment of the duration, the user can remove or insert a processing phase.

Furthermore, the overview of the selected day provides a summary of the central *KPIs*. Thereby, the *KPIs* "*working efficiency*" and "*perceived working efficiency*" are only calculated for completely processed tasks. In addition, the expenditures of time for the self-initiated breaks and the external caused interruptions are displayed.

A detailed reporting of the *KPIs* for a processing phase, a single task or a working day are displayed in the sections "*task overview*" and "*KPI overview*". In these sections, all introduced *KPIs* (Sect. 3.3) are presented to the user for a single task, a single working day or a freely configurable period of time.

5 Field Study

With regard to the absence of empirical research about the individual work methods of an employee during a service provision, this paper addresses the following research question: Which *KPI*s for the assessment of an individual work organization exist, and how do communication delays and interruptions affect the quality of task scheduling and processing? Specifically, we focus on the intended effect of the individual work organization on the productivity of a service. Due to the fact that we were not able to investigate all stakeholders of a knowledge-intensive service provision, we have drawn on, and synthesized prior research on the documentation of the tasks and the interruptions of one employee, as well as the *KPI*s for the individual work organization. We believe that only a field study can further improve our collective understanding of the complexity of using an *APP* to measure and analyze the individual work organization. In order to use such a tool in a company, we present the outcome of a small verification study for the software concept as well as data of a service company.

5.1 Data Sources and Methods of Data Collection

The validation study for the developed software concept was based on two main ethnographic techniques [10]: participant observation and use of short interviews. The level of detail required for the individual calculation of the introduced *KPI*s demanded the capturing of tasks to be performed by the employee of the work system. We decided that a direct observation technique such as "shadowing" observation technique, similar to the methods used in previous studies of [3, 10], would be disruptive and inaccurate. Thus, we asked the employees of the work systems to keep diaries to generate the information needed for the verification of our concept. In our case, each employee of a work system tracked his or her activities during each observed work day and participated in a short follow-up interview. Whenever the employee got a new task or performed an action such as making a phone call or writing an email, he or she had to write down the time and other details regarding the task. Whenever the documentation of an individual work organization was unclear, we asked the observed subject at the end of the day. Inspired by Beuscher-Mackay's [3] structured observation method, we designed a tracking log where we transcribed the generated data (e.g. time stamps, event type, types of cooperation) of activities and interruptions. A total of 9.575 min was recorded in the observation at the field site.

Over a period of three consecutive days, eight people aged between 27 and 36 years (28, 20 years, $SD = 3.53$) were investigated. Each employee documented his/her individual work organization for the whole day. Before the data collection started, a further day was used to introduce the concept of the survey. For the next three following days, a formal data collection was done and the average time of formal observation for each individual was 6:39 h ($SD = 1{:}13$ h).

All eight employees worked in offices; two had an office with two fully-employed persons. Each workplace had a computer with internet connection, phone and resources such as books, binders, stationary, etc. The printer and copier were located at the end of an aisle. Due to the fact, that in general the doors were left open, the employees were able to easily communicate with other colleagues.

5.2 Results

Within the investigation period, the number of tasks processed by all employees was 106. Moreover, 192 interruptions were caused by employees of external work systems or were self-initiated breaks. The employees spent an average of less than 25 min (SD = 5 min) on a task before they switched to another task, or were interrupted. The data collected also indicates, that in fact, the effort required to completely process a task was in average approximately 70 min (*SD* = 10 min). Thereby, a single interruption took quite long (combined mean = 11 min, *SD* = 5 min). But, the most surprising fact was that a person easily spent over 9 min for an informal communication with a colleague before he or she chooses another task.

Taking into account the findings of previous research [3, 10, 17], we structured our data into eight main groups of tasks in order to compare our results with those of earlier studies. Therefore, analogous to Beuscher-Mackay the category "desk work" was defined as the time during which employees were sitting at a desk and worked with a computer or other physical artifacts [3]. Compared to earlier studies [13, 17] with non IT-based work systems our results reveal that more time was spend in deskwork, but that the findings are very similar to the results of Gonzáles and Mark [10]. Yet, the time spent on unscheduled meetings (e.g. going and entering to other offices, meeting at the coffee machine or the copier) was substantially lower (combined mean = 21 min., *SD* = 6 min.) than the level identified by Gonzáles and Mark [10] (98 min). This is not surprising, as the investigated employees mostly worked in their own single office, and not in cubicles (open space office). Therefore, it was not easy to interact and communicate with colleagues without moving. On the other hand, the results for the time spent on planned meetings (7, 8 % of all task categories) confirm the previously identified relationship between a higher hierarchical position, and more time spent in meetings [10].

The task categories are divided as outlined in Fig. 5, with a focus on the relative quantity as well as the proportion of time spent on each category. Looking more closely at the distribution of time across the categories, we can clearly see that also an average short-term duration of a task category (e.g. emails, telephone calls) can causes a greater proportion of work time. A major reason is the high frequency of new tasks (writing emails or telephone calls), which occur unannounced.

The individual work organization can result in a planned or unplanned task switch as well as the occurrence of an interruption. If an interruption occurs, in almost 60 % of the cases employees resume the disturbed sequence. The duration of a non self-initiated interruption can range from 2 min up to 25 min. In Fig. 6 and Table 1 the distribution of the interruptions among the observed employees is shown.

In fact, the large number of interruptions for w_4 resulted from the location of his office within the building—close to the copier—and the occupancy of the room with two fully-employed employees and one temporarily present assistant. This leads to interruptions caused by colleagues within the room and employees who worked at the copier. However, the good values for "*work efficiency*" and

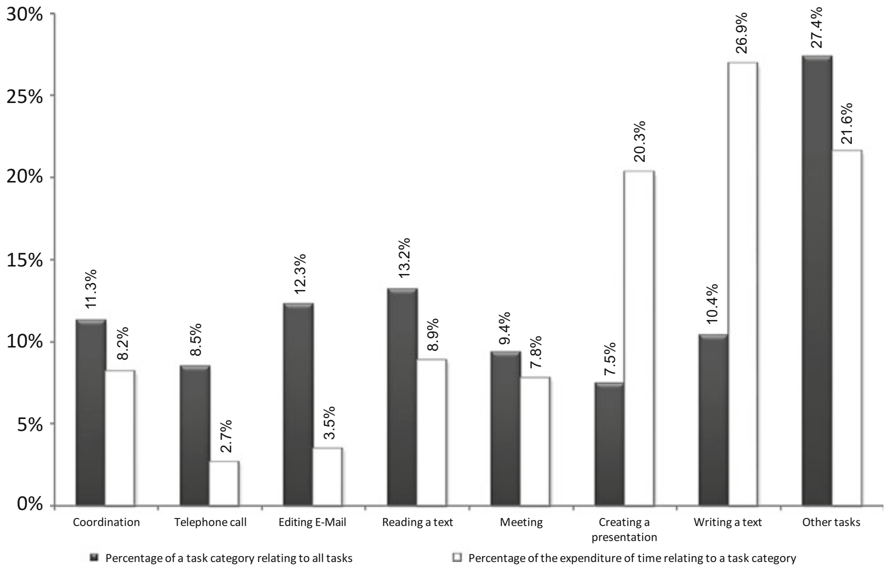


Fig. 5 Proportions of task categories on the observed days

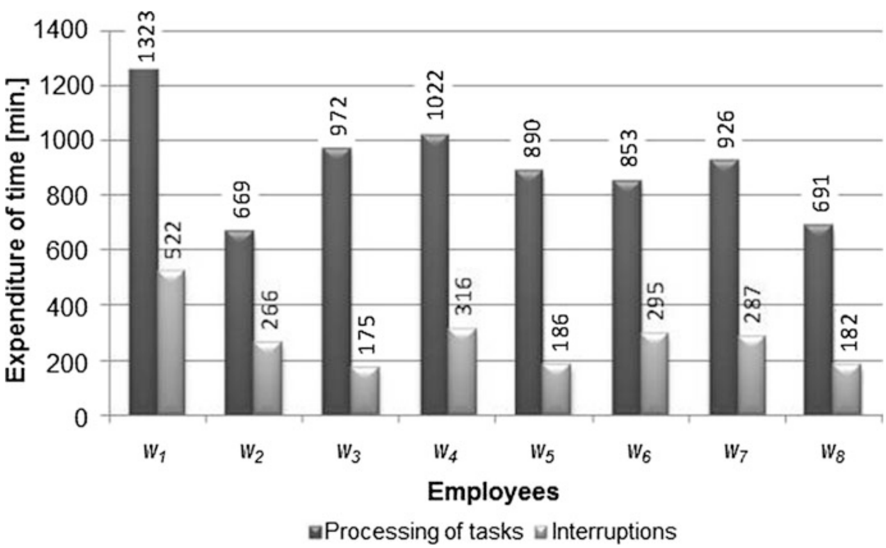


Fig. 6 Expenditure of time for task-processing and interruptions for each observed person

"*perceived work efficiency*" of employee w_5 indicate a productive individual work organization. But, in addition, the content of the tasks determined the continuity of processing and the number of interruptions. Thus, w_5 prepared an experimental setup in a laboratory. Therefore, his availability for other employees was limited.

Table 1 Values of the *KPIs* for the efficiency of the individual work organization

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	∅
WE_{ges}	1.22	1.24	1.19	0.871	0.87	1.21	1.02	1.10	1.10
PE_{ges}	0.92	0.89	0.78	0.57	1.29	0.87	0.66	0.49	0.81
PC_{ges}	0.61	0.55	0.45	0.38	0.91	0.78	0.39	0.42	0.56

The results in Table 1 show that six of the eight employees need a greater expenditure of time than initially planned ($WE_{ges} > 1$). Only the employees w_4 and w_5 achieved a work organization which can be termed as efficient. The further analysis of the data as well as the results of the interviews indicated that employee w_4 focus on a robust planning. Therefore, the values for $ET^p_{i,w}$ were in the majority of the tasks overstated. On the other hand, w_5 required significantly less time for two of his tasks due to a better-than-expected availability of information. The efficient individual work organization and task processing was confirmed by the *KPI* $PE_{ges,w5}$.

The statistical analysis of the key figures determines the presence of a high correlation between the *KPI* "*perceived work efficiency*" and the self-assessment of the employee ($r = 0.85$, $p < 0.05$). For the rating of the overall perceived quality of an individual work organization at the end of the day, a five points Likert scale was used. Further, a significant correlation between the *KPIs* "*work efficiency*" and "*perceived work efficiency*" was observed ($r = 0.92$, $p < 0.05$).

To illustrate the distribution of the interruptions over a regular work day, we use a curve progression of the prorated interruptions, as outlined in Fig. 7. The slope of the curve reflects the probability of an interruption and can help people to effectively manage their work. Within the time periods: 09:00–10:00 h, 11:00–12:00 h and 15:00–16:00 h 43 % of the interruptions occur during a work day. The main reasons for an interruption during the morning are telephone calls and urgent incoming emails. However, in the afternoons the interruptions are caused by conversations initiated by colleagues or self-initiated breaks resulting from drinking coffee or smoking cigarettes.

For instance, in the specific case of an individual work organization explored here, we analyzed only three consecutive working days of a person. Therefore, a feedback from another employee (different work system) regarding a specific work result could not be observed. Thus, no calculation of the *KPI* "*work effectiveness*" was possible to accurately assess the full impact of the individual work organization on the quality of the outcome. Also, the high proportion of "other tasks" indicates that we did not use a complete description of task categories for knowledge-intensive services. Our results suggest that future studies should investigate and extend the task categories of Mintzberg [17] and Mintzberg [19] and Beuscher-Mackay et al. [3] through an intensive observation of persons in the work systems of a service provision. As a result, the tasks and actions may be identified which determine the productivity of a service provision.

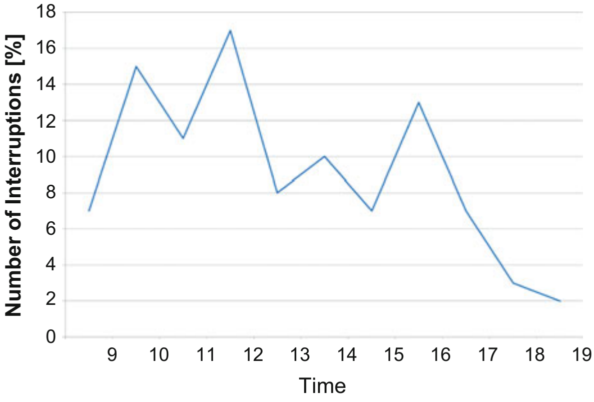


Fig. 7 Mean frequency of interruptions during a working day

6 Conclusion and Prospect

So far, the introduced performance measurement system and the implemented *APP* are restricted to the autonomous evaluation of the individual work organization of one employee. To achieve a realistic bottom-up evaluation of the productivity of a knowledge-intensive service provision, an extension of the proposed work analysis methodology is necessary, especially with regard to the collaboration of employees. Therefore, our future work will focus on the networking of the *APP*, which is then installed on different smartphones and tablets. With such a communication between different smartphones (*APP*-entities) and a master system, tasks can be assigned to different employees as well as a feedback regarding the current process status can be transferred. Based on the current information of the individual *APP*s the master would be able to calculate a forecast concerning the future progression of the service. Furthermore, the data generated by the *APP,* can be used by an algorithm of an assistance system. Such a system should be integrated in a future version of the *APP*, and may assist and improve the individual as well as the collective work organization. This includes the development of a task sequence and the definition of timeslots for processing specific tasks. The algorithm for a decentralized assignment of tasks to employees as well as the requirement to update the current task status via *APP*, will provide the data base for a situation based efficient planning and controlling of services (Fig. 8).

Due to the continual recording of individual actions and the evaluation of the corresponding effectiveness and efficiency, the system allows an individual performance rating. Therefore, the rights of codetermination (in Germany: §87 BetrVG) have to be considered during the conception and implementation phase of such a system.

Depending on the functions of the system, it may provide the benefit that an employee of a knowledge-intensive service company is empowered to document his/her individual work organization and to identify the existing potentials. To fully

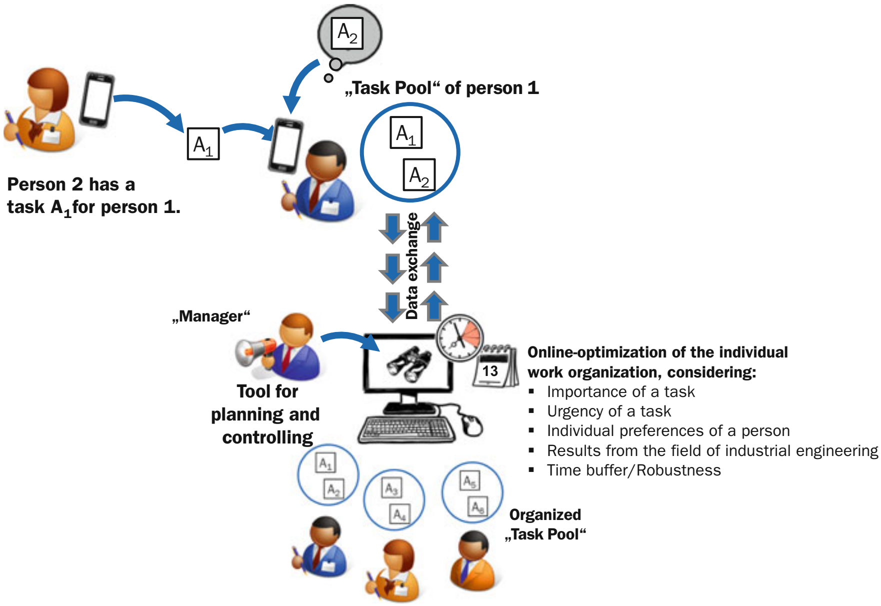


Fig. 8 Integrative concept for the person-centered monitoring and control of services

tap this potential, it will be necessary to ensure sufficient involvement of all actors of a service provision in the assessment to optimize the key figure "*service productivity*". A possible increase of effectiveness and efficiency does not necessarily imply a higher workload and higher amounts of tasks for each employee. It should rather reduce the self- and externally initiated interruptions, in order to ensure focused and continuous working.

References

1. Alvesson M (1995) Management of knowledge-intensive companies. Walter de Gryther, Berlin
2. Baumgärtner M, Bienzeisler B (2006) Dienstleistungsproduktivität – Konzeptionelle Grundlagen am Beispiel interaktiver Dienstleistungen. Fraunhofer IRB Verlag, Stuttgart
3. Beuscher-Mackay S, Jeske T, Steiger P, Hinrichsen S, Schlick CM (2009) Tätigkeitsanalyse und Personalentwicklungskonzepte. In: Schenk M, Schlick CM (Hrsg) Internationale Dienstleistungen und Internationalisierung. Gabler, Wiesbaden, 219–265
4. Cutrell E, Czerwinski M, Horvitz E (2001) Notification, disruption, and memory: effects of messaging interruptions on memory and performance. Paper presented at the human–computer interaction-interact_01, Tokyo
5. Dunkel H (1999) Psychologische Arbeitsanalyse: Verfahrensüberblick und Auswahlkriterien. In: Handbuch psychologischer Arbeitsanalyseverfahren, vdf Hochschulverlag an der ETH. Zürich, pp 9–30

6. Engelhardt S (2006) Neue Mitarbeiter erfolgreich einarbeiten – Erfolgreiche Unternehmen investieren in ihr Humankapital. Verlag W. Kohlhammer, Stuttgart
7. Fandel G (2005) Produktion I. Produktions- und Kostentheorie, 6. Aufl. Springer, Berlin
8. von Garrel J, Tackenberg S, Seidel H, Grandt C (2014) Dienstleistungen produktiv erbringen: Eine empirische Analyse wissensintensiver Unternehmen in Deutschland. Springer, Gabler, Wiesbaden
9. Gillie T, Broadbent D (1989) What makes interruptions disruptive. A study of length, similarity, and complexity. Psychol Res 50:243–250
10. Gonzáles VM, Mark G (2004) "Constant, Constant, Multi-tasking Craziness": managing multiple working spheres. Proceedings of Human Factors in Computing Systems (CHI 2004), pp 113–120
11. Grönroos C, Ojasalo K (2004) Service productivity towards a conceptualization of the transformation of inputs into economic results in services. J Bus Res 57:414–423
12. Hacker W (1984) Psychologische Bewertung von Arbeitsgestaltungsmaßnahmen: Ziele und Bewertungsmaßstäbe, Spezielle Arbeits- und Ingenieurpsychologie in Einzeldarstellungen – Band 1, 2. erw. Aufl. Springer-Verlag, Berlin u. a.
13. Horne JH, Lupton T (1965) The work activities of 'middle' managers? An exploratory study. J Manag Stud 2(1):14–33
14. Kamiske GF, Brauer J-P (2008) Qualitätsmanagement von A bis Z: Erläuterungen moderner Begriffe des Qualitätsmanagements. Hanser, München
15. Kirsch G (1983) Effizienz und Effektivität: Zielbestimmung in der Sozialpolitik. In: Pfaff M (Hrsg) Public transfers and some private alternatives during the recession, Duncker und Humblot. Berlin, pp 9–26
16. Lasshof B (2006) Produktivität von Dienstleistungen. Mitwirkung und Einfluss des Kunden. Dt. Univ.-Verlag, Wiesbaden
17. Mintzberg H (1970) Structured observation as a method to study managerial work. J Manag Stud 7:87–104
18. Mintzberg H (1971) Managerial work: analysis from observation. Manag Sci 18(2):B-97–B-110
19. Mintzberg H (1973) The nature of managerial work. Prentice Hall, Englewood Cliffs, NJ
20. O'Conaill B, Frohlich D (1995) Time in the workplace: Dealing with interruptions. Paper presented at the CHI proceedings, short papers
21. Perlow LA (1997) Finding time: How corporations, individuals, and families can benefit from new work practices. ILR Press, Ithaca, NY
22. Perlow LA (1999) The time famine: toward a sociology of work time. Adm Sci Q 44(1):57–81
23. Petz A, Duckwitz S, Mütze-Niewöhner S, Schlick CM (2013) Development and evaluation of a novel service productivity model. In: Meier H (Hrsg) Product-service integration for sustainable solutions. Proceedings of the 5th CIRP international conference on industrial product-service systems. Springer, Berlin, pp 383–394
24. Rennecker J, Godwin L (2005) Delays and interruptions: a self-perpetuating paradox of communication technology use. Inform Organ 15:247–266
25. Rubinstein JS, Meyer DE, Evans JE (2001) Executive control of cognitive processes in task switching. J Exp Psychol Hum Percept Perform 27(4):763–797
26. Schäfer E, Buch M, Pahls I, Pfitzmann J (2007) Arbeitsleben!: Arbeitsanalyse - Arbeitsgestaltung - Kompetenzentwicklung. Kassel University Press, Kassel
27. Schlick CM, Bruder R, Luczak H (2010) Arbeitswissenschaft, 3rd edn. Springer, Berlin
28. Wilson MA, Dierdorff (2013) Work analysis methods. In: Wilson MA, Bennett Jr W, Gibson SG, Alliger GM (Hrsg) The handbook of work analysis: methods, systems, applications and science of work measurement in organizations. Routledge, New York, NY, pp 1–2
29. Zülch G (1997) Industrial engineering. In: Luczak H, Volpert W (Hrsg) Handbuch Arbeitswissenschaft, Schäffer-Poeschel, Stuttgart, pp 183–187

Assessing and Increasing In*nova*tiveness of SMEs in the Context of Their *Demo*graphic Development

The Joint Project NovaDemo

Stefan Wassmann, Claudia Kramer, Sonja Schmicker, Barbara Deml, Sarina Töpperwien, and Marcel Förster

Abstract

The joint project NovaDemo focuses on assessing and increasing innovativeness of SMEs in the context of their demographic development. Based on analyses of age structure and innovation process analyses, a project-specific diagnostic instrument, the NovaDemo assessment tool, was developed. It is used to determine individual and group-specific innovativeness. At the individual level, four innovativeness types were identified: the "averagely creative innovation driver", the "slightly creative team worker", the "highly creative balanced type" and the "averagely creative reserved type". Based on these insights, the NovaDemo training programme was developed to increase innovativeness. The close proximity of learning and work situations in terms of time and content allows for the optimisation of efficiency thanks to a work-integrated execution adapted to the respective working group.

S. Wassmann (✉) • C. Kramer • M. Förster
Chair of Human Factors Engineering, Otto von Guericke University Magdeburg, Magdeburg, Germany
e-mail: stefan.wassmann@ovgu.de; claudia.kramer@ovgu.de; marcel.foerster@ovgu.de

S. Schmicker
METOP GmbH, Affiliated Institute to the Otto von Guericke University, Magdeburg, Germany
e-mail: sonja.schmicker@metop.de

B. Deml
ifab-Institute of Human and Industrial Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: barbara.deml@kit.edu

S. Töpperwien
Schunk Sintermetalltechnik GmbH, Thale, Germany
e-mail: sarina.toepperwien@schunk-group.com

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_2

Keywords
Innovativeness • Measurement • Assessment tool • Increase • Training programme • Work-integrated • Individual level • Group level • Age structure analysis • Innovation process analysis • Innovation process model • Barriers to innovativeness • Innovation-relevant personality traits • Innovativeness types

1 NovaDemo Joint Project Content and Process Description

As part of the demographic development in Germany, many companies have to face the challenge of maintaining consistently high quality of results with an increasingly aging workforce [1]. To nevertheless ensure or even increase the competitiveness of small and medium sized enterprises (SME), it is necessary to determine and subsequently increase their innovativeness at individual and group level.

The NovaDemo joint project explores this issue. The detailed requirement and baseline analyses comprise in addition to the age structure analysis of east German SMEs a comprehensive innovation process analysis in the form of qualitative interviews. These analyses form the basis for the derivation of seven innovation-relevant personality traits, which are described in Sect. 2.1. Subsequently, a diagnostic instrument for determining the individual and group-specific innovativeness was developed (refer to Sect. 2.2). The execution (refer to Sect. 2.3) as well as the most significant results at the individual and group-specific levels (refer to Sects. 2.4 and 2.5) are described in more detail. To specifically train and increase innovativeness the work-integrated NovaDemo training programme was subsequently developed (refer to Sects. 3.1 and 3.2). In addition to the sampling description (refer to Sect. 3.3), selected evaluation results of the NovaDemo training programme are presented from the participants' perspective (refer to Sect. 3.4). Furthermore, the insights gained will be made freely accessible to other companies. To achieve this, a short version of the NovaDemo assessment tool and of the NovaDemo training programme was derived. We will not present NovaDemo^{light} in this paper and refer you to the publication by Schmicker et al. [7].

This joint project is composed of three partners. The *Lehrstuhl für Arbeitswissenschaft und Arbeitsgestaltung, Otto-von-Guericke-Universität in Magdeburg* focuses mainly on the theoretical conception of the newly developed instruments and methods. The *METOP GmbH, An-Institut der Otto-von-Guericke-Universität Magdeburg*, acts as a link between science and industry and transfers the project-specific developments to the industry partner. The *Schunk Sintermetalltechnik GmbH* in the German city of Thale is the industry partner for the project. All practical tests took place at their premises. Additionally, the two transfer partners *Innovations- und Gründerzentrum Magdeburg GmbH* and *PMC Pulvermetallurgisches Kompetenz-Centrum Thale GmbH* facilitated the knowledge transfer to other companies and industries.

2 Innovativeness Assessment at Individual and Group Level Using the NovaDemo Assessment Tool

2.1 Objective and Theoretical Background of the NovaDemo Assessment Tool

The objective of the NovaDemo assessment tool is the assessment of the innovativeness of individuals and mixed-aged teams.

A key starting point for the development of the NovaDemo assessment tool is the internal innovation process analysis (in form of qualitatively conducted interviews). In addition to the extraction of innovation-relevant personality traits, one objective was the development of a project-specific innovation process model. As previous innovation process models (for example VDI Guideline 2221) had a strong technical, process and economic focus and did not place much emphasis on individual or social factors in innovation processes, the adaptation of conventional innovation process models seems reasonable.

Based on comprehensive literature research and the 13 conducted qualitative innovation process analyses, it was possible to develop seven innovation-relevant personality traits for the NovaDemo joint project (refer to Table 1). These personality traits illustrate the project-specific meaning of innovativeness and are determined with the NovaDemo assessment tool by the targeted use of an appropriate group exercise as well as tests and questionnaires.

Another significant result of the innovation process analysis is the graphical innovation process model (refer to Fig. 1).

Figure 1 shows firstly the phases of the invention process (upper part of the figure). Following the provision of the idea impulse, the idea generation begins. Subsequently, it is necessary to evaluate and refine these ideas before initiating the planning phase. The final step of the invention process is the implementation (here meaning the creation of a prototype). Due to the recursive loops, all phases are inter-connected and if necessary can be repeated. Secondly, the figure illustrates the relationship between the work required and the frequency of the social interactions during the individual phases (lower part of the figure).

Furthermore, a tabular innovation process model was created. It describes the individual phases in more detail and explains in depths the psychological, social, and ergonomic processes of the innovation process. Table 2 shows a more detailed list of psychological barriers to innovativeness based on the barrier model by Schlick [6], which was developed with the help of innovation process analyses and literature research.

Please refer to the dissertation by Wassmann [8] for a more detailed description of the social and ergonomic processes and further results.

Table 1 Innovation-relevant personality traits [8, p. 186 f.]

Innovation-relevant personality traits	Explanations/examples
Creativity	Creativity is understood as purely intellectual creative power—factors such as usefulness and feasibility have (as yet) no role to play at this level. Creativity is one of the most important elements of the individual innovativeness
Absence of mental barriers during the innovation process	Psychological innovation barriers are hindrances that have a negative impact on the progress of innovation processes. Individuals without psychological innovation barriers are able to protect the sensitive innovation process in all phases from harmful interferences
Social skills	Three important aspects of social skills are: (1) Assertiveness appropriate to the situation, (2) an adequate adaptability to social expectations and values of the respective group and (3) the skill to overcome personal conflicts constructively
Performance motivation	The higher the professional performance motivation of an individual, the more committed to achieving the respective objectives the individual is. Performance motivated individuals are also very confident of success
Methodical expertise	Individuals with a high level of methodical expertise have, for example, sound knowledge about creativity techniques, decision-making methods and moderation techniques. Furthermore, they know how to communicate these methods to a group and which method is useful for which situation
Ability to cooperate	Ability to cooperate describes to what extent an individual is able to cooperate with other individuals within a group and to what extent the individual is happy doing so. Additionally, the ability to cooperate refers to the skill to be able to maintain a common work climate for the whole team
Ability to communicate	Ability to communicate does not only encompass the verbal and non-verbal communication of information, but also what is commonly understood to be a 'good listener'

2.2 Structure of the NovaDemo Assessment Tool

The NovaDemo assessment tool is structured in two blocks, which are organised as illustrated in Table 3. Block I consists of a video-documented structured group exercise. These innovative design tasks illustrate an invention process most succinctly. Next, project-specific procedures are used for the assessment of innovation-relevant personality traits in Block II in addition to standard creativity tests and questionnaires.

The procedures selected for assessing the innovativeness closely reflect the seven innovation-relevant personality traits (please refer to Table 1). This way the individual level as well as the group level may be analysed. To validate the newly developed procedures already established tools were used (refer to Table 4).

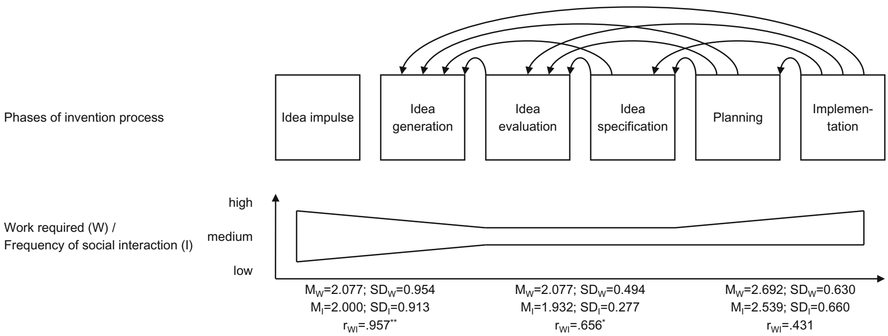


Fig. 1 Graphical innovation process model [8, p. 178]

Table 2 A more detailed list of barriers to innovativeness—psychological barriers (based on [8, p. 182])

Reason for barriers	Explanations/examples
Knowledge and thought patterns	Rejection of new ideas, habitual routines, insisting on ones own opinion, egoism, preconceptions, inability for self-criticism, lacking ability to solve conflicts, blinkered mind-set, forming an opinion too quickly
Inhibiting cognitions	Fear of risks, fear of changes, authority dependence, wanting to fulfil the expectations of others, wanting to give in, search for the perfect solution, low self-confidence, not wanting to impose ones ideas on others, fear to accept responsibility for new ideas
Motivation and activity-orientation related processes	Fear of commitment, indecisiveness, lack of inquisitiveness, flaws when defining objectives, lack of perseverance, satisfaction with the current achievements

2.3 Trialling the NovaDemo Assessment Tool

The NovaDemo assessment tool was tested from July to October 2013 under laboratory testing conditions and within the operational context of the industry partner. In total 94 participants took part in the tests. 32 % of the total sample were women. The age of the participants ranged from 18 to 79. 39.4 % of the participants were aged between 18 and 30, 17 % were between 31 and 50, and 43.6 % were 51 or older.

The ideal group consisted, corresponding to the actual operational situation determined during the baseline analysis, of six participants, with two women and four men as well as two younger and four older working together in the group exercise. Unfortunately, it was not possible to achieve this composition for all tests. To be able to make a statement regarding the relation between the age structure within the group and its innovativeness, five age-homogeneous young and five age-heterogeneous groups were observed in the laboratory.

Table 3 Blocks, content and duration of the NovaDemo assessment tool [8, p. 230]

Block		Content	Duration
Block I	Welcome and preparation	– Mutual introduction of the assessors and the participants – Explanations regarding the experiment process – Informing the participants about data protection and obtaining consent to record video footage	Approx. 10 min
	Structured group exercise	– Participants perform the group exercise based on continuous instructions by the assessors	Approx. 90 min
Break		– Respite for the participants	Approx. 15 min
Block II	Questionnaire and tests	– Participants answer the questionnaire and take the tests – The assessors introduce each procedure and provide (additional) verbal instructions – Freely selectable short break (approx. 5 min)	Approx. 90 min
	Saying goodbye	– Answering any questions the participants may have – Saying goodbye	Approx. 5 min
Total duration			Approx. 3.5 h

Following a short welcome by the assessors, the assignment for the structured group exercise was presented to the participants. The objective was to build a floatable pirate boat with the highest possible load capacity using only provided materials. Here the individual phases of the invention process (refer to Fig. 1) were completed under set time limits (the implementation phase being the exception). The success criteria included (1) the amount of the used materials in grams, (2) the time spent (in minutes) to build the boat during the implementation phase and (3) the maximum load capacity of the boat (measured by continuously loading glass marbles until the boat sunk).

Furthermore, an expert rating regarding the usefulness and originality of the boat as well as regarding the successful cooperation within the group was carried out. Figures 2 and 3 show examples of a "good" and a "bad" pirate boat. Additionally, an objective assessment at individual and group-specific level was made by trained observers using video recordings of the group exercises and structured observation sheets.

Next, the participants answered the questionnaires and performed the tests listed in Table 4.

2.4 Results of the NovaDemo Assessment Tool at Individual Level

Applying the cluster analysis four different innovativeness types were determined, which are presented using their profile paths (refer to Fig. 4). These differ significantly for all seven ascertained personality traits (refer to Table 1).

Table 4 Overview of all the procedures used in the NovaDemo assessment tool as well as the operationalised constructs [8, p. 194]

Title/description of the procedure	Operationalised variables/constructs	Self-development (S) or for validation (V)
Structured group exercise (observation at group level)	Social-sociological barriers, group climate, specific and general evaluation of the invention phases, evaluation of the innovation performance	S
Structured group exercise (observation at individual level)	Personality (Big Five), social skills, performance motivation, methodical expertise, ability to cooperate, ability to communicate	S
Questionnaire on socio-demographic information	Socio-demographic background information	S
Evaluation questionnaire for the group exercise	Reflection on the group exercise from the participants perspective (among others regarding the assignment, the collaboration, the joint approach)	S
Questionnaire for assessing innovation-relevant personality traits	Personality (Big Five), social skills, performance motivation, methodical expertise, ability to cooperate, ability to communicate	S
Tests for determining verbal and figurative and graphic creativity	Verbal and figurative creativity	S
Questionnaire for assessing psychological innovation barriers	Knowledge and thought patterns, inhibiting cognitions, motivation and activity-orientation related processes	S
NEO five factors inventory	Personality (Big Five)	V
Social skills inventory—abbreviated version	Social orientation, offensiveness, self-guidance, reflexibility	V
Performance motivation inventory—abbreviated version	Professional performance motivation	V
Extracts from the Berlin intelligence structure test	Verbal and figurative graphical inventiveness	V

The spread of the found innovativeness types among the test persons were as followed:

– "the averagely creative innovation driver": 27 %
– "the slightly creative team worker": 20 %
– "the highly-creative balanced type": 33 %
– "the averagely creative reserved type": 20 %

No differences regarding the allocation to these clusters and the gender were found [8].

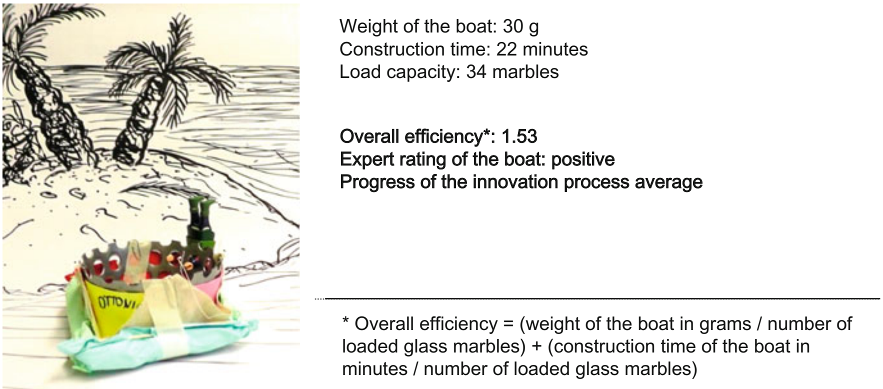


Fig. 2 Example for a “good” result [7]

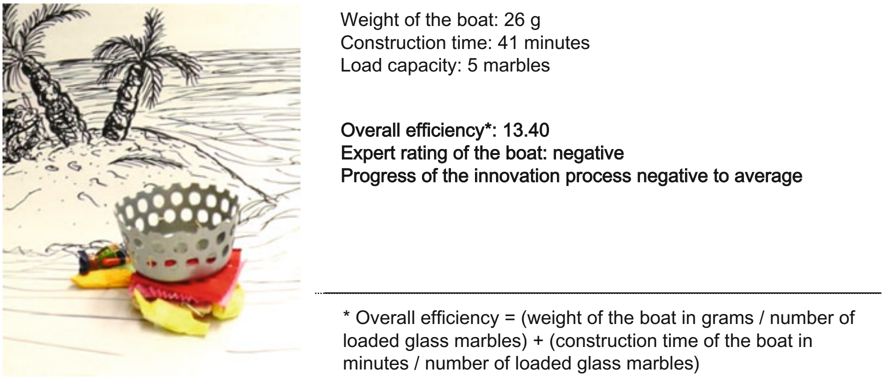


Fig. 3 Example for a “bad” result [7]

The variable ‘age’ showed significant correlations to two of the seven innovation-relevant personality traits. A negative relation between creativity and age and a positive relation between absence of barriers during the innovation process and the age of the participants was determined [8].

2.5 Results of the NovaDemo Assessment Tool at Group Level

Regarding the innovativeness at group level the resulting distribution of the four innovativeness types is shown in Fig. 5. The groups were assessed regarding their overall efficiency and according to the expert ratings (refer to Figs. 2 and 3) and subsequently rated as “little innovative”, “averagely innovate” or “highly innovative”. A balanced mix of the four ascertained innovativeness types seems recommendable for an at least averagely innovative team.

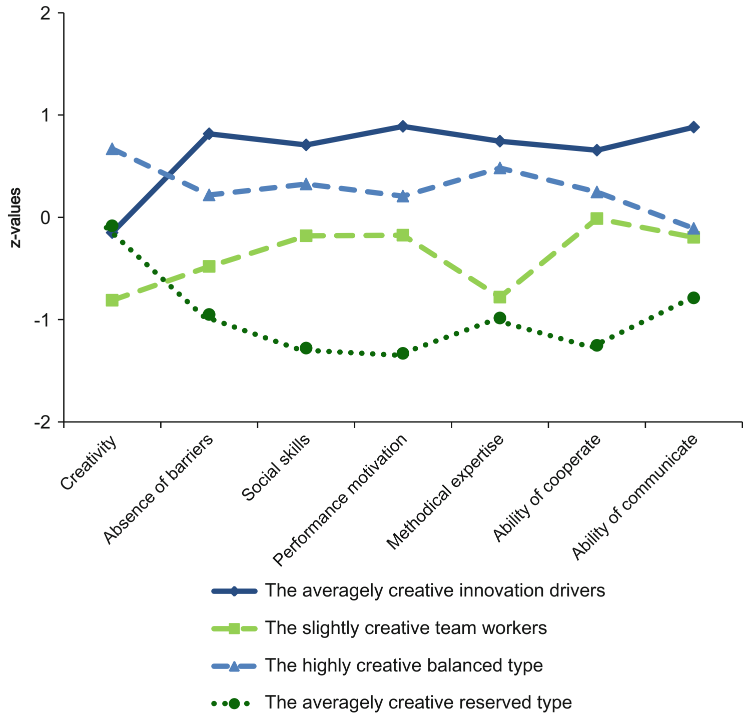


Fig. 4 Profile paths and z-standardised mean values of the four innovativeness types for the seven innovation-relevant personality traits [9, p. 215]

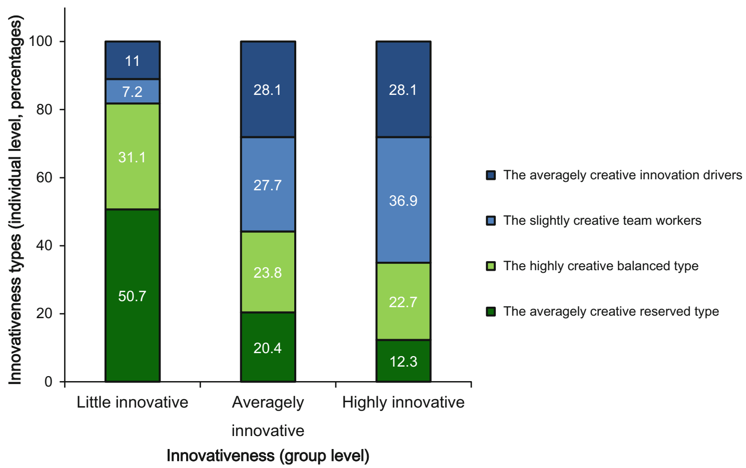


Fig. 5 The spread of the four innovativeness types across three groups with varying levels of innovativeness (in percentage) [8, p. 279])

3 Increasing Innovativeness at Individual and Group Level with the NovaDemo Training Programme

3.1 Objective and Theoretical Background of the NovaDemo Training Programme

The NovaDemo training programme is directly based on the insights gained using the assessment tool and aims to increase the individual and group-specific innovativeness. The personal strengths of the individuals are to be incorporated into the team composition and into the work process in an optimum manner. Here, the maintenance of a balanced mix of the four innovativeness types is of high importance.

To increase efficiency, the training units are conducted in a work-integrated manner (meaning "on the job", based on Conradi [2]). Real meeting situations are accompanied by the trainers to allow for direct knowledge transfer and feedback. The design of the training measures (meaning regarding duration, working hours, proximity to work content, location, and trainer) was developed together with the industry partner applying a participatory approach. Particularly suitable for the execution are meetings with innovation character that take place in regular intervals and with the same participants attending each meeting. Furthermore, the majority of the group members should have taken part in the NovaDemo assessment.

3.2 Structure of the NovaDemo Training Programme

The trained areas of competence within the training programme relate directly to the seven innovation-relevant personality traits, which were ascertained using the NovaDemo assessment tool (refer to Table 1). The training programme focuses on the following competency categories (refer to Table 5):

All training groups take part in the basic module to ensure a common knowledge base. Regarding the elective units, the participants may decide for themselves which training units the trainers should present and which are most suitable for the respective team. To ensure a time-efficient execution, the training units are work-integrated (refer to Fig. 6). This further reduces the spatial, time and content distance between training and work situations, which in turn has a positive effect on the knowledge transfer [3]. If taking part in the basic module plus the four elective modules, the participants invest about four extra hours **Fehler! Verweisquelle konnte nicht gefunden werden.**

In total, the training programme consists of five content-coordinated progressive components. At the start of a meeting, the trainer imparts subject-relevant competencies (= component 1). To make the imparted knowledge accessible at the company for the long-term, so-called "reminders" with concise summaries of the training content are handed out at the end of the training (= component 2). All "reminders" may be accessed at any time in the respective meeting rooms in a dedicated method box by the training participants. Component 3 comprises

Table 5 Training units of the NovaDemo training programme [4]

Basic module
– Communication beneficial for improving the team spirit
Elective module

– Creativity
– Absence of barriers in the innovation process
– Social skills
  – Avoidance and handling of conflicts
  – Assert ones own objectives
– Methodical expertise
  – Decision-making and assessment methods
  – Innovation conducive moderation
  – Structured problem solving
  – Goal-oriented specification
  – Sustainable scheduling
– Ability to cooperate

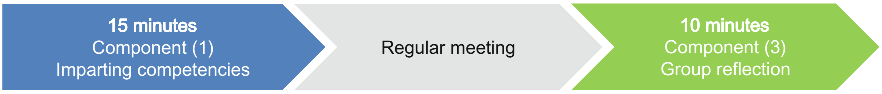


Fig. 6 Work-integrated process of the NovaDemo training programme [4]

worksheets for self- and group reflection. Using the worksheets allows the participants to outline their own progress and reveal their personal improvement potential. During the group reflection, strengths and weaknesses of the group regarding a specific competency category may be determined. The innovation advisor (= component 4; Kramer et al. [5]) is a knowledge resource providing the complete content of the components 1, 2, and 3 to interested parties online and is freely accessible (https://wikis.ovgu.de/iaf-ag/doku.php). Furthermore, during a consultation with the trainer all participants had the opportunity to discuss topics, which were not to be debated in the presence of the team.

3.3 Description of the NovaDemo Training Programme Sample

As an example, the training programme was tested in three meeting groups with representatives of the industry partner. The execution period started in April 2014 and ended in February 2015. Table 6 shows the age structure and the gender ratio. Figure 7 illustrates the spread of the innovativeness types within the training groups.

3.4 Evaluation of the NovaDemo Training Programme

Using a specifically designed evaluation questionnaires, the opinion of the participants regarding the practical feasibility of the training programme was

Table 6 Socio-demographic characteristics of the sample (training programme) [7, p. 221 f.]

Training group	N	Gender	Age group
1	3	♀: 33.3 % ♂: 66.7 %	Younger: 0.0 % Middle aged: 66.7 % Older: 33.3 %
2	12	♀: 8.3 % ♂: 91.7 %	Younger: 0.0 % Middle aged: 41.7 % Older: 58.3 %
3	8	♀: 12.5 % ♂: 87.5 %	Younger: 37.5 % Middle aged: 50.0 % Older: 12.5 %
Total	23	♀: 13.0 % ♂: 87.0 %	Younger: 13.0 % Middle aged: 47.8 % Older: 39.1 %

Notes: Age groups: Younger (up to 30 years of age), middle aged (31–50 years of age), older (from 50 years of age)

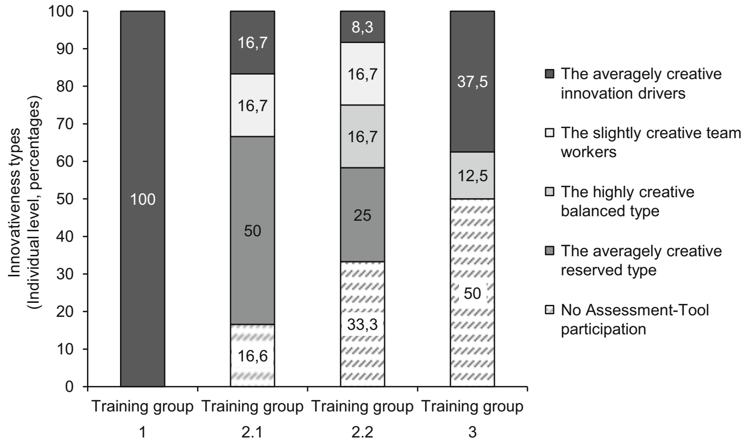


Fig. 7 Spread of the innovativeness types in the individual training groups [7, p. 223]

assessed. The items were rated using a five-point Likert scale, with 1 meaning, "I do not agree" and 5 meaning "I agree". In total, seven training participants evaluated the measures. Table 7 shows general statements on the NovaDemo training programme.

Based on these ratings, it is assumed that the participants' overall impression of the training programme as well as its content, execution, the procedure and the trainers was positive. Selected statements regarding the efficiency of the training programme also indicate concurrence with the relevant statements (refer to Table 8).

Table 7 General statements on the NovaDemo training programme (N = 7) [7, p. 239]

General statements on the NovaDemo training programme	M	SD
Overall I got a positive impression of the training programme	4.0	0.9
In my opinion, the training was well and logically structured	4.1	0.8
In my opinion, the content of the explanations was comprehensible and understandable	4.3	0.7

$M_{Overall} = 4.1$

Table 8 Statements on the efficiency of the NovaDemo training programme (N = 7) [7, p. 239]

Statements on the efficiency of the training programme	M	SD
I liked it that no special preparation was required for the individual training units	4.4	0.7
The scope of the training content was sufficiently informative	4.4	0.7
In my opinion, the ratio between the usefulness of the training programme and time spent and personal efforts made is adequate	3.7	0.7

Table 9 Statements on effectiveness and practical feasibility in general (N = 7) [7, p. 240]

Statements on effectiveness and practical feasibility	M	SD
The training gave me new insights and strengthened my understanding and my competencies	3.4	0.9
I can imagine integrating the insights gained during the training programme into my daily work routine	3.9	0.6
I am hoping that my participation in the training programme will have a positive impact on my personal (not professional) development	3.4	0.9
I expect that participating in the training programme will increase my innovativeness	3.1	0.6
I am hoping that participating in the training programme will increase my innovativeness at group level (meaning within the innovation team)	3.3	1.2
I think that thanks to the integration of the training times into existing meetings, it was possible to focus on real work situations, which in turn created practical relevance	3.6	0.9

$M_{Overall} = 3.2$

With $M_{Overall} = 3.7$ the structure as well as the process of the training programme were found to be good. Furthermore, the participants were able to rate the duration of the training using a three-step scale (too long [1], adequate [2], too short [3]). Due to the mean value of $M = 2.4$ ($SD = 0.5$) it is assumed that the duration of the training was adequate, but tended to be considered as too short. The statements on effectiveness and practical feasibility also received positive ratings (refer to Table 9).

Based on these results, a time-delayed impact of these training measures may be assumed. The participants indicated a personal increase in knowledge through the training programme between 20 % and 70 %. All training groups estimated the increase in overarching competencies to be at 50 %. Table 10 illustrates the overall rating of the NovaDemo training programme.

Table 10 Statements regarding the final rating and satisfaction (N = 7) [7, p. 241]

Statements regarding the final rating and satisfaction	M	SD
In my opinion the training programme is suitable to train the innovativeness of individuals	3.6	0.9
In my opinion the training programme is suitable to train innovativeness at group level	3.9	0.8
I think that participating in this training programme will have a positive long-term impact on the company's success	3.4	0.9

The participants "enjoyed" taking part in the training programme (M = 4.1; SD = 1.0) and would recommend it in its current form (M = 3.7; SD = 0.7). At the end, the participants were asked to apply an overall rating between 1.0 ("very good") and 5.0 ("insufficient") to the training. With a mean value of 2.4 (SD = 0.6) the training was on average rated as good. Overall, these ratings suggest that the applicability of the training programme in SMEs appears appropriate and promising.

Improvement suggestions were only made regarding the organisational framework conditions. No criticism regarding the content of the concepts was made.

4 Summary of the Most Significant Insights of the NovaDemo Joint Project

The NovaDemo joint project delivers significant insights regarding the understanding of innovativeness at individual and group level for both science and industry.

The NovaDemo innovation process model consists of six phases (idea impulse, idea generation, idea evaluation, idea specification, planning, and implementation) that are inter-connected through recursive loops and may be repeated. Furthermore, effort and frequency of social interactions differ depending on the respective phase within the invention process.

Based on the seven innovation-relevant personality traits (creativity, absence of barriers, performance motivation, methodical expertise, ability to cooperate, ability to communicate), the four innovativeness types "the averagely creative innovation driver", "the slightly creative team worker", "the highly-creative balanced type" and "the averagely creative reserved type" were identified at individual level using the NovaDemo assessment tool and cluster analytical calculations. In this present study, a balanced mix of these four types delivered an at least averagely innovative result.

The NovaDemo training programme is intended to increase innovativeness and is carried out in a work-integrated manner. The training units relate closely to the different innovativeness types and are individually adapted to the composition within the respective groups. The five components of the training programme (imparting of competencies, method box, worksheets for self- and group reflection, innovation advisor, consultation) serve a successful knowledge transfer and the sustainability of the training measures. Furthermore, the participants' open-mindedness and acceptance regarding the training programme had a positive impact on its success and effectiveness.

NovaDemo^{light} is the abbreviated version of the presented instruments and methods and is intended to allow professionals to make an initial assessment (in the form of a screening) and to increase the innovativeness independently in their own company.

Acknowledgement This project has been funded by the German Federal Ministry of Education and Research (BMBF) and by the European Social Fund of the European Union.

References

1. Buck H (2006) Lernende Unternehmen zur Bewältigung der Alterung der Belegschaft. In: Statistisches Bundesamt (publisher) Statistik und Wissenschaft. Demographischer Wandel – Auswirkungen auf das Bildungssystem. Contributions to the scientific colloquium from 18 and 19 November 2004 in Wiesbaden. SFG Servicecenter Fachverlage, Reutlingen
2. Conradi W (1983) Personalentwicklung. F. Enke Verlag, Stuttgart
3. Kals E (2009) Arbeits- und Organisationspsychologie kompakt. Weinheim, Beltz
4. Kramer C, Schmicker S, Deml B, Wassmann S, Töpperwien S (2015) Training zur Steigerung der Innovationsfähigkeit von Einzelpersonen und altersheterogenen Arbeitsgruppen – Das Trainingsprogramms NovaDemo. In: Gesellschaft für Arbeitswissenschaft e. V. (Hrsg.) VerANTWORTung für die Arbeit der Zukunft. GfA-Press, Dortmund
5. Kramer C, Schmicker S, Wassmann S, Töpperwien S, Förster M, Deml B (2015) Innovations-ratgeber. Otto-von-Guericke-Universität. Magdeburg. https://wikis.ovgu.de/iaf-ag/doku.php
6. Schlick GH (1995) Innovationen von A-Z: Begriffe, Definitionen, Erläuterungen und Beispiele. Expert. Renningen-Malmsheim
7. Schmicker S, Wassmann S, Kramer C, Töpperwien S, Förster M, Deml B (2015) Erfassung und Steigerung der Innovationsfähigkeit von KMU vor dem Hintergrund der demografischen Entwicklung – Ergebnisse des Verbundprojekts NovaDemo für Wirtschaft und Wissenschaft. In: Bullinger AC, Cernavin O, Richter D, Schmicker S (eds) Innovationspotenziale im demografischen Wandel. Forschung und Praxis für die Arbeitswelt. Rainer Hampp Verlag, München, Mering
8. Wassmann S (2014) Innovationsfähigkeit im demografischen Wandel: Entwicklung und Validierung des Assessment-Tools NovaDemo zur Erfassung der Innovationsfähigkeit von Einzelpersonen und Arbeitsgruppen. Dissertationsschrift. Karlsruher Institut für Technologie (KIT). Karlsruhe. http://nbn-resolving.org/urn:nbn:de:swb:90-436569
9. Wassmann S, Schmicker S, Deml B, Kramer C, Töpperwien S (2014) Ermittlung der Innovationsfähigkeit von Einzelpersonen und altersheterogenen Arbeitsgruppen: Das Assessment-Tool NovaDemo. In: Gesellschaft für Arbeitswissenschaft e. V. (Hrsg.) Gestaltung der Arbeitswelt der Zukunft. GfA-Press, Dortmund

The Quality Culture Inventory (QCI): An Instrument Assessing Quality-Related Aspects of Work

Christine Sattler, Karlheinz Sonntag, and Katja Götzen

Abstract

Quality is considered to be one of the most important competitive factors of profit as well as non-profit organizations. Accordingly, quality assurance as well as quality management procedures represent established tools within the management strategy of numerous organizations. In addition to that, the importance of developing and furthering an organization-specific quality culture is increasingly emphasized. In order to operationalize the quality culture construct we developed the Quality Culture Inventory (QCI), which was administered to N = 93 employees in the framework of a comprehensive pilot study. First results support the validity and reliability of the developed quality culture scales. The QCI represents the first systematical approach allowing organizations to assess the status quo of their quality culture. Respective results provide an empirically-based foundation to reflect on quality culture and derive specific measures for quality development.

Keywords

Quality • Quality culture • Operationalization

1 Theoretical Background

Quality ranks among the most important competitive factors of profit as well as non-profit organizations. Accordingly, quality assurance and quality management procedures constitute central issues within the management strategy of numerous

C. Sattler (✉) • Kh. Sonntag • K. Götzen
Department of Industrial and Organizational Psychology, Heidelberg University, Heidelberg, Germany
e-mail: christine.sattler@psychologie.uni-heidelberg.de; karlheinz.sonntag@psychologie.uni-heidelberg.de; katja.goetzen@psychologie.uni-heidelberg.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_3

organizations [8, 23]. Intensive research over the past decades has led to countless quality assurance concepts and tools allowing organizations not only to monitor but also develop their quality standards.

In this context, DIN EN ISO 9000/9001 and Total Quality Management (TQM) represent established approaches which are applied by a wide range of organizations worldwide. DIN EN ISO 9000 [9] and DIN EN ISO 9001 [10] standards provide fundamentals and requirements for the design and implementation of quality management systems in terms of a structured action guideline, while the concept of TQM stresses the importance of continuous quality development. Accordingly, organizations applying TQM commit themselves to constant quality improvement on all organizational levels, including the quality management system itself [6, 7].

However, apart from emphasizing formal and structural aspects of quality, current discussions extend the understanding of quality assurance by focusing on the actual *realization* of quality efforts, namely the *quality culture* of an organization. The quality culture concept emphasizes the importance of informal and non-structural aspects of quality. These include a high quality awareness of leaders and staff as well as shared attitudes and values towards quality.

The first comprehensive definition of quality culture was developed within the higher education context. According to the European University Association (EUA) [11, 15]:

“Quality Culture refers to an organisational culture that intends to enhance quality permanently and is characterised by two distinct elements: on the one hand, a cultural/psychological element of shared values, beliefs, expectations and commitment towards quality and, on the other hand, a structural/managerial element with defined processes that enhance quality and aim at coordinating individual efforts” [11, p. 10].

According to this definition, quality not only involves the permanent realization of formal quality assurance measures (structural/managerial element) but also the consideration of informal elements such as shared values towards quality issues (cultural/psychological element). Although the definition of the EUA was developed within the higher education context, it is immediately transferable to all other types of organizations dealing with quality assurance issues.

Although there is growing awareness regarding the importance of promoting and developing an organization-specific quality culture, only a small number of empirical studies addressed the operationalization of this construct so far. Consequently, we pursued the objective to contribute to this highly relevant research field by developing an empirically-based Quality Culture Inventory (QCI) within the framework of the heiQUALITY Cultures Project.

2 The heiQUALITY Cultures Project: Main Objectives

The heiQUALITY Cultures Project ("Development and Testing of an Instrument for the Description and Assessment of Quality Cultures in Higher Education Institutions") was initiated in 2012 (funding period: 4/2012–5/2015). The main objectives of the research project were:

1. to develop a comprehensive definition and assessment model of quality culture
2. to develop a feasible and valid Quality Culture Inventory for higher education institutions, which can be easily modified to suit other types of organizations within the profit and non-profit sector
3. to describe and assess the quality culture of organizations on a structural-formal *and* organizational-psychological level
4. to analyze the strengths and weaknesses of quality cultures
5. to derive target-oriented recommendations for action in order to further and develop the current quality culture of organizations

According to these objectives, the Quality Culture Inventory is not only intended to answer a large variety of scientific questions. First and foremost the QCI aims to be a practice-oriented tool, which enables organizations to assess their current quality culture as well as analyze its strengths and weaknesses. Results of respective analyses are supposed to represent an essential basis to develop recommendations for quality development and justify the implementation of quality actions based on empirical evidence.

The following sections provide a detailed overview about our methodology in order to achieve these objectives.

3 Methodology

At the beginning of the heiQUALITY Cultures Project we defined several milestones, which represent essential achievements during the project term (Fig. 1). In a first step we performed a systematic literature review in order to identify previous studies focusing on the operationalization of quality culture. In a second step we conducted $N = 41$ international expert interviews relating to potential dimensions of the quality culture construct. Following the results of the systematic literature review and the international expert interviews we constructed a comprehensive assessment model capturing the essential dimensions of quality culture. The assessment model served as an empirical basis for the development of the Quality Culture Inventory, which was tested concerning its validity and reliability within the framework of a comprehensive pilot study ($N = 93$).

The following subsections provide more detailed information about the methodology and essential results of the respective milestones.

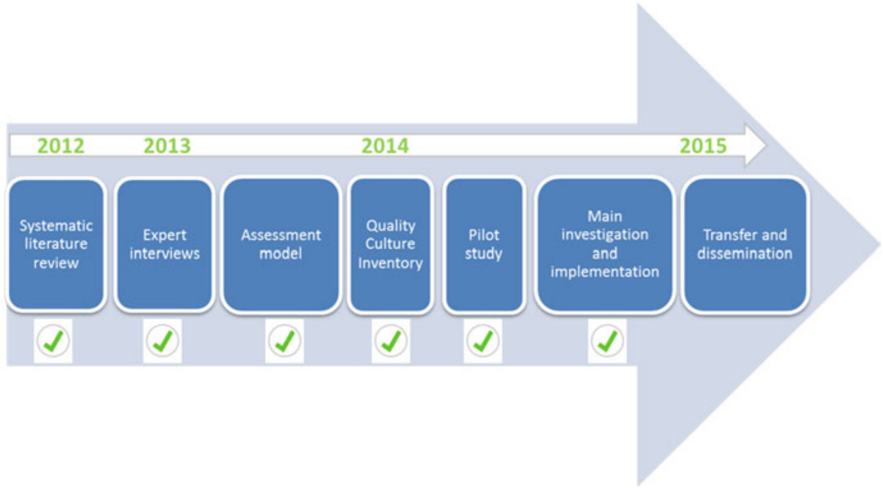


Fig. 1 Milestones of the heiQUALITY cultures project

3.1 Systematic Literature Review

The first step of the heiQUALITY Cultures Project was to perform a systematic literature review on previous studies operationalizing the quality culture construct. A main objective of this review was to identify relevant quality culture dimensions, which need to be considered within the framework of a comprehensive quality culture model. The systematic literature review included three databases: HEIDI (interdisciplinary database of Heidelberg University), PSYNDEX, and PsycINFO. First, we performed a title search on the term "quality culture" and its German equivalent "Qualitätskultur". A second title search focused on the combination of the terms "quality", "higher education" and, "dimension". The systematic literature review was limited to publications within peer-reviewed journals until December 2012.

Our search strategy initially resulted in 786 matches (Table 1). However, only 181 publications were considered relevant to our research questions after reading all abstracts in detail. As little as 21 of these studies applied an empirical approach operationalizing quality and/or quality culture. Another 4 of these studies were excluded from further analyses due to lacking relevance after reading the full paper. Accordingly, 17 research papers were considered appropriate for a detailed evaluation with respect to our research objectives.

Although identified by the same research strategy, the selected papers focus on operationalizing rather heterogeneous aspects of quality. While only 3 out of 17 studies address the operationalization of "quality culture" directly [1, 21], all remaining studies focus on the assessment of general "quality dimensions" (N = 8 [2, 16]) or "service quality" (N = 6 [4, 12]) respectively. With the exception of one paper, all selected studies apply a survey approach in order to operationalize the

Table 1 Results of the systematic literature review

Database	Search terms	Total hits	Empirical studies	Other relevant publications	Exclusion
HEIDI*	Quality Higher education Dimension	98	18	31	49
HEIDI	Qualitätskultur/ qualitaetskultur	92	0	15	77
PSYNDEX*	Quality Higher education Dimension	6	0	3	3
PSYNDEX	Qualitätskultur/ quality culture	57	0	21	36
PsycINFO*	Quality Higher education Dimension	59	[1[a]]	24	35
PsycINFO	Quality culture	474	3	66	405
TOTAL		786	21[b]	160	605

Notes: *Peer-reviewed journals only; [a]This publication was additionally found in the HEIDI data base and is thus not counted in the "Total" row; [b]4 publications were subsequently excluded due to lacking relevance

quality construct of interest. In this context, recipients of quality culture (students) clearly represent the primary and sometimes only target group of the developed surveys (N = 13). Strikingly, only one study includes the view of the leadership level, whereas five studies consider staff as a relevant target group.

In summary, the systematic literature review identified a total number of 17 empirical studies with respect to our research questions. The selected studies focus on the operationalization of various quality dimensions within organizations. In fact, only three out of 17 studies were found to address the quality culture construct directly—a result which supports our approach of promoting more empirical research in this field.

3.2 International Expert Interviews

In order to explore the quality culture construct in more detail we conducted N = 41 international expert interviews (time period: March–May 2013). The main objective of these interviews was to generate additional information on quality culture dimensions, which need to be considered in the construction of the QCI. Another main focus was to explore supportive as well as hindering factors with respect to the development of an organization's quality culture.

Selection criteria for being a subject-matter expert (SME) included a high level of practical experience concerning quality assurance in organizations as well as relevant research publications on quality assurance and/or culture (for a more detailed description of selection criteria see Sattler et al. [17]). The interviews were performed on the basis of a semi-structured interview guideline following a previous approach on operationalizing the concept of "learning culture" by Sonntag et al. [18, 19]. By deliberately choosing an open question format the explorative character of the interviews was taken into account. Sample questions include "What do you associate with the term quality culture?" or "From your point of view, how can a quality culture be furthered sustainably?" The interviews were mainly conducted in a face-to-face situation (N = 35). However, due to economic reasons, 6 interviews were performed via telephone. The interview length was 60 min on average, resulting in an extensive interview material of approximately 41 h in total.

Executive summaries of each interview served as the foundation to identify relevant quality culture dimensions from the experts' point of view. All interviews were analyzed by two trained raters. In case of inconsistencies ratings were put up for discussion in order to reach agreement. The final step of the interview evaluation was to cluster and summarize the obtained results.

Strikingly, a total number of 39 experts (95.1 %) emphasized the importance of *quality-oriented leadership* and *communication* during their interview. Moreover, *commitment*, *participation*, and *quality objectives* were identified as relevant quality culture dimensions by more than 70 % of our experts. Finally, *shared quality values*, *mutual trust*, *individual responsibility*, *recognition* and *information* ranked among the 10 most frequently mentioned dimensions (agreement rate > 65 %).

The respective results provide substantial information concerning specific dimensions, which need to be considered in the assessment of quality culture. This holds especially true for quality-oriented leadership and communication as well as commitment and participation (for a more detailed outline of the expert interview results see Sattler et al. [17]). Sample quotes on the latter dimensions are provided in Table 2.

3.3 Assessment Model of Quality Culture

The results of the systematic literature review and our international expert interviews provided us with empirical evidence for the development of our assessment model (see Fig. 2). According to this model quality culture comprises a structural-formal and an organizational-psychological level.

The structural-formal level is further divided into *normative*, *strategic* and *operative* aspects of quality assurance and quality management. According to Bleicher [3] normative aspects comprise goals and principles which are designed to enable the viability of an organization. With respect to quality culture normative management finds its expression in the presence of officially documented quality goals for instance. Responsibilities for the implementation of quality goals are determined on the strategic level (e.g. a central quality assurance unit). Finally, specific tools and measures for quality assurance and development are located on the operative level (e.g. controlling, evaluation).

Regarding the organizational-psychological level we differentiate between individual and collective quality culture dimensions. The model postulates that every

Table 2 Sample quotes on selected quality culture dimensions

Dimension	Sample quotes
Leadership	"Leadership is absolutely important [as a constituent of quality culture]. If there is no understanding at the leadership level it will not work."
Communication	"If we try to develop something that is very complicated and only understandable by real experts, then you will have problems. So communication and the idea of developing a vocabulary that is understandable for everybody is very important."
Commitment	"Commitment [. . .] [is] the key aspect [..] of organizational life". "Quality culture is something active. It always has something to do with creativity, participation, and commitment of as many stakeholders as possible."
Participation	"Quality culture is a participative thing. It has to be constructed together. It has to be debated and discussed and agreed upon."

Note: Sample quotes from N = 41 international expert interviews conducted from March–May 2013

Fig. 2 Assessment model of quality culture [17]

individual has an impact on the lived quality culture of an organization in terms of individual commitment, sense of responsibility, and engagement towards quality. In contrast to that, collective elements are characterized by interactions between organizational members. The collective dimension is represented by quality-oriented leadership, communication, and participation which are supported by shared values and trust among organizational members.

The model assumes that the quality culture dimensions do not exist independently from each other but are interrelated. As a result, leadership, communication and participation are depicted as an arrow, illustrating a dynamic connection between the structural-formal and the individual level. For example it is presumed that *quality goals* on the normative level need to be *communicated* adequately in order to reach every *individual* of an organization.

In summary we developed a comprehensive assessment model of quality culture which represents an empirical basis for the actual operationalization of the quality culture construct. The model reflects both empirical research and the expertise of international professionals concerning quality assurance and quality culture.

3.4 Quality Culture Inventory (QCI)

Following the creation of the quality culture model we developed the Quality Culture Inventory (QCI) in order to operationalize the quality culture construct. The QCI comprises two questionnaires which refer to the structural-formal and the organizational-psychological level respectively. The following subsections provide a detailed insight to the respective questionnaires including sample items.

3.4.1 Structural-Formal Questionnaire

In order to operationalize the structural-formal level we created the "structural-formal questionnaire". This part of the QCI is targeted at quality assurance experts within an organization. The 73 items of the questionnaire are based on the results of a comprehensive literature review focusing on normative, strategic as well as operative aspects of quality assurance. Within the scope of standardized interviews subject-matter experts are asked to indicate the level of implementation regarding normative (e.g. quality goals), strategic (e.g. responsibilities) and operative elements (e.g. instruments) of quality assurance with respect to various performance areas of an organization. The structural-formal questionnaire pursues the objective to assess the status quo of the organization's quality assurance and quality management in order to reveal potentials concerning structural-formal quality development. The questionnaire was validated by $N = 4$ experts who were asked to evaluate its content as well as its practicability. Sample items are provided in Table 3.

3.4.2 Quality Culture Questionnaire

In order to operationalize the organizational-psychological level we developed the "quality culture questionnaire". In contrast to the structural-formal questionnaire, this part of the QCI addresses all members of an organization. First, we performed a comprehensive literature review aiming to identify previous instruments operationalizing our proposed quality culture dimensions. In the course of this review we identified a total number of 76 potentially relevant questionnaires. In a next step, the corresponding scales and items were thoroughly evaluated regarding their relevance to our research objectives. Suitable items were selected and, if necessary, adapted in terms of content. Moreover, we decided to develop new items for those quality culture aspects which were not covered by previous empirical approaches

Table 3 Sample items of the structural-formal questionnaire covering the normative, strategic, and operative levels of quality assurance within organizations

Quality Goals (normative level)
Has your organization developed quality goals?
☐ Yes ☐ No, but we are currently developing quality goals ☐ No
Responsibilities for Quality Assurance (strategic level)
Has your organization created a special position and / or department which exclusively deals with quality assurance? (Multiple answers possible)
☐ Yes, on the central level (please specify): ____ ☐ Yes, on a decentralized level (please specify): ____ ☐ No, but we are currently planning to introduce such a position and / or department on the central level ☐ No, but we are currently planning to introduce such a position and / or department on a decentralized level ☐ No
Quality assurance instruments (operative level)
Does your organization assure the efficiency of its quality assurance measures by applying a comprehensive quality control system (e. g. feedback loops)?
☐ Yes, *all* our quality assurance measures are interconnected by a corresponding control system ☐ Yes, *some* of our quality assurance measures are interconnected by a corresponding control system ☐ No, but we are currently planning to introduce a corresponding control system ☐ No

(e.g. quality values). Respective items were created on the basis of the results of our systematic literature review and international expert interviews.

The quality culture questionnaire comprises a total number of 53 items. Participants are asked to indicate their level of agreement towards various statements on a 6-point Likert scale ("doesn't apply at all" to "fully applies"). Sample items of the questionnaire are provided in Table 4.

The individual dimensions—commitment, responsibility, and engagement—are represented by 4 items each. The same holds true for three of the collective dimensions, namely participation, shared values and trust. In contrast to that, quality-oriented leadership is operationalized by 12 items, accounting for the proposed multidimensional nature of the construct (e.g. providing appropriate feedback, setting high quality standards). For the same reason, quality-oriented communication is represented by a total number of 9 items (see Table 4).

Table 4 Sample items of the organizational-psychological questionnaire covering the individual and collective levels of quality culture within organizations

Dimension	Definition	Sample item	Source	Items
Commitment	Quality commitment describes the extent to which organizational members identify with the quality goals of their organization and feel emotionally obliged to perform their daily work assignments on a high quality level	"I am particularly concerned to support the quality development of my organization"	Adapted from: Jackson [14] (affective subscale)	4
Engagement	Quality engagement refers to the willingness to make additional efforts in order to enhance the quality of work	"I am willing to make additional efforts in order to meet the quality demands of my work"	Adapted from: Jackson [14] (behavioral subscale)	4
Responsibility	Quality responsibility can be defined as the perceived obligation to care for and actively support quality issues	"I feel that I am jointly responsible for the quality development of my organization"	Adapted from: Jackson [14] (cognitive subscale)	4
Leadership	Quality-oriented leaders set high quality standards and show exemplary behavior regarding the quality of their own work. Moreover, they provide appropriate feedback towards quality-related aspects of work and support the professional development of their staff	Leader Self-Assessment: "It is important to me to appreciate good working results adequately." Employee Assessment: "My supervisor provides constructive feedback concerning the quality of my work"	Adapted from: Heinitz and Rowold [13]	12
Communication	Quality-oriented communication is expressed by a shared use of quality-related information as well as an open dialogue regarding measures of quality development	"Ideas concerning quality improvement are openly discussed in our department"	Adapted from: Brodbeck et al. [5]	9
Participation	Participation is defined as the willingness to take an active part in the ongoing events of an organization	"I keep myself up to date concerning new developments at my organization"	Adapted from: Staufenbiel and Hartz [20]	4

(continued)

Table 4 (continued)

Dimension	Definition	Sample item	Source	Items
Shared values	Shared quality values express themselves as leading principles in the quality development of an organization	"Quality values of my organization are actually put into practice"	Own development	4
Trust	Trust can be defined as a positive social attitude towards someone (e.g. leader, colleague) or something (e.g. quality management strategy)	"I have full confidence in my employee's skills"	Adapted from: Zeitz et al. [22]	4
Global aspects	Evaluation of the whole organization concerning quality-related aspects	"Our organization is characterized by high quality awareness"	Own development	8
Total				53

Additionally, we decided to add 8 "global items" to the questionnaire. These items require an assessment of the quality culture of the whole organization, thereby representing "global aspects" of an organization's quality culture.

The quality culture questionnaire is designed as an online tool. In order to proof its validity and reliability a first version of the questionnaire was tested within the framework of a comprehensive pilot study (N = 93). Potential participants were contacted via email distribution lists provided by three higher education institutions, which took part in the pilot study. According to the content alignment of the heiQUALITY Cultures project, the pilot sample included employees representing heterogeneous status groups at higher education institutions (e.g. professors, administration).

The following section summarizes the main results of the pilot study.

4 Main Results of the Pilot Study

In total N = 93 higher education institution members participated in the pilot study of the quality culture questionnaire. In terms of group affiliation N = 25 (26.9 %) of the participants belonged to the leadership level (management, professors), whereas N = 68 (73.1 %) participants represent various employee groups (i.e. academic, administrative, secretarial and service staff).

The reliability values obtained within the pilot study can be judged as mainly satisfactory (see Table 5). This particularly applies to the scales assessing quality-oriented leadership, participation, shared values, trust as well as global aspects of quality culture. The internal consistencies (Cronbach's Alpha) of the respective

Table 5 Quality culture questionnaire: dimensions, number of items and reliability (r_{it} = corrected item-total correlation, α = Cronbach's Alpha)

Section	Dimensions	Items	r_{it} Min	r_{it} Max	α
Quality culture	Commitment	4	.335	.498	.603
	Engagement	4	.417	.583	.696
	Responsibility	4	.134	.324	.358
	Leadership	12	.502	.839	.935
	Communication	9	.167	.818	.871
	Participation	4	.396	.663	.716
	Shared values	4	.629	.629	.772
	Trust	4	.381	.642	.734
	Global aspects	8	.555	.824	.889
	Total	53			

scales range between α = .716 and α = .935, indicating an acceptable (participation) to excellent (leadership) reliability. Furthermore, corrected item-total correlations (r_{it}) of the respective items show acceptable to very good values.

The scales operationalizing commitment and engagement were found to be in a critical but still acceptable range (α = .603 and α = .696 respectively). In contrast to that, the responsibility scale yielded an inacceptable internal consistency (α = .358), indicating a need for revision. Moreover, the scale operationalizing quality-oriented communication showed a special characteristic: the communication scale obtained a good internal consistency of α = .871—however, the corrected item-total correlations of three corresponding items ranged between $r_{it} = .167$ and $r_{it} = .192$, which has to be judged as inacceptable.

Following the results of the pilot study, items assessing individual commitment, responsibility and engagement towards quality were partially modified for the final version of the quality culture questionnaire in terms of wording. Moreover, three items assessing communication were excluded from the final questionnaire due to insufficient corrected item-total correlations. All remaining items were retained unmodified following their positive evaluation within the pilot study. Consequently, the final version of the quality culture questionnaire consists of N = 50 items in total.

5 Discussion and Future Prospects

The revised version of the Quality Culture Inventory represents the first systematical and empirical approach allowing for a comprehensive assessment of quality culture. The QCI offers organizations the opportunity to assess the status quo of their quality culture, in order to derive specific measures for quality development.

In the meantime, the revised Quality Culture Inventory was administered to all members of the higher education institutions, which took part in the pilot study

(N = 789). In addition to the actual quality culture dimensions, several potentially dependent (e.g. commitment towards the organization) as well as moderating variables (e.g. conscientiousness) were considered in the scope of the main investigation. Consequently, numerous further results regarding the validity, reliability, and practicability of the questionnaire can be expected in the near future.

With respect to limitations it needs to be mentioned that our pilot study exclusively considered organizations of the higher education context. However, it is assumed that the questionnaire can be easily adapted to organizations of the private sector by slightly modifying the wording and context of the quality culture scales and items.

In summary the QCI represents a cost- and time-effective instrument assessing quality culture in organizations. Results offer a sound empirical basis for reflecting on and discussing about quality culture in order to generate target-oriented ideas for its progress and development.

Acknowledgements The heiQUALITY Cultures Project is funded by the "Innovation and Quality Fund (IQF) of the Ministry of Sciences, Research and Arts Baden-Württemberg, Germany" (Innovations- und Qualitätsfonds (IQF) des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg, Deutschland). Term: 04/2012–05/2015.

References

1. Ali HM, Musah MB (2012) Investigation of Malaysian higher education quality culture and workforce performance. Qual Assur Educ 20(3):289–309
2. Ardi R, Hidayatno A, Zagloel TYM (2012) Investigating relationships among quality dimensions in higher education. Qual Assur Educ 20(4):408–428
3. Bleicher K (2011) Das Konzept Integriertes Management. Visionen – Missionen – Programme. Campus, Frankfurt
4. Brochado A (2009) Comparing alternative instruments to measure service quality in higher education. Qual Assur Educ 17(2):174–190
5. Brodbeck FC, Anderson N, West M (2000) TKI: Teamklima-Inventar. Hogrefe, Göttingen
6. Brüggemann H, Bremer P (2012) Grundlagen Qualitätsmanagement. Vieweg & Teubner, Wiesbaden
7. Defeo JA, Juran JM (2010) Juran's quality handbook: the complete guide to performance excellence. McGraw-Hill, New York, NY
8. Defeo JA, Juran JM (2014) Juran's quality essentials: for leaders. McGraw-Hill, New York, NY
9. DIN Deutsches Institut für Normung e. V. (ed) (2005) Qualitätsmanagementsysteme – Grundlagen und Begriffe. Quality Management Systems – Fundamentals and Vocabulary (ISO 9000: 2005). Beuth, Berlin
10. DIN Deutsches Institut für Normung e. V. (ed) (2008) Qualitätsmanagementsysteme – Anforderungen. Quality Management Systems – Requirements (ISO 9001: 2008). Beuth, Berlin
11. European University Association (2006) Quality Culture in European Universities: A Bottom-up Approach. Report on the Three Rounds of the Quality Culture Project 2002–2006. European University Association. Brussels. http://www.eua.be/eua/jsp/en/upload/quality_culture_2002_2003.1150459570109.pdf. Accessed 27 May 2015
12. Gallifa J, Batallé P (2010) Student perceptions of service quality in a multi-campus higher education system in Spain. Qual Assur Educ 18(2):156–170

13. Heinitz K, Rowold J (2007) Gütekriterien einer deutschen Adaptation des Transformational Leadership Inventory (TLI) von Podsakoff. Zeitschrift für Arbeits-und Organisationspsychologie 51(1):1–15
14. Jackson PR (2004) Employee commitment to quality: its conceptualisation and measurement. Int J Qual Reliab Manag 21(7):714–730
15. Loukkola T, Zhang T (2010) Examining Quality Culture: Part 1 – Quality Assurance Processes in Higher Education Institutions. European University Association. Brussels. http://www.eua.be/pubs/Examining_Quality_Culture_Part_1.pdf. Accessed 27 May 2015
16. Murias P, de Miguel JC, Rodríguez D (2008) A composite indicator for university quality assessment: the case of Spanish higher education system. Soc Indic Res 89(1):129–146
17. Sattler C, Götzen K, Sonntag Kh (2013) Assessment of Quality Cultures in Higher Education Institutions – First Results from the heiQUALITY Cultures Project. Paper presented at the 8th European Quality Assurance Forum (EQAF), Gothenburg, Sweden. http://www.eua.be/Libraries/EQAF_2013/1b_4_Sattler.sflb.ashx. Accessed 27 May 2015
18. Sonntag Kh, Schaper N, Friebe J (2005) Erfassung und Bewertung von Merkmalen unternehmensbezogener Lernkulturen. In: Arbeitsgemeinschaft Betriebliche Weiterbildungsforschung e.V./Projekt Qualifikations-Entwicklungs-Management (ed) Kompetenzmessung im Unternehmen. Lernkultur- und Kompetenzanalysen im betrieblichen Umfeld. Waxmann. Münster 18:19–340
19. Sonntag Kh, Stegmaier R, Schaper N et al (2004) Learning in organisations – operationalizing learning culture. Zeitschrift für Lernforschung 32(2):104–127
20. Staufenbiel T, Hartz C (2000) Organizational Citizenship Behavior: Entwicklung und erste Validierung eines Messinstruments. Diagnostica 46(2):73–83
21. Trivellas P, Dargenidou D (2009) Organisational culture, job satisfaction and higher education service quality: The Case of Technological Educational Institute of Larissa. TQM J 21 (4):382–399
22. Zeitz G, Johannesson R, Ritchie JE (1997) An employee survey measuring total quality management practices and culture: development and validation. Group Organ Manag 22 (4):414–444
23. Zollondz HD (2011) Grundlagen Qualitätsmanagement. Einführung in Geschichte, Begriffe, Systeme und Konzepte. Oldenbourg, München

Team Work and Leadership in an Aging Workforce: Results of an Intervention Project

Franziska Jungmann, Frank Hilgenberg, Susanne Porzelt, Michael Fischbach, and Jürgen Wegge

Abstract

The demographic change in most industrialized countries has increased both the average age of employees and the age diversity of the workforce. As a consequence, organizations and managers are keen to find new strategies that help promoting health and performance of older workers as well as the effective cooperation between young and old employees in increasingly age-diverse work teams. We briefly present a theoretical model that explains why and under which conditions age diversity is indeed a problem for effective team work. Next, we summarize key findings of the project TED that examined several of these factors in the manufacturing industry. Job tasks in this industry are often characterized by high physical work load and restricted degrees of autonomy which makes work highly susceptible to the effects of demo graphic change. In this project, we evaluated the effectiveness of a newly developed intervention for leaders. In cooperation with a German car manufacturing company, 1600 employees and their leaders were investigated over 3 years. The results show, inter alia, the expected positive impact of a training that promotes age-differentiated leadership on performance and health of employees.

Keywords

Demographic change • Age-differentiated leadership • Age diversity • Teamwork • Aging and work • Training • Intervention • Manufacturing industry

---

F. Jungmann (✉) • S. Porzelt • J. Wegge
Professur für Arbeits- und Organisationspsychologie, Technische Universität Dresden, Dresden, Germany
e-mail: franziska.jungmann@tu-dresden.de

F. Hilgenberg • M. Fischbach
Volkswagen AG, Standortoptimierung Kassel, Baunatal, Germany

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_4

1 Introduction

The population is aging in most industrialized countries. While life expectancy is rising, birth rates are lower than ever before for many years. These developments have an influence on the age composition and structure of the workforce. As a result of the demographic change, at the moment (in 2015) in Germany, for example, every third employee is 50 years and older [26]. By 2060, the participation rate of elderly employees (50-years and older) in the workforce will have increased by 10–20 % in most industrialized countries [25]. Additionally, lower birth rates will lead to severe shortages in the labor pool for many branches [1]. In an attempt to buffer the unfavorable effects of these demographic developments, many governments have decided to increase the official retirement age and shorten the duration of education (e.g., Bologna reform). Overall this leads to an increased age diversity in the workforce [e.g., 1, 29]. Hence, organizations are challenged with older people working longer and at the same time with people starting to work earlier in their life. Among many organizations and their managers the assumption prevails that the aging workforce may impair productivity due to increased absenteeism, lower innovation and reduced performance [25, 27]. Therefore, organizations and the human resource management are called upon to find strategies for the successful integration of older employees and for developing an age-differentiated work design that ensures a healthy aging at work for both younger as well as older employees across the working life span. In the following chapter, we first describe well supported theories that indeed predict negative effects of high age diversity in work teams. Next, we present a study conducted in the manufacturing industry aiming at analyzing and improving age-diverse teamwork and age-differentiated leadership in production teams.

2 Theoretical Background

2.1 Age Diversity and Age-Diverse Teams

The composition of a team has several consequences for team functioning. Theories of social categorization [28] and models of information elaboration and decision making [e.g., 15, 32] predict that age diversity has rather negative effects. In the process of social comparison, team members will distinguish between similar and dissimilar team members. As people tend to favor, trust and cooperate more with people who are similar, diversity within a team may lead to biases, emotional conflicts or discriminative behavior and, thus, hinder team functioning and performance. In contrast, models of information elaboration and decision making in teams propose that teams consisting of diverse team members with different experiences, perspectives, and knowledge show improved team functioning and performance based on the exchange and elaboration of information. Thus, there are good arguments in favor and against high age diversity in teams. Not surprisingly,

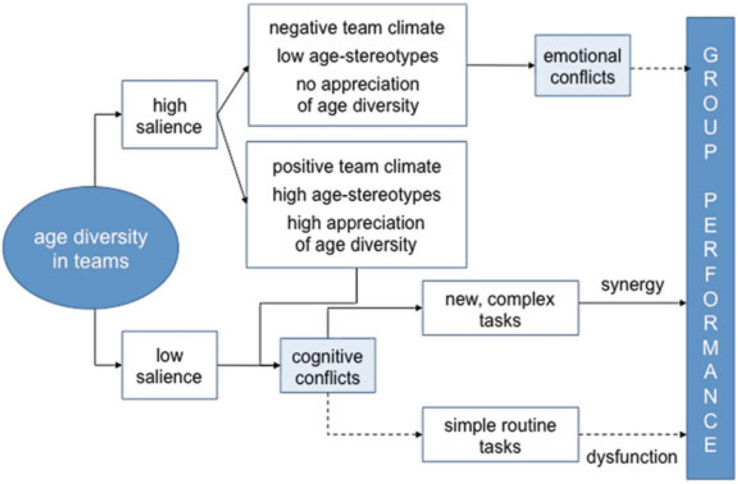


Fig. 1 Model of group effectivity in age-diverse teams [34]

results of empirical studies on this question are rather inconsistent (for an overview see [20, 35]).

However, recent review articles and meta-analytic studies show that negative effects of age diversity typically are stronger than positive effects [13, 22, 38]. Hence, it is crucial to identify and implement optimal conditions for effective group functioning and performance in age diverse teams. One model that summarizes the different paths linking team composition in terms of age with group performance as well as indicating the main moderating and mediating variables regarding potential benefits and risks related to age diversity was developed by Wegge and colleagues [34; see Fig. 1]. This model is based on a 6-year research program in which data from more than 745 natural teams with 8848 employees in different fields were collected. Moreover, central assumptions of this model were tested with a representative survey of the German workforce. The model proposes that—ceteris paribus—age diversity in work groups will have negative effects on group performance, motivation, and health of group members because objective age diversity in teams increases the salience of age diversity (i.e. whether diversity is indeed observed by group members and therefore triggers social categorization). When the salience of age diversity is high, conflicts increase and this yields low team effectiveness. However, it is further postulated that, under favorable conditions, beneficial effects should be observed, too. High appreciation of age diversity (i.e. positive judgments regarding the value of age diversity in team work), a positive team climate and high team task complexity (novelty) are considered as favorable moderating variables. Thus, in testing the model, corresponding measures of age diversity salience, age diversity appreciation, and team climate were developed [34].

Current research supports this model [see also 7, 35]. Thus, it can be concluded that a balanced team composition with regard to age, high appreciation of age diversity a positive team climate and low levels of age stereotypes among team members as well as an age-differentiated work design are relevant for age diversity to exert its positive effects. Based on these insights, a new training for supervisors which addresses the aforementioned aspects and seeks to improve team performance and health of team members by implementing age-differentiated leadership was developed.

2.2 Age-Differentiated Leadership

Leaders' attitudes and behaviors have a significant impact on the health and performance of employees [8, 39]. This is illustrated, for example, by a Finnish longitudinal study [12, 30] revealing that age-differentiated leadership is the only significant factor that maintained and fostered work ability among elderly employees. A low level of age-stereotypes, an open communication, the willingness to cooperate, and creating possibilities for individual work design characterize this style of leadership. However, previous studies also have shown that leaders tend to have high levels of age stereotypes and show more age-discriminative behaviors [24] resulting in negative effects for employees health. Thus, training for leaders reducing such stereotypes and promoting those behaviors that support work ability across the working life span are warranted. In addition to the findings from Ilmarinen and Tempel [12] as well as Tuomi and colleagues [30], it is also important that leaders consider the needs and motives of every age group, especially younger and older employees, as well as pay attention to the specific aspects of age-diverse teams [see above; 37]. For example, for elderly employees it is more important to share their knowledge and experiences than for younger employees [11]. If elderly employees can act upon this so-called generativity motive, they experience positive emotions and receive appreciation for their knowledge and experience [41]. As for younger employees, leaders should support learning, skill development, and career development by offering varying tasks and feedback on a regular basis [23, 41]. Positive effects of this extended model of age-differentiated leadership could be found in several studies [14, 35, 37]. Age-differentiated leadership was associated with increased team performance in terms of fewer conflicts among teams as well as less emotional exhaustion, improved well being, and fewer turnover intentions among employees. So far, however, it remains unclear whether similar effects can be implemented in the area of production.

2.3 Challenges in the Manufacturing Industry

Job tasks in production are characterized by short cycle times, high rate of repetitions, low decision latitude, shift-work, and high physical workloads [25]. These jobs show a rather negative age trend as physical capabilities tend to decrease with age [e.g., 6, 17] and these age related changes can hardly be compensated [25, 33]. Consequently, increased absenteeism and reduced performance among elderly workers can be observed in the manufacturing industry [e.g. 7]. Thus, it is particularly important to design healthy work places for employees to prevent health impairments and maintain work ability of employees across their working life span [5].

3 Objectives of the Project

Due to above mentioned high physical and psychological job demands, the car manufacturing industry is highly susceptible to the demographic challenges [i.a. 6, 7]. The project TED (short for the German project title "Teamarbeit und Fuehrung im demographischen Wandel [teamwork and leadership in times of demographic change]") was designed to make a valuable contribution to existing approaches within the cooperating organization seeking to promote health of young and old employees. Within the project, we studied the impact of an aging workforce on outcomes related to health and performance as well as conceptualized and evaluated the effectiveness of a leadership training focusing on the specific demands in this industry. The newly developed intervention aimed at raising awareness of age, aging, and the demographic change as well as maintaining and fostering the performance of age-diverse teams and promoting health and work ability of aging production workers. In the following sections we describe the procedure and study design of the project as well as the conceptualization of the training and its evaluation.

4 Methodological Design

4.1 Procedure, Study Design and Sample

The project was carried out in cooperation with a German car manufacturing company, producing car components and genuine parts. About 90 teams with about 1600 blue-collar workers and their 135 respective leaders of two production lines were studied over a 2-year time span. The employees worked on paced assembly and manufacturing lines or work stations in shift work. Like other car manufacturing companies in Germany, this organization prevented lay-offs during the economic crisis in 2008 and 2009 by granting older employees (55 years and older) partial retirement models with financial compensation [see 6]. This intervention intensively altered the age structure of the work force and had an impact on the way elderly employees were seen, seemingly as "replaceable". In the year of 2008

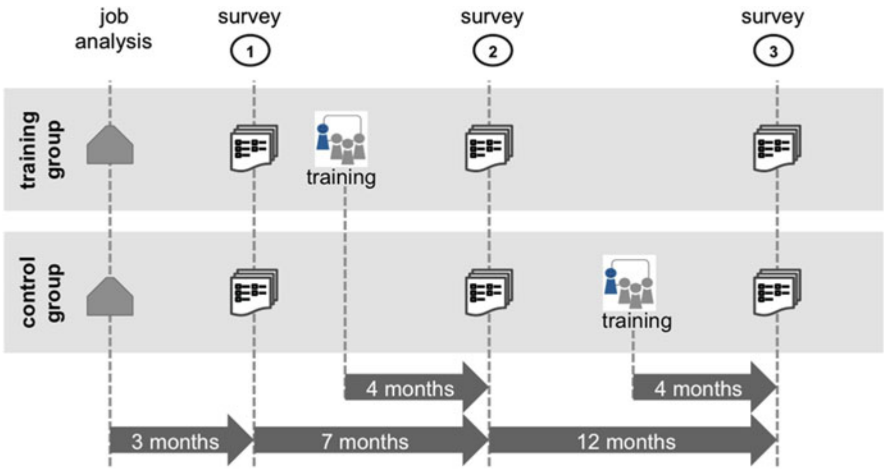


Fig. 2 Schematic design of the project

the average age of the workforce was 42.2 years. However, it is anticipated that the average will increased by about 5 years until 2018. In the light of this prognosis, the car manufacturing company is looking for strategies to cope with the demographic challenges, especially how to maintain and foster health among all groups in the workforce and how to optimize the teamwork of younger and older employees. To achieve the overall project objective, several methods were used which are described below. Figure 2 summarizes both, the procedure and the evaluation design.

4.1.1 Objective Job Analysis

At first, we conducted a job analysis to (1) learn about the work tasks and organizational aspects of the work (e.g., shift work, job rotation, task interdependence, communication, breaks) and (2) identify age-critical work stations. The results were used for the training intervention as well (see Sect. 4.2). To assess the psychological aspects of work tasks, work organization and work environment, we adapted standardized methods [e.g., 9, 19, 40] to the specific situation of the organization. The resulting observational interviews were conducted by two trained psychologists for all different job tasks of employees and teams on the production line.

4.1.2 Survey Among Employees and Leaders

All employees of the 90 teams (a total of 1600 employees) in the participating production lines and their respective leaders were asked to complete a survey that included all the relevant study variables. In the first stage of the project, the survey was used to assess the current status of teamwork and leadership in regard to aging, team work, and leadership as well as health and performance. In total 973 employees and team leaders from all teams completed the survey (response rate of 65 %). The employee's average age was 42.4 years (+/− 9.5 years). The sample was 2.8 % female. The teams consisted of about 9.4 (+/− 7.2) team members. With regard to age, the mean age among the teams ranged from 30.7 to 51.8 years. Following Harrison and Kleins [10] terminology, age diversity was measured utilizing the standard deviation of the individual age to measure age diversity as separation and the Blau-Index to measure age diversity as variety.

Working conditions (e.g., physical job demands, autonomy), various process variables of teamwork and leadership (e.g., information elaboration, emotional conflicts within teams, age stereotypes, age-differentiated leadership) and health as well as performance-related outcomes (e.g., innovative behaviors, work ability, physical and psychological health complaints, job satisfaction) were assessed using standardized and scientifically validated questionnaires. The survey data was complemented with objective data on team composition, including age, gender, and ethnicity of the employees within the teams. Furthermore, we were able to include objective data on absenteeism and productivity as indicators of health and performance on team level.

4.1.3 Evaluation Design

A training-group waiting-control-group design was utilized to assess the effectiveness of the training intervention. Using the evaluation framework of Kirkpatrick [16], we assessed the effectiveness on four different levels. Immediately after the training, the participating leaders were asked how satisfied they were with the training, the trainers, and organizational aspects of the training. Furthermore, we tested their knowledge of training related topics before and after the training. To assess changes in behavior, we surveyed all employees and leaders again after 4 and 16 months after the training (see Fig. 2). The data was again collected using standardized scales. This longitudinal design enabled us to compare the short and long term effects of the intervention as well as its impact on the training and control group. Changes on an organizational level were assessed by ratings of team effectiveness of both employees and leaders as well as by analyzing objective data on absenteeism.

4.1.4 Documentation of Organizational Change

As this project was conducted with natural teams of the organization, organizational changes (such as additional shifts, restructuring, fluctuation among leaders and team members) occurred on a regular basis as part of an everyday routine. In order to register and later on control for possible changes, we documented alterations utilizing a standardized form assessing all changes affecting the teams on a monthly basis (not shown in Fig. 2; based on [21]). This enabled us to estimate the impact of parallel organizational change, to comprehensively evaluate the short and long-lasting effectiveness of the intervention, and to identify important moderating and mediating variables.

4.2 Conceptualization of the Training Intervention

Similar to other diversity trainings, the intervention developed for this study aims at increasing awareness for diversity, reducing discrimination and stereotyping and thus changing the respective behavior [3]. Such "(age) awareness workshops" are considered to be an important approach when working towards the reduction of prejudices and stereotyping [24]. The training was, in particular, designed to raise participants' awareness for the topic of age-diverse teams, its effects and for the underlying processes. Another objective was to provide leaders with different options to handle and reduce the occurrence of age stereotypes and to support appreciative and age-differentiated leadership behavior. In order to achieve this objective, we applied a variety of methodology [2] that goes beyond the mere transfer of knowledge and information. Instead it allowed the trainees to discuss and practice critical situations and incidents. In line with recent findings on changing people's attitudes [see 18], the training differentiates between implicit attitudes that are subtle and only take effect on a person's behavior due to activation or salience and explicit attitudes that are consciously accessible. Both types are explained and discussed during the training sessions. Table 1 provides on overview of the training structure, the modules with their objectives and content.

The training aimed at increasing the awareness of the participating leaders for changes with age, challenges with an aging work force as well as optimizing age diverse team work. Furthermore, leaders were ought to learn, discuss and implement options for handling age stereotypes, for a rewarding/appreciative and age-differentiated leadership behavior, facilitating a synergetic collaboration among young and old team members, and designing age-differentiated organization and contents of work tasks. Thus, with the training we intended to elicit changes in behaviors (e.g., learning and practicing skills and abilities) and attitudes (e.g., increasing awareness and knowledge, see [14, 35]). To ease the learning process and transferring training content into everyday routine, a diverse methodology and an appealing didactical approach was applied (e.g., discussions, small group activities, role plays, case studies). By involving other internal stakeholders (e.g., from the human resource department or health management) a permanent and sustainable implementation of the training was fostered. The final training intervention consisted of six modules (see Table 1).

Table 1 Overview over contents of modules in the training intervention

Objectives	Contents
(1) Demographic change, age and aging	
– Being aware of the impact of demographic change, aging at work, and age stereotypes	– Changes with aging (i.e., health, motivation, learning, social competencies) – Age stereotypes: processes, advantages and disadvantages
(2) Leadership of an aging workforce	
– Understanding the importance of leadership for employees health and performance – Knowing and practicing principles of age-differentiated leadership, especially communication and reward/appreciation	– Age-differentiated leadership: definition, principles, effects – Basis principles of communication – Reward/appreciation of employees performance
(3) Age-diverse teams and change processes	
– Being aware of age diversity at work – Knowing and understanding the value of age diversity and increased appreciation – Knowing about impact of organizational change on individual and team performance	– Teamwork in companies and organizations – Age diverse teams: definition, processes, risks and opportunities – Change processes: definition, effects
(4) Age differentiated work design	
– Knowing basic principles of age-related aspects of work design and organization	– Basic principles of age-differentiated work design (including short breaks)
(5) Action planning	
– Developing an action plan and strategies how to implement it	– Summary of all modules – Discussion of possible implementation actions and drawing an action plan
(6) Review	
– Refreshing central topics of previous modules – Increasing the effectiveness and sustainability of planned actions	– Review of central topics – Discussion of current state of implementation of planned action – Group discussion and/or role plays on problem solving in process of implementing action plans

5 Selected Results

5.1 Impact of Age-Diverse Team Work and Age-Differentiated Leadership

We used the first survey wave to analyze the current status among all teams and identify fostering and hindering conditions of team work and leadership. The results show that in teams with high age diversity (in terms of increased separation), team members reported more emotional conflicts ($r = +.22$; $p < .10$) and reduced innovative behavior ($r = -.29$; $p < .01$). In line with this, teams with higher levels of age diversity show elevated levels of (objectively measured) absenteeism ($r = +.14$; $p < .10$).

Furthermore, the more team members experience their leaders to show behaviors of age-differentiated leadership, the more they report improved collaboration within the team (more elaboration of information: $r = +.55$; $p < .01$; fewer age stereotypes: $r = -.43$; $p < .01$), increased performance (enhanced innovative behavior: $r = +.36$; $p < .01$) and better health (increased recovery experience after work: $r = +.39$; $p < .01$; elevated work ability: $r = +.39$; $p < .01$; reduced absenteeism: $r = -.26$; $p < .01$). These findings support once again basic propositions of the underling research model (Fig. 1) and indicate—as expected—that age-differentiated leadership is beneficial for the work ability of employees.

5.2 Evaluation of the Training Intervention

As mentioned above, the training aimed at increasing awareness of topics associated with demographic change and an aging workforce and its implications for team performance and employees' health. The participating leaders developed an action plan for both their leadership behavior and for changes within their team, including aspects of work design. We evaluated the training on four levels following the evaluation approach of Kirkpatrick [16].

Besides a great interest in the topics of the training intervention the participants reported a high level of satisfaction with its content and stated that their expectations had been met (see Fig. 3). In addition, the participating leaders reported increased knowledge on all the relevant topics after the training (see Fig. 4).

Moreover, the training had a positive impact on topics covered in the training as well as on the cooperation within in the teams in relevant variables. Regarding the team leaders' attitudes towards the additional value of age-diverse teamwork, the expected interaction effect could be found. Before the training there was no measurable difference between the team leaders. However, leaders that participated in the training intervention reported more appreciation for age diversity in teams than leaders in the control group. Similarly, team members of trained leaders showed higher levels of appreciation for age diversity in teams (see Fig. 5, left). This was especially important since the appreciation of age diversity in teams is associated with higher job satisfaction, lower conflicts within the teams, and more innovative behavior [4, 31, 36]. Furthermore, we were able to show effects of the

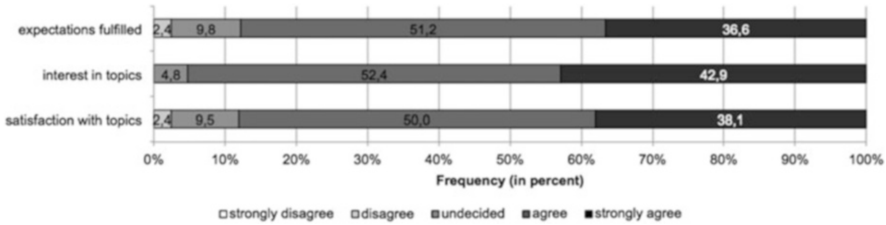


Fig. 3 Training evaluation: satisfaction with the training of participating leaders

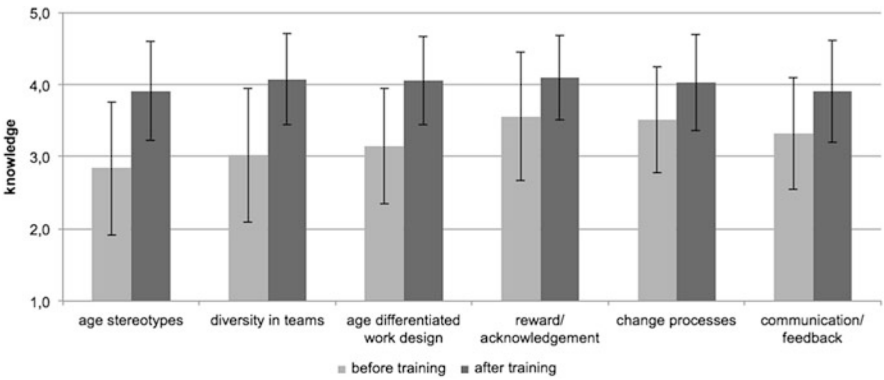


Fig. 4 Training evaluation: knowledge on training relevant topics of participating leaders before (*light*) and after (*dark*) the training (scale: 1 = no knowledge at all to 5 = very good knowledge)

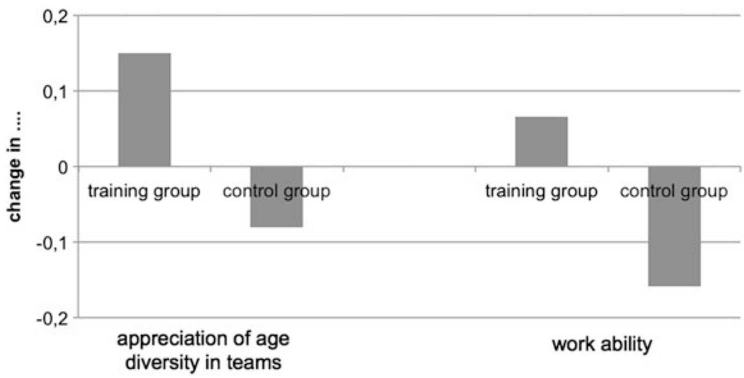


Fig. 5 Training evaluation: effect of training on changes in appreciation of age diversity in teams (*left*) and work ability (*right*) from first to second survey wave among employees of trained or untrained leaders (positive values indicating an increase, negative values indicating a decrease)

training on team performance and health. Team members of trained leaders reported an enhanced exchange of information in the teams (no figure) and better subjective work ability (see Fig. 5, right). These findings indicate that the leaders' attitudes changed during the intervention, which in turn improved the team members' health as well as the way they interact and cooperate with each other.

6 Discussion

The project TED examined the beneficial and hindering conditions of teamwork and leadership in age-diverse teams in the car manufacturing industry with a longitudinal study design. The results confirmed the positive impact of age-differentiated leadership on cooperation within teams (e.g., conflicts) and the subjective health status of team members (e.g., work ability). Thus, promoting age-differentiated leadership seems to be an effective intervention when coping with demographic challenges also in high risk and susceptible industries tasks such as the manufacturing industry [12, 35, 37]. The participants were highly satisfied with the structure and the content of the training and reported a considerable increase of knowledge on the training topics. In addition, the training intervention had a positive impact on leaders' behaviors and attitudes and further on the attitudes and performance of the employees. Overall, the results once again demonstrate the need for an age-differentiated leadership in age-diverse teams. Within this project we developed a scientifically evaluated intervention that helps to raise awareness for age diversity in teams. It provides leaders with a set of actions which help them to successfully deal with the possible problems that occur in an aging workforce.

References

1. Baltes BB, Finkelstein LM (2011) Contemporary empirical advancements in the study of aging in the workplace. J Organ Behav 32(2):151–154
2. Bezrukova K, Jehn KA, Spell CS (2012) Reviewing diversity training: where we have been and where we should go. Acad Manag Learn Educ 11(2):207–227
3. Chesler M, Delgado H (1987) Race relations training and organizational change. Strategies for improving race relations: The Anglo American experience, 182–204
4. Ellwart T, Rack O (2009) Änderung von Einstellungen zur Altersdiversität durch Informationsmanagement. In: Korunka C, Kirchler E, Ulferts U (Hrsg) Entscheidungen und Veränderungen in Arbeit, Organisation und Wirtschaft, Wien, S, 121
5. Esslinger AS, Emmert M, Schöffski O (eds) (2010) Erfolgreiches Betriebliches Gesundheitsmanagement in Organisationen: Grundlagen und Best Practices. Gabler Verlag, Wiesbaden
6. Frieling E, Kotzab D, Diaz JAE, Sytch A (2013) Assembly tasks in the automotive industry: a challenge for older employees. In: Schlick CM, Frieling E, Wegge J (eds) Age-differentiated work systems. Springer, Berlin, pp 201–226
7. Fritzsche L, Wegge J, Schmauder M, Kliegel M, Schmidt KH (2014) Good ergonomics and team diversity reduce absenteeism and errors in car manufacturing. Ergonomics 57(2): 148–161
8. Gregersen S, Kuhnert S, Zimber A, Nienhaus A (2011) Führungsverhalten und Gesundheit – Zum Stand der Forschung. Gesundheitswesen 73:3–12
9. Hacker W, Fritsche B, Richter P, Iwanowa A (1995). Tätigkeitsbewertungssystem TBS. Verfahren zur Analyse, Bewertung und Gestaltung von Arbeitstätigkeiten. In: Ulich E (Hrsg) Mensch-Technik-Organisation (Bd. 7). Zürich: vdf Hochschulverlag
10. Harrison DA, Klein KJ (2007) What's the difference? Diversity constructs as separation, variety, or disparity in organizations. Acad Manage Rev 32(4):1199–1228
11. Hertel G, Thielgen M, Rauschenbach C, Grube A, Stamov-Roßnagel C, Krumm S (2013) Age differences in motivation and stress at work, Age-differentiated work systems. Springer, Berlin, pp 119–147
12. Ilmarinen J, Tempel J (2002) Arbeitsfähigkeit 2010: Was können wir tun, damit Sie gesund bleiben? VSA-Verlag Hamburg, Hamburg
13. Joshi A, Roh H (2009) The role of context in work team diversity research: a meta-analytic review. Acad Manage J 52:599–627

14. Jungmann F, Wegge J, Liebermann SC, Ries BC, Schmidt KH (submitted). Improving team functioning and performance: conceptualization and evaluation of a leadership training on age diverse teams and age differentiated leadership
15. Kerschreiter R, Mojzisch A, Schulz-Hardt S, Brodbeck FC, Frey D (2003) Informationsaustausch bei Entscheidungsfindungsprozessen in Gruppen: Theorie, Empirie und Implikationen für die Praxis. Teamarbeit und Teamentwicklung, 85–118
16. Kirkpatrick DL (1976) Evaluation of training. In: Craig RL (ed) Training and development handbook. McGraw-Hill, New York, NY, pp 301–319
17. Mohren DCL, Jansen NWH, Kant IJ (2010) Need for recovery from work in relation to age: a prospective cohort study. Int Arch Occup Environ Health 83(5):553–561
18. Paluck EL, Green DP (2009) Prejudice reduction: what works? A review and assessment of research and practice. Annu Rev Psychol 60:339–367
19. Pohlandt A, Richter P, Jordan P, Schulze F (1999) Rechnergestütztes Dialogverfahren zur psychologischen Bewertung von Arbeitsinhalten (REBA). In: Dunckel H (Hrsg) Handbuch psychologischer Arbeitsanalyseverfahren. In: Ulich E (Hrsg) Mensch-Technik-Organisation (Bd. 14). Zürich: vdf Hochschulverlag
20. Pytlovany AC, Truxillo DM (2015) Age diversity at work. In: Pachana NA (ed) Encyclopedia of geropsychology. Springer, Singapore. doi:10.1007/978-981-287-080-3_21-1
21. Richter P, Nebel C, Wolf S (2010) Ja, mach nur einen Plan! Gesundheitsinterventionen in turbulenten Zeiten. In: Rigotti T, Korek S, Otto K (eds) Gesund mit und ohne Arbeit. Pabst Science Publishers, Lengerich, pp 73–90
22. Ries BC, Diestel S, Shemla M, Liebermann SC, Jungmann F, Wegge J, Schmidt KH (2013) Age diversity and team effectiveness. In: Schlick CM, Wegge J, Frieling E (eds) Age-differentiated work systems. Springer, Berlin, pp 89–118
23. Roßnagel C, Hertel G (2006) Altersbedingte Unterschiede in Inhalten und im Zustandekommen von Arbeitsmotivation und Arbeitszufriedenheit. Z Arbeitswissenschaft 60 (3):181–186. doi:10.3239/9783640689750
24. Roth C, Wegge J, Schmidt KH (2007) Konsequenzen des demographischen Wandels für das Management von Humanressourcen. Z Personalpsychologie 6:99–116
25. Schlick CM, Frieling E, Wegge J (2013) Age-differentiated work systems. Springer, Berlin
26. Statistisches Bundesamt (2009) 12. koordinierte Bevölkerungsvorausberechnung des Statistischen Bundesamtes. Eigenverlag: Wiesbaden
27. Streb C, Voelpel S, Leibold M (2009) Aging workforce management in the automobile industry: defining the concept and its constituting elements. Zeitschrift für Personalforschung/German Journal of Research in Human Resource Management 23(1):8–27
28. Tajfel H, Turner JC (1986) The social identity theory of intergroup behavior. In: Austin S, Austin WG (eds) Psychology of intergroup relations. Nelson Hall, Chicago, pp 7–24
29. Truxillo DM, Fraccaroli F (2013) Research themes on age and work: introduction to the special issue. Eur J Work Organ Psychol 22(3):249–252
30. Tuomi K, Ilmarinen J, Seitsamo J, Huuhtanen P, Martikainen R, Nygard CH, Klockars M (1997) Summary of the Finnish research project (1981–1992) to promote the health and work ability of aging workers. Scand J Work Environ Health 23:66–71
31. Van Dick R, Van Knippenberg D, Hägele S, Guillaume YR, Brodbeck FC (2008) Group diversity and group identification: The moderating role of diversity beliefs. Human Relat 61(10): 1463–1492
32. van Knippenberg D, de Dreu CKW, Homan AC (2004) Work group diversity and group performance: an integrative model and research agenda. J Appl Psychol 89:1008–1022
33. Warr P (2001) Age and work behaviour: physical attributes, cognitive abilities, knowledge, personality traits and motives. Int Rev Ind Organ Psychol 16:1–36
34. Wegge J, Jungmann F, Liebermann SC, Shemla M, Ries BC, Diestel S, Schmidt KH (2012) What makes age-diverse teams effective? Results from a 6-year research-program. Work 41: 5145–5151

35. Wegge J, Schmidt KH (2015) Diversity Management – Generationenübergreifende Zusammenarbeit fördern. Hogrefe, Göttingen
36. Wegge J, Schmidt KH, Liebermann S, van Knippenberg D (2011) Jung und alt in einem Team? Altersgemischte Teamarbeit erfordert Wertschätzung von Altersdiversität [Young and old in one team? Mixed-age teamwork requires appreciation of age diversity]. In: Gelléri P, Winter C (eds) Potentiale Der Personalpsychologie. Göttingen, Hogrefe, pp 35–46
37. Wegge J, Schmidt KH, Piecha A, Ellwart T, Jungmann F, Liebermann SC (2012) Führung im demografischen Wandel. Report Psychologie 37:344–354
38. Wegge J, Shemla M (2013) Diversity-Management. In: Sarges W (ed) Management Diagnostik, 4th edn. Hogrefe, Göttingen, pp 147–155
39. Wegge J, Shemla M, Haslam SA (2014) Leader behavior as a determinant of health at work: specification and evidence of five key pathways. Ger J Res Hum Resour Manage 28:6–23
40. Wendsche J (2014) Das Pausencheck-Verfahren. TU Dresden, Dresden
41. Zacher H, Degner M, Seevaldt R, Frese M, Luedde J (2009) What do younger and older workers want to accomplish? Age-related differences in content and characteristics of occupational goals. Zeitschrift fur personalpsychologie 8(4):191–200

Beneficial Effects of Servant Leadership on Short- and Long-Term Indicators of Employees' Psychological Health

Wladislaw Rivkin and Klaus-Helmut Schmidt

Abstract

Servant leadership is conceptualized as a leadership style, which encourages employees to behave in a social responsible way. In the present study, we examine the positive relationship between servant leadership and employees' psychological health. We propose that this beneficial relationship results from the potential of servant leaders to shape employees' needs and to create work environments that fulfil these needs. We examine the proposed relationships of servant leadership (a) competing for variance with job ambiguity as a well-known job-stressor, and (b) in relation to long- and short-term indicators of psychological health. In a sample of $N = 443$ employees, we tested the relationships of servant leadership and job ambiguity to ego depletion, need for recovery, emotional exhaustion, and depersonalization as indicators of psychological health. Our results demonstrate that servant leadership is positively related to employees' psychological health and accounts for unique amounts of variance in the examined short- and long-term indicators of psychological health over and above that explained by job ambiguity. Accordingly, servant leadership can be regarded as an important determinant of employees' psychological health.

Keywords

Servant leadership • Employees psychological health • Occupational stressors • Burnout

W. Rivkin (✉) • K.-H. Schmidt
Leibniz Research Centre for Working Environment and Human Factors (IfADo), Dortmund, Germany
e-mail: Rivkin@ifado.de; schmidtkh@ifado.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_5

1 Introduction

Due to corporate scandals (e.g., Enron, WorldCom or Anglo Irish Bank) and the associated unethical behaviors of leaders and employees, public confidence into leader behavior has declined in the recent years [35]. Indeed, some studies have argued that certain leadership styles may be associated with unethical behaviors [6, 29, 32, 41, 54]. Thus, recent research has focused on leadership styles that encourage socially responsible and moral behaviors such as servant leadership, ethical leadership, and authentic leadership [2, 3, 11].

Among other outcomes, previous research has also identified leadership as a determinant of employees' psychological health [33]. For example, in their 2011 report, the Federal Institute for Occupational Safety and Health in Germany reported a steady decrease of work-related psychological health in the past years and predicted that in the upcoming years this trend will result in high absenteeism rates in Germany [12]. Thus, from an organizational perspective, threats to employees' psychological health can be expected to account for major productivity losses in the near future. To prevent these losses, research needs to identify factors that can improve employees' psychological health.

Previous research has provided inconsistent results regarding the relationship between different leadership styles such as transformational leadership, consideration, initiating structure and employees' psychological health [1, 39, 43]. In addition, only a few studies have examined the relationship of leadership styles that have been proposed to promote socially responsible and moral behaviors to employees' psychological health [30]. In the present study, we aim to address this gap in leadership research by examining the relationship between servant leadership and employees' psychological health.

Servant leadership is a leadership style that focuses on serving multiple stakeholders of the organization. Hale and Fields [26] define servant leadership as "an understanding and practice of leadership that places the good of those led over the self-interest of the leader, emphasizing leader behaviors that focus on follower development, and de-emphasizing glorification of the leader" (p. 397). On the basis of this definition, we predict that the proposed beneficial relationships between servant leadership and health should become manifest in negative relationships between servant leadership and short- and long-term indicators of psychological strain, which are thought to reflect overall employees' psychological health. These predictions draw on organizational fit theory [15–17] and social identity theory [48]. Organizational fit theory proposes that psychological health arises from a high fit between employees' needs and organizational provisions and affordances. According to social identity theory, individuals define themselves as members of groups, and thereby internalize social identities that serve to structure both organizational perceptions and behavior. Amongst other things, Haslam et al. [27] argue that shared social identity serves as a basis for feelings of trust, support, and belongingness—feelings that in turn are expected to improve employees' psychological health. In the present paper, we integrate both theories and propose that servant leaders can create a shared social identity among followers and thus fulfill followers' needs. Consequently, we expect servant leadership to be positively related to health because it speaks to followers' needs for a sense of shared social identity and hence provides a high needs-supply fit.

In the present study, we examine the proposed relationships in regard to long and short-term indicators of job-strain (ego depletion, need for recovery, emotional exhaustion and depersonalization) which reflect employees' psychological health. Furthermore, we simultaneously test the relationship of servant leadership and job ambiguity a well-known job stressor, which has been repeatedly found to predict strain. In this way, we may provide evidence that servant leadership accounts for unique amounts of variance in indicators of strain over and above that explained by job stressors such as job ambiguity.

We believe that our research has the capacity to provide several contributions to the literature on leadership and health. First, it may provide initial evidence about the nature of the relationship between servant leadership and various indicators of psychological health. Second, it examines whether servant leadership may explain unique parts of variance in indicators employees' psychological health over and above job ambiguity a major job-stressor.

In the following, we will first elaborate on the construct of servant leadership and distinguish it from other leadership styles. Next, we will focus on the relationship between servant leadership and employees' psychological health, and present a theoretical foundation for the proposed beneficial relationships. Finally, we will present the details of our analyses.

1.1 Servant Leadership

The previously described characteristics of servant leadership such as behaving ethically, helping followers grow and succeed, putting followers first [35] reflect that servant leadership focuses on multiple stakeholders of the organization such as shareholders, the community, customers, and especially followers. The core idea of servant leadership is that managers set aside their personal self-interest for the benefit of collective interests [9, 23]. Thus, servant leaders do not lead for their own benefit or for the benefit of their organizations, but to integrate the interests of multiple stakeholders, and especially their employees. Consequently, servant leaders do not lead through formal authority or charisma as other leadership styles such as transformational leadership [13, 52] suggest, but instead rely on "one-on-one" communication to understand the abilities, needs, desires, goals and the potential of their employees [35, p. 162]. Additionally, servant leaders shape their employees' views and values to encourage them to become servants and servant leaders themselves [25].

Even though servant leadership overlaps with other leadership styles such as transformational, ethical and authentic leadership [5, 11, 51], it also differs from these constructs in certain key aspects. First, Graham [24] argues that leadership styles such as transformational leadership fail to consider the importance of a moral compass, which constitutes a crucial aspect of servant leadership. Second, in contrast to leadership styles that include ethical aspects (e.g., authentic leadership and ethical leadership), servant leadership focuses on the success of multiple stakeholders of the organization. Third, servant leadership is especially focused on the interests, and competencies of followers. Thus, servant leaders aim to develop their employees, and to support their growth and success [46].

Recent research provides strong support for the idea that servant leadership exerts unique beneficial effects on various job attitudes, fairness perceptions, and also job performance. For example, Ehrhart [22] demonstrated that servant leadership accounts for additional variance in commitment (5 %), job satisfaction (7 %), perceived supervisor support (4 %) and procedural justice (8 %) over and above leader member exchange (LMX) and transformational leadership. These results are also supported by further research that provided evidence for beneficial effects of servant leadership on employee work outcomes over and above other leadership styles (e.g., transformational leadership, and LMX) [21, 35, 40, 44]. In conclusion, servant leadership is characterized by unique behavioral patterns and attitudinal aspects, which are distinct from other related leadership concepts and thus account for a broad spectrum of positive outcomes, even after controlling for other aspects of leadership. Yet despite its various beneficial outcomes, to the best of our knowledge, no previous studies have examined the relationship between servant leadership and employees' psychological health.

1.2 Servant Leadership as a Predictor of Employees' Psychological Health

As previously described the relationship between servant leadership and employees' psychological health is grounded in principles of organizational fit theory [15–17] and social identity theory [48]. On the dyadic level, we argue that servant leaders create work environments that have affordances that fulfill employees' individual needs. According to organizational fit theory, a high fit between a person and its environment reduces strain while a low fit is expected to increase strain. In this regard, previous fit research has identified two types of person-environment fit. The first type is commonly referred to as demands-abilities fit. It describes whether situational demands can be met by a person's abilities. The second type is referred to as needs-supply fit and relates to the match between a person's needs and provisions, and affordances in a given environment. We expect that the beneficial relationship between servant leadership and employees' health results primarily from a high degree of needs-supply fit.

At the same time, on a group level, we argue that a shared social identity influences employees' psychological health by creating an atmosphere of trust, support, justice, and belongingness [27]. Social identity theory suggests that in various contexts individuals define themselves as members of social groups (e.g., as 'us' family, friends, and colleagues). These social groups provide individuals with personal security, and emotional bonding. At the same time though, individuals tend to experience negative psychological consequences if they lack or lose social identity (e.g., if they are rejected from groups) and positive consequences if they maintain or gain a sense of shared social identity (e.g., if they identify with groups [15] see also Cruwys et al. [20]). The core aspects of servant leadership speak to issues of social identity. For example, by promoting ethical behaviors among their followers, servant leaders establish norms that are embraced by all members of their group, and followers' enactment of these norms helps to establish a shared sense of positive social identity. At the same time, followers' sense that leaders are 'doing it for us' (rather than for themselves) should help to cultivate both followership and psychological health [28].

Integrating both theories, we argue that through its focus on helping followers grow and achieve agency by acting in a manner consistent with commonly agreed views and values, servant leaders establish a shared social identity among group members. Thereby, servant leaders shape their followers' needs and help to create work-environments that fulfill these needs. For example, a servant leader may emphasize the importance of giving back to the community and induce the need to be involved in community service or volunteer activities. Then again, the servant leader may fulfill this need by introducing community service activities at corporate events. In short, we argue that the beneficial relationship between servant leadership and employees' psychological health reflects processes occurring at both group and dyadic levels. At the group level, it arises from feelings of trust, support, and belongingness that result from the leader's cultivation of a sense of shared social identity; at the dyadic level, it results from the leader's enhancement of needs-supply fit.

1.3 The Present Research

We conducted a study to provide empirical evidence for the proposed beneficial relationships between servant leadership and employee's psychological health operationalized as indicators of strain. To demonstrate that servant leadership shares unique proportions of variance with the examined indicators of strain, we simultaneously tested servant leadership in combination with job ambiguity a well-known job stressor. Job ambiguity involves a perceived lack of job-related information and reflects employees' perceptions of uncertainty concerning various aspects of their jobs [10]. The negative impact of job ambiguity on, for example, burnout has been observed in multiple studies. Schwab and Iwanicki [45] reported that among teachers, role conflict and job ambiguity account for considerable proportions of variance in emotional exhaustion (23 %) and depersonalization (20 %). Furthermore, in their meta-analysis, Lee and Ashforth [34] found moderate to strong correlations of job ambiguity to both burnout symptoms.

Furthermore, we test the relationship of servant leadership with different indicators of strain. We examine emotional exhaustion and depersonalization as long-term indicators of strain. Emotional exhaustion is considered to be the main component of burnout and is defined as a chronic state of depletion and fatigue resulting from one's work [36, 53]. Depersonalization, another dimension of burnout, refers to negative and cynical attitudes towards people at work [15]. As short-term indicators of strain, we examine ego depletion and need for recovery. Ego depletion refers to a momentary state of regulatory resource depletion following regulatory demands [7]. According to resource-based conceptualizations of psychological strain [38], ego depletion can be thought as a short-term correlate of impairments in psychological health. Need for recovery reflects the need to recuperate from work tasks that is strongest in the last hours of work and directly after work [50].

In sum, we test the impact of servant leadership on employee health (a) on both long- and short-term indicators of strain and (b) competing for variance with a well-known stressor. We argue that the proposed approach should allow us to test whether the relationship between servant leadership and employees' health is invariant to different boundary conditions such as individual indicators of strain, and whether servant leadership explains unique amounts of variance in employees' psychological health. Thus, our hypothesis is:

Servant leadership is beneficially related to psychological health over and above job ambiguity. In view of the examined indicators of strain servant leadership is negatively related to (a) ego depletion, (b) need for recovery, (c) emotional exhaustion and (d) depersonalization.

2 Method

2.1 Participants

The sample for our study was recruited from a major bank in Germany. In their work environment, job ambiguity constitutes one of the main stressors because employees have to balance customers' needs with organizational priorities [18]. For example, even though employees need to sell financial products to customers, they also have to consider customers' interests. Thus, we decided to control for job ambiguity as a stressor in this study.

All participants were contacted via e-mail and received an online survey which was completed during regular working hours. Participation was voluntary, and all participants were assured that their responses would remain confidential. Out of 705 contacted persons, we received 443 responses (63 % response rate). Participants' age ranged from 18 to 60 years ($M = 39.22$; $SD = 10.68$). Of these participants, 56 % were female, and 23 % worked part-time.

2.2 Measures

Job ambiguity was assessed with nine items from a scale developed by Breaugh and Colihan [10], which was translated and validated in German by Sodenkamp and Schmidt [47]. Here participants indicated their perceived lack of job-related information on a 7-point Likert-scale (1 = not at all, 7 = a great deal). Sample items are

"I know how to get my work done (what procedures to use)" (work method ambiguity) and "I know when I should be doing a particular aspect (part) of my job" (scheduling ambiguity). All items were recoded so that higher scores reflect greater job ambiguity and were then averaged to a single scale score.

We used Ehrhart's [22] scale to measure servant leadership. On the basis of a literature review, Ehrhart [22] identified seven major categories of servant leader behaviors (forming relationships with followers, empowering followers, helping followers grow and succeed, behaving ethically, having conceptual skills, putting followers first, and creating value for those outside of the organization). Afterwards, he developed two items for each category resulting in a 14-item measure of servant leadership. Factor analyses revealed that this measure had a one-dimensional structure [22, 49]. For the present study, the items were translated into German through a three-step procedure. This involved the original items being translated into German, then back into English, and then compared. In our study, participants rated the behavior of their leader on these items using 5-point Likert-scales (1 = not at all, 5 = a great deal). Table 1 gives an overview of the original items and the German translations.

The measurement of ego-depletion was based on five items that assessed the current experience of resource depletion and low will-power (e.g., "At the moment, I feel increasingly less able to focus on something." or "At the moment, I feel as if I have no willpower left"). The scale was originally developed and validated by Bertrams et al. [8] and all items were scored on 5-point Likert-scales (1 = not at all, 5 = a great deal). Day-specific need for recovery was assessed with five items (e.g., "After the present day's work I feel so tired that I cannot get involved in other activities." or "My job causes me to feel rather exhausted." [50]). All items were scored on the same 5-point Likert-scales.

The burnout dimensions of emotional exhaustion (eight items) and depersonalization (five items) were assessed with a German version [14] of the Maslach and Jackson [36] Burnout Inventory. Sample items are "I feel emotionally drained from my work" (emotional exhaustion) and "I have become more callous toward people since I took this job" (depersonalization). The items were rated on a 6-point Likert-scale (1 = not at all, 6 = very strong).

2.3 Factor Structure

Before testing our hypotheses, we examined whether the factor structure of our translated measurement of servant leadership resembles the factor structure of the original measurement. Thus, we conducted an exploratory factor analysis (EFA). Results of the EFA (principal component analysis with orthogonal varimax rotation) provided a two-factor structure of servant leadership. The amount of variance explained by both factors was 62.7 %. The first factor includes the first 12 Items (cf. Table 1). These items reflect the prioritization of subordinates concerns. The first factor accounts for the largest proportion of variance (53.7 %). The factor loadings range from .83 to .51. Items 13 and 14 represent the second factor. This

Table 1 Servant leadership: Original items and German translations

	Original items	German translations
1	My department manager spends the time to form quality relationships with department employees	Mein unmittelbarer Vorgesetzter investiert viel Zeit, um gute Beziehungen zu den Mitarbeitern aufzubauen
2	My department manager creates a sense of community among department employees	Mein unmittelbarer Vorgesetzter erzeugt ein Zusammengehörigkeitsgefühl unter den Mitarbeitern
3	My department manager's decisions are influenced by department employees' input	Mein unmittelbarer Vorgesetzter lässt sich in seinen Entscheidungen von den Ansichten der Mitarbeiter beeinflussen
4	My department manager tries to reach consensus among department employees on important decisions	Mein unmittelbarer Vorgesetzter versucht bei wichtigen Entscheidungen, einen Konsens unter den Mitarbeitern herzustellen
5	My department manager is sensitive to department employees' responsibilities outside the work place	Mein unmittelbarer Vorgesetzter nimmt auf die außerberufliche Lebenssituation der Mitarbeiter Rücksicht
6	My department manager makes the personal development of department employees a priority	Für meinen unmittelbaren Vorgesetzten ist die persönliche Weiterentwicklung der Mitarbeiter ein vorrangiges Ziel
7	My department manager holds department employees to high ethical standards	Mein unmittelbarer Vorgesetzter hält die Mitarbeiter zur Einhaltung hoher moralischer Standards an
8	My department manager does what she or he promises to do	Mein unmittelbarer Vorgesetzter hält, was er verspricht
9	My department manager balances concern for day-to-day details with projections for the future	Mein unmittelbarer Vorgesetzter verknüpft Alltagsangelegenheiten mit langfristigen Plänen für die Zukunft
10	My department manager displays wide-ranging knowledge and interests in finding solutions to work problems	Mein unmittelbarer Vorgesetzter verfügt über weitreichende Kenntnisse bei der Bewältigung von Arbeitsproblemen
11	My department manager makes me feel like I work with him/her, not for him/her	Mein unmittelbarer Vorgesetzter gibt mir das Gefühl, dass ich mit ihm und nicht für ihn arbeite
12	My department manager works hard at finding ways to help others be the best they can be	Mein unmittelbarer Vorgesetzter arbeitet hart daran, andere dabei zu unterstützen, ihr Bestes zu geben
13	My department manager encourages department employees to be involved in community service and volunteer activities outside of work	Mein unmittelbarer Vorgesetzter ermutigt die Mitarbeiter, sich an gemeinnützigen und ehrenamtlichen Aktivitäten außerhalb der Arbeit zu beteiligen
14	My department manager emphasizes the importance of giving back to the community	Mein unmittelbarer Vorgesetzter betont die Notwendigkeit, für das gesellschaftliche Wohl einen Beitrag zu leisten

factor reflects encouragement of ethical and prosocial behavior. It accounts for 8.9 % of variance. The factor loadings range from .73 to .67. Because Costello and Osborne [19] argue that factors with less than three items tend to be unstable, we conducted another EFA and specified the extraction of only one factor as suggested by the original scale [22]. This factor accounts for 53.7 % of the total variance. The factor loadings ranged between .83 and .51. Because the one-factor model resembled the original scale and the factor loadings did not fall below the critical value of .30 [19], we decided to use a composite measure of servant leadership. Thus, as suggested by Ehrhart [22], this involved averaging responses to all 14 servant leadership items (cf. Table 1).

2.4 Data Analysis

Before testing the hypotheses, we conducted Harman's one-factor test [42] in order to analyse potential confounding effects due to common method variance. The results of this test suggested that a common method factor accounted for 31.3 % of variance. On this basis, we infer that our results are not seriously biased by high common method variance. Subsequently, we analysed our data using three-step hierarchical linear regression analyses with ego depletion, need for recovery, emotional exhaustion and depersonalization as outcomes. In Step 1, we entered the control variables of age, gender, and working time into the regression to control for their potential confounding influences on the relationships under examination [31, 37]. In Step 2, we introduced job ambiguity. In Step 3, servant leadership was added into the regressions.

3 Results

Table 2 presents the means, standard deviations, and reliabilities of study variables.

Our hypothesis proposed that servant leadership is negatively related to ego depletion, need for recovery emotional exhaustion and depersonalization over and above job ambiguity. Results of multiple regression analyses relating to this hypothesis are presented in Table 3. These indicate that, after controlling for demographic variables, job ambiguity is positively related to ego depletion, need for recovery (cf. Table 3), emotional exhaustion and depersonalization (cf. Table 4). Moreover and theoretically more important, servant leadership was negatively related to all four outcomes (ego depletion [$\beta = -.15$; $p < .01$], need for recovery [$\beta = -.18$; $p < .01$], emotional exhaustion [$\beta = -.25$; $p < .01$] and depersonalization [$\beta = -.25$; $p < .01$]). The incremental variance explained by servant leadership was respectively 2 % for ego depletion, 2 % for need for recovery 6 % for emotional exhaustion and 5 % for depersonalization (cf. Table 3). These results thus support our hypothesis.

Table 2 Means, standard deviations, internal consistencies (Cronbach's alpha) and intercorrelations

Variable		1	2	3	4	5	6	7	8	9
1.	Age	–								
2.	Gender[a]	**.11**	–							
3.	Working time[b]	**−.19**	**.42**	–						
4.	Job ambiguity	**−.15**	.01	.04	*(.90)*					
5.	Servant leadership	−.08	.01	.05	**−.37**	*(.93)*				
6.	Ego depletion	**.11**	−.01	.06	**.34**	**−.29**	*(.89)*			
7.	Need for recovery	.08	−.02	**.16**	**.21**	**−.25**	**.67**	*(.87)*		
8.	Emotional exhaustion	.05	.04	**.12**	**.37**	**−.37**	**.76**	**.71**	*(.89)*	
9.	Depersonalization	−.07	**.15**	**.12**	**.37**	**−.36**	**.54**	**.48**	**.66**	*(.79)*
	M	39.22	1.44	1.78	2.80	3.37	1.77	2.14	2.82	2.04
	SD	10.68	0.50	0.42	0.95	0.76	0.56	0.69	0.98	0.81

Note. [a]Gender (1 = female, 2 = male), [b]Working time (1 = part-time, 2 = full-time). Internal consistency estimates (Cronbach's alpha) are in parentheses in the diagonal. Numbers in bold $p < .05$. $N = 443$

Table 3 Regression results (β values) for ego depletion and need for recovery

	Ego depletion			Need for recovery		
Variable	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
Age	.14**	.20**	.18**	.14**	.17**	.15**
Gender	−.07	−.08	−.08	−.14**	−.14**	−.14**
Working time	.12*	.12*	.12*	.24**	.24**	.24**
Job ambiguity		.37**	.31**		.23**	.17**
Servant leadership			−.15**			−.18**
$R^2(\Delta R^2)$	.02(.02)	.15(.13)	.17(.02)	.04(.04)	.10(.06)	.12(.02)
F for change in R^2	3.54*	67.98**	10.28**	7.73**	26.00**	13.18**

Note. $^{*}p < .05$. $^{**}p < .01$. $N = 443$

Table 4 Regression results (β values) for emotional exhaustion and depersonalization

	Emotional exhaustion			Depersonalization		
Variable	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
Age	.08	.14**	.11*	−.07	−.01	−.05
Gender	−.03	−.04	−.04	.14*	.13*	.13**
Working time	.15**	.15**	.15**	.05	.05	.05
Job ambiguity		.38**	.29**		.37**	.27**
Servant leadership			−.25**			−.25**
$R^2(\Delta R^2)$	.01(.01)	.15(.14)	.21(.06)	.02(.02)	.16(.14)	.21(.05)
F for change in R^2	2.93*	74.44**	30.41**	4.67**	69.14**	30.32**

Note. $^{*}p < .05$. $^{**}p < .01$. $N = 443$

4 Discussion

The aim of the present research was to provide evidence for the hypothesized beneficial relationships between servant leadership and employees' psychological health. We demonstrated that servant leadership accounts for additional variance in short- and long-term indicators of strain (ego depletion, need for recovery, emotional exhaustion and depersonalization) over and above job ambiguity, a well-known job stressor. These results support our claims that servant leadership is (a) beneficially related to short- and long-term indicators of psychological health and (b) that it accounts for unique variance in these indicators of psychological health over and above job-specific stressors.

4.1 Theoretical Implications

Our study provides several important contributions to research on servant leadership. First, we know of no previous research that has examined the relationship between servant leadership and employees' psychological health. Integrating our results and previous research on servant leadership demonstrates that servant leadership does not only improve outcomes such as job attitudes and job performance [49], but also that these improvements may also have benefits for employees' health. This is especially important because research on other leadership styles such as transformational leadership indicates that these leadership styles are primarily beneficial for the organization or the leader regardless of potential negative consequences for employees [37].

Second, our findings demonstrate that servant leadership accounts for variance in indicators of employees' psychological health over and above job ambiguity as a prominent job-stressor. These results indicate that the effects of servant leadership are unique.

4.2 Limitations and Suggestions for Further Research

It remains the case, however, that our study also has several limitations that need to be discussed. First, our study variables were all operationalized by means of self-report. Thus, common method variance or a self-report bias might have contaminated the observed relationships [42]. However, Harman's one-factor test indicated that common method influences were not a severe biasing factor in our data.

Second, although we assumed a particular causal order of the variables, the correlational design of our research does not permit causal conclusions. Other causal directions or even reciprocal relations could be possible as well. For example, psychologically healthy employees might have made the choice to work for servant leaders. Additionally, as we suggested in our theoretical argument, shared social identity might have been a third factor that influenced both the perception of servant leadership on the part of employees as well as indicators of employees' psychological health. While we cannot rule out this possibility, research on alternative leadership styles and health has also demonstrated the health enhancing effects of leadership [39]. Additionally, given that we controlled for the effects of job ambiguity (as a strong predictor of psychological health) on the relationship between servant leadership and psychological health, alternative causal paths seem rather unlikely. However, further research should focus on disentangling the relationship between servant leadership and psychological health. In particular, it is necessary to examine more closely the mechanisms (e.g., shared social identity, needs-supply fit [16, 27]) that may account for the beneficial relationship between servant leadership and employee health.

4.3 Practical Implications

The results of our study indicate that servant leadership is beneficially related to employees' psychological health. Thus, it appears that organizations that seek to improve the psychological health of employees should consider encouraging their leaders to lead on the basis of the principles of servant leadership. This can be achieved through leadership training and through the role modeling of servant leadership by current leaders.

More specifically, in the first instance, practitioners might design leadership training programs that elaborate on the basic principles of servant leadership such as forming relationships with followers, empowering followers, helping followers grow and succeed, and behaving ethically. Part of such training involves discussing different ethical perspectives and reflecting on how a servant leader might engage constructively with these perspectives. Another useful exercise may be to simulate and record annual appraisals of employees and evaluate these appraisals according to principles of servant leadership (e.g., appreciation of employees). Additionally, in this training program, trainees may also learn about the benefits of servant leadership over and above other leadership styles.

Second, research on servant leadership argues that this leadership style can be best taught by setting examples [4]. Thus, being led by a servant leader is expected to increase the likelihood of a follower becoming a servant leader. Accordingly, leaders should provide examples and lead in accordance with the principles of servant leadership with a view to encourage employees to follow in their footsteps. Indeed, in so far as the present evidence suggests, broad acceptance of servant leadership will have positive consequences for employees' psychological health. Thus, it appears that there is much to be gained through the transmission of this approach from one generation of leaders to the next.

References

1. Arnold KA, Turner N, Barling J, Kelloway EK, McKee MC (2007) Transformational leadership and psychological well-being: The mediating role of meaningful work. J Occup Health Psychol 12(3):193–203
2. Avolio BJ, Gardner WL (2005) Authentic leadership development: Getting to the root of positive forms of leadership. Leadersh Q 16(3):315–338
3. Avolio BJ, Walumbwa FO, Weber TJ (2009) Leadership: Current theories, research, and future directions. Annu Rev Psychol 60:421–449. doi:10.1146/annurev.psych.60.110707.163621
4. Barling J, Christie A, Turner N (2008) Pseudo-transformational leadership: Towards the development and test of a model. J Bus Ethics 81(4):851–861. doi:10.1007/s10551-007-9552-8
5. Bass BM, Avolio BJ (1990) Transformational leadership development: Manual for the multi-factor leadership questionnaire. Consulting Psychologists Press, Palo Alto, CA
6. Bass BM, Steidlmeier P (1999) Ethics, character, and authentic transformational leadership behavior. Leadersh Q 10(2):181–217. doi:10.1016/S1048-9843(99)00016-8
7. Baumeister RF, Bratslavsky E, Muraven M, Tice DM (1998) Ego depletion: Is the active self a limited resource? J Pers Soc Psychol 74(5):1252–1265. doi:10.1037/0022-3514.74.5.1252
8. Bertrams A, Unger A, Dickhäuser O (2011) Momentan verfügbare Selbstkontrollkraft – Vorstellung eines Messinstruments und erste Befunde aus pädagogisch-psychologischen Kontexten. [Momentarily Available Self-Control Strength – Introduction of a Measure and First Findings from Educational-Psychological Contexts]. Zeitschrift für Pädagogische Psychologie 25:185–196

9. Boyatzis RE, McKee A (2005) Resonant leadership: Renewing yourself and connecting with others through mindfulness, hope, and compassion. Harvard Business School Press, Boston, MA
10. Breaugh JA, Colihan JP (1994) Measuring facets of job ambiguity: Construct validity evidence. J Appl Psychol 79(2):191–202. doi:10.1037/0021-9010.79.2.191
11. Brown ME, Treviño LK (2006) Ethical leadership: A review and future directions. Leadersh Q 17(6):595–616
12. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) und BMAS (2012) Stressreport Deutschland 2012 [Stressreport Germany 2012]. Unfallverhütungsbericht Arbeit. Dortmund/Berlin/Dresden
13. Burns JM (1978) Leadership. Harper & Row, New York, NY
14. Büssing A, Perrar K-M (1992) Die Messung von Burnout. Untersuchung einer deutschen Fassung des Maslach Burnout Inventory (MBI-D). [Measuring burnout: A study of a German version of the Maslach Burnout Inventory (MBI-D).]. Diagnostica 38:328–353
15. Caplan RD (1983) Person-environment fit: Past, present, and future. In: Cooper CL (ed) Stress research. Wiley, New York, NY, pp 35–78
16. Caplan RD (1987) Person-environment fit in organizations: Theories, facts, and values. In: Riley AW, Zaccaro SJ (eds) Occupational stress and organizational effectiveness. Praeger, New York, NY, pp 103–140
17. Caplan RD (1987) Person-environment fit theory and organizations: Commensurate dimensions, time perspectives, and mechanisms. J Vocat Behav 31:248–267
18. Chebat J-C, Kollias P (2000) The impact of empowerment on customer contact employees' roles in service organizations. J Serv Res 3(1):66–81. doi:10.1177/109467050031005
19. Costello AB, Osborne JW (2005) Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Pract Assess Res Eval 10(7)
20. Cruwys T, Haslam SA, Dingle GA, Haslam C, Jetten J (2014) Depression and social identity: An integrative review. Pers Soc Psychol Rev 18(3):215–238
21. Dinh JE, Lord RG, Gardner W, Meuser JD, Liden RC, Hu J (2014) Leadership theory and research in the new millennium: Current theoretical trends and changing perspectives. Leadersh Q 25(1):36–62
22. Ehrhart MG (2004) Leadership and procedural justice climate as antecedents of unit-level organizational citizenship behavior. Pers Psychol 57(1):61–94
23. George B (2003) Authentic leadership: Rediscovering the secrets to creating lasting value, 1st edn, Warren Bennis signature series. Jossey-Bass, San Francisco
24. Graham JW (1991) Servant-leadership in organizations: Inspirational and moral. Leadersh Q 2 (2):105–119. doi:10.1016/1048-9843(91)90025-W
25. Greenleaf RK (1977) Servant leadership: A journey into the nature of legitimate power and greatness. Paulist Press, New York, NY
26. Hale JR, Fields DL (2007) Exploring servant leadership across cultures: A study of followers in Ghana and the USA. Leadership 3(4):397–417
27. Haslam SA, Jetten J, Postmes T, Haslam C (2009) Social identity, health and well-being: An emerging agenda for applied psychology. Appl Psychol 58(1):1–23
28. Haslam SA, Reicher SD, Platow MJ (2013) The new psychology of leadership: Identity, influence and power. Psychology Press, Hove
29. Howell JM, Shamir B (2005) The role of followers in the charismatic leadership process: Relationships and their consequences. Acad Manage Rev 30(1):96–112. doi:10.5465/AMR.2005.15281435
30. Ilies R, Morgeson FP, Nahrgang JD (2005) Authentic leadership and eudaemonic well-being: Understanding leader–follower outcomes. Leadersh Q 16(3):373–394
31. Indik B, Seashore SE, Slesinger J (1964) Demographic correlates of psychological strain. J Abnorm Soc Psychol 69(1):26–38. doi:10.1037/h0040300
32. Kanungo RN, Mendonca M (1996) Ethical dimensions of leadership, Sage series in business ethics. Sage, Thousand Oaks, CA

33. Kuoppala J, Lamminpää A, Liira J, Vainio H (2008) Leadership, job well-being, and health effects - A systematic review and a meta-analysis. J Occup Environ Med 50(8):904–915
34. Lee RT, Ashforth BE (1996) A meta-analytic examination of the correlates of the three dimensions of job burnout. J Appl Psychol 81(2):123–1333
35. Liden RC, Wayne SJ, Zhao H, Henderson D (2008) Servant leadership: Development of a multidimensional measure and multi-level assessment. Leadersh Q 19(2):161–177. doi:10.1016/j.leaqua.2008.01.006
36. Maslach C, Jackson SE (1986) Maslach Burnout Inventory, 2nd edn. Consulting Psychologists Press, Palo Alto, CA
37. Moen P, Dempster-McClain DI (1987) Employed parents: Role strain, work time, and preferences for working less. J Marriage Fam 49(3):579. doi:10.2307/352203
38. Muraven M, Baumeister RF (2000) Self-regulation and depletion of limited resources: Does self-control resemble a muscle? Psychol Bull 126(2):247
39. Nielsen K, Randall R, Yarker J, Brenner S-O (2008) The effects of transformational leadership on followers' perceived work characteristics and psychological well-being: A longitudinal study. Work Stress 22(1):16–32. doi:10.1080/02678370801979430
40. Peterson SJ, Galvin BM, Lange D (2012) CEO servant leadership: Exploring executive characteristics and firm performance. Pers Psychol 65(3):565–596. doi:10.1111/j.1744-6570.2012.01253.x
41. Peus C, Kerschreiter R, Frey D, Traut-Mattausch E (2010) What is the value? Zeitschrift für Psychologie/Journal of Psychology 218(4):198–212. doi:10.1027/0044-3409/a000030
42. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP (2003) Common method biases in behavioral research: A critical review of the literature and recommended remedies. J Appl Psychol 88(5):879–903. doi:10.1037/0021-9010.88.5.879
43. Rowold J, Heinitz K (2008) Führungsstile als Stressbarrieren [Leadership Styles as Stress Barriers]. Zeitschrift für Personalpsychologie 7(3):129–140
44. Schneider SK, George WM (2011) Servant leadership versus transformational leadership in voluntary service organizations. Leadersh Org Dev J 32(1):60–77. doi:10.1108/01437731111099283
45. Schwab RL, Iwanicki EF (1982) Perceived role conflict, role ambiguity, and teacher burnout. Educ Adm Q 18(1):60–74
46. Smith BN, Montagno RV, Kuzmenko TN (2004) Transformational and servant leadership: Content and contextual comparisons. J Leadersh Org Stud 10(4):80–91
47. Sodenkamp D, Schmidt K-H (2000) Weiterentwicklung und Konstrukt-Validierung eines Verfahrens zur Messung erlebter Rollenambiguität bei der Arbeit. [Development and Construct Validation of a Method to Measure Role Ambiguity at Work]. Zeitschrift für Arbeitswissenschaft 54:37–43
48. Tajfel H, Turner JC (1979) An integrative theory of intergroup conflict. The social psychology of intergroup relations, 33–47
49. van Dierendonck D (2011) Servant leadership: A review and synthesis. J Manag 37(4):1228–1261. doi:10.1177/0149206310380462
50. van Veldhoven M (2003) Measurement quality and validity of the "need for recovery scale". Occup Environ Med 60(Suppl I):i3–i9. doi:10.1136/oem.60.suppl_1.i3
51. Walumbwa FO, Avolio BJ, Gardner WL, Wernsing TS, Peterson SJ (2008) Authentic leadership: Development and validation of a theory-based measure? J Manag 34(1):89–126
52. Weber M (1921/1946) The sociology of charismatic authority. In: Gerth HH, Milles CW (trans., eds), Max Weber: Essays in sociology. Oxford University Press, New York, NY, pp 245–252
53. Wright TA, Cropanzano R (1998) Emotional exhaustion as a predictor of job performance and voluntary turnover. J Appl Psychol 83(3):486–493. doi:10.1037/0021-9010.83.3.486
54. Yukl GA (2013) Leadership in organizations, 8th edn. Pearson, Boston, MA

Time Management Requirements for Holistic Working Time Organisation

Patricia Stock

Abstract

Designing holistic working time systems is a highly complex planning issue that places high demands on operational planners. Comprehensive information is required to align the personnel inventory and personnel requirement to the best possible standard. The previously common method of entering information in a table in which the number of employees required and available is specified for each time slice is often no longer sufficient because this is simply a static review that cannot adequately take into account employees' individual working time restrictions and wishes. This means that, in future, the previously usual time management methods—such as time recording or systems with predefined times—may no longer provide sufficiently precise data for working time organisation, and new methods have to be developed to determine and align the personnel requirement and personnel inventory.

Keywords

Working time • Personnel requirement • Personnel inventory • Time management

1 Need for Employee-Orientated Working Time Organisation

The world of work is currently undergoing a fundamental change initiated by various megatrends. These megatrends can be divided into economic developments (e.g. globalisation, evolution from an industrial to a knowledge-based society, progress in information and communication technology), demographic developments

P. Stock (✉)
REFA-Institut e.V., Dortmund, Germany
e-mail: patricia.stock@refa-institut.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_6

(e.g. ageing and declining population) and social developments (e.g. individualisation, value shift, feminisation) [21, p. 13]. The resulting changes mean that companies must fulfil an increased need for flexibility [3, p. 44] to continue their quick response to increasingly individualised customer wishes. Measures to exploit flexibility potential can be attributed to working time organisation, work organisation, operating means usage or a combination thereof [see 2, p. 72]. For example, a call centre's operating hours can be extended through innovative working time models with which the working times of (local) employees can be flexibly adapted to best meet the order-related personnel requirement. Alternatively, operating times can also be realised through globalised work organisation in which the use of modern information and communication technology distributes service times across sites in different time zones, where employees have fixed working hours and the information and communication technology used provides constant access to the latest processing status. The instruments that can be used vary considerably, which is why there will be no blanket recommendations on how they are to be used.

Demographic change and its consequences have been the topic of political, scientific and economic discussion for many years now. On 25 April 2012, the German government approved its demographic strategy, which, alongside federal measures, contains proposals to be realised in collaboration with states, local authorities, associations, social partners and other stakeholders in civil society (for more detailed information, see BMBF [4] [German Federal Ministry of Education and Research], BMFSFJ [5] [German Federal Ministry for Family Affairs, Senior Citizens, Women and Youth]). As a result of this demographic change, there is growing evidence of a skills shortage that is forcing companies to actively seek out new employees and retain them through attractive working conditions, while also giving employees a new sense of confidence.

Social and individual perceptions of work have also changed significantly in the past few years. In the twenty-first century working world, the 'old deal' of simply exchanging working time for a salary therefore no longer functions unconditionally. Employees also want the time they spend at work to be meaningful, fulfilling and stimulating. A person's job should not compete with his or her private life but, where possible, harmonise with it [30, p. 56]. While the rule of the industrial society was 'People Follow Jobs', today's slogan is 'Jobs Follow People': Highly qualified career starters and young professionals in particular are ever less willing to adapt to external conditions imposed by companies and employers that do not meet their expectations, wishes and needs. In addition to professional self-fulfilment and income security, the focus is shifting to the work/life balance, the compatibility of work and family [30, p. 56]. In a representative FORSA study (Institute for Social Research and Statistical Analysis in Germany) of employees and civil servants who hold a degree from a university or technical college and work as skilled employees or managers, 88 % of those surveyed stated that jobs should "be suited to them and their own needs", and 82 % stated "that they would like more time for their family, partner or friends in their current job" [29]. All in all, skilled workers and managers in Germany value flexible working hours: 64 % of participants could not even imagine joining a company that did not offer this option

[29]. In light of this, companies must always consider their employees' wishes and needs.

In contrast, the consequences of the progress made in information and communication technology (also known worldwide as the "fourth industrial revolution" and in Germany as "Industrie 4.0") have only recently become the subject of discussion. In future, evolving information and communication technology will inform the way that people live and work through virtual networking and flexible forms of cooperation and organisation, as well as open up new business potential for companies [21, p. 16]. This will fundamentally change the way in which companies are organised, for example through the optimisation of internal and intercompany processes via cyber-physical systems or the support and control of business processes via knowledge-based systems [21, pp. 16]. This will also open up new potential for work structuring and organisation as well as for working time organisation by primarily simplifying or simply providing access to intelligent tools, instruments and automation, production and networking technologies as well as to globally distributed information, knowledge, skills, resources, work partners and markets [19, p. 2]. For companies, this presents considerable opportunities and potential for the flexible structuring of work and value creation processes [19, p. 5], which they can use to secure a decisive competitive advantage. Likewise, as a result, employees are also presented with an increased blurring of the lines between working time and free time with all its associated opportunities (e.g. greater degree of freedom to structure one's life as desired) and risks (e.g. pressure to be constantly available) [19, p. 5].

The phenomenon of constant availability is causally related to the development of modern information and communication technologies. Availability can take the form of personal contact, telephone calls, e-mails, text messages (SMS) or newer forms of communication such as instant messaging services (Skype, MSN, ICQ), chat functions or social networks (Facebook, MySpace, Google etc.) [24, p. 8]. At the same time, the activities of users of new communication methods can also be tracked, for example their online activities on social networks or real-world activities when they post photos or location data. A meta study by the Initiative Gesundheit und Arbeit (Health and Work Initiative, iga) analysed various studies on the topic of constant availability [24]. The studies analysed showed that many employees can be contacted regarding professional matters outside their working hours. However, the number of people actually affected fluctuates significantly—for example, according to the DGB (Confederation of German Trade Unions) "Gute Arbeit" (Good Work) 2011 index, 27 % of those surveyed expected that they would be required to be available for work during their free time as well, while a 2010 study by the federal association of company health insurance funds (BKK) stated that 84 % of employees surveyed were available for professional matters outside their regular working hours [24, p. 8]. The meta analysis therefore concluded that there has not yet been a comprehensive and representative study on constant availability and its health consequences [24, p. 8]. A qualitative expert survey subsequently carried out by iga identified the negative effects (e.g. shorter relaxation times, poor planning capability, loss of structure and control) and positive effects (e.g. more flexibility and mobility, identity and status gains) of constant availability and derived recommendations for dealing well with constant availability [24, pp. 19]. On the other hand, a survey performed as part of the fourth wave of the iga (Initiative for Health and Work in Germany) barometer 2013 showed that almost half of participants had the opportunity to either deal with or be available for private matters during working hours [13, p. 11].

Digitalisation therefore presents the opportunity to satisfy the flexibility needs of companies, customers and employees because work no longer has to be performed in the traditional manner—at a workstation during normal working hours—but can potentially be liberated from a specific space and time. This therefore allows customers and employees to better combine their work and private lives. This also strengthens Germany as a business location because demographic change can be countered and hidden sources of labour can be mobilised in that members of the working population can once again join the world of work who were previously prevented from taking on employment—despite being suitably qualified—due to conflicts in their social environment (e.g. children or parents with care duties).

However, digitalisation also presents a risk because the blurring of the lines between work and private life can increase the work load on employees. According to Section 5 of the German Occupational Safety and Health Act (Arbeitsschutzgesetz), employers are required to carry out a risk assessment to identify suitable occupational safety and health measures. This means that, as part of their duty of care, employers must also carry out a risk assessment for work digitalisation measures and take suitable measures to prevent an excessive work load for employees. However, this is not just the company's responsibility; rather, individual employees must also be able to take responsibility and look after themselves in organising their interaction with digital technologies.

Although many companies are now aware of the profound change currently underway, they do not yet have suitable tools to embed this in their company organisation and to create a particularly flexible work structure. This article outlines their time management requirements.

2 Holistic Working Time Organisation as a Complex Planning Task

The desire for more flexible operating and working times is expressed more and more frequently by both companies [e.g. 8, p. 162] and employees [e.g. 14, p. 36]. Working time organisation places great demands on operational planners because, in addition to legal and salary-related provisions and ergonomic recommendations, the specific situation of the company and employees must also be adequately taken into account (see [17, p. 52]; for more detailed information, see [23]). These requirements sometimes conflict with one another. Recently, there has been an increasing focus on the compatibility of work and private life: For example, the "Familienbewusste Arbeitszeiten" (Family-Conscious Working Times) initiative was launched in Germany in February 2011 by the German government, trade associations and the Deutscher Gewerkschaftsbund (German Confederation of Trade Unions) [6].

The aim of holistic working time organisation is, therefore, to consider the objectives of all relevant interest groups—particularly employers, employees, social partners and customers—to an equal extent when selecting and structuring working time systems (i.e. all working time models practised in a department or company) and during the later planning of specific working times for employees. It follows that there are no standard solutions for working time organisation; instead, the specific operational situation must always be analysed to find an individual solution. Because a working time model or system will often remain valid in practical use for several years, the wrong decision can have very far-reaching consequences.

There are a range of working time models with different levels of flexibility for employees and companies. To better distinguish between these two dimensions, the term "working time sovereignty" is used when working hours are controlled by the employee, and "working time flexibility" when the company has the right of disposition [see 17, p. 53]. In 1995, Knauth [16, p. 210] estimated that around 10,000 different working time models were already being practised around the world. In recent times, further innovative working time models have been developed to better cover increasingly dynamic customer requests on the one hand and to enable employees to strike a better balance between their work and private lives on the other hand.

This is often impeded by the fact that many decision makers—particularly in small and medium-sized companies—have only very little understanding of the possibilities and effects of working time organisation [following 10, p. 245]. There are generally very few concrete findings about the effect of specific working time models on private life. Existing studies focus on more general aspects such as how free time is used [18] or the health effects and consequences of long working hours [27].

Aligning the personnel inventory and personnel requirement is considered one of the most important management tasks [8, p. 162]. Against the backdrop described, a systematic approach is required to planning and implementing working time systems that adequately considers the range of alternatives. The broad spectrum of working time models means that the organisational process is highly complex, cannot be resolved in a routine manner and takes a great deal of time [1, p. 184]. Due to the resulting complexity, the use of software is recommended when structuring working time systems that often includes an ergonomic assessment of working time models. Commercial software for working time organisation generally include only a static assessment (independent of the dynamic operational timeline) in which the available and required capacities are compared [see 23, p. 360]. In addition, a check is usually performed to determine whether statutory framework conditions and ergonomic recommendations for the department under observation are being observed. Böker [7] tested and presented the four commercial software procedures most commonly used for shift planning in Germany that allow such a static assessment to be performed. Existing procedures for structuring working time models take into account working time preferences at most but do not consider the strain placed on an individual by their work and private life. Only the online risk assessment procedure of the "Initiative Neue Qualität der Arbeit" (Initiative New Quality of Work, INQA) includes a blanket test for the probability of restricted participation in social life [see 9, p. 154].

A variety of input data not generally available as standard within companies is therefore always required for the systematic structuring of working time systems. The planning result depends heavily on the quality of this data. The following section first presents the input data required for working time organisation, which is then used to derive time management requirements.

3 Input Data for Working Time Organisation

During working time organisation, information is required about the personnel inventory and personnel requirement to ensure that these factors are aligned to the best possible standard. Generally, suitable operating and working times must first be determined for the company by means of complex analyses [see 26, pp. 107].

3.1 Personnel Requirement Dimensions

The personnel requirement is the personnel capacity necessary in qualitative and quantitative terms to perform work tasks. It is described by the type (qualification) and number of personnel required as well as the time or deadline, the duration and, if applicable, the location of the work [20, p. 270]. The personnel requirement therefore represents the operational requirements for personnel deployment. In the simplest case, the personnel requirement can be shown as a table in which the number of employees required is specified for each period (often by hour). Many computer-based procedures for structuring working time models use this as the basic data for determining the best working time model, for example [see 7, p. 30]:

– BASS 4.0 [12] uses the workforce or shift requirement and the potential shift types and break information as its basic data. The ergonomic assessment criteria can be weighted individually and encompass a definition pool for legal specifications and ergonomic recommendations.
– The workforce requirement recorded by e-Shift-Design [15] encompasses the different shift types to be planned, the target workforce per shift type and weekday, and the breaks. Six criteria are used for the assessment (e.g. number of working days or number of free days in succession) for which the planner can configure the weighting.
– Optischicht [25] gives 12 criteria (e.g. free time on weekends, prohibited or unfavourable shift sequences) for which the planner can define an order of

priority. The shift requirement, shift workforce and weekend working time are calculated based on the workforce requirement and the shift types.

– Ximes-OPA/-SPA [28] is also based on the personnel requirement, whereby the Time Intelligence Solutions [TIS] module supports the evaluation of a wide range of datasets and their aggregation to form an overall workforce requirement [see 11, pp. 63]. The requirement can therefore also be calculated for each task. Building on this, possible shift groups, reserve personnel requirements, full time, part time and individual assignments are determined. The solutions found are assessed with the aid of legal and ergonomic criteria.

However, a static assessment of this type is often insufficient because dynamic effects can occur in operations, for example due to stochastic execution times for work tasks or periodic fluctuations in capacity requirements that are generally noted in the average, but in reality are subject to stochastic effects.

During working time organisation, further information should therefore be available regarding the different dimensions of the quantitative and qualitative personnel requirement.

3.1.1 Quantitative Personnel Requirement

The *quantitative personnel requirement* C_{MBr} refers to the amount of time required to perform tasks at a specific point in time and for a specific duration [20, p. 270] and comprises the following two partial requirements (Fig. 1, right side):

– The *operational personnel requirement* C_{ME} represents what is required to perform all tasks that arise. This includes the *order-related personnel requirement* C_{MA} for processing orders and/or customers and the *additional personnel requirement* C_{MZ} for performing additional tasks not included in the order processing, such as maintenance or handovers between shifts [20, p. 271].

  Both the order-related requirement and additional personnel requirement are generally subject to mostly stochastic effects. The volume of the order-related personnel requirement can vary depending on stochastic execution or transport times, for example. The timing of an order and its activities can change due to fluctuations in order receipt or unplanned waiting times, which means that the resulting order-related personnel requirement can change in form. Due to increasing turbulence and customers' growing flexibility requirements, the order-related personnel requirement should be viewed as dynamic.

  By its very nature, the additional personnel requirement cannot be planned and depends strongly on work organisation. Stable and controlled processes can help to reduce the additional personnel requirement incurred, but generally cannot eliminate it entirely. To determine the personnel requirement, it is therefore extremely important to identify the additional tasks that arise and to obtain information about their predicted frequency and duration. Time recording, for example, may prove useful here. However, time recording usually needs to be agreed on by the work council therefore the employer and the work council have to work closely together.

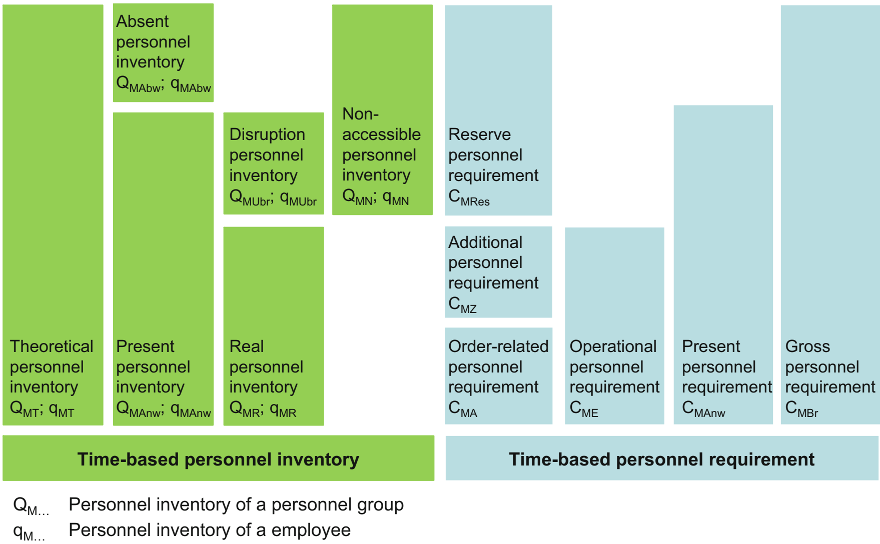


Fig. 1 Time-based personnel requirement and personnel inventory where covered [20, p. 276]

– The *reserve personnel requirement* C_{MRes} is necessary due to the reduction in capacity caused by the non-accessible personnel inventory (see Sect. 3.2.1), for example to compensate for personnel illness or holiday. In addition, it is generally difficult to plan the reserve personnel requirement, which can change at short notice. Key figures (e.g. illness rate) can help to improve the reserve personnel requirement forecast but must be kept up to date. In particular, it must also be checked whether key figures of this type can be recorded for the entire company or whether they must be determined for specific departments.

Furthermore, a distinction must be made between the *time-based personnel requirement C* and the *numerical personnel requirement n*: The time-based personnel requirement describes the personnel time required for a period, while the numerical personnel requirement specifies the number of people needed [20, pp. 206]. Commercial software for working time organisation (see Sect. 3.1) works with the numerical personnel requirement.

In the simplest case, the numerical personnel requirement can be determined by dividing the time-based personnel requirement by the duration of the period concerned. This figure is generally to be rounded up (i.e. if the computer calculates a numerical personnel requirement of 1.1, two people are required) because it is often not possible to flexibly deploy part-time employees, particularly in the case of smaller periods within a shift (e.g. 1.1 employees are required only in the period from 10.00 am to 11.00 am and two employees are required the rest of the time). This figure must also be rounded up if the tasks are not scheduled at a specific time and may arise dynamically within the period concerned.

Furthermore, differences may arise between the time-based and numerical personnel requirement that fundamentally depend on the requirements of the task in question: For example, in an observed period of 1 h, a time-based personnel requirement of 30 min may require more than one person to complete the tasks, for instance if a task requires two people or if two customers arrive at the same time and refuse to wait. Other aspects may also play a role: In retail, for example, a minimum of two people are often deployed in the evening for safety reasons, even if the time-based personnel requirement would require only one person to be present. In light of this, both the time-based and numerical personnel requirement should be recorded when determining the personnel requirement.

Ultimately, a distinction must also be made between the *task-related* and *position-related personnel requirement*: The definition of position, i.e. the assignment of tasks to a task bearer, can—depending on the assignment—further increase the position-related personnel requirement compared with the task-related personnel requirement determined. If tasks are assigned to more than one position, this makes capacity alignment between the personnel requirement and personnel inventory considerably more complex.

Depending on the respective operational conditions, various methods can be used to determine the quantitative personnel requirement, such as deterministic methods, stochastic methods (creation of key figures, regression analysis, correlation analysis, exponential smoothing), econometric methods, simulation methods, estimation methods. For more detailed information, see REFA [20, pp. 280]. However, these methods do not always adequately consider the outlined dimensions and, in particular, the dynamic effects of the work system.

3.1.2 Qualitative Personnel Requirement

The *qualitative personnel requirement* takes into account the qualification of the personnel required to fulfil a task. The objective of qualitative personnel planning is to align the requirements of the existing or planned future work systems with the services offered by the existing personnel or personnel to be recruited so that the requirements and service offering can be harmonised as far as possible [20, p. 270].

The performance requirements that a work system places on a worker can vary in severity and difficulty and result from the following aspects:

– The *work task* defines the skills, knowledge and abilities that a person requires to process the task. The work task also results in mental and physical requirements for the person. The work task is generally the key influential factor for the performance requirements.
– The *work equipment* can also require the worker to have special skills, knowledge and abilities.
– The *work environment* can also influence the performance requirements, i.e. due to environmental factors and the social setting. For example, hot or cold workstations are relevant to working time organisation because they can place a considerable physical burden on the worker. Various studies have shown that

an individual's private social sphere also plays a role in the structuring of shift models.
– Finally, *work organisation* is also relevant, for example a worker's responsibilities or the operational structure.

To achieve targeted personnel deployment, it is important to define the necessary requirements for the tasks that arise. Requirements determination identifies the requirements placed on a worker by a work task. There are various procedures for determining requirements such as the REFA requirement analysis [for more detailed information, see, for example, 22, pp. 638], which consider different types of requirement.

Often, only the quantitative personnel requirement is taken into account in working time organisation. However, holistic working time organisation should also adequately consider the qualitative personnel requirement; otherwise, a worker's performance may be impaired or quality may be lacking when the order is executed.

3.1.3 Local Personnel Requirement

Finally, the local dimension of the personnel requirement describes the spatial characteristics and/or location of the work performed. Increasing digitalisation means that some work tasks no longer need to be executed at a company workstation (see Sect. 1), but can also be performed in other locations such as a home office, co-working space or public space (café, train etc.). This gives employees a new degree of freedom but also requires a corresponding corporate, management and trust culture. Communication outlay can also increase if employees are not on the company premises.

To identify the local personnel requirement, each work task must be checked to determine whether it is tied to a company workstation and its specific requirements. Legal aspects such as occupational safety and health or data protection must also be taken into account here because these will naturally apply to flexible working too. There is currently no method to systematically determine the local personnel requirement.

3.2 Personnel Inventory Dimensions

The personnel inventory is the personnel capacity available in qualitative and quantitative terms to perform work tasks. It comprises the people employed at the company [20, p. 249]. Determining the personnel inventory sounds simple at first because a company should generally be aware of the people it employs. Nevertheless, the personnel inventory is not necessarily deterministic and may be subject to dynamic influences.

As with the determination of the personnel requirement, further information should therefore be available regarding the different dimensions of the quantitative and qualitative personnel inventory.

3.2.1 Quantitative Personnel Inventory

In the quantitative assessment of the personnel inventory, a distinction must be made between the following factors (Fig. 1, left side):

- The *theoretical personnel inventory* Q_{MT} results from the employees' contractually defined working times [20, p. 209]. This is usually specified as a number of hours per week, whereby the contract may go into greater detail regarding how the contractually agreed working time is to be fulfilled (e.g. at fixed or variable times). The theoretical personnel inventory is therefore always known.
- However, the theoretical personnel inventory is not available in full because there is always a *non-accessible personnel inventory* Q_{MN} [20, p. 250]. This consists of the absent personnel inventory and disruption personnel inventory in the department or company:
  - The *absent personnel inventory* Q_{MAbw} is calculated from times at which employees are not present during their working times defined by the company. Absences can be regularly scheduled in the form of employee holidays. Absences that cannot be scheduled may result from illness, unexcused absence, late arrival, doctor's appointments or appointments with authorities [20, pp. 254]. Companies are frequently aware of the *level of absence*, but this can fluctuate considerably. An aggregated level of absence is of only limited use in working time organisation because it obscures dynamic fluctuations in the personnel inventory. A detailed analysis of absences by means of time management is not only necessary for working time organisation; it can also uncover problems in the work system that can be eliminated through suitable working time organisation (or work structuring) and thus reduce absences.
  - The *disruption personnel inventory* Q_{MUbr} is calculated from times at which employees are present at the company but are not working. This may be caused by, for example, work equipment faults, lack of material, waiting for means of transport or works meetings [20, pp. 262]. The disruption personnel inventory cannot be scheduled and may indicate work organisation problems. As with the level of absence, companies are often aware of a *level of disruption*. Here too, this is not generally sufficient for holistic working time organisation because dynamic influences are not taken into account.
- The *real personnel inventory* Q_{MR} is therefore the difference between the theoretical personnel inventory and the non-accessible personnel inventory. It describes the accessible personnel inventory for performing the tasks that arise.

As with the personnel requirement, a time-based and numerical personnel inventory can be determined. If the capacity required fluctuates greatly over the course of a day or week, part-time personnel can be employed to cover particular peaks in capacity. If working time accounts are used, the time-based personnel inventory can also fluctuate. These fluctuations cannot be scheduled by the company if the working time sovereignty for the time account lies solely with the employee.

The time-based personnel inventory can also be hampered by employees' health restrictions, and may cover only certain periods or activities. In Germany, relevant regulations can be found in a variety of laws: For example, a ban on night work can reduce the accessible personnel inventory at night (11.00 pm to 6.00 am). The Youth Employment Protection Act (Jugendarbeitsschutzgesetz) significantly reduces permitted employment times for young people (under the age of 18) compared with those for adults. In addition, the Maternity Protection Law (Mutterschutzgesetz) considerably restricts permitted employment times and activities for expectant mothers.

Finally, employees' private lives will naturally determine their own working time restrictions and wishes. This can be due to a wide variety of reasons such as childcare, caring for relatives, secondary employment or voluntary roles. If existing restrictions on working times are not taken into account when planning working times, this can even mean that an employee is no longer available to the company or to the job market as a whole.

Therefore, the accessible personnel inventory can definitely fluctuate depending on a department's HR structure. If appropriate data is provided, time management can help to predict these fluctuations. At the moment, however, these aspects are not yet taken into account systematically in working time organisation, but are at best considered intuitively by planners.

3.2.2 Qualitative Personnel Inventory

The qualitative personnel inventory results from the services that employees can provide. A distinction can be made here between two dimensions of human capability:

– *Ability to perform* describes those characteristics that physiologically or psychologically influence a worker's performance. A further distinction can be made here between unalterable constitutional characteristics (e.g. sex, physique), disposition characteristics (e.g. intelligence, age, weight, health, rhythmological influences) and alterable qualification and competence characteristics (e.g. experience, knowledge, abilities, skills, education) [22, pp. 87]. The ability to perform is generally to be regarded as constant across at least one shift, but usually across a longer period. A person's health and rhythmological influences may be exceptions here, but may also play practically no role in working time organisation.
– *Willingness to perform* describes a worker's willingness to use his or her ability to perform at the current point in time. Aspects such as tiredness, motivation, satisfaction and mood play a key role here [22, pp. 87]. Willingness to perform can be influenced by external factors. It can also fluctuate within a shift. This means that this factor is practically incalculable in working time organisation, and is therefore not examined more closely in practice.

Both the ability and willingness to perform are subject to fluctuations. Performance differs on an interindividual basis—between different people—and within a single person (on an intraindividual basis), with the potential performance depending on a person's current state of health or motivation, for example [22, p. 87]. While the ability to perform can at least be evaluated in principle, there is barely any valid method of determining the willingness to perform.

In day-to-day operations, qualitative assessments of the personnel inventory are usually based solely on the creation of personnel types who have the same qualifications and are defined for specific capacity requirements. However, this causes problems if employees have very different qualifications and lots of personnel types have to be taken into account in working time organisation. There is currently no systematic procedure for determining the dynamic quantitative and qualitative personnel inventory.

3.3 Complex Determination of Personnel Requirement and Personnel Inventory

When the outlined dimensions are taken into account for the personnel requirement and personnel inventory, a number of requirement and inventory profiles are produced that the working time system to be designed must fulfil to the best possible standard. Although this increases the complexity of the planning problem, it also increases the degree of freedom in order to better respond to employees' wishes to work at specific times or locations.

When working time systems are assessed, the cost of personnel deployment is usually also incorporated in working time organisation. This cost can differ in particular if an employee receives individual compensation or bonuses are contractually agreed (e.g. for night work, paid overtime). In this case, the cost of personnel deployment adds another condition to the planning problem.

Determining the personnel requirement and personnel inventory is thus by no means trivial because comprehensive data must be determined that is usually not held by companies as standard. However, time recording usually needs to be agreed on by the work council therefore the employer, the work council and the employees have to work closely together. This process needs a culture of trusts in which every party acts caring for the wishes of the other party.

4 Time Management Requirements for Working Time Organisation

The aim of time management is to manage all times required in a company for workers, work equipment and work objects, and also in particular to gather and prepare data for working time organisation [22, p. 665]. Time management is therefore of considerable importance for designing working time systems and planning personnel schedules. In day-to-day operations, however, it is often underestimated.

Working time data can be determined via continuous monitoring (e.g. time recording based on REFA), sample monitoring (e.g. multi-moment recording) and computer-based analytical procedures (e.g. systems with predefined times such as MTM or the work factor system) (for detailed information, see [22, pp. 671]). There are also some special ways to determine time data (e.g. surveys, comparisons and estimations; [22, p. 671]. The methods mentioned are a deterministic way of recording a company's personnel requirement. Time management also provides some methods for the stochastic determination of capacity requirements that examine relationships between the personnel requirement and other variables, namely the creation of key figures, regression analysis, correlation analysis and exponential smoothing [see 20, pp. 290]. In future, these methods will become increasingly important for working time organisation in order to adequately map the dynamic effects in a company. Furthermore, other methods of exploratory data analysis and inductive statistics must be incorporated in time management to describe a company's work system and the resulting personnel requirement more precisely.

While various methods of determining the personnel requirement are available and used to varying degrees in day-to-day operations, there are practically no standardised methods for the differentiated analysis of the dynamical personnel inventory. In particular, employees' wishes regarding working times (and, if applicable, locations) have not as yet been systematically incorporated in the determination of the personnel inventory. In future, time management must therefore develop corresponding methods based on employee surveys and employ these in day-to-day operations. The relevant legal requirements must be observed here. In Germany, the Works Constitution Act (Betriebsverfassungsgesetz) regarding data determination and the Federal Data Protection Act (Bundesdatenschutzgesetz) regarding data usage and management are particularly relevant here.

Ultimately, new methods must also be developed for aligning the personnel requirement and personnel inventory that can take into account stochastic influences. Due to the complexity of the planning problem, these methods will also become increasingly complex, meaning that greater use could be made of heuristic approaches, for example. This increases the need for training—both for operational planners and the works councils involved—so that the most suitable methods can be selected and applied to the company in question.

5 Discussion

The holistic assessment of working time systems that give equal weight to the interests of companies and employees requires new time management methods to be used that adequately map the various dimensions of the personnel requirement and personnel inventory.

Furthermore, to assess alternative solutions, a holistic key figure system is therefore required (for more detailed information, see [23, p. 360], which includes key figures to assess the following:

– Compliance with legal, salary-related and individual contract specifications
– The implementation of ergonomic recommendations
– The compatibility of work and private life
– The achievement of monetary and production logistics objectives

Up to now, commercial tools for working time organisation generally have not provided options for modelling the outlined profiles for the personnel inventory and personnel requirement or for multi-criteria assessment. This is due to several reasons, including the associated outlay for determining data.

On the other hand, personnel-orientated simulation provides the option to map and distinguish between an extended personnel requirement and personnel inventory via the detailed modelling of the work system and thus to take these into account in working time organisation. In contrast to the commercial tools, personnel-orientated simulation enables the behaviour of complex work systems to be analysed and therefore provides further options for finding solutions. However, this is an expert procedure that requires extensive time management data about the work system in question in order to create the simulation model; this data is not always available in companies and may first have to be collected before a simulation study can be conducted. Nonetheless, the fact that working time systems often remain valid for very long periods means that this outlay can pay off.

References

1. Ackermann K-F (1990) Prozeßstandardisierung des Arbeitszeitmanagements. In: Ackermann K-F, Hofmann M (eds) Innovatives Arbeits- und Betriebszeitmanagement. Campus Verlag, Frankfurt/M, pp 183–212
2. Baszenski N, Stowasser S (2013) Flexibilität in Unternehmen der Metall- und Elektroindustrie. In: Bornewasser M, Zülch G (eds) Arbeitszeit – Zeitarbeit: Flexibilisierung der Arbeit als Antwort auf die Globalisierung. Springer Gabler, Wiesbaden, pp 61–76
3. Bellmann L, Crimmann A, Evers K (2013) Betriebliche Dynamik und Flexibilisierung auf dem deutschen Arbeitsmarkt. In: Bornewasser M, Zülch G (eds) Arbeitszeit – Zeitarbeit: Flexibilisierung der Arbeit als Antwort auf die Globalisierung. Springer Gabler, Wiesbaden, pp 43–60
4. BMBF – Bundesministerium für Bildung und Forschung: Wissenschaftsjahr 2013 – Die demografische Chance. http://www.demografische-chance.de/. Retrieved 30.05.2015
5. BMFSFJ – Bundesministerium für Familie, Senioren, Frauen und Jugend: Demografischer Wandel, 24. September 2014. http://www.bmfsfj.de/BMFSFJ/Familie/demografischer-wandel.htm. Retrieved 30.05.2015
6. BMFSFJ – Bundesministerium für Familie, Senioren, Frauen und Jugend (2014) Material zur Initiative 'Familienbewusste Arbeitszeiten'. http://www.erfolgsfaktor-familie.de/default.asp?id=518. Retrieved 30.05.2015
7. Böker K-H (2011) Spezielle Software für 'Gute Schichtarbeit'. Schichtplan-Modellierungs- und Bewertungs-Software im Vergleichtstest. Computer und Arbeit 1:28–32
8. Brützel U (1998) Aktive Personalsteuerung für mehr Produktivität. Der Arbeitgeber 50(6): 162–164
9. Dittmar O, Schomann C, Nachreiner F (2010) Online-Risikobewertung von Arbeitszeiten. In: Zülch G, Stock P (eds) Einfluss der Arbeitszeitgestaltung auf die Work-Life-Balance. Karlsruher Institut für Technologie, Karlsruhe, pp 151–165

10. Ferreira Y, Landau K (2001) Umsetzungen flexibler Arbeitszeitregime am Beispiel eines deutschen Großflughafens. In: Gesellschaft für Arbeitswissenschaft (ed) Arbeitsgestaltung, Flexibilisierung, Kompetenzentwicklung. GfA-Press, Dortmund, p 245–250
11. Gärtner J, Boonstra-Hörwein K (2012) Sozialverträgliche Schichtplan- und Arbeitszeitgestaltung durch XIMES. In: Zülch G, Stock P (eds) Arbeitszeit – Familienzeit – Freizeit. Karlsruher Institut für Technologie, Karlsruhe, pp 60–83
12. GAWO – Gesellschaft für Arbeits-, Wirtschafts- und Organisationspsychologische Forschung e.V. (2007) *BASS 4 (Version Dezember 2007)*. http://www.gawo-ev.de/cms2/index.php?page=bass-4&phpMyAdmin=3ff319a86a720ca633f7d8bfe08952c0&phpMyAdmin=8b6ed5803bbabc8d5f96599c9c6997ad. Retrieved 30.05.2015
13. Hessenmöller A-M, Pangert B, Pieper C, Schiml N, Schröer S, Schüpbach H (2014) iga.Barometer 4. Welle 2013: Die Arbeitssituation in Unternehmen. Eine repräsentative Befragung der Erwerbsbevölkerung in Deutschland. Flexibilität, Life-Domain-Balance und Gesundheit; Auswirkungen von Erwerbslosigkeitserfahrungen. AOK-Bundesverband, BKK-Bundesverband, Deutsche Gesetzliche Unfallversicherung, Verband der Ersatzkassen. Juli 2014. http://www.iga-info.de/fileadmin/redakteur/Veroeffentlichungen/iga_Reporte/Dokumente/iga-Report_27_Barometer_Flexibilitaetsanforderungen_Erwerbslosigkeitserfahrungen.pdf. Retrieved 30.05.2015
14. IDA – Institut für Demoskopie Allensbach (2010): Monitor Familienleben 2010. http://www.bmfsfj.de/RedaktionBMFSFJ/Abteilung2/Pdf-Anlagen/familienmonitor-2010,property=pdf,bereich=bmfsfj,sprache=de,rwb=true.pdf. Retrieved 30.05.2015
15. ifaa – Institut für angewandte Arbeitswissenschaft e.V. (2009) *e-Shift-Design – Die Praxishilfe zur Schichtplanung*. http://www.arbeitswissenschaft.net/e-Shift-Design-Die-Praxishil.766.0.html. Retrieved 01.03.2014
16. Knauth P (1995) Was kann das betriebliche Zeitmanagement zur sozialverträglichen Gestaltung von Arbeitszeiten beitragen? In: Büssing A, Seifert H (eds) Sozialverträgliche Arbeitszeitgestaltung. Rainer Hampp Verlag, München, pp 209–220
17. Knauth P (2002) Arbeitszeitflexiblisierung aus arbeitswissenschaftlicher Sicht. In: Zülch G, Stock P, Bogus T (eds) Arbeitszeitflexibilisierung im Dienstleistungsbereich. Shaker Verlag, Aachen, pp 51–74
18. Knauth P, Hornberger S (1997) Schichtarbeit und Nachtarbeit: Probleme – Formen – Empfehlungen, 4th edn. Bayerisches Staatsministerium für Arbeit und Sozialordnung, Frauen und Gesundheit, München
19. Picot A, Neuburger R (2013) Arbeit in der Digitalen Welt. Zusammenfassung der Ergebnisse der AG1-Projektgruppe anlässlich des IT-Gipfels-Prozesses 2013. München, Oktober 2013. http://www.bmwi.de/BMWi/Redaktion/PDF/A/arbeit-in-der-digitalen-welt,property=pdf,bereich=bmwi2012,sprache=de,rwb=true.pdf. Retrieved 30.05.2015
20. REFA (1985) Methodenlehre der Planung und Steuerung – Teil 2, 4th edn. Carl Hanser Verlag, München
21. Rump J, Walter N (eds) (2013) Arbeitswelt 2030: Trends, Prognosen, Gestaltungsmöglichkeiten. Schäffer-Poeschel, Stuttgart
22. Schlick C, Bruder R, Luczak H (2010) Arbeitswissenschaft, 3rd edn. Springer, Heidelberg
23. Stock P, Deml B (2014) Ergonomic design of working time models. In: Trzcielinski S, Karwowski W (eds) Advances in the ergonomics in manufacturing: managing the enterprise of the future. CRC Press, Boca Raton, FL, pp 357–367
24. Strobel H (2013) Auswirkungen von ständiger Erreichbarkeit und Präventionsmöglichkeiten. Ed.: AOK-Bundesverband, BKK-Bundesverband, Deutsche Gesetzliche Unfallversicherung, Verband der Ersatzkassen. May 2013. (iga.Report 23). http://www.iga-info.de/fileadmin/redakteur/Veroeffentlichungen/iga_Reporte/Dokumente/iga-Report_23_Staendige_Erreichbarkeit_Teil1.pdf. Retrieved 30.05.2015
25. TÜV Nord (2012) Optischicht: Schichtpläne gesundheitsgerecht und sozialverträglich gestalten. http://www.tuev-nord.de/de/optischicht-99912.htm. Retrieved 30.05.2015
26. Wehrhahn B, Holzbach W, Heinen E (2001) Flexible Arbeitszeitgestaltung (eds) Institut für Technik der Betriebsführung. Verlag Dr. Jochem Heizmann, Gifthorn

27. Wirtz AK (2010) Lange Arbeitszeiten – Untersuchungen zu den gesundheitlichen und sozialen Auswirkungen langer Arbeitszeiten. Universität Oldenburg. http://oops.uni-oldenburg.de/928/1/wirlan10.pdf. Retrieved 30.05.2015
28. Ximes (2014) Personalplanung. http://www.ximes.com/. Retrieved 30.05.2015
29. XING (Hrsg.): Pressemitteilung: Forsa-Studie zu Wünschen deutscher Berufstätiger: Frauen suchen Sinn in der Arbeit, Männer ein gutes Gehalt. https://corporate.xing.com/no_cache/deutsch/presse/pressemitteilungen/pressemitteilungen-detailansicht/article/pressemitteilungbrforsa-studie-zu-wuenschen-deu/. Retrieved 30.05.2015
30. Zukunftsinstitut (ed) (2011) Heidelberger Leben Trendmonitor 2011. http://www.eaf-bund.de/documents/Paare_Navipunkte/Studie_Heidelberger_Leben.Trendmonito_2011_final.pdf. Retrieved 30.05.2015

Money Is Not Everything! Or: The Importance of Working Time Characteristics and Appreciation for the Recommendation of One's Own Driving Profession

Patricia Tegtmeier, Ulrike Hellert, and Bianca Krol

Abstract

The already existing shortage of qualified bus, tram and truck drivers will continue to increase in the coming years. Transport companies face a major challenge to ensure their human assets and competitiveness. Above all, suitable trainees must be obtained for the driving professions. The currently more than 800,000 professional drivers are playing a significant role as multipliers in this context. Research group DO.WERT conducted a survey among experienced professionals and trainees in various driving careers. Results indicate that only a quarter of the experienced workers would recommend their own driving profession. Crucial to a (non-)recommendation of the profession are experienced appreciation and working hours.

Keywords

Skill shortage • Career choice • Working time • Time pressure • Appreciation

P. Tegtmeier (✉)
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund, Germany
e-mail: tegtmeier.patricia@baua.bund.de

U. Hellert
iap - Institut für Arbeit & Personal, FOM Hochschule für Oekonomie & Management, Essen, Germany
e-mail: ulrike.hellert@fom.de

B. Krol
ifes - Institut für Empirie & Statistik, FOM Hochschule für Oekonomie & Management, Essen, Germany
e-mail: bianca.krol@fom.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_7

1 Introduction

According to TÜV Rheinland [23, p. 12], there is already a shortage of qualified drivers. As a result of demographic change in combination with the high and specific workloads (e.g. long-lasting, security-related attentional demands, shift work, night work) this current staff shortage is going to spread significantly in the next few years. In part this development is based on the known overall social effect of demographic change: the baby boomers are increasingly reaching retirement age and simultaneously the group of people that enters into the workforce is sinking. Only about 2040 when the Baby Boomer generation has retired from the labor force, the age structure stabilizes [8, p. 4]. Thus, the proportion of over-50s employees among professional drivers in 10 years from about a quarter in 2001 has increased to just over 39 % in 2011. During the same period, the proportion of 25- to 35-year-olds has decreased from almost 22 % to only 14 % [2, p. 101]. Parallel to this development further increases in transport services are expected [4, p. 35]. Therefore, to secure their human assets and capacity to compete transport companies should increasingly seek younger employees and especially suitable trainees for the driving professions.

In addition to the overall demographic effects, specific logistical aspects exacerbate the shortage of qualified drivers. In the past career changers were usually hired to fill vacancies in driving occupations. The main qualification being a valid driving license, that often was acquired—free of charge for the companies—via the Bundeswehr [16, p. 21]. By the end of conscription the option to recruit these externally trained drivers ceased to exist. Although the end of conscription was foreseeable, the transport sector only after a considerable delay began, to train drivers itself. Currently, only one in three commercial transport operation is providing opportunities for apprenticships in the dual system [23, p. 17]. Especially small companies seldom meet all standards required to train apprentices themselves [2, p. 122]. However, more than 80 % of companies in the traffic and logistics sector have fewer than 20 employees [16, p. 21]. Simultaneously, the requirements in this profession segment increased gradually over the past years. Besides additional technical knowledge, the demands of personal, activity-related and social-communicative skills have increased [7, p. 17]. Drivers, with the capacity and commitment to use their knowledge and skills actively, as well as suitable trainees are needed.

However, the declining population of school-leavers combined with falling numbers of apprentices due to a general academization [4, p. 28] lead to increased competition for suitable trainees between training establishments and various industries. In addition the career aspirations of potential apprentices are not evenly spread across the various professions. Professions are chosen according to their potentials for social approval and recognition. If the image of a profession appears unsuitable to achieve social recognition, it becomes unattractive irrespective of the actual professional content and the interests of potential trainees [6, p. 10]. A study researching the social prestige of different occupations found professional drivers ranking [22, p. 266] in the lowest-eighth of 128 assessed professions. The image of driving professions is rather poor according to industry experts as well [2, p. 123, 16, p. 35]. Therefore, getting potential apprentices interested in the driving professions based on this image, seems rather unlikely. So even training-ready companies may not win enough young talent for apprenticeships. In 2011 seven percent of reported training slots remained vacant [2, p. 123].

Seeking competent advice plays an important part in choosing an occupation. In a longitudinal study on the transition from school to work "Hauptschüler/innen" (secondary school pupils) most frequently consulted family members and friends [13, p. 18]. Companies in logistics predominantly engage school leavers holding a "Hauptschulabschluss" (secondary general school certificate) as apprentices. Thus, the currently more than 800,000 professional drivers hold an important multiplier function as a source of information on the advantages and disadvantages of their own profession.

Within the framework of the research group DO.WERT[1] an empirical survey was performed focusing employability, health at work, organization of work, working time and qualification in various driving occupations. Shown below are the analysis of factors correlated with a positive recommendation of the own driving profession.

2 Method

The overall survey on professional conditions of the driving profession was conducted between September 2013 and May 2014. It consisted of a mix of quantitative and qualitative data collection with different groups of people: A quantitative questionnaire study was aimed directly at experienced professional drivers and trainees for transportation of cargo as well as apprentices for passenger transportation (FiF).[2] Semi-structured expert interviews[3] were a further data source. Experts being persons that could provide information about various aspects of (driving) work life based on a professionally close contact with drivers of different divisions.

[1] DO.WERT—Demography oriented organization and design in the logistics and transport sector—more value for smart, green and integrated transport is an interdisciplinary research group of the FOM University of Applied Sciences, funded 2013–2014 under the program FH structure by the Ministry of Innovation, Science and Research of North Rhine-Westphalia (NRW MIWF).

[2] FiF (Fachkraft im Fahrbetrieb)—Skilled Transport Employee (m/f) is a recognized public transport professional with 3 years of training. In addition to driver certification for buses, trams and underground trains also technical as well as business training content is conveyed.

[3] As experts persons are included that can be seen with regard to Gläser and Laudel "as a source of specialist knowledge about the social issues to be explored" [9, p. 12].

2.1 Questionnaire Survey

The survey of experienced professional drivers took place: in the context of qualification schemes at ongoing training providers, within the framework of regular truck meets, at local trucking companies, by sending questionnaires to trucking companies, and via an online questionnaire. Overall, drivers from different areas of transport could be reached. Access to the trainees for professional truck drivers as well as FiF apprentices was established via vocational colleges. The pupils had the opportunity to fill in the questionnaires during lessons.

In addition to general socio-demographic questions, items about work experience and the company the questionnaire consisted of items to different topics. Shown below are only the questionnaire aspects, with results later referred to are shown below.

Time spent working in 2013 was queried as a numeric in hours. Comparable questions were asked, e.g. in the context of the BIBB/BAuA employment survey[4] in 2012 [18]. Point of reference for information on the contractual and actual time of work was the weekly working hours. Further queried was the number of contractual and taken leave days, as well as the number of Saturdays and Sundays worked, number of days off that were occupationally not spent at home and sick leave days for 2013.

Items to *workload* were taken from the ISTA [21] and Hellert [11]. Six items were used to assess time requirements (e.g. "How often does it happen that you skip a 30-min break because of too much work?") on a five-point Likert scale (1 = never, 5 = always). An item was added concerning problems based on working hours with family and friends on the same scale.

Predictability of working hours was assessed through two questions: a six-level item for knowledge of schedules in advance (1 = 0 to 1 day in advance, 6 = longer than 30 days in advance) and a four-level item on the frequency of short-term changes (1 = less than once a month, 4 = more than 4 times a month).

Decision latitude at work was determined through five items (e.g. "You have influence on the planning of working hours (shifts and overtime)") from the ISTA scales [21]. An assessment was made on five-point Likert scales (1 = always, 5 = never).

Job satisfaction and strain were one item assessment each. Satisfaction with the existing working conditions as a whole was judged on a five-point scale (1 = very satisfied to 5 = very dissatisfied) [11]. Strain was surveyed through, whether the own current activity may be pursued until retirement under the current requirements, on a three-step scale (1 = yes probably, 3 = no, probably not). This item was supplemented by a possibility to an open response citing the terms and conditions that would be necessary from the perspective of the respondents.

[4] BIBB—Bundesinstitut für Berufsbildung, BAuA—Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.

Three items were assigned to *economic results* of work. Asked was who bears the costs for the mandatory training required by the professional drivers-qualification law (BKrFQG), the payment of premiums and the amount of the income—the latter as a six-level classification in 500 € increments (1 = less than 1000 €, 6 = 3000 € and more).

Appreciation has been obtained for three different groups (society in general, clients and dispatchers) on five-step Likert scales (1 = very much, 5 = very little/no).

Advantages and disadvantages of the own driving profession was queried by means of open-ended questions. A dichotomous question (Yes/No) was linked to whether the own profession would be recommended.

2.2 Expert Interviews

For the qualitative survey 13 expert interviews were conducted. The average interview length was 52 min. The shortest interview took 25 min, the longest interview comprised 84 min. The interview guide contained the same topics as the questionnaire. Additional information was asked about the employment structure in the driving profession and the recruitment of career changers vs. trainees in the context of the need for skilled workers. Four employees of different transport police directorates (prevention) were interviewed. Another four interviews were conducted with managers from different companies, including two from public transport and two cargo companies. Three interviews were conducted with OSH "working professional drivers" contacts and two teachers from education and training in the driving professions.

The evaluation of the interviews took place in the form of a qualitative content analysis according to Gläser and Laudel [9, p. 197–260]. The performance profile on social rating of occupations according to Goesmann [10, p. 40–60, see Table 1] served as a basis of content analysis as well as the open answers of the questionnaire.

The physical and mental effort to be provided is an aspect of the *effort category*. Details of both inhibiting and promoting working conditions at driving jobs are associated with this category.

Statements about the skills and qualifications needed are assigned to the *category resources*. The qualifications correspond to the formal requirements for the purposes of education and training as well as the acquired knowledge needed to perform the job.

The *factual result* includes the quality of task performance. This is sometimes difficult to distinguish from the resources mentioned above, if, as here, the quality of the result is closely linked to professional knowledge and skills [10, p. 50]. Allocated to this category were in particular, statements about concrete measurable skills. A key characteristic here was the possible visibility of performance and quality.

Table 1 Profile on social rating of occupations

Dimension	Category
Effort	Resources needed: skill, competencies, qualifications Demands: physic or psychic
Result	Performance: workmanship Social: importance for society in general Economic: income, market demand
Person	Self-realization, authenticity

Source: Goesmann, C. (2010), p. 37 [10]

The problem solution for individual clients as well as the need for the goods and services for the population as a whole is represented in the category of *social result*.

The central aspect of the category *economic result* is the income as a financial indicator of a job well done. Other paid services such as paid premiums and reimbursement of training costs can be added as an expression of appreciation to the employees. Although not monetary job security can be assigned to the economic result as well.

The aspects of self-determination and self-realization in the profession are associated with the *person* dimension. Occasions for self-development, motivation and personal responsibility as well as obstructive factors were recorded. Also included were the reasons for the career choice in the driving professions.

3 Results and Discussion

The questionnaire respondents with a total of 436 participants divided into approximately 52 % experienced professional drivers and nearly 48 % apprentices (69 % trainee truck drivers and 31 % trainee FiF). The average age of experienced drivers was 47 years and ranged from 23 to 72 years—both the average age and the distribution corresponds to the socio-demographic data from other surveys concerning professional drivers [16]. According to the experts the partially high age of the driver shows three things: good new drivers lacking, companies now like to engage proven remaining forces back. There are drivers who not only can but also want to drive beyond the official retirement age—however, partly also because the pension has to be bolstered: "Sometimes you wonder at what age people do the job, yet this is certainly due to a certain necessity".[5] The apprentices on average were about 22 years old with a span from 16 to 45 years. Their rather higher average age is likely due to the fact that companies prefer apprentices who at least possess a car driving license.

[5] Since a translation always includes an interpretation, the original quotations are provided in footnotes. "teilweise wundert man sich, in welchem Alter die Leute den Job noch machen, das ist sicherlich auch aus einer gewissen Not heraus".

Approximately 4 % of experienced drivers were female. The ratio of female trainees was slightly higher with about 6 %. In detail the proportion of women was 17 % for FiF trainees, whereas female prospective truck drivers accounted for less than 2 %. Lohre et al. [16] show a similar distribution between the sexes for driving occupations. The interviewed experts suspected a general lack of interest on part of women. Specifically with regard to training, technically interested women would rather choose professions with a better image (for example, mechatronics). The experts, however, saw no significant reason in lack of physical strength. By using the existing technology for the driving professions this is no longer necessary in large scale usually. At the same time some of the interviewees reported enthusiastically partially coupled with amazement on the expertise and the technical and driving skills shown by women. Despite these abilities and partly attested positive impact on the working atmosphere and the general handling of the vehicles, the recruitment of women in the driving professions is not actively promoted. In the interviews a rather traditional idea of the distribution of employment and family work prevailed. As primarily seen responsible for family work, women have even greater problems than men concerning the dominant working hours in the driving profession: "Especially for younger women, our schedule is of course highly uninteresting. We just need to be very flexible and you can't do that, if you regularly take children to school".[6] Especially the (imputed) fertility was seen as an obstacle to hire female drivers: "the other (reason) is, of course, that you have to consider additional problems concerning women [. . .], I once had one, was really great that one, until she got a child."[7]

The note on working hours being unfavorable primarily for women seems too short-sighted. Already women in other service occupations, especially in retail and in clinical and home care, work similar hours and shift models. It is also questionable whether a division of work and family life on the basis of sex is continued to be practiced by younger generations. Work to family reconciliation is far more important for young professionals of both sexes than for previous generations (Hoffmann-Lun et al. 2005, p. 18). "I don't take every extra shift—no one can pay me the quality time with my kids",[8] as a male trainee expressed on a workshops, where result of the study were presented.

[6] "Gerade für jüngere Frauen ist der Fahrdienst natürlich hochgradig uninteressant. Wir müssen halt sehr flexibel sein und das geht nicht, wenn man regelmäßig Kinder in die Schule bringt".

[7] "das andere ist natürlich, dass bei Frauen da halt eben andere Problemstellungen noch dazu kommen, die man berücksichtigen muss [. . .] ja, ich hatte mal eine gehabt, die ist supertoll hier gefahren, bis sie dann ein Kind bekommen hat".

[8] "ich übernehm' nicht jede Zusatzschicht—die Zeit mit meinen Kindern kann mir gar keiner bezahlen!".

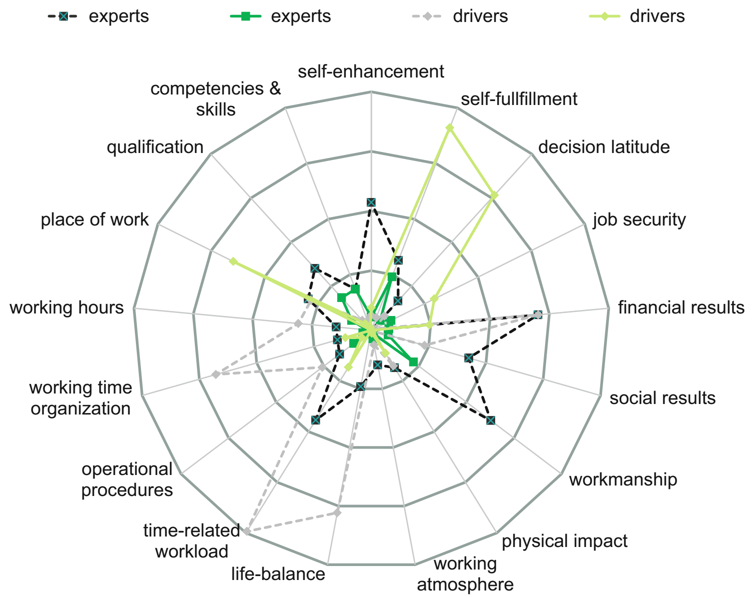


Fig. 1 Job profile evaluation (open answers)

3.1 Assets and Drawbacks of Professional Driving

The drivers themselves see the benefits of driving occupations, first and foremost in the opportunities for self-development and self-fulfillment (Fig. 1). In the corresponding open questionnaire part answers indicative the fun of driving and the variety of activities were mentioned most frequently. Also frequently stated were the opportunities for independent and self-responsible work as one's own boss in the driver's cab. Furthermore, the specific workplace is a pro for the drivers—both the technical aspects as well as to be out on the road, to experience different places and people and especially not to be stuck in an office.

The interviewed experts were overall significantly more problem-centered in their statements. But the positive aspects were also seen especially in categories concerning self-development, authenticity and fulfillment: "I know many drivers who put their heart and soul into the business."[9] They described the driving profession as best suited for individualists with a taste for single workstations, a high willingness to take on responsibility and an interest in the driving itself. According to those interviewed, the professional drivers can be mainly divided into two groups: one group deliberately chose the profession because of a fascination with the automobile and the fun in driving. Other important reasons for this group are, the realistic assessed chances to autonomy, variety and independence,

[9] "ich kenne viele Fahrer, die hängen mit Herzblut an der Sache".

which the driving profession can provide: “to have a certain freedom that you’re not constantly sitting in the office but to be on the move”.[10] In many cases, members of this group have in advance knowledge about the driving profession and the associated requirements. Often these were mediated by family members in the same business. Women, who take a driving career, were more associated with this group by the interviewees. The other group was described by the experts as the “last resort”.[11] The driving jobs are the last chance of a job or a training place for these. The main reason to drive is purely financial. Within this group a rather resigned attitude towards the profession and the possibilities for self-realization prevails. The experts agreed less about the amounts of professional drivers in the two aforementioned groups. With regard to those undertaking the 3-year apprenticeship a 50:50 distribution was adopted. Career changers, so the assessment, disproportionately belong to the second group. The interviewees suspected regardless of group membership that missing advancement and growth opportunities in the long run lead to problems for job-related (self-)motivation.

Both experts and the drivers questioned saw the disadvantages of driving occupations, primarily in the form of physical and psychological requirements as well as the (visible) result of the work done. Both aspects including the respective quantitative data will be discussed in detail below.

3.1.1 Workload, Temporal Working Characteristics and Life-Balance

Working conditions in the driving professions were assessed as rather unfavorable by all experts. Mainly, the long to overlong working hours and the widespread shift work were mentioned in the interviews. A look at the questionnaire data shows that the contractual hours of respondents with work experience on average was 44.5 h per week. A quarter reported working hours beyond 48 h per week—a contractual basis inconsistent with the German working time laws.

With an average 51.5 h per week the actual hours worked by the experienced drivers are significantly higher than the agreed upon. More than 13 % of the respondents indicated that they actually work more than 60 h a week. Matched against IAB data of 2012 mean actual working hours exceeded that of the average male full-time worker by about 7.5 h per week [5]. The trainees stated an averaged 40.2 contractual hours per week. About 4 % indicated contractual working hours exceeding 48 h per week. The average hours worked by trainees was 46.5 h per week. Differentiated according to the two groups of trainees the actual working hours per week among trainees FIF ranged on an average of 39.9 h, while the trainees for BKF worked about 49.7 h. The interviewed experts linked especially these long hours both with insufficient rest and with significant impairment in regard to the compatibility of private interests and obligations and the driving

[10] “ne gewisse Freiheit zu haben, dass man nicht ständig im Büro sitzt, sondern auch draußen unterwegs ist”.

[11] “Auffangbecken Kraftfahrer”.

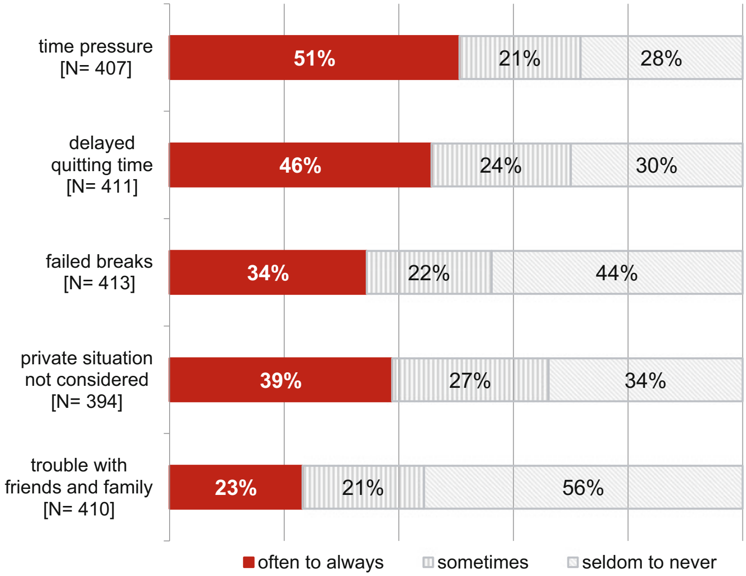


Fig. 2 Time-related requirements and workload

profession. In particular, this poor compatibility was mentioned in the interviews as a main factor for the low attractiveness of the profession.

The vast majority of respondents (about 79 %) stated that they obtain their schedules with a very short lead of zero to one days. Furthermore, 38 % of drivers indicated that short-term changes in the work plan occur more than four times a month. Asked to what extent consideration is given to working time with regard to their personal situation almost 40 % of the experienced drivers and trainees replied rarely or never. Coordination with private appointments (e.g. for health care, administrative procedures, hobbies as well as with the family and friends) is a major challenge in this context. Not surprisingly about 23 % of the drivers reported often or always having problems with family or friends due to their working hours (Fig. 2). A significant correlation to the actual hours worked could be observed ($r_s = 0.40$, $p \leq 0.001$)

The high concentration requirements and the very frequent time pressure have also been mentioned as stressful for drivers by the experts. In the open questionnaire stress and time pressure alongside permanent concentration and high responsibility ranged at the top of the mentioned disadvantages. Congested roads not only lead to conflicts with other road users. Combined with tight tour plans, and small time windows for loading and unloading at the ramp high flexibility and time requirements arise [20, p. 123]. In passenger transport increased time pressure results from very tight turning times, fast cycles and ever shorter rest breaks [17, pp. 53, 19, pp. 17]. About half of the sample stated that they are always or often under time pressure due to deadlines. Only one third was seldom or never under time pressure because of deadlines (Fig. 2). Compared to other professions the perceived stress through time and deadline pressure is significantly higher [20, p. 126]. So the drivers are at a high risk as especially experienced time pressure is related to physical and emotional exhaustion and complaints such as depression, insomnia, nervousness and irritability [14, p. 107]. In addition one-third of all respondents also failed to take breaks (very) often because of too much work, and about 46 % of the total sample finished work later than intended always or often due to the amount of work. Thus, the employees lack the necessary regeneration from work in form of pre-planned rest periods. According to the Job Demands-Resources Model of Bakker and Demerouti [3, pp. 314, 2014: pp. 8] personal and work resources can act as a stress-reducing buffer for work requirements. A self-directed working time is such a factor to deal with job demands. For most professional drivers this resource is not available. Only about 25 % often have influence on the planning of their working hours. As might be expected in this context, 68 % of experienced and 47 % of drivers in training rated their workload as a high or very high—significantly correlated with time pressure ($r_s = 0.47$, $p \leq 0.001$), going without breaks ($r_s = 0.47$, $p \leq 0.001$) and delayed end of work ($r_s = 0.45$, $p \leq 0.001$).

In many interviews the unfavorable traffic and parking situation was cited as a further significant stress factor. Lack of parking for trucks and buses promotes exceedances of driving times and so affects recovery. According to experts, the alignment of the existing parking (cab to the motorway, lack of noise prevention, unfavorable parking lot lighting) contributes significantly to the lack of recovery and poor sleep quality.

The proportion of experienced drivers who declared that under current conditions they probably cannot drive until retirement added to 38 %. In addition to a connection to a high subjective work load ($r_s = 0.41$, $p \leq 0.001$) corresponding links are found in the data between the assessment of reaching retirement age in good health and aspects of the reconciliation of professional and private life. Drivers who often have problems with friends and family because of working hours, are also less likely to expect to drive up to retirement ($r_s = 0.40$, $p \leq 0.001$). If on the other hand their personal situation is taken into account in the organization of working time, they are also more optimistic to reach retirement still being professional drivers ($r_s = 0.40$, $p \leq 0.001$). In addition to the relationship between job demands and resources described above, an independent path between resources and (work) motivation exists [3, pp. 312]. This combination of resources and motivation is especially intensified by high job demands [3, p. 315] as could be found with professional drivers. Taking in mind the increased desire for compatibility a working time supporting the reconciliation of work and family can be an important motivational resource to try staying in the profession. At the same time stress-reducing effects can be assumed: Most drivers have little direct contact with colleagues. Exchange of experiences and peer support cannot be spontaneous, but only with organizational overhead. Also breaks are usually spent alone or with "unfamiliar" individuals [12, pp. 18]. Drivers therefore receive less help and support from colleagues as employees in other industries. This also might reduce the manageability, with negative impact on the salutogenisis [1]. They also less often experience themselves as part of a community at work [15, p. 83]. Therefore, friends and family are all the more important as a resource to reduce strain and maintaining employability. A third of the trainees also did not think to reach retirement age as a driver. Since these are just at the beginning of their professional career, this assessment is all the more alarming.

3.1.2 The (Visible) Results of Work

Depending on existing qualifications and operationally acquired specialized knowledge and skills, the experts currently saw an already high job security for professional drivers. Increasing recruitment problems due to increasing shortage of qualified drivers expand the guarantee, to find and keep a job even further. Aware of the ongoing shortage, the apprentices in particular not only positively highlighted the foreseeable job security positively in the open answers of the questionnaire. They also foresaw an increase in wages.

At the same time, the experts and drivers alike were unanimous in their assessment that the wages paid customary in the market are too low. To ensure lower transport costs, cutting wages has been used in the industry as cost reduction for a long time. As a result, wages have dropped steadily for years. With 43 %, most of the experienced indicated to earnings between 2000 € and 2499 € gross per month. Another 23 % assigned themselves to the next smaller income category from 1500 € to 1999 €. Especially in the light of the time demands identified above, the de facto paid hourly wage is put in perspective considerably. Both professionals and experts were of the opinion that the payment is not adequate to the displayed high-quality performance and the requirements. The experts saw an expression of lack of appreciation in a payment this low.

Both the surveyed apprentices as well as experienced professionals felt little valued in their work. Over half of the respondents indicated that their activity is met with little to no appreciation by the society in general. The experts interviewed saw the reason for this in the missing knowledge about the importance and necessity of driving professions in the general population. The driving professionals share this problem with other service professions, whose importance for society is at most visible when expected services are not provided. “For customers, it’s invisible, that one tries to be punctual—which is sometimes impossible because of the traffic”.[12] As portrayed by the media, particularly traffic news, trucks are seen as the cause of high street wear, crowded freeways and traffic jams by many parts of society. In the open answers drivers several times described a resulting assessment by others as “being fools”[13] as a disadvantage of their profession. Almost all interviewees stated that the driving professions have significantly lower prestige in comparison with other (neighboring) EU countries.

[12] “Für die Kunden ist es unsichtbar, das seiner versucht pünktlich an Ort und Stelle zu sein—was manchmal aus verkehrstechnischen Gründen einfach nicht geht”.

[13] “Trottel der Nation”.

The professional environment has more direct information to assess the quality of performance und services provided by the drivers. Therefore, the recognition of mastery and effort shown by drivers through their customers and their dispatchers appears all the more important. The respondents felt valued a bit more by their direct customers than by society in general. After all, almost a quarter got quite or very much appreciation from this direction. On the part of the dispatchers under a third of the drivers was met with quite or very much appreciation. More than a third, however, felt little or unappreciated by this clientele. The experienced appreciation can be described as another important resource within of the above-described job-demand resources model (Sect. 3.1.1). The path to motivation seems rather obvious but there seems to be a direct connection to strain reduction as well. Interesting in this context is the relationship between the appreciation by the customers and dispatchers and temporal work characteristics for the respondents. A low predictability of working time (schedule known in advance $r_s = -0.31$, $p \leq 0.001$; short term changes $r_s = -0.37$, $p \leq 0.001$) went hand in hand with a lower sense of appreciation on the part of these two groups. Employees who often were under time pressure ($r_s = -0.37$, $p \leq 0.001$), had to skip breaks ($r_s = -0.32$, $p \leq 0.001$) or had a belated end of work ($r_s = -0.36$, $p \leq 0.001$) were also less likely to feel valued by their clients and/or dispatchers. So respondents saw a link between a very specific appreciation of their work and a well-designed work schedule. Likewise the interviewed experts linked time pressure due to deadlines to a very job-related and not enough on the driver-oriented schedule. In a successful scheduling and a concomitant communication between drivers and dispatch, the experts saw an important resource for the health of those employees and their loyalty to the company.

3.2 Recommending the Own Driving Profession

With 55 %, slightly more than half of the drivers would not recommend their profession. Only a quarter of the experienced employees would recommend joining their respective driving profession. In contrast, a total of approximately 65 % of the trainees supported a recommendation.

Drivers with lower subjective workload are also more likely to recommend their profession ($r_s = -0.38$, $p \leq 0.001$). Highest correlations were found for the work-related temporal factors time pressure ($r_s = -0.33$, $p \leq 0.001$) and going without breaks because of a high workload ($r_s = -0.33$, $p \leq 0.001$). If these time requirements were perceived less frequently, the drivers would be more likely to advise prospective apprentices to choose their profession. To a lesser extent fewer hours actually worked ($r_s = 0.28$, $p \leq 0.001$) and seldom having problems with family or friends due to working hours ($r_s = 0.28$, $p \leq 0.001$) were linked to a recommendation.

As an aspect of the visible result of the work, the economic result is considered first, because in addition to the requirements the income was listed as a disadvantage in particular. Suggestions to become a driver and wages were associated ($r_s = 0.36$, $p \leq 0.001$), however the direction of this relationship seemed rather

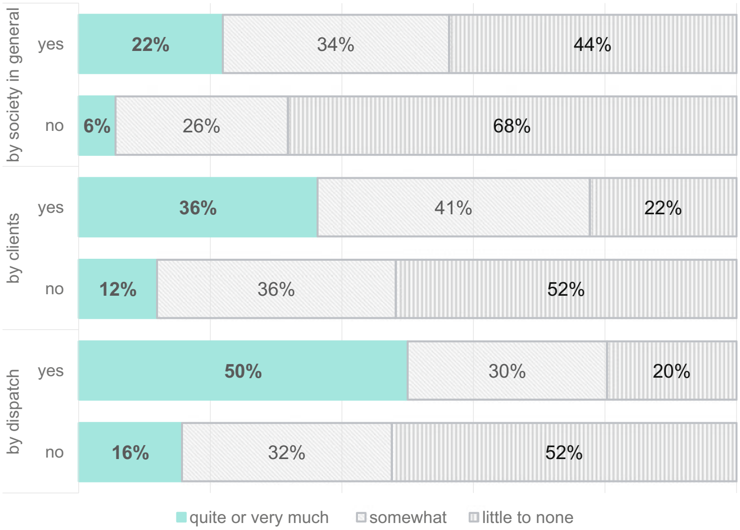


Fig. 3 Recommendation of the own driving profession depending on the perceived appreciation

odd. Positive recommendations were combined with lower wages. As this might be due to more trainees with lower wages compared to experienced drivers recommending their profession this was further analyzed as part of a multiple logistic regression described below. Furthermore, mainly those would vouch for their profession who see their services valued (Fig. 2) by their clients ($r_s = 0.37$, $p \leq 0.001$) and most pronounced dispatchers ($r_s = 0.40$, $p \leq 0.001$). The relationship between appreciation by society in general and a positive recommendation of the own driving profession was also substantial but definitively smaller ($r_s = 0.28$, $p \leq 0.001$) (Fig. 3).

A logistic regression was completed to determine the important factors for not recommending one's own driving profession. Based on the correlations results described above, three work-related temporal factors (pressure, failed breaks and trouble with friends and family due to working time), the actual working hours, all three appreciation indicators and the income were included in the analysis. Since the focus was placed on individual stress factors, the overall workload was not part of the regression. The model includes dummy variables for each temporal factor contrasted against the highest category (often to always).[14] A dummy variable for the actual weekly working hours was build based on five categories used by the BAuA [18] and compared to the second lowest (representing the most common contractual working hours). Appreciation was contrasted against the highest option (rather much to plenty) and income to the lowest (below 1000 € per month). Table 2 shows the resulting coefficients and corresponding odds ratios in model 1.

[14] The five point scales for temporal factors and appreciation used in the questionnaire were collapsed into three point scales in order to enhance the contrasts.

Despite being a workload, the hours actually worked per week and problems with family and friends based on working time are no suitable predictor of the direction of the professional recommendation. The esteem by customers and society in general were not good indicators likewise. The expectations on the part of the drivers for these are probably very low, so that the absence fails to have a great influence. Appreciation by the dispatcher on the other hand is clearly an appropriate indicator of the career choice. As the lack of appreciation increases the chance that the own profession is considered not advisable increases by nearly 200 % ($p < 0.05$, odds ratio: 2.96). Missed breaks as a consequence of too much work ($p < 0.001$, odds ratio: 0.20) and time pressure ($p < 0.05$, odds ratio: 0.28) also have a good predictive potential for the direction of the professional recommendation. For respondents located in the reference group (high frequencies of time-related workload), the chance of non-recommendation increased dramatically. Income had a significant association with the professional recommendation in the four highest categories as well. However, this link is rather counterproductive intuitively—for the higher wage categories, the chance of not-recommendation is increased compared to the lowest. There were no significant interaction terms (for purpose of clarity interactions were shown in the table only if significant).

As described above, the trainees, who largely belonged into the lowest income category, recommended their driving profession more frequently. Therefore, group of respondents (dichotomous trainees and professionals) was added as a factor in model 2 (Table 2). Time pressure ($p < 0.01$, odds ratio: 0.06), failed breaks ($p < 0.01$, odds ratio: 0.15) and appreciation by the dispatch ($p < 0.01$, odds ratio: 4.90) remained significant predictors in Model 2. The direction of the link described above remained, and was even more pronounced in size. As can be seen, income was no longer connected to willingness to recommend after controlling for group membership. So, although many respondents were not satisfied with their payment, the income was no significant factor for a recommendation of the driving profession. It is crucial, whether the respondent was either an experienced driver or a trainee ($p < 0.05$, odds ratio: 194.33), with an increased chance that the profession was not recommended by the experienced. This distinctive group difference may be due to several factors. Differences between the groups in the reasons of career choice may be important. According to the experts, a lot more of the trainees had actively chosen this profession. In this case, these mostly knew in advance about benefits and disadvantages, leaving less potential for disillusionment. Furthermore, in case of an active decision for the profession it is more beneficial for the self-image of students to look at one's own driving career as desirable. Another possible cause may be that the trainees are not subject to the workload to the same degree compared to experienced drivers. However, both groups differed little in their data according to the frequency of time pressure, missed breaks and family problems due to work. Contrary to the experienced drivers, however, the vast majority of trainees reported actual working hours between 30 and 40 h per week. It can be assumed that in the case of apprentices, compliance with the law is ensured more strongly, so this might be another important factor influencing the group difference. In this context the significant interaction between the actual working week and the

Table 2 Logistic regression of work-related temporal factors, actual weekly working hours, appreciation, income and the group questioned on intention not to recommend one's own driving profession to prospective trainees

	Model 1		Model 2	
Variable	B	Odds ratio	B	Odds ratio
Constant	−0.15		−0.81	
Time pressure (sometimes)	−0.78	0.46	−0.82	0.44
Time pressure (less often to seldom/ never)	−1.29*	0.28	−2.80**	0.06
Failed breaks (sometimes)	−0.27	0.77	−0.47	0.62
Failed breaks (less often to seldom/ never)	−1.61***	0.20	−1.93**	0.15
Trouble with family and friends due to working hours(sometimes)	−0.56	0.57	−0.71	0.49
Trouble with family and friends due to working hours (less often to seldom/ never)	−0.66	0.52	−0.74	0.48
Hours actually worked (up to 30 h)	1.46	4.29	2.99	19.98
Hours actually worked (41 h to 48 h)	−0.19	0.83	1.04	2.84
Hours actually worked (49 h to 60 h)	−0.27	0.76	0.15	1.16
Hours actually worked (60 h and more)	−0.32	0.72	0.65	1.92
Appreciation society in general (medium)	−0.09	0.91	−0.19	0.82
Appreciation society in general (low to very low/none)	0.94	2.55	0.78	2.18
Appreciation clients (medium)	0.08	1.08	0.20	1.22
Appreciation clients (low to very low/none)	0.63	1.88	0.97	2.65
Appreciation dispatchers (medium)	0.47	1.60	1.09*	2.97
Appreciation dispatchers (low to very low/none)	1.09*	2.96	1.59**	4.90
Income 1000 € to 1499 €	0.93	2.54	−0.97	0.38
Income 1500 € to 1999 €	2.46***	11.72	−0.49	0.61
Income 2000 € to 2499 €	2.05***	7.78	−0.79	0.45
Income 2500 € to 2999 €	1.98**	7.26	−0.96	0.38
Income 3000 € and above	2.66*	14.26	0.20	1.23
Experienced vs. apprentices			5.27*	194.33
experienced: Hours actually worked (up to 30 h)			4.47	87.01
experienced: Hours actually worked (41 h to 48 h)			−3.65**	0.03
experienced: Hours actually worked (49 h to 60 h)			−2.07	0.13
experienced: Hours actually worked (60 h and more)			−3.66**	0.03
Nagelkerke pseudo r-square	0.452		0.525	

(continued)

Table 2 (continued)

	Model 1		Model 2	
Variable	B	Odds ratio	B	Odds ratio
Chi-square	$\chi^2 = 118.97$, df = 21, p < .0001		$\chi^2 = 141.86$, df = 32, p < .0001	

Notes: The coefficients for the work-related temporal strains are contrasted with the highest frequency (often to always). The appreciation coefficients are contrasted with the high to very high category. The coefficients for actual weekly working hours are contrasted with the category (30–40 h) as this reflects the contractual hours rated mostly. The income coefficients are contrasted with the lowest category (below 1000 € gross per month). *p < .05; **p < .01; ***p < .001

group in the model 2 is interesting. Those who, despite their status as trainees, had actual weekly working hours between 41 and 48 ($p < 0.01$, odds ratio: 1.53) or more than 60 h ($p < 0.01$, odds ratio: 1.53) were even less inclined to recommend their profession to others in comparison with their experienced colleagues.

4 Conclusion

Above all it is the organizational and individual appreciation and work-related temporal factors that determine a (non-)recommendation of the profession. It appears that on account of the perceived appreciation it's not a "well done" and not basically a financial value, which causes a (non-)recommendation of ones driving profession. Rather, first and foremost it is the recognition of the demands and stresses of work and relevant support by employers and dispatchers. In light of the reported working time conditions and the esteem partly perceived as very low on part of the companies, an external image campaign appears ill-suited to attract potential trainees and career changers for the driving profession. At worst, such an image in retrospect might be perceived as a sham and lead to an increased dropout rate. Instead the support by the company as well as the dispatch is in demand. Caught between clients, profitability and drivers, especially dispatchers can only plan driver oriented, if the company management advocates this. Thus, a well-organized working time not only affects the workload and job satisfaction of current drivers to a great extent. It also has the potential to enhance the prospects to attract trainees based on the recommendations of the current drivers and such to prevent a shortage of skilled workers in the long run.

References

1. Antonovsky A (1997) Salutogenese. dgtv-Verlag, Tübingen
2. BAG (2012) Marktbeobachtung Güterverkehr – Auswertung der Arbeitsbedingungen in Güterverkehr und Logistik 2012. BAG Bund, Köln

3. Bakker AB, Demerouti E (2007) The Job Demands-Resources model: state of the art. J Manag Psychol 22(3):309–328
4. Bioly S, Klumpp M (2014) Statusanalyse der Rahmenbedingungen für Fahrberufe in Logistik und Verkehr. ild Schriftenreihe Logistikforschung, Bd. 39. MA Akademie, Essen
5. Boeckler impuls (2014) Fremdbestimmt ist ungesund. 3/2014. geprüft am 9. July 2014. http://goo.gl/E8BBKW
6. Eberhard V, Scholz S, Ulrich JG (2009) Image als Berufswahlkriterium. Bedeutung für Berufe mit Nachwuchsmangel. Bibb 3/2009, 9–13
7. FOM Hochschule für Oekonomie & Management (ed) (2014) Logistik und demografischer Wandel. Arbeitsmarkt, Arbeitsbedingungen, Entwicklungsrisiken und Handlungsoptionen der Fahrberufe in Deutschland. DVV Media Group, Hamburg
8. Fuchs J, Söhnlein D, Weber B (2011) Projektion des Arbeitskräfteangebots bis 2050: Rückgang und Alterung sind nicht mehr aufzuhalten. IAB-Kurzbericht 16/2011. Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg
9. Gläser J, Laudel G (2007) Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen, 3rd edn. Verlag für Sozialwissenschaften, Wiesbaden
10. Goesmann C (2010) Verbandliche Arbeit als Kampf um Anerkennung. Die Erhöhung der sozialen Wertschätzung für die Altenpflege als Ziel intermediärer Organisationen. Der Andere Verlag, Tönning
11. Hellert, Ulrike (2001) Humane Arbeitszeiten. Hagener Arbeiten zur Organisationspsychologie 1. Lit, Münster
12. Hess K (2010) Gestaltung mobiler Arbeit. In: Brandt C (Hrsg), Mobile Arbeit – Gute Arbeit? (S. 17–32). ver.di., Berlin
13. Hofmann-Lun I, Gaupp N, Lex T, Reißig B (2005) Hauptschülerinnen und Hauptschüler – engagiert, motiviert, flexibel? DJI Bull 73:16–18
14. Junghans G (2013) Termin- und Leistungsdruck. In: Lohmann-Haislah A (ed) Stressreport Deutschland 2012. Psychische Anforderungen, Ressourcen und Befinden. Dortmund, BAuA, pp 107–112
15. Lohmann-Haislah A (2013). Stressreport Deutschland 2012. Psychische Anforderungen, Ressourcen und Befinden. BAuA, Dortmund
16. Lohre D, Bernecker T, Stock W, Düsseldorf K (2012) ZF-Zukunftsstudie Fernfahrer – Der Mensch im Transport- und Logistikmarkt. EuroTransportMedia: Verlag, Friedrichshafen. http://goo.gl/Halp5O, 2. Mai 2014
17. Nachreiner R, Arlinghaus A, Bockelmann M, Bruchhagen A (2013) Gesundheit im Fahrdienst. Zwischenauswertung der Online-Umfragen. Gawo, Oldenburg. http://goo.gl/nTNXDp, 2. Mai 2014
18. Nöllenheidt C, Wittig P, Brenscheidt S (2014) Grundauswertung der BIBB/BAuA-Erwerbstätigenbefragung 2012. Vergleich zur Grundauswertung 2006. BAuA, Dortmund
19. Resch H (2012) Arbeitsverdichtung im Fahrdienst als Folge der Restrukturierung im ÖPNV. Arbeitspapier 2012. Hans Böckler Stiftung, Düsseldorf
20. Roth J, Schygulla M, Dürholt H, Nachreiner F, Pankonin C (2004) Betriebs- und Arbeitszeiten beim Gütertransport und bei der Personenbeförderung. BAuA, Dortmund
21. Semmer NK, Zapf D, Dunckel C (1998) Instrument zur stressbezogenen Tätigkeitsanalyse ISTA. In: Dunckel C (ed) Handbuch psychologischer Arbeitsanalyseverfahren. Zürich, Hochschulverlag, pp 179–204
22. Tomasik MJ, Heckhausen J (2006) Sozialprestige von Ausbildungsberufen aus der Sicht von Realschüler/-innen. Zeitschrift für Sozialpsychologie 37(4):259–273
23. TÜV Rheinland (2012) Aktuelle Studie: Fahrermangel bei LKW bedroht auch die Verkehrssicherheit. http://goo.gl/nk36L1. 4 Apr 2014

Analyzing Motivation-Enhancing Features in Work Orders

A Methodical Procedure for Analyzing Motivation-Enhancing Features in Written Work Orders

Ralf Kassirra and Herbert Rausch

Abstract

In order to find evidence of effective optimization of motivation enhancement within a task culture, a content analysis method is applied. This method takes the nature and extent of motivation-enhancing features into account. Based on the results of motivational research (e.g. the self-determination theory of motivation), an ordered category system of indicators has been developed. In the same manner as content analysis, a manual containing examples and operational descriptions was developed in order to analyze tasks using a combined theory-based deductive and inductive tasks-based approach. The main categories were expanded along with the motivation-enhancing features (e.g. task format, required cognitive operation, didactic intent). At present, written work orders in a select subject (Arbeit-Wirtschaft-Technik, a subject taught at Bavarian secondary schools) are coded and analyzed descriptively. This paper describes a systematic approach, the main categories and their selected indicators, selected results and their objectivity. Finally, methodological problems and limitations are discussed.

Keywords

Self-determination theory • Motivation-enhancing features • Work orders • Instructions • Content analysis • Interest theory • Work motivation • Mittelschule • Secondary school

R. Kassirra (✉) • H. Rausch
Lehrstuhl für Ergonomie, Technische Universität München, Munich, Germany
e-mail: kassirra@lfe.mw.tum.de; rausch@lfe.mw.tum.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_8

1 Superordinate Context

The presented approach to the analysis of work orders is embedded in a higher-level research process; in the process work aids are developed for teachers to improve the motivational design of written work orders. A survey was already carried out among teachers at Bavarian secondary schools (n = 150, 24 questions). The survey revealed facts concerning the practice of optimizing written work orders as it relates to motivation and the use of instructional media (e.g. textbooks, teacher's guides, printed collections of work orders for teachers etc.). This survey together with an analysis of existing tasks serve to highlight the prevailing culture in terms of motivational task-related features. In a subsequent step, the features that are not present or are only partially present within the dominant culture of written work orders are tested for motivational effects. On this basis, developed work aids such as design guidelines are ultimately tested and optimized. The research team selected written work orders in the field of Arbeit-Wirtschaft-Technik at grade level seven and eight at Bavarian secondary schools for the analysis. AWT is an important subject that is taught in secondary schools in Bavaria and is highly regarded due to its importance in preparing for professional and private life. Pupils learn about the importance of technology, business and work. A non-mathematical and non-scientific subject was chosen because intense research has been conducted in this area in the field of task culture and design of teaching, classroom settings and whole environments for several years now [see 7, 8, 13, 17]. Another argument in favor of this research is the imminent change in the curricula. As of 2017 new curricula will be gradually introduced for secondary schools in Bavaria over a 5-year period. One feature of the new curricula will be a shift towards descriptions of competencies. Predetermined subjects will no longer hold sway; teachers will therefore be given greater scope to design their own lessons. This, in turn, means that teachers will have to design written work orders if they are to teach autonomously. Guidelines for a motivation-enhancing design of work orders may be of use.

Finally, selected typical examples of written work orders found in textbooks suited to the eighth grade are provided:

– Just remember this schema of the three basic functions when answering the questions. It serves as a guide when learning about plants. (Wege zum Beruf 8, p. 6/2; [5])
– Individually crafted pieces of furniture are more expensive than customized pieces. Explain why this area of interior design is suited to higher-level needs. (Wege zum Beruf 8, p. 19/1; [5])
– A restaurant needs fresh produce daily. List the advantages and disadvantages of timely storage! (AWT aktuell 8, p. 13/2; [9])

2 Research Questions

Based on the results of the previous survey, the analysis of the features in work orders focuses on written work orders in the most widely used textbooks. Its main objective is to show how and the extent to which motivation-enhancing features are taken into account in AWT lessons and textbooks. These features are intended to contribute to the highest possible level of intrinsic motivation. In addition, correlations between significant changes in the written work orders and the year of task creation or between motivation-enhancing features, the openness of the results of the didactic intention or thematic integration are the subject of research. Accordingly, the following issues are primarily addressed in this study: What motivation-enhancing features are included in written work orders in AWT? To what extent are motivation-enhancing features written in the task settings of AWT?

3 Theoretical Background

The basis for the following reflections on the motivation-enhancing design of work orders are selected theories and studies that explain motivation and that often address the design of whole classroom settings [e.g. 10, 15, 17, 18]. The focus on the specific work order as a written set of instructions or a work instruction to motivate students in the context of practical teaching work is new in non-scientific and non-mathematical education research. As a result, the theoretical foundations presented here had to be applied to the situation of dealing with written work instructions and be used to derive writable motivation-enhancing characteristics in instructions. The self-determination theory posited by Deci and Ryan was extensively consulted [e.g. 2, 3]. Of equal significance was the classification created by Manfred Prenzel which pertains to vocational training [e.g. 15, 16]. Manfred Prenzel's appraisal provides significant support, in particular with regard to the motivation-enhancing categories in the manual used for the coding process (see Sect. 5).

3.1 Self-Determination Theory

With their self-determination theory [e.g. 2, 3], Deci and Ryan posit different qualitative characteristics of motivated action [see 3: p. 224]. Their theory is that the quality of motivation depends significantly on the extent of the experienced self-determination and the (anticipated) self-efficacy. They assume "[dass] sich motivierte Handlungen nach dem Grad ihrer Selbstbestimmung bzw. nach dem Ausmaß ihrer Kontrolliertheit unterscheiden lassen" [3: p. 225]. "When the environments become more controlling, children lose intrinsic motivation and self-esteem" [2: p. 270]. Motivation levels are differentiated: as the degree of self-determination increases, "external", "introjected", "identified" and "integrated regulation" are differentiated. With regard to motivation-enhancing tasks, it is essential that a "graduated" increase in the quality of motivation or readiness to act is accompanied by an increase in the degree of experienced self-determination. The motivation to perform well is greatest when control conditions are minimized and support for autonomy is optimized [see 3: p. 235]. The theory postulates three basic psychological needs: the need for self-determination, the need for self-efficacy and the need for social relatedness [see 10: p. 18], which are equally relevant to intrinsic and extrinsic motivation [see 3: p. 229]. The model also assumes that humans have the innate motivational tendency to feel connected to other people in a social setting, to feel effective in this environment and thereby feel personally autonomous and pro-active [see 3: p. 229]. The basic needs described can help to create an intentional design of objects or situations that are likely to instill intrinsic motivation. They provide evidence of factors that are responsible for instilling intrinsic motivation and extrinsic motivation development (ibid.). In conclusion, tasks can be more intrinsically motivating when they ensure the highest possible level of experienced self-determination, the experience of self-competence and social relatedness.

3.2 Other Relevant Models and Theories

In addition, the concepts formulated by Manfred Prenzel [15, 16], in particular, are used to systematize motivation-enhancing categories and develop individual indicators. In his studies he explored the conditions required for self-determined, motivated and interested learning. His systematization of these conditions on a much broader scale is important in terms of their application to written tasks or work orders. By reference to the self-determination theory, a theoretical framework can be deductively modeled for the classification of individual characteristics of written work orders. In addition, individual motivation-enhancing indicators are based on the "Job Characteristic Model" [6] or to a large extent on the theory of interest postulated by the Munich group whose members included Hans Schiefele, Andreas Krapp and Manfred Prenzel [e.g. 11, 12, 19]. Likewise, aspects of flow theory were considered, e.g. the suitability of the requirement level [e.g. 1].

4 Method

The method of content analysis [e.g. 4, 14] in the field of empirical social research forms the framework for the general approach (Table 1). Following Werner Früh [4] a basic system of categories which was initially deductive, i.e. purely theory-based, was developed. Individual subcategories were formed and, in some cases, initial indicators were described. In the second step the research team proceeded inductively. In other words we tested the system of categories using a number of written work orders. The work orders were not derived from the later sample. Other indicators to individual categories were found in this sample. A few additional

Table 1 Steps involved in the methodological approach

Planning	– Formulation of the research questions and hypotheses – Extracting the main categories
Developing	– Setting the sample, the analysis unit and the context unit – Deductive development of categories (based on theory) – Inductive development of categories (based on testing) – Operational description and examples – Development of a table to transfer data
Testing	– Training of coders – Sample coding – Validity and reliability tests – Clarifying the operational description
Using	– Coding the sample – Meetings with the encoders (approximation) – Checking and validating data – Validity and reliability tests
Evaluation	– Transferring the data to SPSS – Presentation and interpretation

Based on Früh 2004 and adapted

subcategories were also found. Descriptions of each category were written for all levels and characteristic examples were collected and displayed. A manual for the future work of the coder was thereby gradually created based on a series of sample runs and clarifications. The advantage of this combined deductive and inductive approach is that an analysis result can show whether all types of indicators exist in reality as well as those that can be derived from the theory but were not used. In this way you can see what is but you can also get an idea of what could be.

4.1 Deductive Development of Categories and Manual

After specifying the research questions (see Sect. 3) I set the sample, the analysis unit and the context unit. The sample comprises the written work orders in the four most widely used textbooks for seventh and eighth grade at secondary schools in Bavaria. The aforementioned survey of teachers provided the basis for deciding on the four textbooks to be used in the study. The analysis unit was the complete written work order. Written work orders can either be embedded in the text or printed outside the text and immediately recognizable as a work order. To assess the characteristics of a work order the coders used the entire previous contents of a chapter that deals with the same topic (context unit).

A schema of five main categories was then developed based on theoretical foundations (Table 2): format of solution (open, semi-open, closed), required cognitive operation, didactic intention, motivation-enhancing features and formal characteristics (e.g. year of publication, curriculum reference, type of publication,

Table 2 Theory-based categories of the manual

Category	Subcategories
Format of solution	Open, semi-open, closed
Required cognitive operation	e.g. reproduction/recalling, transfer, problem-solving
Didactic intention	e.g. elaboration, repeating, testing
Motivation-enhancing features	e.g. perceived support for self-determination, perceived support for competence, perceived interest in the content prepared the author of the work order
Formal characteristics	e.g. year of publication, curriculum reference, type of publication

etc.). These categories were divided into further subcategories filled directly with appropriate indicators (at the level of writable characteristics).

The question as to the theoretical feasibility of each category always had to be considered, e.g. Manfred Prenzel's categories of motivation-enhancing features had to be selected and reduced because they were based on the design of entire learning environments. One question always had to be considered: What motivation-enhancing features can be integrated into written work orders? The decisive criterion for selection was the ease with which they could be incorporated into written work orders. I used this criterion to analyze a whole series of usable results from research into motivation in learning environments and classroom settings [e.g. 8, 10, 15, 17, 18].

These categories were presented with the first versions of operational descriptions and hand-crafted examples of work orders in a manual that helps coders understand and detect subtle differences. The numerical codes of each indicator are also included in the manual. The codes help to convert features or characteristics of work orders into data. Excel was to be used for data transfer as it is already installed on most computers and the research team could safely assume that subsequent encoders were also familiar with this program. Thus, I had to design a simple table for data transfer to Excel.

4.2 Inductive Category Formation and the Training of Coders

In the course of the first rounds of the sample coding (five test runs in total) of written work orders outside the sample, indicators were added to each category, partly grouped together or separated, operational descriptions were refined and illustrated using examples until such time as reliability and validity tests yielded satisfactory results (see chapter "Time Management Requirements for Holistic Working Time Organisation"). The manual derived from the initially deductive and subsequently inductive process of categorizing includes 52 individual categories at the lowest level, each of them with three to ten different indicators. By way of example, subcategories of motivation-enhancing features with their respective number of individual indicators (Table 3), largely based on the concepts developed by Manfred Prenzel [15, 16] are shown in the following:

Table 3 Subcategories of "motivation-enhancing features" and number of indicators

D	Motivation-enhancing feature	Number of indicators
D.1	Perceived support for self-determination	24
D.2	Perceived support for competence	18
D.3	Perceived social relatedness	3
D.4	Perceived interest in the content developed by the author of the work order	3
D.5	Perceived relevance of the learning material/of the action	19
D.6	Quality of the instruction	21

The indicators were gradually supplemented; I added examples of tasks in schoolbooks suited to other grades. The first two test runs were carried out by coders, both of whom would be subsequently barred from working with the sample. Reliability tests (formula of Holsti; see [4: p. 179]) and interviews conducted with the coders clearly showed those operational descriptions still requiring clarification. Once the reliability scores achieved were around the 0.75 mark, coders were selected to work with the sample. Two trainee teachers were hired as new coders. The main reason for doing so was that it is nearly impossible to recruit teachers for this type of work. It would entail higher costs and qualified teachers are not always available for such a long period. Another alternative would have to been to engage the help of pupils, but the complexity of the manual and the limited time available militated against such an option. Accordingly, the two female coders familiarized themselves with the manual, the terms of reference and the table for data transfer (Excel table) during three training sessions. To that end, a sample of work orders was encoded by both. Different decisions were discussed in detail at each session. Based on these discussions, the manual was expanded and adapted to the capabilities of the coders. In the end the manual ran to more than 30 pages. The manual covers all the basic rules of coding operation, information on sample size, analysis unit and context unit as well as all categories filled with subcategories at all levels right down to the indicators. The indicator sequence in the manual is strictly consistent with the sequence in the Excel table to enable the coder to work as quickly as possible. For the same reason operational descriptions and examples were to be as short as possible but also as precise as possible. Once the validity and reliability tests had exceeded 0.75 in more than 80 % of the indicators, the research team started to code the sample. Selected work orders were occasionally coded by the two coders simultaneously. It soon became apparent that the validity and reliability scores of certain indicators were still unsatisfactory. The list of results for the tests aids an appraisal of the results of each category (see chapter "Time Management Requirements for Holistic Working Time Organisation").

The manual lists more than 51 indicators which means that more than 51 decisions must be made during coding for each work order. In order to provide an idea of the scope manual and the category system, I have presented a few selected categories at different levels belonging to the main category of

Table 4 Subcategories of "perceived support for competence"

	Subcategory	Number of indicators
D.2.1	Feedback	6
D.2.2	Sense of discovery/discovery of new things	4
D.2.3	Competitive situation/scope for comparison with others	2
D.2.4	Orientation of individual reference standard	4

Table 5 Selected indicators of the subcategory "feedback"

D.2.1.1	Concrete indication of revisions (in text, pictures etc.)
D.2.1.2	. . .
D.2.1.5	Possibility of review by revisions (in text, pictures etc.)
D.2.1.6	Call for revisions
. . .	

Table 6 The indicator "concrete indication of revisions"

D.2.1.1	Concrete indication of revisions
Operational description	A warning is issued to check the quality of your work order-based process, if necessary by reading in text or representations
Example	"[. . .]If you are unsure, take another look at the chart on page 12!"

motivation-enhancing features and the subcategory of perceived support for competence (see Table 4). Based on the theoretical background, the research team proceed on the assumption that the perceived support for competence is contingent on individual feedback, the feeling of discovering new things, the scope for comparison with others or the orientation of an individual reference standard.

The last level of subcategories is represented by various indicators. The indicators are descriptive features of written work orders that can be clearly identified by the coders. Based on the theoretical background, the coders and the author described and assigned features likely to have a motivation-enhancing effect. Thus, the subcategory "feedback" is described for example by the indicators "possibility of review by revisions", "concrete indication of revisions" or "call for revisions" (see Table 5).

"Possibility of review by revisions" means that you can check the result of your work order-based processing by reading texts in the book, or using images and graphics at any time. In the case of a "concrete indication of revisions" the work order clearly specifies that this can be done (see Table 6) and a "call for revisions" means that one should check the quality of one's own results by reading information in the textbook. The information can be represented as text, picture, diagram or chart on the same page as the work order or even on the previous page.

The last number in the indicator code is the number the coders must code for the specified indicator. In the example provided they have to award a "1" if they think it constitutes a "concrete indication of revisions". If a specified indicator belonging to

Table 7 Agenda of coder meetings (fixed sequence of coder meetings)

1	Transfer of new data
2	Update of the working time accounts
3	Discussion of reliability and validity tests
4	Discussion of decision-making problems
5	Agreement on and explanation of new work orders
6	Date to be scheduled for the next meeting

a lower level subcategory is not applicable, they are required to enter a "0" in the row of the written work order.

Modification of the manual during coding of the sample is prohibited in order to create a stable coding system. The author discusses the results with the two coders at regular monthly meetings. The coders are instructed to record all decision-making problems as precisely as possible. A clear exchange of ideas helps to ensure a minimum level of reliability and validity. In addition, small samples are coded again simultaneously. All decisions and their settlement are recorded accurately and must be considered in the subsequent analysis and interpretation of data. In addition, a fixed sequence of all meetings with the coders is recorded to ensure quality (see Table 7). The first item on the agenda was the transfer of new data. New data was transferred and saved on the computer and server at the Institute of Ergonomics. After updating the working time accounts, the research team discussed the results of the previous tests of reliability and validity, and whether we had previously coded identical sample selections. These discussions often last more than 2 h due to serious differences of opinion. Where differences of opinions could not be set aside, guidelines for future decisions were agreed or different points of view were respected and ascribed potential significance. In addition, the coders and the author often chose to encode individual data series or features once more. The research team then discussed the listed decision-making problems by reference to concrete examples. It was extremely important that the discussion be initiated by the two coders to see whether they had a similar conceptual understanding. Some of these discussions also led to modifications in the manual and repeat coding. The next step involved agreeing on and explaining new work orders that had to be completed before the next meeting. Finally, the research team scheduled a suitable date for the next meeting.

5 Results

5.1 The Sample

At present (May 2015), 438 written work orders in the textbooks "Wege zum Beruf 8" (Bildungshaus Schulbuchverlag), "Praxis 8" (Westermann Schulbuchverlag) and "AWT aktuell 8" (Oldenbourg Schulbuchverlag) have already been converted into data format by our coders. The sections of these textbooks that have already been coded are uneven due to the ongoing coding operation (see Table 8). Our aim

Table 8 Current categorization of coded work orders by textbook (May 2015)

Textbook: title (publishing house)	Number of coded work orders	Proportion in the sample (%)
Wege zum Beruf 8 (Bildungshaus)	182	41.6
Praxis 8 (Westermann)	113	25.8
AWT aktuell 8 (Oldenbourg)	143	32.6
Arbeit-Wirtschaft-Technik 8 (Ernst Klett)	0	0
Total	438	100

is to conduct a full survey of four textbooks for the eighth grade and code each fifth written work order in four textbooks for the seventh grade. As a result, the research team will have to encode more than 350 additional written work orders. It will also be necessary to have a look at the textbooks for the seventh grade as they are produced by a different team of authors. Therefore, the author will be able to see whether the design of written work orders depends on the team of authors or not. Since this selected sample is limited, the results presented are not applicable to all textbooks. Nevertheless, certain interesting trends are already apparent.

5.2 Selected Results

All the results presented here achieved satisfactory results in all three reliability tests; in other words, a reliability of more than 0.7 for the particular subcategory. Some features were never identified by our coders or rarely appear in the sample of 438 written work orders in the three different textbooks. Our coders never came across written work orders that allowed a subtask or task to be selected. This means that all written orders and subtasks must always be processed in full.

Similarly, the attractive effects of processing were not hinted at in written work orders in 98.6 % of the sample. The written work order only hinted at attractive consequences for the editors in three instances, e.g. "...if you solve this problem, you will be able to produce ice by yourself." The work order stated the intended learning outcomes in two instances (0.5 % of the actual sample).

Generally speaking, new insights were not obtained from studying a work order. In other words, new knowledge or new information was not gained by editing a written work order within the previous sample. A typical example is a survey on the reasons for a company closure in the vicinity of the school.

Only 3.5 % of the written work orders analyzed require a comparison with others or creates playful competition. In other words, pupils are rarely required to present the own results or to compare their work with that of other pupils.

The coding of the indicators in the subcategory "perceived support for self-determination" creates the impression that there are often other choices, but explicit reference to such choices is rare. 83.7 % of the work orders generally provide the option to choose the social situation for implementation of the work order. Thus,

Table 9 Results of the subcategory D.2.1 feedback

Indicators in the subcategory "feedback"		Number of coded work orders	Coded work orders in percent (%)
D.2.1.1	Concrete indication of revisions (in text, pictures etc.)	0	0
D.2.1.2	Call for assessment by others	1	0.2
D.2.1.3	Clear criteria for success	1	0.2
D.2.1.4	Available control data (e.g. interim results, solutions etc.)	0	0
D.2.1.5	Possibility of review by revisions (in text, pictures etc.)	130	30.3
D.2.1.6	Call for revisions	0	0
D.2.1.7	Another form	1	0.2
D.2.1.8	Several indicators	1	0.2
D.2.1.0	No possibility of feedback available	295	68.8
Total		429	100.0

pupils can decide whether they will work alone, in pairs or in larger groups, but they are never prompted to do so. At least 8.6 % of the work orders provide a number of different social forms based on the wording, e.g. "Consider working with your neighbor or alone!".

The coding of the subcategory "feedback" provides comparably interesting results (see Table 9). The described feedback options are offered in 31.2 % of the analyzed written work orders. In other words, 68.8 % of the work orders do not provide an opportunity to check own results by feedback. In 30.3 % of cases, our coders found the indicator "possibility of review by revisions". However, no work order gives a concrete indication of revisions or a call for revisions. Only 0.2 % (in other words only one work order per indicator) of the work orders requests assessment by others, presents clear criteria for success (i.e. the features of a good solution are described) or includes more of the indicators described.

The substantive interest of the author is recognizable in only one written work order. In other words, 99.8 % of the work orders are not worded in such a way that would suggest that the corresponding work is interesting.

There are striking differences between the various textbooks in terms of the format of solutions. For example, 73 % of the coded work orders in "AWT aktuell" are assignable to the subcategory open tasks. However, only 20 % of the work orders in "Wege zum Beruf" are coded as open tasks ($n = 145$). "Open tasks" denotes complete openness as regards the scope of the solution. Thus, a defined plurality of different solutions does not exist. There can be no binding solution; there is no clear right or wrong way.

More detailed analyzes follow completion of the coding process for the total sample. Other differences between the textbooks can then be ironed out.

6 Discussion

6.1 Reliability and Validity of Codings

Chapter "Beneficial Effects of Servant Leadership on Short- and Long-Term Indicators of Employees' Psychological Health" described persistent problems associated with reliability and validity in certain categories. This will be expanded by referring to a concrete example. In repeated two-time codings the coders repeatedly produced poor reliability and validity scores (formula of Holsti; see [4: p. 179]) for several subcategories in over 25 work orders when coding the sample. One of these problematic subcategories was "required cognitive operation" (category B), which is further broken down into the subcategories (at indicator level) coding, recall, reorganization, transfer, reflection and problem-solving. The category is intended to specify the cognitive operations required. Initial reliability scores were always less than 0.7 in this category. Therefore, the coders and the author discussed the differences in the coding at regular meetings. The first step the research team took was an attempt to achieve a higher reliability score by modifying the manual. We sought to refine the operational descriptions in this category and in 11 other categories. Following that, the research team needed to recode the entire sample for the modified categories. The result for "required cognitive operation" was not much better so we decided to merge some of the subcategories. The author did this because the coders had the impression that descriptions of the various subcategories were not selective. A discussion of individual examples confirmed our suspicions. Hence the research team formed new subcategories as follows: coding (1), recall and reorganization (2), transfer and reflection (3) and problem-solving (4). The Author and the coders applied the same method to the sub-categories "A.2 semi-open format of solution", "D.1.1.4 choice of equipment" and "D.5.2 original approach". Validity was checked by comparing those results recorded by the experimenter and those results recorded by the coders. Result scores should be more than 0.7 for reliability and validity if the results are to form the basis for further interpretation and research.

6.2 Empirical Evidence

In order to make reliable statements about the actual motivating effect on pupils, all the indicators described would have had to be tested. This requires an enormous amount of time and manpower. The author did not have the resources required for such a task. That is why the researcher chose a more inductive approach to this stage of the research. First, the author wanted to determine which potential motivation-enhancing features are seldom or never taken into account. This then presents an opportunity to determine those features with the potential to make a big impact requiring little effort. Those features that are not taken into account should be selectively tested in a next step.

6.3 Transferability to Other Contexts

There are many other contexts unrelated to the field of academia that might make use of a manual for analyzing motivation-enhancing features in work orders. In principle, all those fields in which people are assigned written as well as oral work orders or instructions could benefit from the use of a manual. Similar situations are found in a business context, for example. The manual can be used to determine when employees are instructed in the handling of new machines or issued guidelines for staying healthy. Furthermore, work orders are issued in training, retraining and continuing education. Even one's private life can be determined by written orders and guidelines for action. In addition, instructions for using new products should be designed in a motivation-enhancing way.

7 Next Steps

Once coding has been completed and evaluated in accordance with the research questions, the author will use the results to organize the testing of selected features. To that end, tasks will be designed using those features of the manual which have yet to be seen to work in practice the author can then determine which features actually affect student motivation. Based on these results, the research team would then focus on the features of the categories "perceived support for competence" and "perceived support for self-determination". This would be followed by an attempt to develop guidelines for designing work orders for the subject "Arbeit-Wirtschaft-Technik" which also be applied to other subjects. However, this analysis is only an initial step in the research of the motivation-enhancing design of written work orders. In my opinion, the resulting manual with its concisely described indicators could be considered a valuable synopsis of many factors and features with the potential to significantly increase motivation in the context of work orders. The effect of the various features would then have to be verified progressively by empirical means. In conclusion, a culture of work orders could be created that not only supports improved academic performance but also ensures that our children really like attending school.

References

1. Csikszentmihalyi M (1993) Das Flow-Erlebnis. Jenseits von Angst und Langeweile: im Tun aufgehen. 5. Aufl. Klett-Cotta (Konzepte der Humanwissenschaften), Stuttgart
2. Deci EL, Ryan RM (1985) Intrinsic motivation and self-determination in human behavior. Plenum Press, New York, NY
3. Deci EL, Ryan RM (1993) Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. Zeitschrift für Pädagogik 39(2):223–238
4. Früh W (2004) Inhaltsanalyse, 5th edn. UVK Verlagsgesellschaft, Konstanz
5. Frauenknecht et al (2007) Wege zum Beruf 8. Bildungshaus Schulbuchverlage, Braunschweig

6. Hackman JR, Oldham GR (1975) Development of the job diagnostic survey. J Appl Psychol 60:159–170
7. Stadler M (2009) Modul 1: Weiterentwicklung der Aufgabenkultur. In: Prenzel M, Friedrich A, Stadler M (Hg) Von Sinus lernen. Wie Unterrichtsentwicklung gelingt. 1. Aufl. Klett, Kallmeyer, Seelze-Velber, pp 18–22
8. Kiemer K, Seidel T, Gröschner A, Pehmer A-K (2014) Effects of a productive classroom discourse intervention on students' motivation to learn mathematics and science. Learn Instr 35:94–103
9. Klose et al (2007) AWT aktuell. Oldenbourg Schulbuchverlag, München
10. Kramer K (2002) Die Förderung von motivationsunterstützendem Unterricht. Ansatzpunkte und Barrieren. Dissertation. Christian-Albrechts-Universität, Kiel. Philosophische Fakultät
11. Krapp A (1992) Das Interessenkonstrukt. Bestimmungsmerkmale der Interessenhandlung und des individuellen Interesses aus der Sicht einer Person-Gegenstands-Konzeption. In: Krapp A, Prenzel M (Hg) Interesse, Lernen, Leistung. Neuere Ansätze der pädagogisch-psychologischen Interessenforschung. Aschendorffsche Verlagsbuchhandlung (Arbeiten zur sozialwissenschaftlichen Psychologie, 26), Münster, pp 297–330
12. Krapp A (1998) Entwicklung und Förderung von Interessen im Unterricht. Psychologie in Erziehung und Unterricht 44:185–201
13. Krebs I (2009) Wie kooperieren Lehrkräfte? In: Prenzel M, Friedrich A, Stadler M (Hg) Von Sinus lernen. Wie Unterrichtsentwicklung gelingt. 1. Aufl. Klett, Kallmeyer, Seelze-Velber, pp 160–163
14. Mayring P (2010) Qualitative Inhaltsanalyse. Grundlagen und Techniken, 11th edn. Beltz, Weinheim
15. Prenzel M, Drechsel B (1996) Ein Jahr kaufmännische Erstausbildung. Veränderungen in Lernmotivation und Interesse. In: Unterrichtswissenschaft. Zeitschrift für Lernforschung 3, pp 217–234
16. Prenzel M, Drechsel B, Kliewe A, Kramer K, Röber N (2000) Lernmotivation in der Aus- und Weiterbildung: Merkmale und Bedingungen. In: Harteis C et al Kompendium Weiterbildung – Aspekte und Perspektiven betrieblicher Personal- und Organisationsentwicklung. Leske + Budrich, Opladen
17. Prenzel M, Friedrich A, Stadler M (Hg) (2009) Von Sinus lernen. Wie Unterrichtsentwicklung gelingt. 1. Aufl. Klett, Kallmeyer, Seelze-Velber
18. Röder B (2009) Selbstwirksamkeitsförderung durch Motivierung von Schülern. Dissertation. Freie Universität Berlin, Berlin. Fachbereich Erziehungswissenschaft und Psychologie
19. Schiefele U, Wild K-P (eds) (2000) Interesse und Lernmotivation. Untersuchungen zu Entwicklung, Förderung und Wirkung. Waxmann, Münster

Disability for Service in Public Transport Operations: Risk Factors and Interventions

Martina Bockelmann, Anna Arlinghaus, and Friedhelm Nachreiner

Abstract

In comparison with other jobs public transport drivers show an increased risk in job related health problems. The consequences are frequent and long absences from work as well as temporary and permanent disabilities for driving services. In order to collect evidence about the current state of the prevalence and the conditions contributing to or preventing this disability for service we conducted an online survey with active and former drivers in the Federal Republic of Germany. Data were analyzed using different kinds of survival analysis techniques, with temporary or permanent disability as the event and working conditions and preventive interventions as covariates. While no substantial evidence was found for the long-term effectiveness of person-oriented interventions, interventions addressing the work load of the drivers directly by improving their working conditions seem to be more effective, supporting the importance of structural prevention over behavioral prevention.

Keywords

Shift work • Fitness for duty • Life time exposure • Survival • Hazards

M. Bockelmann • F. Nachreiner (✉)
GAWO e.V., Oldenburg, Germany
e-mail: friedhelm.nachreiner@gawo-ev.de

A. Arlinghaus
GAWO e.V., Oldenburg, Germany

Ximes, Vienna, Austria

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_9

1 Problem

Driving in public transport operations is among others characterized by irregular work hours, hypokinesia and restricted opportunities for social interaction. Increasing economic pressures on the public transport companies, as well as increased traffic density and passenger volumes over the last years have been resulting in higher work load, work intensity, work stress and work strain for the drivers. The consequences of this increasing work load for the drivers are elevated risks of health and social impairments [3–7], high absence rates in comparison with other jobs [e.g. 9] as well as **t**emporary (TDS) or **p**ermanent (PDS) **d**isability for driving **s**ervices, as diagnosed by an occupational physician [1–3, 8]. In order to maintain health and driving ability of their drivers, public transport companies have been and are introducing different kinds of structural and/or behavioral preventive interventions.

One problem, however, is the lacking evidence about the current state of the prevalence of especially temporary or permanent disability for service in the Federal Republic of Germany (FRG). This makes it impossible to determine neither the actual risk of either TDS or PDS, the factors contributing to the risk, nor the effectiveness and efficiency of preventive interventions. Therefore, the aims of this study were to (a) estimate the present risks of temporary and permanent disabilities for service, i.e. how often temporary (reversible) or permanent disabilities for service occur in companies providing public local transport operations in Germany, and (b) which factors or conditions are related to these risks, in order to develop and suggest effective and efficient means for the prevention of such disabilities across the working lives of the drivers and thus to prevent any health based early retirements or job losses. This becomes especially relevant with a view to the increasing retirement age in the FRG, and thus to extended exposure towards occupational work stress. The results should therefore also be of a more general interest with regard to the extension of life time working time, beyond the occupational group of public transport drivers.

2 Methods

Data were collected by an online survey addressed to current and former drivers—including bus, tram and underground/subway drivers—occupied in or retired from local transport operations in the FRG. After a short pre-test the survey was accessible for drivers from May until November 2013. The link resp. web address to the survey was distributed via public transport companies, their management and work councils, as well as via relevant internet blogs and homepages. The survey contained questions on

– demographic characteristics, e.g. gender, age, family status
– biographical data, e.g. date of entering driving services
– job design, e.g. operational area, design of working conditions, work hours, driving hours, rest breaks, shift system

– information on the occurrence of a TDS or PDS, e.g. year of the event(s), diagnosis
– information on health related absence periods, e.g. frequency and duration during the last year
– participation in several preventive interventions, e.g. kind and point in time

In total 1419 (German speaking) drivers took part in this study. The proportion of those either retired or no more driving due to a PDS and/or now working in other parts of the companies was 12.6 % (n = 179). The sample consisted of n = 1247 male (=87.9 %) and n = 172 female (=12.1 %) participants. The mean age was ca. 48 years (active drivers: 47.1, former drivers: 53.6 years), indicating a rather old population, which is in agreement with data from the companies or the workers' compensation board.

Besides descriptive analyses, univariate analyses and various regression analytic procedures including survival analyses were conducted. Survival analyses were used in order to deliver information on any time related structures, i.e. the development of TDS and PDS based on years of driving service, indicated by time to failure (i.e. occurrence of the event). Effects for different working conditions were tested and estimated using Kaplan-Meyer analyses or Cox-regression analyses via group comparisons (e.g. drivers with vs. without participation in preventive interventions or drivers with vs. without ergonomic design of driver seats), while at the same time controlling for potential confounders, such as age, which in general seemed to provide a buffering effect.

3 Results

3.1 Prevalence and Diagnoses of Temporary or Permanent Disabilities for Service

Figure 1 shows the prevalence of TDS and PDS in the sample of this study. As can be seen from Fig. 1, 68.3 %, and thus roughly 2/3 of the sample, never experienced such a disability. 26.6 %, however, experienced one or more TDS, 2.1 % both one or more TDS and a PDS, and 3 % a PDS.

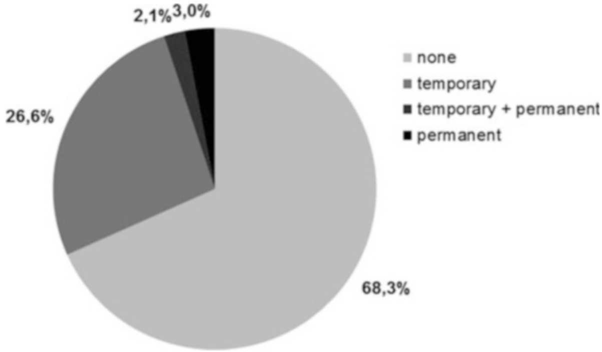


Fig. 1 Distribution of temporary and permanent disabilities for service

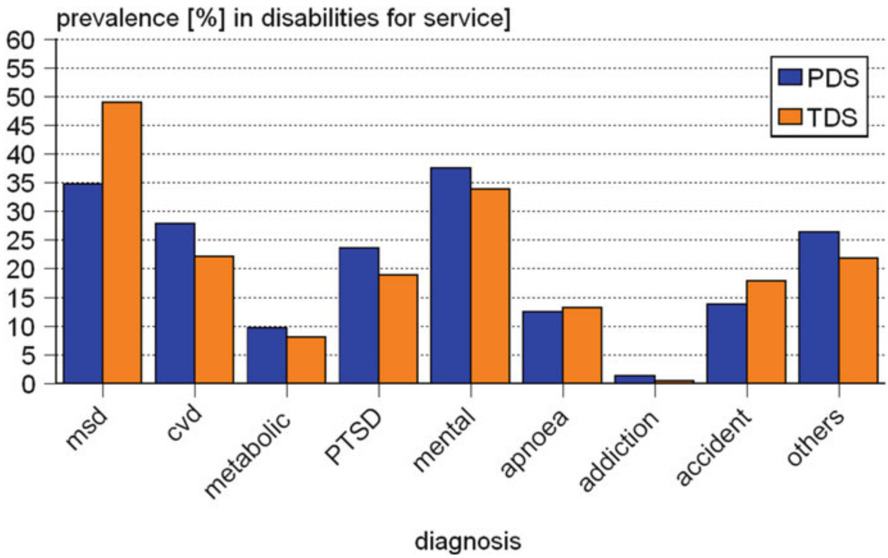


Fig. 2 Diagnoses for disabilities for driving [msd = musculo-skeletal disorders, cvd = cardio-vascular desease, PTSD = posttraumatic disorders]

In general this would mean that the risk of a PDS is rather low, with $p \approx .05$. On the other hand the risk of a TDS with $p \approx .29$ is rather high, indicating that ca. 1/4 of the drivers experience such a TDS once or more in their working life, indicating the above mentioned high risk of health effects for public transport drivers.

The main reasons for being diagnosed as disabled for driving services are shown in Fig. 2, i.e. mental health problems, musculo-skeletal problems and cardio-vascular diseases. This finding corresponds to earlier studies, but shows an increase in the importance of mental health problems. It seems further remarkable that addiction obviously plays a minor role in the development of a disability for service, neither for TDS nor for PDS.

3.2 Absenteeism from Work

Also in accordance with earlier findings the results concerning absenteeism in this study show frequent and especially long periods of illness absence, as can be seen from Fig. 3. The mean of the number of lost days due to health problems is 23.3 days per driver in the last 12 months before the study, which is definitely higher than in most other job categories [e.g. 9]. Nearly one third of the drivers (=31.2 %) reported very long absence periods (>20 days during the past 12 months) which can be regarded as an indicator of severely impaired health among these drivers.

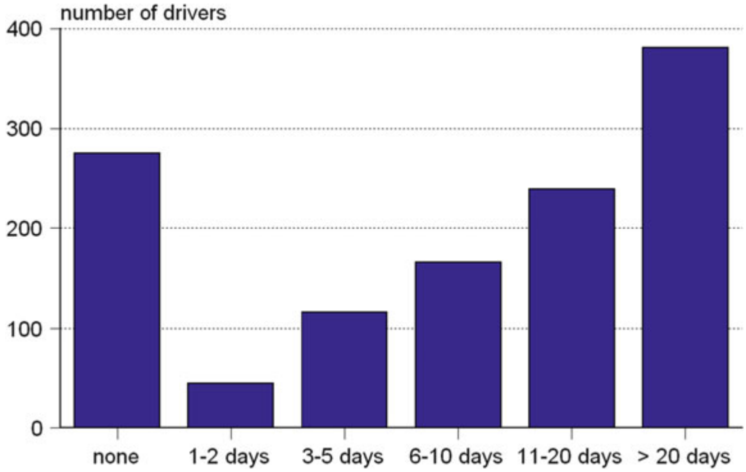


Fig. 3 Days lost due to health impairments

3.3 Time Related Aspects of the Development of Disabilities for Service

Figure 4 shows the survival of drivers without a (first) TDS or a PDS. While in the beginning of the working life all drivers are healthy, TDS and PDS begin to set in gradually and decrease the proportion of healthy individuals over time. As can be seen from Fig. 4, TDS starts earlier in the careers of the drivers, with a nearly linear trend over time of service which ends roughly after 35 years of service, leaving about 72 % of the drivers without any temporary disability. PDS starts later (after about 10–15 years of service) and continues less steeply until after about 30 years of service (with the exception of some few cases) when about 10 % have experienced a PDS. So the risk of a PDS shows a slower increase and a lower extent than that for a TDS. In general it would seem acceptable to assume that the risk of disability for service increases in a non-linear fashion after about 15 years of service and reaches its maximum after about 30 years.

3.4 Factors Affecting the Development of the Time to Failure

Investigating the effects of various working conditions and preventive interventions shows that positive effects can be found in conditions/interventions which address the actual work load of the driver. Several Kaplan-Meier-Analyses indicate that drivers, whose working conditions and work hours were ergonomically designed, could work longer before the first TDS or a PDS was diagnosed than drivers without ergonomic working conditions. Positive effects on the development of the cumulated hazard rates have been found, for example, for an ergonomic design of the driver seats or the air conditioning of the drivers' work areas and control over

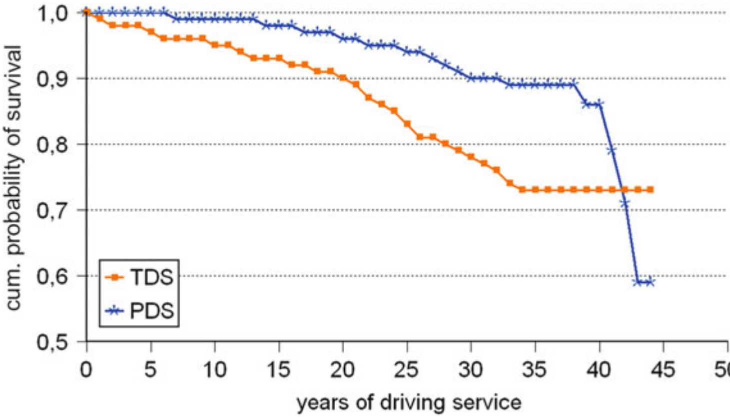


Fig. 4 Cumulated probability of surviving without temporary or permanent disability for service

working hours by drivers or schedules without split shifts. Figure 5 shows the results for ergonomically designed driver seats.

In the group of drivers without seats following ergonomic design principles, the risk of a PDS increases earlier, i.e. after only 12 years of service, as opposed to drivers whose seats uniformly follow ergonomic recommendations. This group shows a rather low risk until ca. 40 years of service. The difference in the mean times to failure between both groups amounts to 5.14 years, which can be regarded as a rather substantial effect. When interpreting these figures, however, it should be kept in mind that beyond 35–40 years into service the numbers of drivers with and without PDS as a matter of fact decrease quite substantially (at least in part due to reaching the normal retirement age) so that these estimates (beyond 40 years of service) must become less reliable and thus should be treated with care.

Looking at the effects of the direction of shift rotation on the risk of a TDS shows that also temporal conditions of work have an impact, as can be seen from Fig. 6. This figure shows that a forward rotation of the shift schedule is associated with a lower risk and which increases later in the career; also here the difference in mean times to failure is substantial, in this case 6.05 years between forward rotating systems and permanent (or so called stabilized) shifts.

In general, we found positive effects for structural interventions, e.g. air conditioning of the driver area in the vehicle, work schedules following ergonomic principles, participation in the design of shift schedules, and the absence of split shifts. It should be mentioned that these effects were statistically controlled for potential confounders, e.g. age, gender, and other working conditions, and are thus not affected by these confounders. We also found a positive effect of work enlargement, achieved via a mixture of driving activities with structurally different activities not related to driving, an approach that in fact changes some of the typical impairing conditions of pure driving, and thus the relatively one-sided structure of the work load of driving. However, since we had not enough cases for that kind of

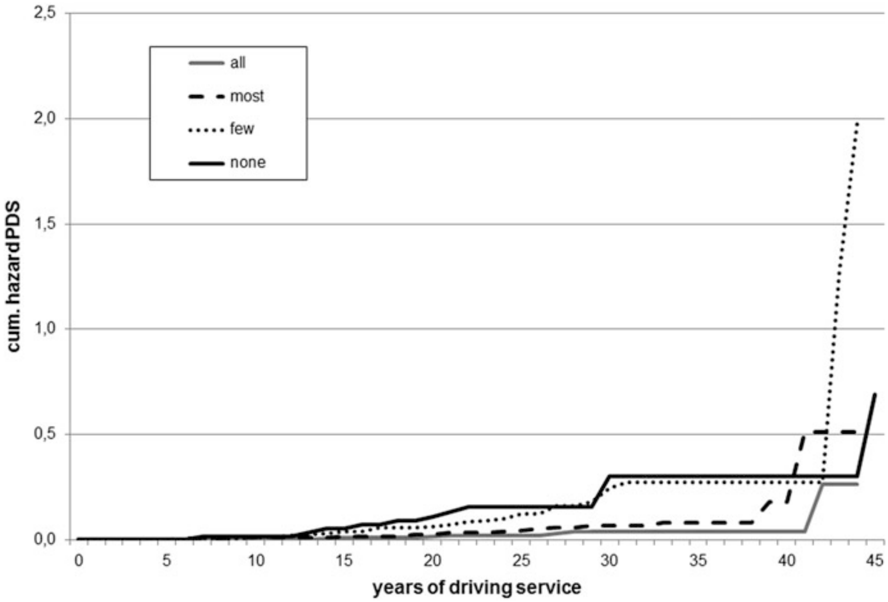


Fig. 5 Cumulated hazard rates for PDS—ergonomic seats

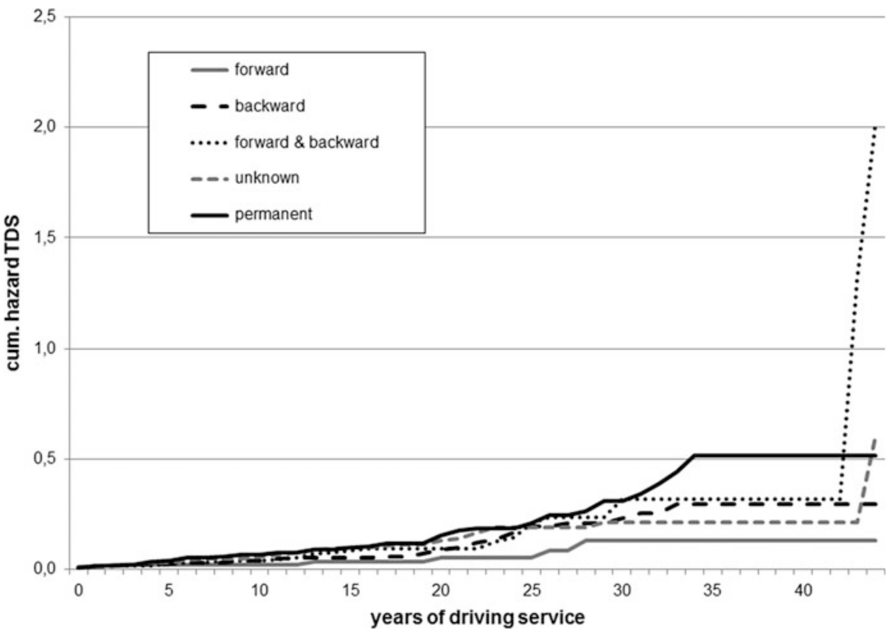


Fig. 6 Cumulated hazard rates for TDS—direction of shift rotation

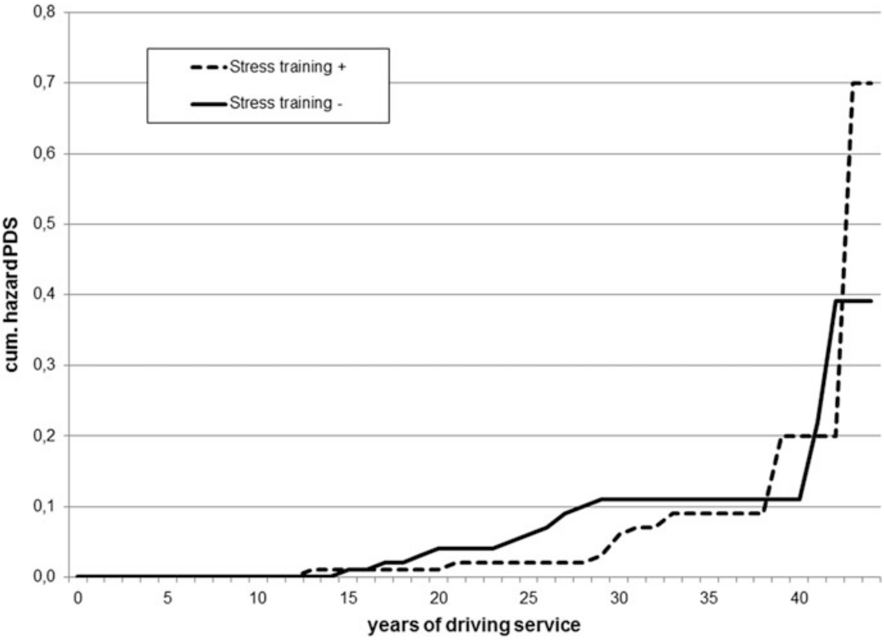


Fig. 7 Cumulated hazard risks of PDS—training on stress reduction and conflict management

enlarged work (i.e. 72 drivers) and only six of those with TDS and/or PDS the results were not statistically significant.

In contrast, behavioral interventions, e.g. training on "stress" reduction or coping with "stress" (with "stress" conceived as an individual response) or conflict management, showed no substantial or sustainable effect. Figure 7 shows a typical result from one of these analyses.

The development of the risk across years on service between participants and non-participants of such training is not statistically significant, nor is the mean difference in time to failure (Δmtf = 0.31 years, wrong direction) between those who attended such training and those who did not. Behavioral interventions that showed some marginal (mostly insignificant) effects on TDS and PDS were those that addressed specific work related skills (e.g. in driving or traffic simulations for developing special skills), whereas generalized approaches, e.g. training aiming at increasing physical activity, or offering opportunities for fitness training (during and after work hours) did not show any effect.

A comparison of mean times to failure for different intervention approaches/working conditions is shown in Fig. 8. As can be seen from Fig. 8, mean times to failure differ substantially for structural interventions, whereas they do not for behavioral interventions. These are selected, typical results, which can be extended to other specific variables, not shown here due to space limitations. Differences in the mean times to failure of about 5–6 years as shown in Fig. 8 between the best and the worst condition and a clear dose response relationship—as with ergonomic

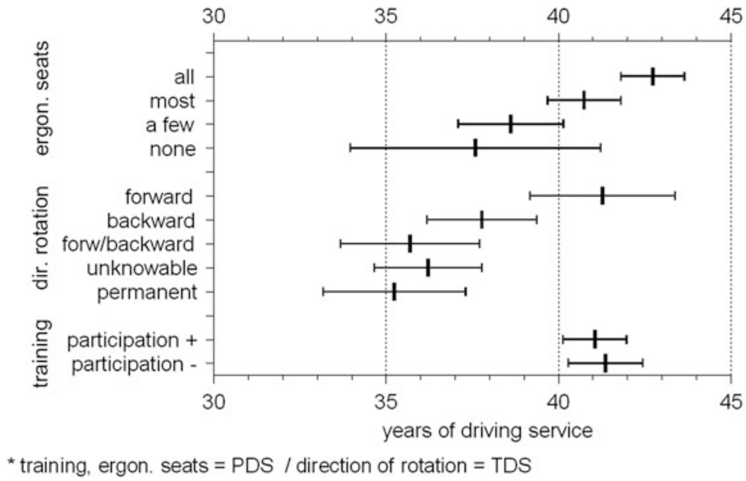


Fig. 8 Mean times to failure (TDS/PDS) for different working conditions

seats—strongly argue for considering ergonomic approaches in the pursuit of reducing the risk of a TDS or a PDS.

Behavioral interventions, on the other hand, seem to offer no substantial chances for reducing or postponing the risk of a TDS or PDS, based on our results.

4 Discussion and Conclusions

The study clearly shows that the prevalence of PDS is rather low in the sample analyzed, although any PDS is one PDS too many, both for the company, the individual and the social system as a whole. It is thus necessary to try to reduce such disabilities for service as much as possible. The results presented have shown that there are chances to do so by using promising preventive interventions and that these are comprised of ergonomic approaches to optimizing the kind and intensity of the resulting work load and work stress for the drivers. It is this direct approach to the improvement of the working conditions for drivers that is obviously promising for risk reduction and postponement.

It should further be mentioned that TDS and PDS have a lot of conditions in common, especially mental health problems, muscular-skeletal impairments, cardio-vascular diseases, and their provoking conditions, although there are, of course, also some specific conditions. It would thus seem promising to concentrate on such common causes, since further analyses (not shown here) demonstrate a link from temporary to permanent disability. In addition a further association has been found between sick leave and both TDS and PDS, indicating common causes for both outcomes which should be further analyzed and addressed in the future. In any case, the results of this study clearly demonstrate possibilities and effective means for reducing drivers' health risks. The high number of days lost to illness, which is only in part due to the high age of the population and which seems to increase in the years to come, would argue for an approach aimed at reducing this proximal outcome, which is related to the working conditions and the resulting temporally delayed effects (not shown here). This problem should thus have a high priority, also from the point of view that impaired health is structurally related to TDS and PDS.

As our results show, the increase in the hazard function starts rather early for TDS, while the risk of PDS increases after about 10–15 years of service. This would argue for primary prevention approaches that start and become effective right at the beginning of the careers of the drivers (e.g. by an ergonomic design of the working conditions which affects all drivers right from the beginning of their time in the job) and not to try to postpone any interventions until the first symptoms show up, and then trying to influence the (further) development of the risk by behavioral interventions. This will become extremely important in connection with the aging work force and the current extension of the normal retirement age in the FRG. It will thus be most crucial for companies to invest in trainin effective and efficient preventive interventions, in order to be able to keep their work force fit for duty until the (extended) retirement age, both from an ergonomic and from an economic point of view.

Ongoing "solutions" where public transport companies are outsourcing the driving services to pure driver companies and thus seem to get rid of the problem, in fact seem to be no solutions at all, since the risk of TDS and PDS should increase and accelerate in such companies, due to the even more one-sided structure of the workload. Laying off drivers with a PDS, because there are no other jobs or activities for them in such companies, shifts the risk to the individual driver and the society and its social security systems. Such solutions are no solutions to the basic problem, but tricky economic maneuvers to avoid the costs of an inadequate design of jobs and working conditions.

Real solutions could thus most probably be found following an approach to change the jobs and the work load of the drivers, e.g. by redesigning the drivers' jobs by a composition of different tasks within a public transport company, where driving is only one part of the job (job enlargement, job enrichment, job rotation, etc.). Although this approach has not shown to be statistically significant superior to pure driving activities, the tendency was in the direction to be expected [8]. In our opinion this lack of significance is not due to a lacking effect but to the small numbers of drivers in enlarged and/or enriched jobs, and especially doing this for a sufficient amount of time. It could thus be that a different design in the analyses, e.g. a more specific control group design with a sufficient number of drivers/ companies using such jobs and comparing the development of the risk across groups over time could yield more convincing results. This, however, was not the purpose of the study reported here, but it should be taken into consideration in a separate study to test for exactly this hypothesis. Such a study would urgently be needed.

Acknowledgements The authors would like to thank Jana Greubel and Alexandra Wassermann for their helpful support in conducting this study, the analysis of the data, and the graphical presentation of the results.

References

1. Bailer H, Tränkle U (1992) Fahrdienstuntauglichkeit – ein unabwendbares Schicksal? Der Nahverkehr 10:11–14
2. Bailer H, Tränkle U (1993) Fahrerarbeit als Lebensperspektive. Wirtschaftsverlag NW, Bremerhaven
3. Garbe C (1981) Gesundheitszustand und gesundheitliche Risiken von Linienbusfahrern in Berlin (West). Reimer, Berlin
4. Kompier M (1996) Bus drivers: occupational stress and stress prevention. International Labour Office, Geneva
5. Kompier M, Aust B, van den Berg A, Siegrist J (2000) Stress prevention in bus drivers: evaluation of 13 natural experiments. J Occup Health Psychol 5(1):11–31
6. Michaelis M (2008) Gesundheitsschutz und Gesundheitsförderung von Berufskraftfahrern. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund
7. Roth JJ, Schygulla M, Dürholt H, Nachreiner F, Pankonin C (2004) Betriebs- und Arbeitszeiten beim Gütertransport und bei der Personenbeförderung. (Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, FB 1033). Wirtschaftsverlag NW, Bremerhaven
8. Tränkle U, Bailer H (1996) Aufgabengestaltung im Fahrdienst. Überlegungen und Untersuchungen zur Verbesserung der Arbeitssituation von Linienbusfahrerinnen und Linienbusfahrern im öffentlichen Personennahverkehr (ÖPNV). Wirtschaftsverlag NW, Bremerhaven
9. Zoike E, Bungard S, Ganske V, Grothaus F, Hertle D, Tewes C, Trümner A, Kliner K (2011) BKK Gesundheitsreport. BKK Bundesverband, Essen

Approaches to Strengthen Behavioral Prevention in a German Medium-Sized Enterprise

Lisa Rücker and Johannes Brombach

Abstract
The improvement of working environment and in particular behavioral prevention has a long tradition in the maintenance of industrial health and safety standards in Germany. Despite further automation of the value process accompanying industry 4.0, human behavior and its impacts in terms of occupational health and safety must not be neglected. In the context of safety regulations and "unsafe behavior", this contribution deals with targeted analysis, the development of practical implementation measures for the reunification of behavioral and situational prevention, as well as the involvement of employees in the continuous improvement process.

Keywords
Behavior-based work safety • Maintenance of industrial health and safety standards • Continuous improvement processes (CIP) • Behavior based safety (BBS)

1 Introduction

Occupational safety continually gets more attention in industry during the last few years. In the age of automation, the fallacy arises that machines and robots are mainly responsible for producing products and services. However, even today, still 70 % of the national product are generated by manpower [12]. So "health and safety" is an intensely discussed topic and not only under cost-related aspects.

L. Rücker (✉) • J. Brombach
Faculty of Engineering and Management, Munich University of Applied Sciences, Munich, Germany
e-mail: ruecker.lisa@hm.edu; johannes.brombach@hm.edu

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_10

Especially for today's company, occupational safety has become a fundamental basis for successful enterprises.

A majority of work accidents are based on "unsafe behavior" of the employees (also called unsafe acts) [10]. Already in Loafman [9] talked about as much as 94 %. According to the German Federal Statistical Office, in 2013 almost 90 % of all work accidents could be explained by unsafe behavior of the employees [8].

▶ Unsafe behavior is defined as the performance of a task or other activity that is conducted in a manner that threatens health and safety of workers or other people.

Germany traditionally focuses on the technical and organizational occupational safety and health (OSH) [11], driven continuously by activities of legislators, companies, trade unions and accident insurances. The English space, however, is marked by behavioral-based prevention approaches [2]. One of these approaches is the "Behavior Based Safety—BBS".

2 Methodology

"Men act upon the world, and change it, and are changed in turn by the consequences of their action." [13, p. 1]. In the first instance the question is considered, why do people show a certain behavior. For this purpose, the Antecedent-Behavior-Consequence model (ABC-model, Fig. 1) by Skinner [13], which he used in the linguistic behavior analysis, can be consulted. People are influenced directly or indirectly by events in their environment.

This model describes that the behavior ("B"), i.e. an act or omission, is triggered by the "antecedents" ("A", preliminary conditions). Antecedents take place prior to the behavior. Relating to the occupational safety these preliminary conditions are e.g. trainings, manuals or the working environment [2]. However, preliminary conditions have only a low impact on behavior (share of approx. 15 %).

On the other side are the "consequences" ("C"), which occur as a result of a certain behavior. They can have a reinforcing or weakening effect and influence the behavior up to approx. 85 %. If a behavior is reinforced by its consequences, the likelihood of recurrence of this displayed behavior increases. Conversely, in the case of a weakening, the likelihood that a displayed behavior will be repeated, declines [2].

Possible consequences include, i.a., punishment, positive or negative reinforcement. While a behavior decreases by punishment, it increases by positive or negative reinforcement. By negative reinforcement, the individual is encouraged to a certain behavior in order to prevent something undesirable (e.g. adherence to the road traffic regulations in order to avoid fines). In contrast, by positive reinforcement, an individual shows a certain behavior for being recognized, praised or rewarded materially [4].

Fig. 1 The Antecedent-Behavior-Consequence-model

While most consequences only have temporary effects, positive reinforcement can sustain a certain behavior permanently. The method "BBS" is based on these principles of positive reinforcement [10].

BBS orientates itself on the security pyramid of H.W. Heinrich, which describes that an accident at work is preceded by a variety of unsafe behaviors. This method is promoted by employees. Its aim is to draw attention on safe and unsafe behavior by mutually observing and giving feedback. Thereby, on the one hand, safe behavior is encouraged by positive feedback, on the other hand, unsafe behavior can be reduced through specifically derived measures. In the long term accidents and injuries can be prevented. While most approaches are guided by supervisors, BBS focusses on the employees steering this method (bottom-up). Employees accept responsibility for their own occupational safety as well as those of their colleagues. They are supported by their superiors.

2.1 The Five Steps of BBS

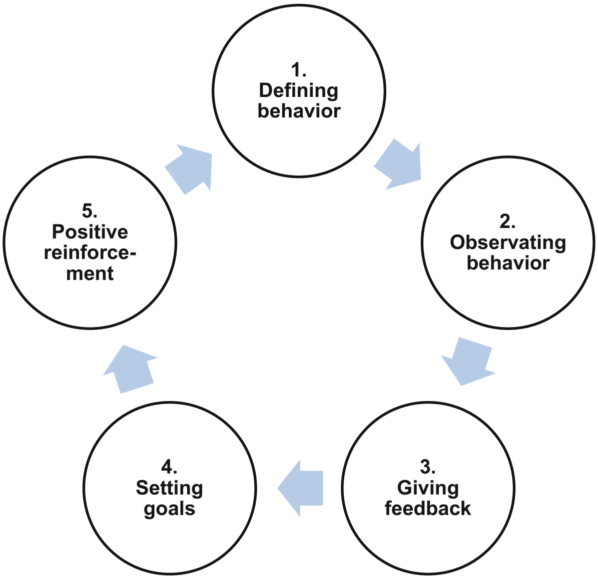
A founder of behavioral safety, E. Scott Geller, ascertained: "It is much more cost-effective to 'act' an employee into safe thinking than it is to 'think' a person into safe acting." [6, p. 309].

As a continuous improvement process BBS is divided into five steps, in the form of a circuit (Fig. 2).

In the first step, a behavior is defined, which should be displayed by an employee in a certain situation. There are special requirements for the definition of a behavior, e.g. clarity, measurability, observability etc. [5] (e.g. "The employee wears safety glasses while drilling"). Information to define a behavior may be found in accident reports, first-aid books, near-accidents and employee interviews etc. Afterwards, these defined behaviors are summarized in the form of an observation card (cf. Fig. 5) [10].

In the second step, the corresponding behavior is observed by using the observation card. It is individually designable when observations are made, whether there are specific fixed periods or typical situations in use. There are moments, for example, where an employee has to wait for the work process of a machine. The observations are documented (anonymously) in the form of a tally sheet (cf. Fig. 5),

Fig. 2 The five steps of BBS



e.g. there is no name on the tally sheet, neither of the observer nor of the observed person. Through this we can avoid the distress of possible negative consequences [2].

Besides observing the employees, feedback is given about the observed behavior (third step). Depending on whether a safe or unsafe behavior has been observed, they get a positive or a constructive feedback including an explanation. Examples: “I saw that you were wearing protective clothing while deburring. I like that, keep it up.”; “I saw that you weren’t wearing protective gloves while handling raw material. Please use them next time, so you do not cut yourself on the sharp ridges and edges.”). The most common form is an oral feedback, which should occur straight after the observation and is only intended for the observed person. There is also the possibility of graphical feedback, often depicted as line graphics. Here, however, are only published results of groups, departments or the entire company, but not of an individual. In very rare cases written feedback is also used [2].

In the fourth step, goals or interim targets are set, which should be reached in a certain period (e.g. increase of the share of safe behavior by 20 %). Thus, the basis for determining the objective is the current situation, i.e. the current share of safe behavior [5, 10].

Finally, positive reinforcement is placed in the fifth step. It is a key aspect of the BBS and differentiates this system from most traditional approaches. An event is seen as a positive amplifier, if according to BBS a safe behavior occurs frequently in the future. This phenomenon is called “law of effect” [14], i.e. “responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation, and responses that produce a discomforting effect become less likely to occur again in that situation.” [7, pp. 108–109]. When correctly used, giving feedback is often already a positive reinforcement. With the positive feedback (as a result of shown safe behavior) having a satisfying effect, this consequence will contribute to a continuous occurrence of safe behavior. In this context, also setting goals and interim targets can be used as a further amplifier [2]. In addition, a variety of studies indicates that employees receiving positive reinforcement are more communicative, performance oriented and creative [3]. Besides these described social amplifiers (e.g. praise, recognition, achievement of objectives), in some cases material amplifiers (e.g. food voucher) exist. However, the focus of the BBS lies on social amplifiers [2, 5].

2.2 Implementation of the BBS in a Medium-Sized Enterprise

As shown in Fig. 3 the possibility of introducing the BBS in a German medium-sized metalworking company has been explored in three phases within a period of 6 months.

2.2.1 Safety Assessment

Within a safety assessment (phase 1), the corporate and occupational safety culture of the enterprise are analyzed.

The analysis of the results of an employee survey in 2013 (participation rate 74.5 %) showed, i.a., that current working conditions are considered to be very good (e.g. all required tools are available). Furthermore, open and trustful communication, as well as mutually given feedback, are frequently used in practice and there is an established culture of error tolerance. This analysis shows that the necessary basis for the implementation of BBS is a present topic.

The number of accidents were first compared with the industry average of the corresponding professional association. With regard to the total number of work accidents in 2013, the analyzed company is well below average. These work accidents were subsequently statistically evaluated concerning their technical/organizational and behavioral (personal) causes, on basis of accident reports and analysis, as well as on an accident data table. As Fig. 4 shows, 83.9 % of all work accidents are associated with behavioral causes. These results reflect a high level of technical and organizational maturity of the enterprise. According to this, the introduction of BBS for long-term reduction of work accidents promises high chances of success.

2.2.2 Planning and Implementation Phase

During the preparations of the test phase (phase 2), two pilot departments of one plant were selected (one production and one training area). The employees were systematically introduced to BBS through the following individual training sessions.

Beginning with a survey, the general willingness of employees regarding work safety should be carried out. Moreover, possible potential danger of the particular department should be identified (e.g. “What activities in your workspace would you

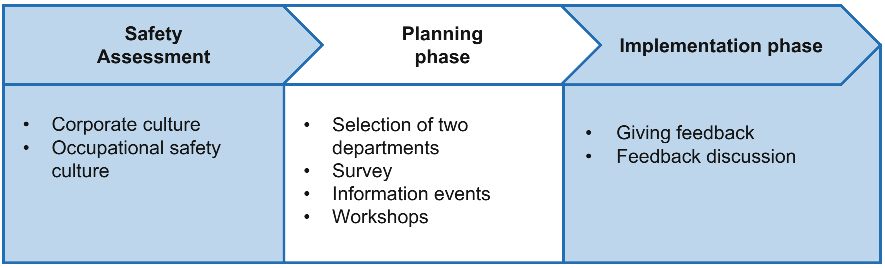


Fig. 3 Implementation plan of the BBS

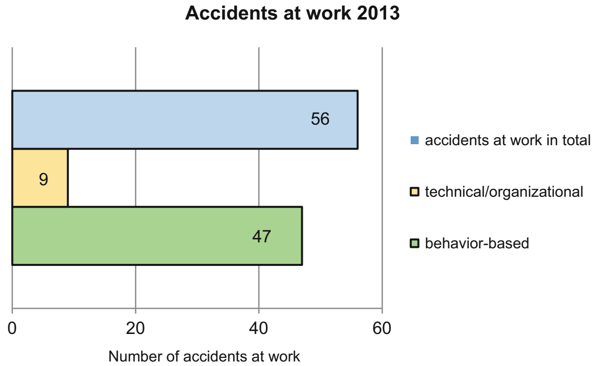


Fig. 4 Accidents at work in 2013

consider as particularly risky?"). The latter is especially interesting, because an employee is usually more familiar with his own workstation and therefore possible dangers are estimated very well.

The survey was followed by a one and a half hour information event preparing all employees for the upcoming test phase. Beginning with an introduction to general occupational safety (including the results of accident statistics), information to behavior-based occupational accidents and unsafe behavior were given. The employees got a detailed explanation of the method of BBS, including definition, its purpose and their own role in this system. Finally, further proceedings with regard to the test phase were discussed.

Specific trainings for a small group for each department (four employees) about observing and giving feedback, as well as the definition of behaviors could successfully be realized by these workshops. For practice, real accident analysis and first-aid books of other plants of the enterprise could be used exemplarily. Subsequently, four to five specific behaviors were identified for each area by using behavioral accident analysis, first-aid books, an evaluation of the prepared questionnaire and a typical course of a day and its workflow, which the employees prepared in advance. The behaviors defined in the workshop formed the basis of an observation card (cf. Fig. 5), which is used to document the observations during the

Department: ____________
Date: ____________

Sheet P-01

Nr.	Behavior	Safe	Unsafe
1.1	While hand deburring and handling raw materials/workpieces, the employee takes care of the work safety aspects. E.g. weight, using PPE (especially safety gloves and safety goggles), etc.		
	Causes for "unsafe": Possible countermeasures:		
1.2	The employee pays attention to safety, while dealing with oils and cooling lubricants, e.g. using personal protective equipment (PPE), order / cleanliness, etc.		
	Causes for "unsafe": Possible countermeasures:		

Further observed hazards:
- e.g. near-accidents
- technical and organizational, as well as behavioral causes

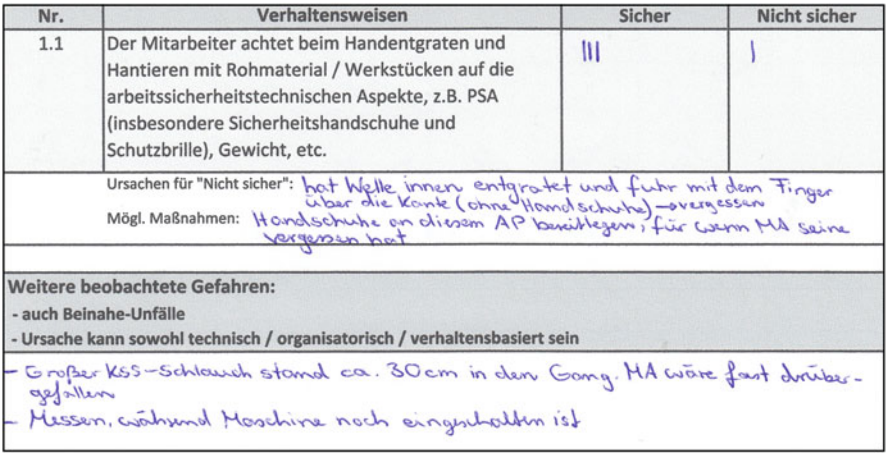
Fig. 5 Part of an observation card

test phase. As Fig. 5 shows, safe and unsafe behaviors are recognized by tallying up the totals. This observation card was prepared accordingly to the theoretical BBS approach, but extended during the investigation by the following specific fields:

a. Employees can add causes for “unsafe” executed behaviors and also define countermeasures (elimination/improvement), if possible.
b. Additionally, employees can document undefined, unsafe behaviors and risks, e.g. near-accidents. Besides the behavioral causes, also the technical and organizational deficiencies should be listed.

Thus, this addition offers the opportunity to identify any technical and organizational defects at an early stage and also to avoid accidents at work.

During the subsequent 4-week test phase (phase 3) observations with daily documentation were conducted. The first week started with concealed observations



Nr.	Verhaltensweisen	Sicher	Nicht sicher
1.1	Der Mitarbeiter achtet beim Handentgraten und Hantieren mit Rohmaterial / Werkstücken auf die arbeitssicherheitstechnischen Aspekte, z.B. PSA (insbesondere Sicherheitshandschuhe und Schutzbrille), Gewicht, etc.	III	I
	Ursachen für "Nicht sicher": hat Welle innen entgratet und fuhr mit dem Finger über die Kante (ohne Handschuhe) → vergessen Mögl. Maßnahmen: Handschuhe an diesem AP bereitlegen, für wenn MA seine vergessen hat		

Weitere beobachtete Gefahren:
- auch Beinahe-Unfälle
- Ursache kann sowohl technisch / organisatorisch / verhaltensbasiert sein

- Großer KSS-Schlauch stand ca. 30cm in den Gang. MA wäre fast drüber-gefallen
- Messen, während Maschine noch eingeschaltet ist

Fig. 6 A small excerpt of an originally filled-in observation card in German

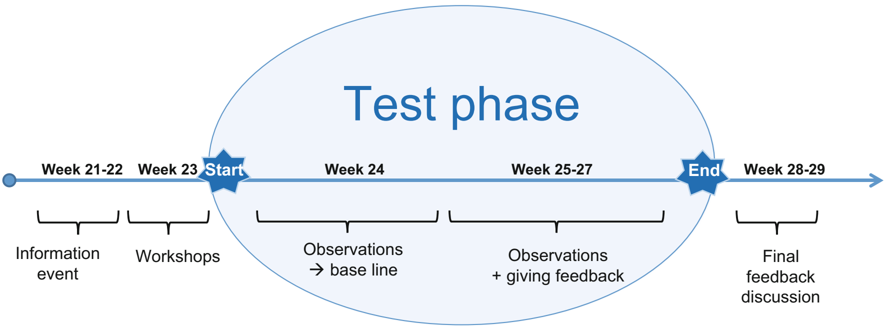


Fig. 7 Time line of the project

to determine the share of safe behavior before starting the BBS ("base line"). In this week, the observers didn't give feedback. The observed neither knew the fact that they were observed, nor which behaviors were observed.

Subsequently observations occurred with immediate oral feedback. Figure 6 shows an excerpt of an originally filled-in observation card in German. The employee noticed for example, that in the concerning workplace safety gloves should be provided, a tripping hazard existed and the machine was not turned off during the measurement process.

Weekly feedback meetings during the whole test phase supported both, the two active observer groups, as well as the regular employees in their daily execution of their work. Figure 7 shows the time line of the whole project.

3 Results

Already after this short test phase, the systematical use led to a better hazard perception and accident avoidance of the employees. Especially for young participants (trainees) a positive change in behavior could be determined. In addition to a higher sense of responsibility towards the safe behavior of their colleagues, they paid more attention with respect to occupational safety issues. Despite initial fluctuations, due to the training and implementation phase of the employees, the evaluation of the observation cards reveals a positive trend. For example, the share of the defined safe behavior "The employee pays attention to safety while dealing with oils and cooling lubricants, e.g. using personal protective equipment (PPE), order/cleanliness, etc." raised from 58.3 % (before) to 67.1 % during this short test phase (Fig. 8). It was observed in both departments (in total eight observers).

Due to low number of responses (eight observers for the above-described behavior) the Wilcoxon signed-rank test was used to verify the significance. Compared to the *t*-test, this non-parametric method doesn't require normal distribution. The test confirmed with a significance level of 5.0 %, that the application of BBS has led to a significant improvement of safe behavior. Nevertheless, the number of test subjects should be increased to get more reliable results.

It can be assumed that a continuing use of BBS will demonstrate a sustainable success in behavior-based injury prevention and the establishment of long-term health and safety measures.

During the implementation phase and the feedback discussions also some serious security problems could be identified and successfully resolved:

a. A used lift truck for high-level racks did not comply with the relevant safety standards anymore. The transported material could tip and fall down, so the machine was shut down immediately.
b. The fork-lift trucks are retrofitted with acoustic signals, in order to reduce the risk of collisions with employees.
c. While withdrawing machine tools, employees often get injuries by inadvertently touching neighboring tools, according to the narrow storage. Within a trainee project the old storage was optimized by a new design (cf. Fig. 9a). Now, the machine tools are placed on different levels and have a larger distance to each other. Figure 9b shows a similar form of storage from another company. A trainee has gone one step further and, as seen in Fig. 9b, developed magnetic protective caps for the milling heads. This idea has been awarded by the professional association of raw materials and chemical industry ("Berufsgenossenschaft Rohstoffe und chemische Industrie") with the special prize in 2015 in the mining sector ("Sonderpreis 2015 Bergbau" [1]).

Due to the success and the positive responses of the employees, BBS will be continued in the training area and should be extended to other plants in the medium term.

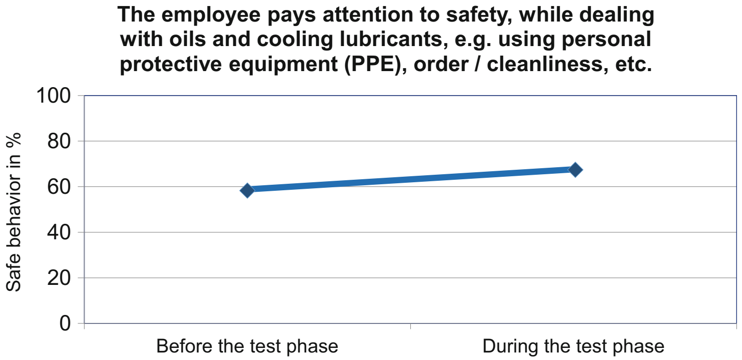


Fig. 8 Graphical analysis of observation before and during the test phase

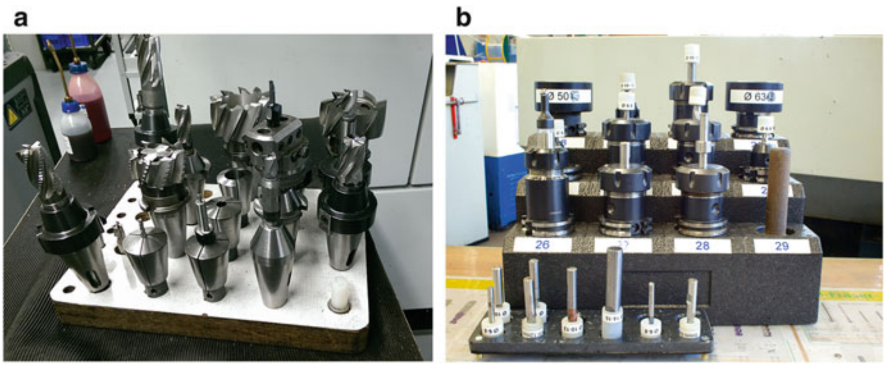


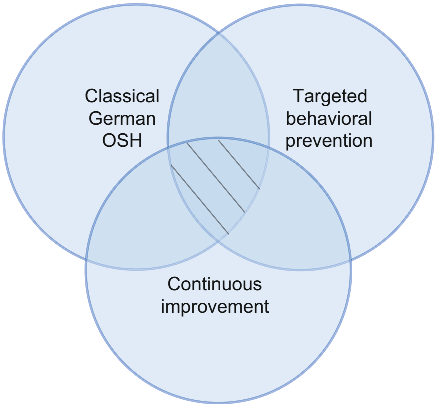
Fig. 9 Narrow storage for machine tools (**a**), magnetic protective caps for the milling heads (**b**) [1]

4 Discussion

The results suggest that a combination of targeted behavioral prevention and the classical German occupational safety and health (OSH) should be implemented in a uniform overall concept together with the idea of continuous improvement (cf. Fig. 10).

In addition to the German occupational health and safety system (i.e. mainly the technical and organizational accident prevention), by using the BBS approach, behavioral accident prevention can be operated. The extension of the observation card, by adding the documentation of further hazards (including technical and organizational causes), provides the ability to bring the company to an optimized position (in all three safety areas). Only a combination of both systems can lead to an effective and lasting reduction of accidents at work.

Fig. 10 The intersection of the work-scientific improvement methods



It is particularly important that the legal regulations of occupational safety are involved in the BBS process and not neglected. As the test phase points out, it is advisable to conduct prevention systems during the training period in an integrative framework. In this project especially the intersecting set of the existing approaches was interesting (cf. Fig. 10). In this context, the employees were motivated by the BBS to continually identify hazards and avoid technical/organizational, but also behavior-based accidents.

References

1. Berufsgenossenschaft Rohstoffe und chemische Industrie (BG RCI) (2015) Förderpreis 2015 Arbeit-Sicherheit-Gesundheit. http://www.bgrci-foerderpreis.de/foerderpreis/makepdf.php?nr=1807. Accessed 06 May 2015
2. Bördlein C (2009) Faktor Mensch in der Arbeitssicherheit – BBS. Erich Schmidt Verlag, Berlin
3. Daniels AC, Daniels JE (2004) Performance management: changing behavior that drives organizational effectiveness. Performance Management Publications
4. Edelmann W (1993) Lernpsychologie. Psychologie Verlags Union, Weinheim
5. Fleming M, Lardner R (2002) Strategies to promote safe behaviour as part of a health and safety management system. Health and Safety Executive, Edinburgh
6. Geller ES (2001) Actively caring for occupational safety: extending the performance management paradigm. In: Johnson CM, Redmon WK, Mawhinney TC (eds) Handbook of organizational performance: behavior analysis and management. Haworth Press, New York, NY, pp 303–326
7. Gray PO (2010) Psychology. Worth Publisher, New York, NY
8. Liersch A (2014) Arbeitsunfälle und arbeitsbedingte Gesundheitsprobleme. Wirtschaft und Statistik (Sept): 561–574
9. Loafman B (1996) Rescue from the safety plateau. Perform Manag Mag 14(3):3–10
10. McSween TE (2003) The values-based safety process. Wiley, Hoboken, NJ
11. Merdian J (2011) Arbeitssicherheitsaudits - Leitfaden für die betriebliche Praxis. Beuth Verlag, Berlin
12. Schraub EM et al (2009) Bestimmung des ökonomischen Nutzens eines ganzheitlichen Gesundheitsmanagements. In: Badura B, Schröder H, Vetter C (eds) Fehlzeiten-Report 2008

– Betriebliches Gesundheitsmanagement: Kosten und Nutzen. Springer Medizin, Heidelberg, pp 101–110
13. Skinner BF (1957) Verbal behavior. Copley Publishing Group, Acton
14. Thorndike EL (1898) The Psychological Review – Animal intelligence: an experimental study of the associative process in animals. Macmillan, New York, NY

Recreational Noise: Objective Measurement and Subjective Assessment of Sound Exposures in a Non-professional Symphonic Wind Orchestra

Mario Penzkofer, Florian Finé, and Karsten Kluth

Abstract

While risks to the hearing of professional musicians were part of several studies in recent years, little attention was paid to the sound pressure exposure faced by amateur musicians. Thus, around 500,000 amateur musicians in Germany are hardly aware of the risks of high sound pressure levels during regular rehearsals or concerts. With the aim to raise the awareness of the musicians to take possible protective measures into account, e.g., use of personal hearing protectors or less loud performances, several sound immission measurements were carried out during the rehearsals of a non-professionally acting symphonic wind orchestra. Furthermore, in order to determine potential hearing threshold shifts, audiometric measurements were performed. In addition to the objective evaluation, the musicians were asked about their personal feelings during the rehearsals, about their knowledge on possible risks to the hearing and on appropriate protection means. The results show that all members of non-professionally acting orchestras can be exposed to sound pressure levels up to 117 dB(A). With an average rehearsal duration of 2 h, these exposures resulted in equivalent continuous sound pressure levels above 92 dB(A). Additionally, substantial temporary hearing threshold shifts could be monitored after the rehearsals. The evaluation of the subjective survey showed that most of the musicians were not aware of the actual sound pressure levels during the rehearsals, and the associated hearing risk. Similar is true for the necessity of the hearing protection devices which were mainly refused.

M. Penzkofer (✉) • F. Finé • K. Kluth
Department of Mechanical Engineering, Ergonomics Division, University of Siegen, Siegen, Germany
e-mail: penzkofer@ergonomie.uni-siegen.de; florian.fine@student.uni-siegen.de; kluth@ergonomie.uni-siegen.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_11

Keywords
Orchestra • Musicians • Sound exposure • Sound exposure topography • Audiometry • Temporary hearing threshold shift

1 Introduction

Music is generally perceived as pleasant but sometimes music, to be fully effective, needs to be loud [31]. Especially orchestral music has become louder over the past centuries. To be properly prepared for playing on the biggest concert stages, the instruments construction is developed in such a way that they are suitable for larger premises and stages. Modern instruments are designed to manage and perform higher sound pressure levels than the previous ones. The use of steel strings began in the middle of the nineteenth century, and resulted in a more distinct and harsher timbre from the bowed string instruments than the earlier ones, which had strings made from animal gut. The wood used for construction of modern bowed stringed instruments is another reason for their higher sound levels; the wood used today is harder than that used in older instruments and, therefore, produces higher sound levels and pitch. Modern brass instruments have a wider drilled interior than do older ones, and percussion instruments use modern materials that also contribute to higher sound levels than before [20]. However, this fact carries some risks. Thus, tinnitus, hyperacusis, distortion problems [1, 23, 29] but also temporary hearing threshold shifts (TTS) can occur at musicians, which can affect their work more severely than a hearing loss [23].

Even though, the restitution of temporary hearing threshold shifts is completed after a few hours of rest in most cases, all sound exposures are summed up by the body in the course of its life [31]. However, there is still a disagreement about the fact how harmful high sound pressure levels for the musicians in the orchestra are. Researchers have studied the sound exposure of classical musicians for more than 40 years, and they still do not agree if the measured peak sound pressure levels in symphony orchestras are as harmful as industrial noise of the same energy [5, 19, 21]. In comparative studies of different sound exposures, classical music was associated with the least severe temporary hearing threshold shifts of less than 10 dB which disappeared rather quickly [37]. The substantially lower physiological costs while hearing classical music apparently indicate a decisive influence of the type of sound exposures [36]. According to Brusis [9] current studies are also less clear in terms of actually monitored hearing impairments based on audiometric examinations, although several studies have assessed hearing threshold shifts of musicians as well [1, 18, 20–22, 29, 33, 34].

Based on the evaluation of current studies of the last 15 years it is also concluded by Richter et al. [32], that performing as a musician in a symphonic orchestra does not involve an increased risk of hearing impairments. Thus, also in studies of Toppila et al. [38] less hearing impairments could be found in comparison to the expectations based on ISO 1999 [16] and ISO 7029 [15]. In this context, also Brusis [9] and Lenzen-Schulte [24] have different opinions. While, according to Brusis [9], a noise-induced hearing loss is more of a rarity for an orchestral musician, since only in rare cases permanent hearing threshold shift were found—especially in the high frequency area—that match an incipient deafness, Lenzen-Schulte [24] considers that the hearing loss is the most common occupational disease among professional musicians.

In other studies e.g. Billeter and Hohmann [6], Haider and Groll-Knapp [12], Jansson and Karlsson [17] sound pressure levels were registered during rehearsals and concerts as well as private lessons which—according to the Noise and Vibrations Occupational Safety and Health Ordinance (LärmVibrationsArbSchV) [28]—exceeded the lower (80 dB(A) respectively 135 dB(C)) and upper exposure action values (85 dB(A) respectively 137 dB(C)) of the daily noise exposure level $L_{EX,8h}$ and the peak sound pressure level $L_{pC,peak}$.

When exceeding the lower exposure action values, the employer shall provide suitable individual hearing protection and is also obligated to inform and instruct the employees who are entitled to carry out preventive audiometric testing. If one of the upper exposure action values are exceeded, the noise area must be marked and a noise reduction program must be created. The employees are also obliged to wear personal hearing protectors and, furthermore, have a claim on occupational health examinations. According to the Directive 2003/10/EC [11] immediate measures are taken if the daily noise exposure level exceeds the limit of 87 dB(A). Therefore, professional musicians, whose number could be quantified in Germany in 2009 to approximately 120,000 [24], should also experience protection against high noise levels [8, 27, 30], either by technical (e.g., improvement of the acoustic absorption in the room, soundproof walls), organizational (e.g., rostering arrangements, changing of the orchestra stage plan) or personal measures (e.g., hearing protectors, cp. left part of Fig. 1).

More than 500,000 non-professional musicians in Germany, who are naturally not considered by the above-mentioned regulations, are hardly aware of the risks of high sound exposures during regular rehearsals or concerts. This problem becomes even more pressing through the fact that rehearsals take place at premises that are suitable for the size of the orchestra quite rare (cp. right part of Fig. 1). Moreover, limited space between the musicians reduces the distance between the ears of one musician and the instrument of the musician sitting next to him/her, which leads to an increased exposure level—triggered by increased direct sound [20, 25]. Although amateur musicians do not reach the weekly amount of time, professional musicians playing their instruments, are often additionally exposed to high sound levels, e.g., from their professional environment, which can lead to hearing impairments in the long run. Therefore, it was the aim of the study to get knowledge about the sound pressure levels amateur musicians are generally exposed to. Furthermore, the musicians acting for pleasure rather than for financial gain should be informed about potential risks to the hearing and preventive measures.

Fig. 1 Hearing protectors (*left part*) and room situation at the rehearsal (*right part*)

2 Methods

The objective of the study was a symphonic amateur wind orchestra which rehearses several times a week and performs regularly at concerts. Symphonic wind orchestras usually have a large cast of over 50 musicians and are characterized by a wide range of soundscape, due to the interaction of many different instruments, such as trumpets, trombones, clarinets, saxophones and several horns. In order to develop the desired tonal characteristics and to meet the demands of a critical audience, regular rehearsals are necessary.

2.1 Measurement of Sound Exposures

Initially, this study must be seen as a pilot project. Therefore, only a total of ten sound pressure level measurements were carried out during the rehearsals so far. The sound level meter XL2 by NTi Audio was used as measurement device which exceeded—in conjunction with the appropriate microphone—the accuracy class 1 of IEC 61672-1 [13]. Attempts were made, to take several impact factors into account during the sound measurements. Thus, not only the position of the sound level meter was varied in order to detect all different groups of instruments, also measurements in different rehearsal premises were carried out. Since the measurement results, as is well known, also depend on each performed work and its individual interpretation [9], furthermore, attempts were made to cover a wide range of various compositions from classic to contemporary.

Due to the fact that the entire rehearsal was recorded as a wave audio file, furthermore, it was possible to create not only a time study, but also to reproduce the entire rehearsal. Thus, the points of particularly high sound exposures could be found in the later sound analysis and, therefore, a connection could be established to the played work respectively its section (bar) as well as to the most involved group of instruments.

2.2 Audiometry

In order to determine potential hearing threshold shifts, audiometric measurements were carried out prior to and after the rehearsals on those musicians who are located at positions in the orchestra where the highest sound exposures were reached or on those with the loudest instruments. Thus, pure-tone audiometry was performed on six musicians aged between 16 and 33 years 30 min prior to and 10 min after the rehearsals, respectively. Since the audiometry could not be performed under laboratory conditions, an experimental chamber was chosen, which had a good shielding against outside noise.

Sound with a frequency of 125 Hz and a level of 55 dB was presented at first on the subjects' right ear, and later on the left ear. As soon as the sound was perceived, the answer pushbutton had to be pressed over the entire duration of the sound presented. As a result of a correct response the volume of the predetermined tone was reduced in steps of 5 dB. If the response was delayed for quieter sounds, given at the wrong time or not at all, the last perceived volume level was recorded in an audiogram card. This sequence was continuously used for the frequencies of 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 6000 and 8000 Hz. After recording all the values in the audiogram card and joining the points, a hearing curve and, therefore, a possible hearing threshold shift or even a hearing impairment could be identified [7]. For further details on audiometry see Irle et al. [14].

2.3 Subjective Assessments

In addition to the objective measurements, a subjective evaluation was carried out by means of a questionnaire comprising 18 items. Besides the general information about the musicians such as age, gender, job and potential pre-existing hearing impairments, the musicians were asked about their personal feelings regarding the sound exposures during the rehearsals arising from their own instrument and the instrument of the musician sitting next to him/her. Furthermore, the musicians were asked whether the reached sound levels in the orchestra were experienced as noise which might even cause discomfort. In the final part of the questionnaire, the musicians should indicate whether they are even aware of potential risks of hearing impairments while playing in the wind orchestra and whether they have been thinking about appropriate means of protection. In addition, reasons were mentioned why, unfortunately, the easiest way of protection, namely the wearing of individual hearing protectors, is often not applied.

3 Results of the Objective and Subjective Evaluation

3.1 Results of the Sound Exposure Measurements

Table 1 shows the measurement values registered during the ten rehearsals, containing information about the location, the measuring points, and the duration of the rehearsal as well as the different sound levels. L_{AFmax} describes the maximum A-weighted sound pressure level with a "Fast" time constant recorded over the period stated. L_{AFmax} is often used as a measure of the most obtrusive facet of the noise, even though it may only occur for a very short time and is the level of the maximum Root Mean Square reading. L_{Aeq} is defined as the notional steady sound pressure level which, over a stated period of time, would contain the same amount of acoustical energy as the A-weighted fluctuating sound exposure measured over that period (cp. DIN 45641 [10]). The C-weighted peak immission sound pressure level $L_{pC,peak}$ corresponds to the absolute peak value of the sound pressure of a sound event and serves as a measure of the mechanical hearing load caused by a single sound event [26].

It can be seen that in each rehearsal a maximum A-weighted sound pressure level L_{AFmax} of at least 107 dB(A) was achieved at the ear of the respective musician. Considering the peak sound pressure level $L_{pC,peak}$, generally more than 120 dB(C) were recorded in each rehearsal. During the first measurement even a $L_{pC,peak}$ of 142 dB(C) was registered at the ear of the musician who played the bass clarinet directly in front of the drums. The determined equivalent continuous sound pressure level L_{Aeq} also takes alarming values of more than 90 dB(A) for all measurements. A more detailed temporal analysis of the detected sound pressure levels shows, on the one hand, that the musicians were exposed to loud music for at least half of the time of the rehearsal (cp. left part of Fig. 2). On the other hand, sound pressure values of 100 dB(A) were reached or even exceeded in an average of 20 % of the exposure time (cp. right part of Fig. 2).

Even if the term of the daily noise exposure level $L_{EX,8h}$ is not defined for activities in the leisure sector, it is worthwhile to exemplarily convert the registered

Table 1 Overview of the registered sound pressure levels during the rehearsals

Location	Measuring point	Duration	L_{AFmax}	L_{Aeq}	$L_{pC,peak}$
Guesthouse	Bass Clarinet	01:37:01 h	117 dB(A)	98 dB(A)	142 dB(A)
Guesthouse	1. Trumpet	00:43:54 h	111 dB(A)	97 dB(A)	132 dB(A)
Gallery	Bass Clarinet	01:43:32 h	110 dB(A)	95 dB(A)	123 dB(A)
Gallery	1. Clarinet	01:47:32 h	107 dB(A)	92 dB(A)	125 dB(A)
Gallery	1. Flute	01:49:32 h	108 dB(A)	95 dB(A)	131 dB(A)
Guesthouse	1. Horn	02:19:42 h	111 dB(A)	96 dB(A)	126 dB(A)
Guesthouse	Alto Saxophone	02:11:43 h	110 dB(A)	94 dB(A)	129 dB(A)
Guesthouse	3. Clarinet	01:40:30 h	107 dB(A)	93 dB(A)	126 dB(A)
Concert-stage	1. Trumpet	02:30:59 h	102 dB(A)	96 dB(A)	132 dB(A)
Concert-stage	Alto Saxophone	02:34:48 h	110 dB(A)	96 dB(A)	130 dB(A)

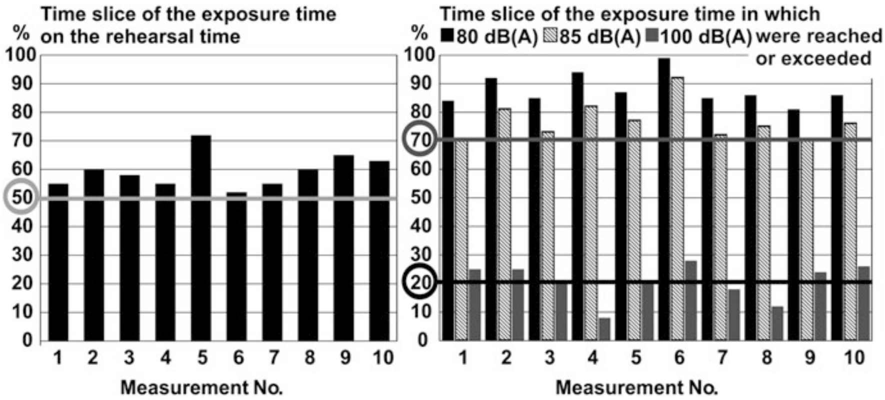


Fig. 2 Time slice of the exposure time on the rehearsal time (*left part*) and time slice of the exposure time in which different sound pressure levels were reached or exceeded (*right part*)

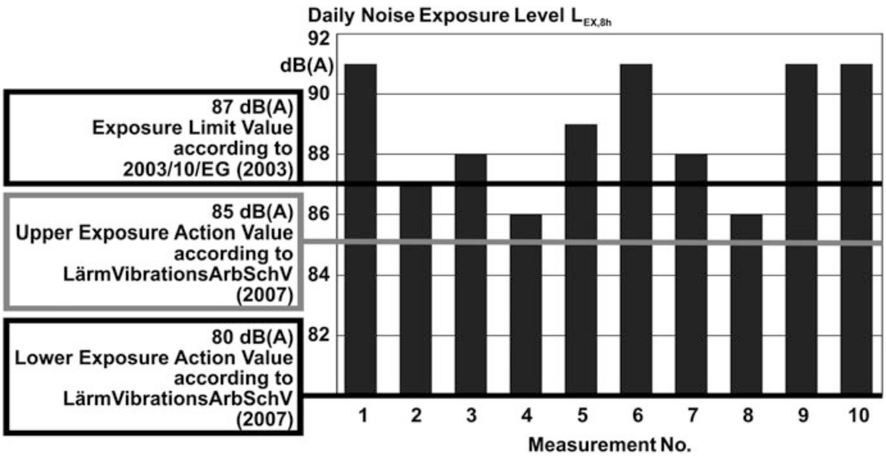


Fig. 3 Calculated daily noise exposure levels $L_{EX,8h}$ with identification of the lower and upper exposure action value (according to the Noise and Vibrations Occupational Safety and Health Ordinance [28]) as well as the exposure limit value (according to Directive 2003/10/EC [11])

sound pressure levels in $L_{EX,8h}$. Assuming that no further exposures are present in an “eight-hour shift”, the daily noise exposure level shown in Fig. 3 would result. The sample calculation shows that according to the Noise and Vibrations Occupational Safety and Health Ordinance [28] both the lower and the upper exposure action level even were exceeded considerably with values of up to 91 dB(A) in most cases.

In order to identify possible noise areas in the wind orchestra, sound pressure level measurements were carried out at various positions during the rehearsals.

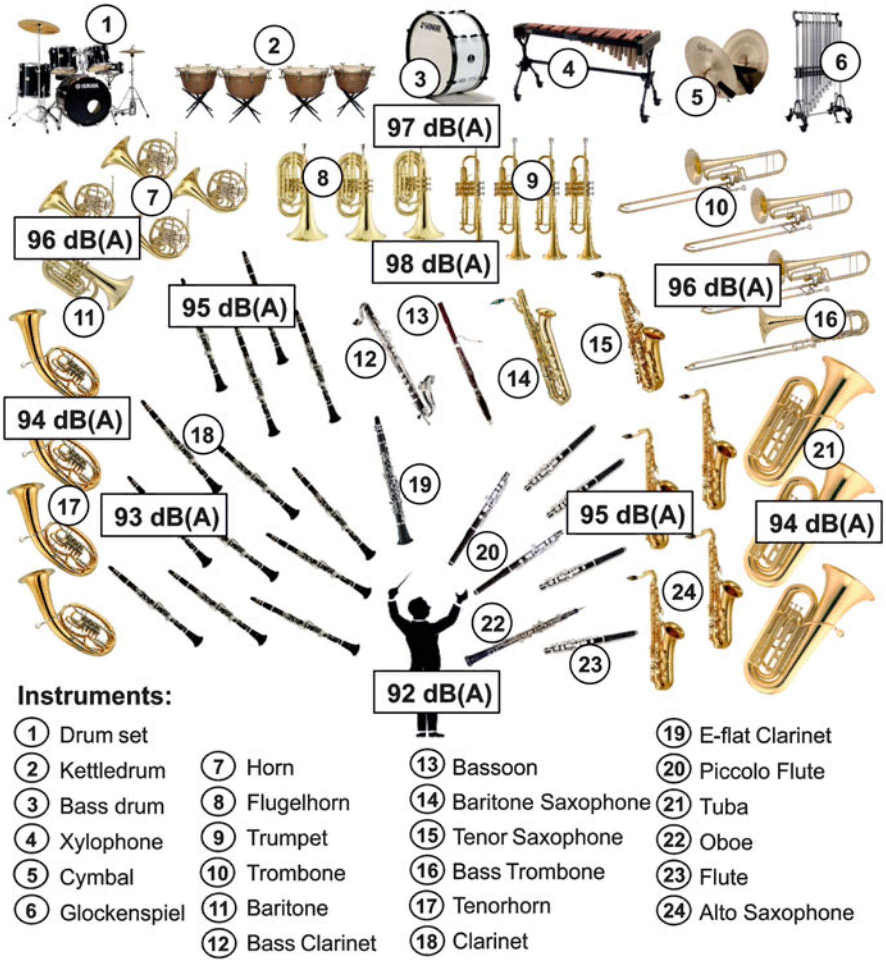


Fig. 4 Sound exposure topography of the symphonic amateur wind orchestra (based on the various measurements given in Table 1)

As illustrated in Fig. 4, a noise topography could be created based on the average equivalent continuous sound pressure level L_{Aeq}.

As can be seen, the highest levels were registered directly in front of the trumpets, particularly at the position of the woodwind players, such as bass clarinet, bassoon and baritone saxophone. Also hornists and trombonists are exposed to high levels. The hornists are probably affected by their own registers and the trombonists decisively by the drums which cause very high sound levels. Therefore, the drummers endanger not only themselves but also the trumpeters, whose heads are usually only a few centimeters away from the cymbal, the kettledrums or the glockenspiel. The saxophonists, flutists and trombonists are exposed to high L_{Aeq}

levels of 96 dB(A), too. The lowest equivalent continuous levels with values of 92 dB(A) were measured in the front area of the orchestra, close to the E-flat clarinet, the piccolo flute, the oboe and the conductor. This is probably due to their distance to the loudest instruments and a partial sound absorption by the musicians sitting in the sound field. The same is probably true for the 2nd and 3rd clarinets, whose L_{Aeq} is just one dB(A) higher. Tubists and tenor horn players sitting on the periphery of the orchestra, therefore, are mostly affected by sound pressure levels caused by their own instruments.

3.2 Results of the Audiometric Measurements

Figure 5 shows the hearing threshold of an 18-year-old drummer (upper part) and of a 22-year-old trumpeter (lower part). The examined drummer played the snare drum during the 2 h rehearsal. The concert cymbal was positioned in a short distance to his right ear. In the blue hearing curve, i.e., the curve, which was recorded prior to the rehearsal, a sharp decrease of up to 20 dB is visible in the high frequency area for the 18-year-old subject in both ears. After the rehearsal a hearing threshold shift of 5 dB—compared to the initial value—could be registered in the area between 125 Hz and 1.5 kHz. At higher frequencies (above 2 kHz) even

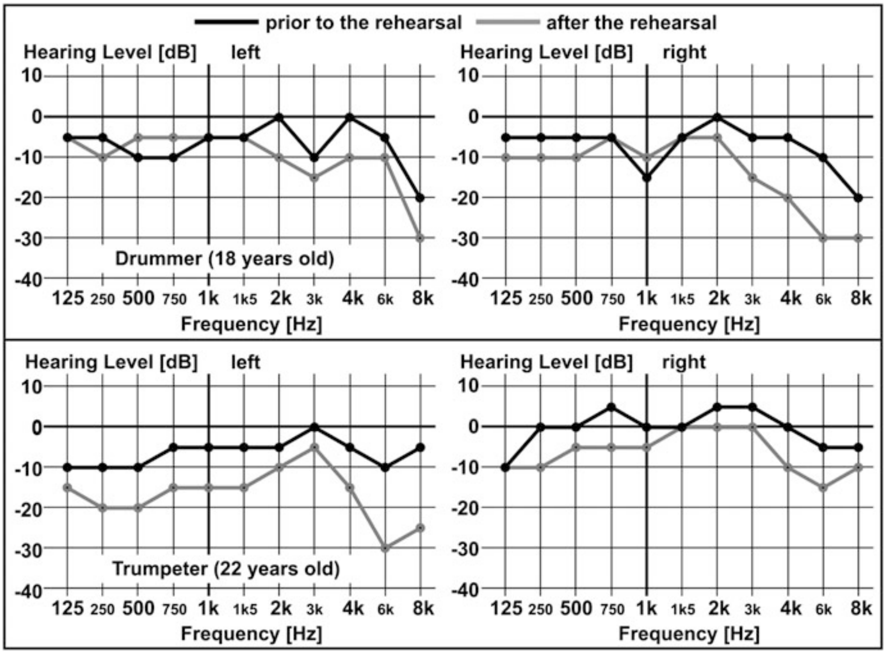


Fig. 5 Results of the audiometric measurements for an 18-year-old drummer (*upper part*) and a 22-year-old trumpeter (*lower part*) prior to and after the 2 h rehearsal

a somewhat stronger hearing threshold shift with a maximum of 10 dB was measured on the left ear. On the right ear—especially at frequencies above 2 kHz—the threshold shift was significantly higher with a maximum of 20 dB at 6 kHz. Thus, in this frequency area the examined drummer listens at least 20 dB less well than prior to the rehearsal. Overall, he barely could experience sound with a volume of 30 dB here.

In the lower part of Fig. 5 it can be seen that the trumpeter exhibits a very good hearing prior to the rehearsal in his right ear (blue curve). After the rehearsal (red curve) threshold shifts are observed in both ears. On the right ear a maximum of 10 dB was measured at frequencies of 250 Hz and 6 kHz. Clearly increased threshold shifts could be registered on the left ear with values of 10 dB for the entire area of low and mid frequencies—except to 125 Hz—, and compared to the original state with values of also 20 dB in the area of 6–8 kHz. Again, the proximity to the concert cymbals which are played only a few centimeters behind the left ear of the examined trumpeter, constitute a cause for the serious shift in the area of high frequencies.

3.3 Results of the Subjective Evaluation

Figure 6 shows the distribution of age and gender of the 56 interviewed musicians of the symphonic wind orchestra. At the time of the survey the musicians were aged between 14 and 64 years, with the major part accounted for the age group of 14–24 years. Furthermore, the orchestra comprised more male than female musicians.

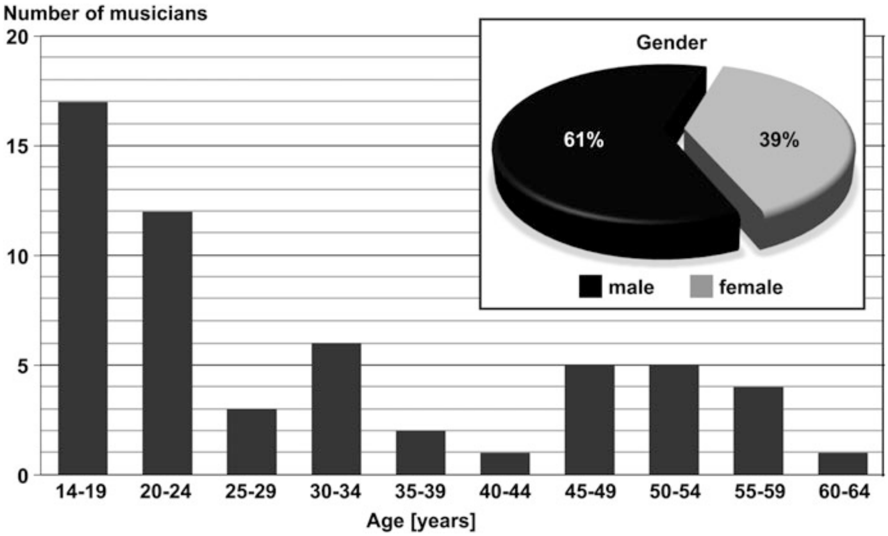


Fig. 6 Distribution of age and gender of the amateur wind orchestra

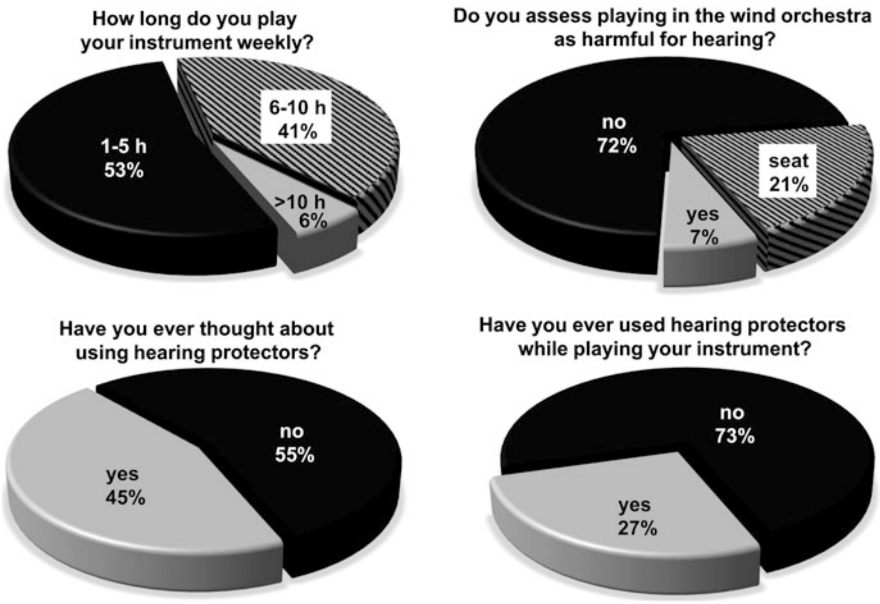


Fig. 7 Results of the subjective assessments: playing time of the instruments (*upper left part*), assessment of potential hearing risks (*upper right part*), use of hearing protection devices (*lower part*)

The upper left part of Fig. 7 shows the relative frequencies of the weekly playing time of the instrument. 41 % of the musicians play music up to 10 h a week. 6 % of the musicians even rehearse more than 10 up to even 30 h a week. As can be seen from the upper right part of Fig. 7, most of the 56 musicians assessed playing in the wind ensemble as probably not harmful for the hearing, although as many as 79 % of the respondents had experienced hearing impairments earlier (not shown here). However, for one fifth of the musicians the position in the orchestra plays an important role in this context. Taking the registered equivalent sound pressure level of more than 92 dB(A) during the rehearsals into account, it seems that the musicians are not really aware of the high levels that prevail on almost any seating position, and the potential hearing risks. Only 7 % of the respondents assessed playing music in the wind ensemble as harmful for the hearing. Due to this result it is remarkable that, after all, already 45 % of the musicians have at least thought about wearing hearing protectors, although this was implemented by only 27 % once before (cp. lower part of Fig. 7). At present, only four of the interviewed musicians wear hearing protectors while playing music.

4 Discussion

The results show that even in orchestras playing music for pleasure rather than for financial gain, sound pressure levels can be achieved, which are quite similar to those of professional ensembles. Thus, in almost every rehearsal, levels above 110 dB(A) were registered, which according to Bank [2] correspond with the noise emission of a jackhammer at a distance of 10 m.

The high levels lead to the—in some cases substantial—exceedance of the upper exposure action values of the Noise and Vibrations Occupational Safety and Health Ordinance [28] not only for the daily noise exposure level $L_{EX,8h}$ of 85 dB(A), but also for the peak sound pressure level $L_{pC,peak}$ of 137 dB(C). With an average $L_{pC,peak}$ of 130 dB(C), the peak sound pressure values for brass players of 115 dB(C) measured by the Federal Institute for Occupational Safety and Health [3] were also widely exceeded. Therefore, a clear hearing risk exists for all musicians and the conductor of a non-professional wind ensemble. With daily noise exposure levels above 85 dB(A), therefore, the musicians are located in an exposure area during their leisure time. While in the professional sector this sound pressure level would have led to applying of a corresponding catalog of protection measures, these measures are not binding for the responsible persons and the musicians in the amateur sector. However, the implementation of protective measures, such as dividing walls between the musicians or the covering of ceilings and walls with reflectors or absorbers—usually entailing high costs—is not realistic in the case of amateur music orchestras, which usually do not receive great financial support. But the use of hearing protectors or the less powerful and, therefore, at least during the rehearsals quieter interpretation of various compositions should be taken into consideration.

According to the audiometry analysis, it can be observed that temporary hearing threshold shifts could be detected in the examined musicians immediately after the rehearsals. Thereby, the characteristic of the threshold shift seems to be dependent on the proximity to strenuous high frequency sound sources and their different degrees of directional sound propagation. Due to the fact that no soundproof room was available, it should be noted that the audiometric measurements were carried out under non-ideal conditions. Therefore, low-level ambient noise could not be avoided. Despite the not controlled measurement conditions, there is a tendency, that after a correspondingly long rehearsal of the wind orchestra, clearly detectable threshold shifts occur, which can have negative impacts on the hearing at least in the long run. Therefore, the amateur musicians should provide sufficient time to recovery for their hearing after rehearsals and concerts. In the subjective evaluation, however, it could be established that to some extent the musicians are also affected by high sound pressure levels in their daily work. Rest periods, which are urgently needed for the regeneration of the hearing and the restitution of potential temporary threshold shifts (TTS), could often not be guaranteed. A problem occurs, however, when the restitution cannot be completed totally prior to the next exposure, and when marginal threshold shifts will be summed up over months and years. In the worst case, permanent threshold shifts (PTS) can be the result. For details see Strasser [35].

The evaluation of the survey showed that most of the 56 musicians were not aware of a potential hearing risk and the achieved sound pressure levels during the rehearsals. Due to the fact that currently 52 musicians permanently remain unprotected in a sound area for up to 3 h, which can lead to hearing impairments in the long run, there is a cause for concern. However, there are reasons for this negative behavior. A conventional hearing protector leads to major changes in the musicians' perception of sound and to difficulties in controlling the volume of its own instrument [27]. Theoretically, an improvement can be achieved by using professional hearing protection devices. However, due to the occlusion of the ear canal, this leads to an increased perception of bone conduction while playing music. This effect is considered as the main obstacle while playing music with ear protection, and is especially problematic for the brass section [8].

Another particularity which become obvious as a result of the questionnaire is the fact that more than half of the musicians of the amateur orchestra were younger than 25 years. This is significant because due to frequent sound exposures, a slowly and insidiously hearing impairment can adjust in young musicians. This impairment may be further increased with advancing age by the natural age-related hearing loss. Therefore, communication with the social environment may not be longer possible without restrictions. Thus, as already required by Becher et al. [4] educational measures should be implemented to raise especially younger musician's awareness for the risks of hearing impairments arising from noisy leisure activities (e.g. playing music, listening to loud music via headphones).

5 Limitations

In general, it has to be stressed that any test results might vary depending on the nature of the music work and the acoustic characteristics of the rehearsal premises. While acoustic stress could generally be observed in this study, its exact nature depending on different variables and any possible long-term effects cannot be stated at present and need to be examined in further studies.

6 Conclusions

Only professional musicians in Germany are protected by appropriate laws and regulations of the risks of high sound pressures. In contrast, however, amateur musicians receive no protection although sound pressure levels above 92 dB(A) could be determined in each rehearsal as well as at every position in an amateur orchestra. These levels are quite similar to those of professional ensembles.

Due to the duration of the rehearsal of up to 3 h, daily noise exposure levels of more than 85 dB(A) result, which means that the upper exposure action value of the Noise and Vibrations Occupational Safety and Health Ordinance [28] is exceeded.

However, the limit of 85 dB(A) only applies for the protection of employees against actual or potential risks to their health and safety through noise at their workplace. For leisure time activities, these limits are not legally binding and, therefore, no measures must be taken.

In audiometric measurements—carried out after the rehearsals—temporary hearing threshold shifts are detected which are quite obvious especially in the high-frequency area with values of 20 dB. These hearing threshold shifts can have negative impacts on the hearing in the long run if sufficient time to recovery for their hearing is not provided after rehearsals and concerts.

The results of the subjective evaluation show that due to lack of information, especially many young amateur musicians are hardly aware of the risks of loud music. Therefore, protective measures, such as e.g. the wearing of hearing protectors are mostly refused.

References

1. Axelsson A, Lindgren F (1981) Hearing in symphony classical orchestral musicians. Acta Otolaryngol 377:3–74
2. Bank M (1995) Basiswissen Umwelttechnik: Wasser, Luft, Abfall, Lärm, Umweltrecht. 3. aktualisierte und erweiterte Aufl, Vogel Buchverlag, Würzburg
3. BAuA (2010) Safe and Sound. Ratgeber zur Gehörerhaltung in der Musik- und Entertainmentbranche. 2. Aufl, Federal Institute for Occupational Safety and Health, Dortmund
4. Becher S, Struwe F, Schwenzer C, Weber K, Berger R (1931) Zur Geschichte der Lärmabwehrbewegung. Schalltechnik 1:1–9
5. Behar A, Wong W, Kunov H (2006) Risk of hearing loss in orchestra musicians: Review of the literature. Med Probl Perform Artists 21:164–168
6. Billeter T, Hohmann B (2001) Gehörbelastung von Orchestermusikern. Fortschritte der Akustik 27:386–387
7. Böhme G, Welzl-Müller K (1984) Audiometrie. Hörprüfungen im Erwachsenen- und Kindesalter. Hans Huber, Bern/Stuttgart/Wien
8. Brockt G (2004) Schallschutz bei Orchestermusikern – 87 dB(A) Expositionspegel als Grenzwert. Proceedings of the Joint Congress CFA/DAGA'04, 1:99–100
9. Brusis T (2010) Akuter Hörverlust beim Orchestermusiker. HNO 59:664–673
10. DIN 45641 (1990) Averaging of sound levels. Beuth Verlag, Berlin
11. Directive 2003/10/EC of the European Parliament and the Council of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise) (Seventeenth individual Directive within the meaning of Article 16(1) of Directive 89/391/EEC)
12. Haider M, Groll-Knapp E (1981) Psychophysiological investigations into the stress experienced by musicians in a symphony orchestra. In: Piperek M (ed) Stress and music. Wilhelm Braumüller, Vienna
13. IEC 61672 (2002) Electroacoustics – Sound level meters – Part 1: Specifications
14. Irle H, Hesse JM, Strasser H (2006) Noise: Physiological costs of noise exposures assessed via measurements of temporary threshold shifts. In: Karwowski W (ed) International encyclopedia of ergonomics and human factors, Volume III, Part 12, Methods and techniques, 2nd edn. Taylor & Francis, London, p 3115–3118

15. ISO, International Organization for Standardization (2000) Acoustics – Statistical distribution of hearing thresholds as a function of age. ISO 7029
16. ISO, International Organization for Standardization (2013) Acoustics – Determination of occupational noise exposure and estimation of noise-induced hearing impairment. ISO 1999
17. Jansson E, Karlsson K (1983) Sound levels recorded within the symphony orchestra and risk for hearing loss. Scand Audiol 12:215–221
18. Johnson D, Sherman R, Aldrige J, Lorraine A (1985) Effects of instrument type and orchestral position on hearing sensitivity for 0.25 to 20 kHz in the orchestral musician. Scand Audiol 14:215–221
19. Kähäri K (2002) The influence of music on hearing. A study in classical and rock/jazz musicians. PhD thesis, Department of Otolaryngology, Faculty of Medicine, University of Götheborg, Sweden
20. Kähäri K, Axelsson A, Hellstöm PA, Zachau G (2001) Hearing assessment of classical musicians. Scand Audiol 30:13–23
21. Kähäri K, Zachau G, Eklöf M, Sandsjö L, Möller C (2003) Assessment of hearing and hearing disorders in rock/jazz musicians. Int J Audiol 42:279–288
22. Karlsson K, Lundquist P, Olaussen T (1983) The hearing of symphony orchestra musicians. Scand Audiol 12:257–264
23. Laitinen H, Poulsen T (2008) Questionnaire investigation of musicians' use of hearing protectors, self reported hearing disorders, and their experience of their working environment. Int J Audiol 47:160–168
24. Lenzen-Schulte M (2014) Schwerhörigkeit bei Berufsmusikern. Dem Ohr ist Musik auch Lärm. FAZ (16.05.2014)
25. Lorenz-Kierakiewitz KH, Lautenbach M, Vercammen M (2011) Schallpegel auf Orchesterpodien und Möglichkeiten der Schallpegelreduktion. Musikphysiologie und Musikermedizin 18(3):79–81
26. Maue JH (2009) Bedeutung des Spitzenschalldruckpegels für die Beurteilung industrieller Arbeitsplätze. Sicherheitsingenieur 8:52–55
27. Oberdanner H, Reintges F, Welzl-Müller K (2002) Persönlicher Gehörschutz für Musiker. Proceedings of DAGA'02, Fortschritte der Akustik, Oldenburg, p 56–57
28. Ordinance on the protection of workers against the risks arising from exposure to noise and vibration (2007) Noise and Vibrations Occupational Safety and Health Ordinance (LärmVibrationsArbSchV) of 06 March 2007 (Federal Law Gazette I, p 261), last amended by Article 3 of the Ordinance of 19. July 2010 (Federal Law Gazette I, p 960)
29. Ostri B, Eler N, Dahlin E, Skylv G (1989) Hearing impairment in orchestral musicians. Scand Audiol 18:243–249
30. Palonen J, Huttunen K, Sorri M (2001) Plexiglass barriers used in hearing protection proved to serve as sound reflectors in military band. Proceedings of the 5th EFAS Congress, Bordeaux
31. Pangert R, Loock F (2004) Musikermedizin, Musikerarbeitsplätze. Eine Einführung für Orchestermusiker, Musikpädagogen und Studenten. GUV-Informationen 8626, Bundesverband der Unfallkassen
32. Richter B, Zander M, Spahn C (2007) Gehörschutz im Orchester. In: Spahn C (ed) Freiburger Beiträge zur Musikermedizin, Bd 4. Projektverlag, Freiburg/Bochum
33. Royster J, Royster L, Killion M (1991) Sound exposures and hearing thresholds of symphony orchestra musicians. J Acoust Soc Am 89:2793–2803
34. Schmidt JM, Verschuure J, Brocaar MP (1994) Hearing loss in students at a conservatory. Audiology 33:185–194
35. Strasser H (ed) (2005) Traditional rating of noise versus physiological costs of sound exposures to the hearing. Ios Press, Amsterdam

36. Strasser H, Irle H, Legler R (2005) Temporary hearing threshold shifts and restitution associated with exposures to industrial noise and classical music of 94 dB(A) for 1 hour and 91 dB(A) for 2 hours. In: Strasser H (ed) Traditional rating of noise versus physiological costs of sound exposures to the hearing. Ios Press, Amsterdam, pp 137–148
37. Strasser H, Irle H, Scholz R (1999) Physiological cost of energy-equivalent exposures to white noise, industrial noise, heavy metal, and classical music. Noise Control Eng J 47(5):187–192
38. Toppila E, Laitinen H, Pyykkö I (2005) Die Auswirkungen von Lärm auf klassische Musiker. Magazine 8 – Lärm bei der Arbeit (Magazin der europäischen Agentur für Sicherheit und Gesundheitsschutz am Arbeitsplatz), p 21–22

Effects of an Exercise Intervention on Metabolism and Quality of Life in Patients with Diabetes Mellitus

Andree Hillebrecht, Pascal Bauer, Torsten Frech, Rüdiger Walscheid, Silvia Linnenweber, Deborah Wyss, Claudia Barthelmes, Karsten Krüger, Gerd Willmund, Frank-C. Mooren, Reinhard Nöring, and Sven Zeißler

Abstract

Diseases of civilisation, such as type 2 diabetes mellitus, play an increasingly important role in shaping the occupational health management strategy. According to the guidelines of the DDG e.V. [German Diabetes Association],

A. Hillebrecht (✉)
Medical Department, Volkswagen AG, Baunatal, Germany

Department of Sports Medicine, JL University Giessen, Giessen, Germany
e-mail: a.hille@gmx.de

P. Bauer
Medical Clinic I – Cardiology, University Hospital of Giessen and Marburg, Giessen, Germany

T. Frech • K. Krüger • F.-C. Mooren
Department of Sports Medicine, JL University Giessen, Giessen, Germany

R. Walscheid
Laboratory Medical Services Centre, Koblenz, Germany

S. Linnenweber
Hannover Medical School- Department of Nephrology, Hannover, Germany

D. Wyss
Lucerne Psychiatry, Outpatient Service, Lucerne, Switzerland

C. Barthelmes
Occupational Health Department, City of Munich, Munich, Germany

G. Willmund
Bundeswehrkrankenhaus Berlin, Zentrum für Psychiatrie und Psychotraumatologie, Berlin, Germany

R. Nöring
Medical Department, Volkswagen AG, Baunatal, Germany

S. Zeißler
Comenius University, Bratislava, Slovakia

Sport Park Zwickau, Glauchau, Meerane, Germany

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_12

regular physical exercise is the basic therapy for type 2 diabetes mellitus. In this study, in a total of 110 subjects with diabetes mellitus, it was shown that a regular, intensively supervised, exercise intervention over a period of 6 months led to a significant improvement in the glucose metabolism (reduction in hemoglobin A1c). In addition, an improvement in the health-related quality of life was observed. A twice-weekly exercise intervention carried out within the framework of occupational health management [OHM] thus appears to represent a sensible therapeutic option and a sensible supplement to the treatment of this disease of civilisation.

Keywords
Exercise • Diabetes mellitus • Quality of life • Hemoglobin A_{1c} • HbA1c

1 Introduction

Diabetes mellitus is a serious disease of civilisation, and its incidence is increasing throughout the world. In the process, its importance, not just from a sociopolitical and economic viewpoint, but increasingly for companies as well, is rising substantially. In Germany, the prevalence of type 2 diabetes mellitus in adults between 18 and 79 years of age is said to be 7.4 % in women and 7.0 % in men [3]. In the Atlas of the International Diabetes Federation (2011), the prevalence in Germany is said to be 8.0 % (20–79 years of age). Furthermore, it can be assumed that the number of cases of undiagnosed type 2 diabetes is high [8, 11]. In some studies the number of undiagnosed cases was nearly as high as the number of diagnosed patients.

The implementation of secondary prevention e.g. in the form of screening tests for the detection of as yet undiagnosed diabetes mellitus as part of a voluntary health check should, therefore, definitely be considered in the context of occupational health management.

A lifestyle intervention with regular exercise plays a substantial therapeutic role in the treatment of type 2 diabetes mellitus. The beneficial effects on the metabolism and cardiopulmonary exercise capacity have already been demonstrated in several studies [6, 7, 12] and have already been incorporated in the German and American treatment guidelines [1, 2].

The guidelines of the German Diabetes Association (Deutsche Diabetes Gesellschaft e.V.- DDG) recommend 150 min of exercise per week for patients with impaired glucose tolerance and 300 min of exercise per week for patients with diagnosed diabetes mellitus. However, the best type, intensity, and duration of training have not yet been definitively established and still remain unclear. In this context, occupational health management does not just focus on improving physical health, but also on the holistic wellbeing of employees.

This study therefore investigates the question of the extent to which a 6-month exercise intervention can achieve changes in the metabolic variables, the health-related quality of life and selected behaviours of patients with type 2 diabetes mellitus and the extent to which it is possible to derive recommendations for optimal occupational health management from the results.

2 Materials and Methods

110 patients with non-insulin-dependent type 2 diabetes mellitus were randomly assigned to one of three different exercise intervention groups or a waiting control group [WCG]. Participants and trainers could not be feasibly be blinded to group assignment after randomization, but the main study outcomes were measured by blinded technologists using objective methods.

The groups are described in Table 1.

The exercise intervention groups completed 6 months of standardised and intensively supervised training. The nature of the training in the different exercise intervention groups differed. There was an aerobic training group [AT], an equipment-supported resistance training group [RT] and a combination training group [CO]. The amount of training time was the same in all three intervention groups. The training exercises are described in Table 2. The training took place in three different sport parks in Zwickau, Glauchau and Meerane in Germany.

The exercise capacity of the subjects was determined at the start and after 3 months on the basis of clinical examinations (a cardiopulmonary exercise test and a strength test) and exercise recommendations were made on the basis of these values.

The metabolic variable hemoglobin A_{1c} (HbA1c) was determined at the start (MT1), after 3 months (MT2) and after 6 months (MT3) in venous blood.

The health-related quality of life was recorded at the MT1, MT2 and MT3 time points, using both the standardised EQ-5D the SF-12 questionnaires. The SF-12 is a multipurpose short form survey with 12 questions, all selected from the SF-36 Health Survey [9, 13]. The questions were combined, scored, and weighted to create two scales that provide glimpses into mental and physical functioning and overall health-related-quality of life. Both questionnaires included following subdomains: general health, physical functioning, role functioning (physical), bodily pain, vitality, role functioning (emotional), mental health, social functioning. The SF-12 is weighted and summed into well interpretable scales for physical

Table 1 Overview of the number of subjects, dropouts, gender and age in the groups

	Group	N	Dropouts (N)	Gender (m/f)	Age at MT1 (years)
Resistance training		27	7	14/13	60.0 ± 8.7
Aerobic training		30	4	15/15	60.8 ± 8.6
Combination group		25	4	15/10	59.8 ± 8.2
Intervention groups overall		82	15	44/38	60.2 ± 8.4
Waiting control group		28	6	15/13	63.8 ± 8.7

Table 2 Overview of the training content in the individual groups

Time period (weeks)	Resistance training group	Aerobic training group	Combination group
1–4	2× 1 circuit, 8 items of exercise/week Gentle to moderate*	2× 15 min/week 80–100 % of the HR at VT1 (CPX)	1 unit/week each of resistance training and aerobic training
5–13	2× 2 circuits, 8 items of exercise/week Moderate*	2× 30 min/week 95–110 % of the HR at VT1 (CPX)	Half the extent of the resistance training and aerobic training per unit
14–26	2× 3 circuits, 8 items of exercise/week Moderate*	2× 45 min/week 95–110 % of the HR at VT1 (CPX)	1× 2 circuits resistance training (30 min) and 15 min aerobic training 1× 1 circuit resistance training (15 min) and 30 min aerobic training

*According to the seven-level intensity scale for strength training (*Source*: Buskies and Boeckh-Behrens 1996)

and mental health, called Physical and Mental Health Composite Scores (PCS & MCS).

EQ-5D is primarily designed for self-completion by respondents and is suited for use in postal surveys, in clinics and face-to-face interviews [10]. The EQ-5D measures quality of life in five dimensions: mobility, usual activities, self-care, anxiety/depression, pain/discomfort.

All statistical analyses were carried out with the program SPSS, version 19, or the SAS 9.2 software package (SAS Institute Inc., Cary, NY, USA) and a deliberate decision was made to use the full 5 % significance level, i.e. no correction was made for multiple tests and each p-value ≤ 0.05 corresponds to a significant result.

In addition to the descriptive statistics, significance testing was performed. To do this, the variable was first tested for normal distribution and, according to the present normal distribution, significance testing was performed by means of a *T*-test for dependent or independent samples. The inter-group comparison was carried out using a multivariate analysis of variance.

For the variables shown, the intervention group [IG] as a whole was compared with the waiting control group [WCG]. For the variables recorded with the aid of the questionnaires, a comparison of the individual intervention groups RT, AT and CO was also carried out.

3 Results

3.1 Changes in the Metabolic Variable HbA1c

The specific metabolic variable HbA1c, for which the baseline values were low overall in both groups, showed a decrease in the IG and an increase in the WCG after the 6-month intervention (see Table 3).

The inter-group comparison of the IG and WCG in respect of the change in HbA1c between MT1 and MT3 showed a significant difference (see Fig. 1).

3.2 SF-12 Questionnaire

3.2.1 Comparison Between the Intervention Group and the Waiting Control Group

The physical sum scale of the SF-12 questionnaire showed an increase in the IG and a reduction in the WCG between MT1 and MT3 (see Fig. 2). The inter-group comparison showed a significant difference between the two groups ($p = 0.040$).

An increase in the mental sum scale was observed in both the IG and the WCG (see Fig. 3). Overall, the difference between the changes in the two groups was not significant.

3.2.2 Comparison of the Exercise Intervention Groups

The comparison of the changes in the individual exercise intervention groups RT, AT and CO did not reveal any significant differences between the groups on either the physical sum scale or the mental sum scale.

Table 3 Overview of the mean value and standard deviation of HbA1c in mmol/l at MT1 and MT3 and the difference between MT3 and MT1 in the IG and the WCG

	Group	N	Mean	Standard deviation
HbA1c MT1 in mmol/l	Intervention group	66	51.69	10.47
	Waiting control group	22	50.52	8.53
HbA1c MT3 in mmol/l	Intervention group	66	50.96	9.92
	Waiting control group	22	53.11	10.08
Change in HbA1c (mmol/l) MT3 minus MT1	Intervention group	66	−.73	7.00
	Waiting control group	22	2.58	5.79

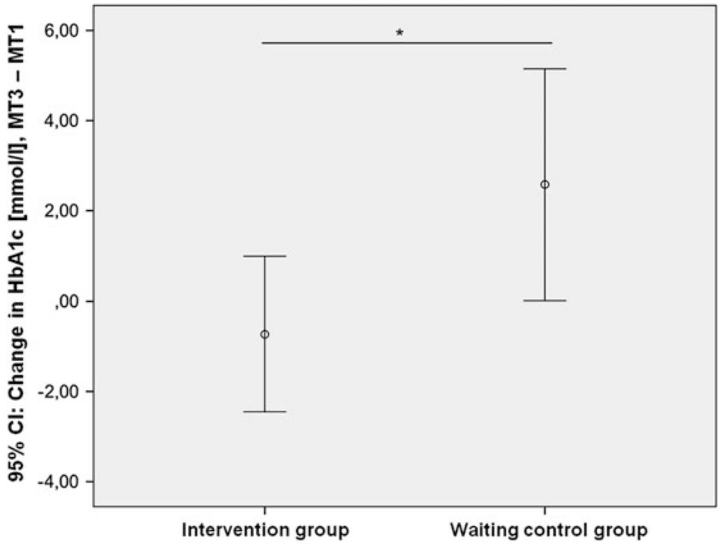


Fig. 1 Change in HbA1c in the IG and the WCG between MT1 and MT3

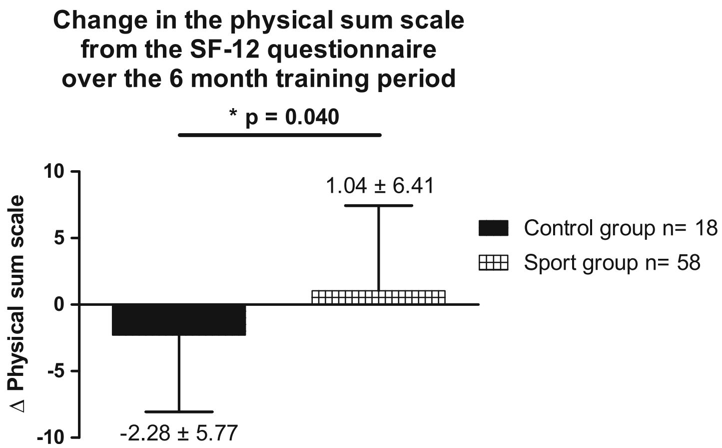


Fig. 2 Change in the SF-12 physical sum scale between MT1 and MT3 in the IG and the WCG

3.3 EQ-5 Questionnaire

3.3.1 Visual Analogue Scale

The results for the EQ-5D VAS (visual analogue scale from 0 to 100 points) variables provide a subjective assessment of the state of health.

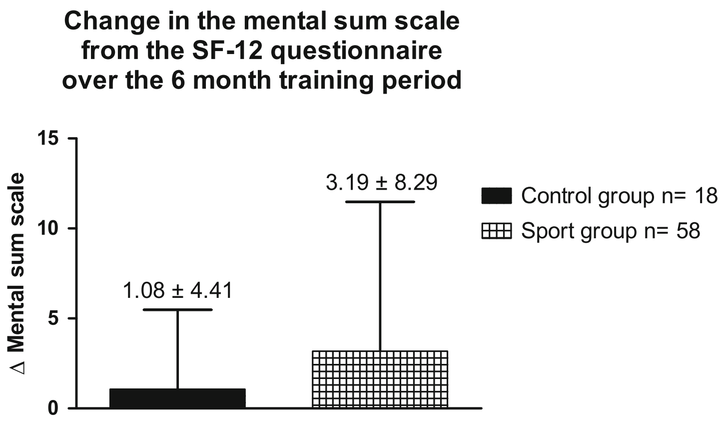


Fig. 3 Change in the SF-12 mental sum scale between MT1 and MT3 in the IG and the WCG

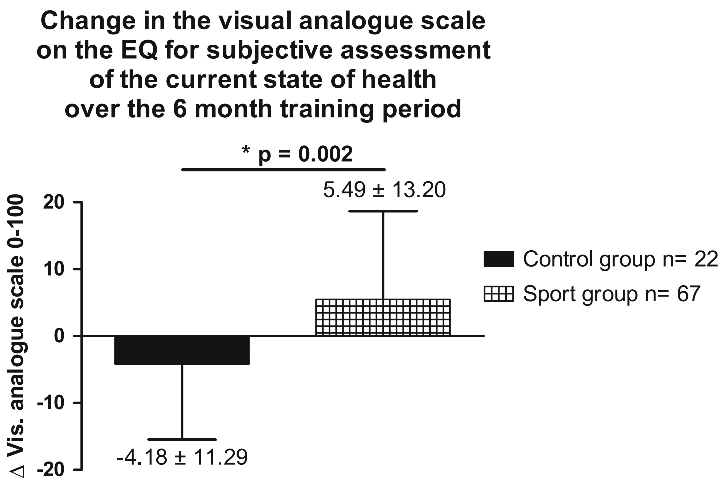


Fig. 4 Change in the VAS on the EQ-5D in the IG and the WCG between MT1 and MT3

Comparison Between the Intervention Group and the Waiting Control Group
In the EQ-5D VAS, there was an increase in the IG and a decrease in the WCG. The change over the intervention period showed a significant difference between the two groups, with $p = 0.002$ (see Fig. 4).

Comparison of the Exercise Intervention Groups
The comparison between the changes in the EQ-5D VAS in the individual exercise intervention groups (RT, AT and CO) did not reveal any significant differences.

3.3.2 Five Dimensions of Quality of Life

The fields covered in EQ-5D, "Mobility", "Self-care", "Usual activities", "Pain/discomfort" and "Anxiety/depression" are shown in detail for all groups at MT1 and MT3 in Tables 4, 5, 6, 7, and 8.

Table 4 Overview of the results of the questionnaire for the mobility field as totals and percentages in the different groups for MT1 and MT3

	MT	IG	AT	RT	CO	WCG
Mobility						
No problems in walking about	*1*	*47 (79.7)*	*16 (76.2)*	*17 (85.0)*	*14 (77.8)*	*15 (83.3)*
	3	51 (86.4)	18 (85.7)	18 (90.0)	15 (83.3)	14 (77.8)
Some problems in walking about	1	12 (20.3)	5 (23.8)	3 (15.0)	4 (22.2)	3 (16.7)
	3	8 (13.6)	3 (14.3)	2 (10)	3 (16.7)	4 (22.2)
Confined to bed	1	0	0	0	0	0
	3	0	0	0	0	0

Table 5 Overview of the results of the questionnaire for the self-care field as totals and percentages in the different groups for MT1 and MT3

	MT	IG	AT	RT	CO	WCG
Self-care						
No problems with self-care	*1*	*58 (98.3)*	*20 (95.2)*	*20 (100)*	*18 (100)*	*18 (100)*
	3	59 (100)	21 (100)	20 (100)	*18 (100)*	*18 (100)*
Some problems with washing and dressing myself	1	1 (1.7)	1 (4.8)	0	0	0
	3	0	0	0	0	0
Unable to wash or dress myself	1	0	0	0	0	0
	3	0	0	0	0	0

Table 6 Overview of the results of the questionnaire for the usual activities field as totals and percentages in the different groups for MT1 and MT3

	MT	IG	AT	RT	CO	WCG
Usual activities						
No problems performing my usual activities	*1*	*48 (81.4)*	*18 (85.7)*	*17 (85.0)*	*13 (72.2)*	*16 (88.9)*
	3	54 (91.5)	19 (90.5)	20 (100)	15 (83.3)	15 (83.3)
Some problems performing my usual activities	1	11 (18.6)	3 (14.3)	3 (15.0)	5 (27.8)	2 (11.1)
	3	5 (8.5)	2 (9.5)	0	3 (16.7)	3 (16.7)
Unable to perform my usual activities	1	0	0	0	0	0
	3	0	0	0	0	0

Table 7 Overview of the results of the questionnaire for the pain/discomfort field as totals and percentages in the different groups for MT1 and MT3

	MT	IG	AT	RT	CO	WCG
Pain/Discomfort						
No pain or discomfort	*1*	*28 (47.5)*	*10 (47.6)*	*9 (45.0)*	*9 (50.0)*	*10 (55.6)*
	3	35 (59.3)	12 (57.1)	13 (65.0)	10 (55.6)	8 (44.4)
Moderate pain or discomfort	1	29 (49.1)	10 (47.6)	10 (50.0)	9 (50.0)	8 (44.4)
	3	23 (39.0)	8 (47.1)	7 (35.0)	8 (44.4)	10 (55.6)
Extreme pain or discomfort	1	2 (3.4)	1 (4.8)	1 (5.0)	0	0
	3	1 (1.7)	1 (4.8)	0	0	0

Table 8 Overview of the results of the questionnaire for the anxiety/depression field as totals and percentages in the different groups for MT1 and MT3

	MT	IG	AT	RT	CO	WCG
Anxiety/Depression						
Not anxious or depressed	*1*	*52 (88.1)*	*18 (85.7)*	*18 (90.0)*	*16 (88.9)*	*15 (83.3)*
	3	50 (84.7)	18 (85.7)	17 (85.0)	15 (83.3)	15 (83.3)
Moderately anxious or depressed	1	7 (11.9)	3 (14.3)	2 (10.0)	2 (11.1)	2 (11.1)
	3	9 (15.3)	3 (14.3)	3 (15.0)	3 (16.7)	3 (16.7)
Extremely anxious or depressed	1	0	0	0	0	1 (5.6)
	3	0	0	0	0	0

4 Discussion

Diseases of civilisation such as diabetes mellitus are increasing throughout the world. There seems to be a large number of undiagnosed cases of type 2 diabetes mellitus in Germany. Secondary prevention within the framework of OHM can therefore be regarded as particularly promising, and OHM is an important element in the fight against diabetes mellitus.

The question of the benefits of OHM in the treatment of diabetes mellitus also needs to be considered. Regular exercise is the basic therapy for type 2 diabetes mellitus, as for many diseases of civilisation. In the case of diabetes mellitus, these recommendations are incorporated in the current guidelines.

The amount of exercise in our study was substantially below the levels recommended in the DDG e.V. guidelines. Despite this, the exercise intervention that was implemented was found to produce an improvement in HbA1c, as an expression of an improved glucose metabolism, after just 6 months.

The reduction in HbA1c was less than in studies with a greater amount of training [6, 7, 12], which could be directly related to the amount of training. However, this effect was also observed in comparison with two studies performed by our study group [4, 5]. In these studies a larger decrease in HbA1c was observed with exercise intervention on only 2 days per week. Thus, the possibility that the relatively small improvement in HbA1c was due to the low baseline level of HbA1c should also be considered. Furthermore, after the 4-week initial training phase, the increase in the exercise capacity in the intervention protocol groups was slow. This could also lead to a smaller reduction in HbA1c.

The positive influence of the exercise intervention on the health-related quality of life was demonstrated in virtually all the variables measured using the EQ-5D and SF-12 questionnaires. No difference between the individual exercise intervention groups was detected. Thus, all the forms of training investigated appear suitable for occupational health management and can be recommended.

In summary, it was found that a moderate, intensively supervised exercise intervention in patients with diabetes mellitus produces positive effects on the metabolism and the quality of life. It seems perfectly feasible to implement exercise therapy of this sort within the framework of OHM. From the viewpoint, in particular, of social responsibility and the long-term use of an optimal occupational health management strategy for the companies as the basis for long-term economic success, the targeted implementation of exercise intervention within OHM should be expedited.

References

1. American Diabetes Association (2012) Standards of Medical Care in Diabetes – 2012. Diabet Care 35:S11–S63
2. Halle M, Kemmer FW, Stumvoll M, Thurm U, Zimmer P (2008) 30. Körperliche Aktivität und Diabetes mellitus. Evidenzbasierte Leitlinie der Deutschen Diabetes-Gesellschaft [Physical activity and diabetes mellitus. Evidence-based guideline of the German Diabetes Association]. Deutsche Diabetes Gesellschaft
3. Heidemann C, Du Y, Schubert I, Rathmann W, Scheidt-Nave C (2013) Prevalence and temporal trend of known diabetes mellitus: results of the German Health Interview and Examination Survey for Adults (DEGS1). Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 56:668–677
4. Hillebrecht A, Zeißler S, Krüger K, Clemm S, Rechner M, Mooren F-C, Hamar D (2011) Effekte einer 6-monatigen Sportintervention bei Patienten mit Diabetes mellitus Typ 2 – Hypertrophietraining versus Kraftausdauertraining – [Effects of a 6-month sport intervention in patients with type 2 diabetes mellitus – hypertrophy training versus muscular endurance training –]. Deutsche Zeitschrift für Sportmedizin 62(7–8):201
5. Hillebrecht A, Zeißler S, Hellmann S, Frech T, Rechner M, Walscheid R, Linnenweber S, Menne J, Hamar D, Mooren F-C (2012) 6-monatige Bewegungsstudie im Rahmen des Reha-Sports für internistische Erkrankungen (Tertiärprävention) bei Typ 2 Diabetikern mittels Untersuchung therapierelevanter Variablen [6-month exercise study within the framework of sports rehabilitation for internal illnesses (tertiary prevention) in type 2 diabetics by investigating treatment-relevant variables]. Diabetologie und Stoffwechsel 7(S1):66
6. Kadoglou NPE, Fotiadis G, Kapelouzou A, Kostakis A, Liapis CD, Vrabas IS (2013) The differential anti-inflammatory effects of exercise modalities and their association with early carotid atherosclerosis progression in patients with Type 2 diabetes. Diabet Med 30:41–50
7. König D, Deibert P, Dickhut HH, Berg A (2011) Resistance exercise and type 2 diabetes mellitus. Deutsche Zeitschrift für Sportmedizin 62(1):5–9

8. Meisinger C, Strassburger K, Heier M, Thorand B, Baumeister SE, Giani G, Rathmann W (2010) Prevalence of undiagnosed diabetes and impaired glucose regulation in 35–59-year-old individuals in Southern Germany: the KORA F4 Study. Diabet Med 27(3):360–362
9. Morfeld M, Kirchberger I, Bullinger M (2011) SF-36. Fragebogen zum Gesundheitszustand. Testmanual. Hogrefe
10. Rabin R, de Charro F (2001) EQ-D5: a measure of health status from the EuroQol Group. Ann Med 33(5):337–343
11. Rathmann W, Haastert B, Icks A, Löwel H, Meisinger C, Holle R, Giani G (2003) High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. Diabetologia 46:182–189
12. Sigal RJ, Kenny GP, Boulé NG et al (2007) Effects of aerobic training, resistance training, or both on glycemic control in type 2 diabetes. Ann Intern Med 147:357–369
13. Ware JE, Kosinski M, Keller SD (1996) A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 34:220–233

Part II

Design of Products

An Integrated Approach of Mental Workload Assessment

Marc Schneider and Barbara Deml

Abstract

There are a lot of methods and instruments trying to assess mental workload reliable and independent from specific use cases or situations. Each method has its own strengths and weaknesses depending on the situation and the on individual person. Attempting to improve each approach separately would indeed allow a further context of use, but do not lead to a global reliability growth. Within this work a new approach is suggested by combining various ocular parameters adapting the individual person. The results of this study show clearly the advantages in mental workload assessment of an individual adapted set of variables in contrast to one variable trying to fit a collectivity of people. This new preparation considering six physiological ocular variables in two different contexts of use can be one step further to a more reliable and global way to assess mental workload.

Keywords

Mental workload • Eye-tracking • Pupil dilations • Nearest neighbor index • Blink rate

M. Schneider (✉)
Institut für Arbeitswissenschaft und Betriebsorganisation, Karlsruher Institut für Technologie, Karlsruhe, Germany
e-mail: marc.schneider3@kit.edu

B. Deml
ifab-Institute of Human and Industrial Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: barbara.deml@kit.edu

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_13

1 Introduction

The reliable assessment of mental workload is both a fundamental and challenging requirement for a user-centered, adaptive workplace design in human-machine-interaction. This is particularly true against the context of older workforce and the increasing complexity of systems. There is already a diversified portfolio of approaches promising the best possible assessment of mental workload which rely on different procedures of subjective, performance as well as physiological measurements. Each of these approaches has certain strengths and weaknesses concerning its reliability [21]. This variability can be assigned to different contexts of use or various experimental settings.

The mental workload itself is a very complex phenomenon which is widely analyzed theoretically as well as empirically for many years until today [13]. There are lots of models and constructs trying to describe cognitive processing and mental workload which are subject to constant expansion, upgrades or modifications [14, 17, 29]. One approach to describe mental workload is based on the resource theory. A human has an individual amount of cognitive capacities to handle different tasks with a specific demand to these resources. On the one hand the demand varies depending on several factors like the time on task, the modalities or the level of complexity and on the other hand the capacity is individually different. Each individual has different resources depending not only on his or her constitution, but also on level of training on the specific task. If the demands exceed the capacity of resources available for a specific task it comes to an overload as well as the other way round too low demands can lead to underload [39, 40]. This approach gives one example of how mental workload and its formation can be described. Besides the above mentioned and often in studies for workload assessment named models for cognitive processing and workload there is another essential theoretical construct considered in this work which finds relatively little application in human factors. In high stress or high workload situations identified by subjective, objective or performance (secondary tasks) measures, the individual workload depends on the particular coping strategy of the human [12, 27, 30]. In general the individual coping can be divided into two modes, the active and the passive coping [30]. The active coping contains conscious adaptions of behavior like changing the amount of effort spent or the prioritized task in a dual-task configuration. Changes in processing strategies or the ignorance of certain task components can also be assigned to the active coping [27]. Passive coping includes the unconscious adaption such as the regulation of the information flow and processing. All in the entire individual coping considers both external (e.g. time pressure, negative feedback) and personality (e.g. experiences, constitution, level of skill) factors [27]. The type of coping depends among others on the subjective perceived level of task-induced demand. Robert and Hockey [30] introduced a two level system. Up to an individual depending amount of demand the human is able to cope with workload passively whereas demand higher than this point can only be regulated by active coping [30]. Nevertheless it is not possible to predict the coping strategies from the task-induced demand alone due to the individual perception and handling with the situation. The individual coping is time-dependent and it can also change over short intervals of time making a reliable prediction of workload additionally difficult [30].

In this work six correlates (Blink Rate, Blink Duration, PERCLOS, Pupil Dilations, Fixation Duration and Nearest Neighbor Index) for mental workload assessment, that are based on eye-tracking data compared in two different scenarios with each three levels of difficulty. Within the first setting a flight simulator task is regarded, whereby the autopilot is defect, so that both the altitude and the heading are to be kept manually. The stress level is varied by the presence of a secondary task: while there was no secondary task in one condition, the participants had to do either an acoustic or a visual 1-back secondary task in the further two conditions. For the 1-back task the subjects have to repeat verbally the last number each time a new one is presented in the both mentioned modalities [19]. Within the second setting a mental rotation task is studied in which two geometries have to be checked for equality. Here the stress level is varied by the number of contour points. As both experimental settings require different cognitive efforts, the workload assessment can be tested in two different contexts of use. The flight simulator task is characterized by a dynamic situation with simultaneous tasks whereas the mental rotation is a sequential and one dimensional setting. Besides procedures regarding the pupil an eye behavior like the pupil dilations and the blink rate the study considers also the gaze strategy operationalized by the nearest neighbor index. The NASA-TLX questionnaire is used to evaluate the subjective perception of the effort needed in the tasks.

The reminder of the article is structured as follows: At first each method is explained theoretically before introducing the experimental setting. Afterwards the results were described before closing with a discussion.

2 Workload Measurement Methods

Commonly three different categories of workload measurement techniques are distinguished: performance measures (e.g. dual-task paradigm) [40], subjective rating scales [31], and psychophysiological measures [21]. Depending on the research question, different measurement instruments may be used for each of the three categories. Within this section, only the last two categories shall be regarded, whereby as self-assessment technique NASA Task Load Index (NASA-TLX) is used [16]. This questionnaire is rather widespread as it is economical, clearly structured and easy to understand [31]. NASA-TLX consists of six items (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration Level), whereas the item “physical demand” is not relevant for our study and thus, it was not regarded here.

Psychophysiological measures (e.g. heart rate, skin conductance and electroencephalography) examine in an objective and continual way what effect mental workload has on an operators’ body. Despite of these advantages the methods are quite often more general indicators of stress and thus, it is of particular importance to assure that they are relevant for the research question at hand. In our study only ocular parameters shall be regarded because of the close linkage of the eyes to the information processing regions of the human [37]. Due to the mentioned problem of sensitivity, we decided to analyze various ocular parameters that differ in terms of temporal sensitivity—the blinking behavior Sect. 2.1, the pupil dilations Sect. 2.3, the fixations duration Sect. 2.4, as well as gaze strategies Sect. 2.5. While, for instance, pupil dilations distinguish by a high temporal resolution, other variables, such as gaze behavior have to be aggregated over time.

Due to the time dependency of the individual coping strategies [30], the following explained parameters are calculated with different resolutions having diverse sensitivities to fluctuations over time. In the case of NNI and pupil dilations there are the mean, median, maximum value and minimum value considered in this study. Especially for the pupil dilations it is additionally possible to calculate the mean of the first values which is equivalent to the first 30 s of the task. On this way it might be possible to detect individual changes with respect to coping strategies.

2.1 Blink Rate

Commonly, an eye blink takes about 70–100 ms [22] and the *blink rate* describes the number of eye closures in a pre-defined period of time. Usually, this time period is set to 1 min, but in the case of very short or variable tasks an interval of 30 s may be analyzed too [41]. The blink rate is often used to measure mental workload. Wilson [41], for example, showed that there is a significant negative relationship between the blink rate and the increasing difficulty of certain temporal segments of a flight in a flight simulator. However, for example, Cardona and Quevedo [7] could not find any significant correlation between the blink rates of car drivers and their performance at different complexity levels. This may also be due to the fact that blink rate is not only influenced by the mental workload, but it is also altered by other user states, such as the degree of sleepiness [38]. There is a seamless continuation between a blink rate of a non-sleepy user and that of a sleepy user, whereby a blink rate of longer than 500 ms may be referred to as micro-sleep [33]. Thus, it is reasonable to combine the blink rate with other indicators of mental workload in order to be more confident that the observed differences are really due to the workload of the user and that they are not caused by other factors.

2.2 Blink Duration and PERCLOS

As shortly mentioned above the *blink duration* can take between 70 and 500 ms. Benedetto et al. [4] suggest that the blink duration is even more sensitive to the cognitive workload than the blink rate. They tested both parameters in a lane change task with two conditions, control condition and dual task condition. Their results showed that the blink duration was shorter for more difficult tasks, whereby no significant effect could have been found for the parameter blink rate [4]. Martins and Carvalho [26], too, analyzed the blink characteristics of users in the context of mental workload and fatigue. In their study the parameter *PERCLOS* (percentage of eye closure) was derived, which is a rather well established parameter to predict fatigue and which is defined as the accumulated blink duration on a pre-defined time interval [1]. Brookhuis and de Waard [6] also assessed PERCLOS and they have shown that this parameter is significantly correlated to the performance of car drivers in a driving simulation. Due to its similarity to the blink duration and the promising results of former studies, the parameter PERCLOS is also taken into account for mental workload assessment in this study.

2.3 Pupil Dilations

The pupil of the human eye may be compared to an optic lens, which changes its size in order to regulate the incidence of light into the eye. The pupil dilations are controlled by two antagonistic muscles surrounding the pupil: the Sphincter Pupillae, which scales down the pupil size para-sympathetically, and the Dilator Pupillae, which scales up the pupil size sympathetically [3]. Some changes in pupil size, the so called light reflexes, adapt the pupil to the light conditions. In bright environments the pupil is small and contracted whereas in dark conditions the pupil is wide and relaxed. This adaption is characterized by smooth and lower frequent fluctuations in pupil size. Beside the lighting conditions, the pupil size is also sensitive to the viewing distance [3]. Furthermore the psycho-sensory reflex is to be mentioned [20]. This reflex is triggered by neurocognitive activities over different muscles depending paths, indicating that the fluctuation in pupil size is also related to mental workload [36]. A huge amount of studies under various conditions have been carried out in order to examine the influence of cognitive workload on the pupil reflex [18]. A major challenge of these studies was to extract the psycho-sensory reflex from the other changes in pupil size like the light reflex. Marshall [22] introduced the Index of Cognitive Activity (ICA), which succeeded the separation of the psycho-sensory reflexes. In the presence of psycho-sensory reflexes the signal, created from the continuous recording of pupil size, is characterized by abrupt discontinuities. In the presence of effortful cognitive processing, the pupil responds rapidly with a reflex reaction [23]. Marshall [22] uses the wavelet analysis which allows analyzing the frequency of a signal without losing the reference of time. There are several recent studies among others in the driver and the aviation context which show explicit the significant relation between the ICA and subjective as well as objective reference measurements [24, 25, 35]. Due to this wide spread positive responses to the promising approach of Marshall [22] the analysis of pupil dilation used in this work is also based on the wavelet analysis to separate the different pupil reflexes. The analysis is built on the Daubechies db 4 wavelet families due to the 60 Hz eye-tracking system used in the experiments [22]. Each type of wavelet families is suitable for a specifically defined frequency range of the output signal [11]. The signal of the pupil diameter orthogonally wavelet transformed and disassembled in a high and a low pass range. In this first step the intervals with high frequencies (detail) are separated from low frequencies intervals (approximate) without losing the reference of time. This transformation is a so called level 1 wavelet transformation [28]. In further levels of transformation the approximate interval can be further divided step by step, but that is not necessary for the pupil analysis. After the transformation, a threshold is used to extract the typical irregular and high frequent changes in pupil diameter caused by cognitive arousal. The thresholding can be done by several types from soft over mini-max to hard thresholds. Each of the mentioned thresholds has its own strengths and weaknesses depending on analyzed signal structure [10, 11, 22, 32]. In this study the hard threshold is used with respect to the chosen signal decomposition technique. After setting a threshold the point of the signal below this border were set to zero, so that only the values corresponding to the criteria of a psycho-sensory reflex remain [22]. The number of reflexes per second is set in relation to a maximum number of 30 per second in order to become scaled value of dilations per second [8].

2.4 Fixation Duration

For human beings sharp seeing with a high resolution is only possible at a relatively small part at the center of the retina, which is known as Fovea Centralis. Due to this restriction, eye movements are needed, which ensure that the retinal image of the interesting object is mapped as accurately as passible on the retina. The alignment of the fovea, the "gaze" on an object is called fixational eye movement. It could have been shown that under high stress the *fixation duration* is directly related to the time, which is needed to either gather information or to understand a certain problem at hand [2]. In other words, if the mean fixation duration is short, a higher amount of time may be dedicated to searching behavior. In this study, too, the fixation duration is considered, whereby the mean value of all fixations in a defined period of time (30 s) is regarded.

2.5 Nearest Neighbor Index (NNI)

The *Nearest Neighbor Index* (NNI) is a spatial statistics algorithm, which in the context of eye-tracking expresses the proximity of each fixation relative to all other surrounding fixations. In other words, it assesses the randomness of fixation patterns. Primarily, this index was used to characterize the natural distribution of plant or animal populations, however, it can also be applied to gaze distribution patterns in order to identify gaze scanning strategies [9]. Thereby a ratio between the average of the observed minimum distances between fixations and the mean distance, which would to be expected for a random distribution of fixations, is derived. In consequence, for a random distribution a ratio that is equal to 1 is to be observed. For values less than 1 the mean of minimum observed distances is smaller than the random one, which indicates a clustering of fixations and which stands for so called informational gaze patterns. For values larger than 1 a regular, non-random gaze behavior is given [9]. According to Di Nocera et al. [9] the value range for NNI lies between 0, which corresponds to maximum clustering, and 2.1491, which accounts for a strictly regular hexagonal pattern. It is assumed that under high stress conditions the gaze behavior does not follow a clear strategy and that it is characterized by searching for information randomly. In consequence, in these situations NNI should account for 1. There is a lot of empirical evidence that NNI reveals a significant correlation with both subjective as well as objective performance parameters of cognitive workload 15].

When applying NNI in a certain study, it is necessary to define the reference area containing all fixations of interest—otherwise neither the random distribution nor the actual distribution of fixation patterns may be derived. According to Di Nocera et al. [9] it is only required that the area is large enough to include all fixations. Within their work a rectangle, set by the lowest and highest values of each x and y coordinates of fixations, is used. This may work for homogeneous conditions without significant outliers, but it causes problems in more heterogeneous conditions, when there are also fixations outside the field of activity. If there is only one fixation, which is far away from the others, the rectangle will be enlarged and a different distribution, with respect to the reference area, will be provided. Against this context it may be reasonable to use a dynamic elliptical area based on a chi-square distribution. Within the work of Schubert and Kirchner [34] this approach was developed successfully evaluated against other statistical distributions in the field of postural analysis; it contained more than 95 % of all points and it adapted very well to the distribution of these points. Due to this dynamic adaption, which is able to cope with heterogeneous conditions, this approach is used to calculate the reference area for the NNI in this work.

3 Experimental Setting

In the experiment 48 subjects (32 males, 16 females) with a mean age of 22.44 years ($SD = 3.26$) took part; most of them were students of our university. In the context of an eye-tracking study it is also of interest to mention that 19 participants (13 glasses, 6 contacts lenses) had a corrected vision. The eye-tracking data were recorded by the 60 Hz binocular DIKABLIS Professional system of the Ergoneers GmbH.

The experiment consisted of two tasks, whereby each task revealed three stress levels. A within-subjects design was chosen and all participants had to pass all six conditions in a randomized order to avoid order effects. After each of the six conditions the participants had to complete NASA-TLX and further on the eye-tracking device was re-calibrated in order to ensure the accuracy of the recordings.

The first task consists of a flight procedure in a flight simulator. The autopilot is defect so that the test person has to control the plane manually. The main task is to keep the heading and the altitude on the target given by the flight director (Alt: 5000 ft.; HDG: 360°) Fig. 1.

The subject controls the plane via stick like in a real setting. On a screen in front of the person the primary flight display with all relevant information is shown. It is depicted greater than normal to focus the concentration on the necessary

Fig. 1 Experimental setting flight task

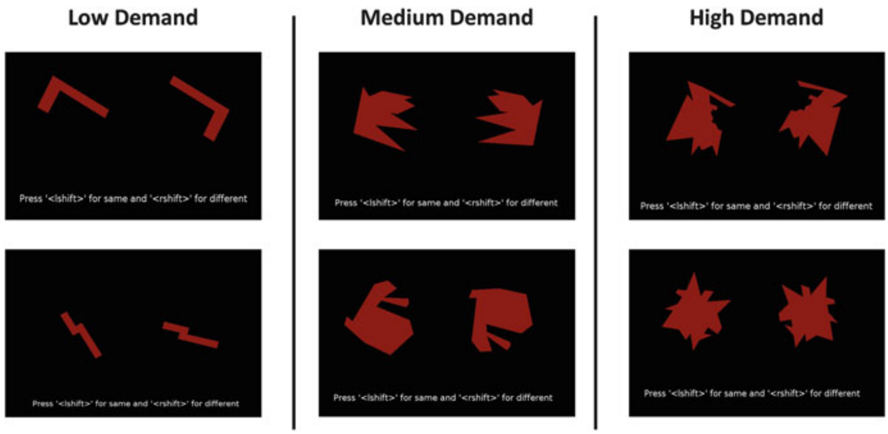


Fig. 2 Experimental setting mental rotation

parameters. Three stress levels have been operationalized by a secondary task: while the additional task was absent in a baseline condition, within two further conditions the participants had to complete a subsidiary task with a cycle time of 3 s, which was either announced acoustically or presented visually on a tablet PC left to the primary flight display Fig. 2. As secondary task a so called 1-back task was chosen. This is a rather widespread type of secondary task, which exists in various versions. Here, the participants had to reproduce verbally numbers on range from zero to ten presented every 3 s like mentioned above. Every time a new number was presented they had to represent the previous one. As an indicator for mental workload the number of mistakes was assessed [40].

The second task consists of a mental rotation in which two geometries had to be checked for equality. Here the three levels of stress (low, medium, high demanding) were defined by the number contour points of the figures Fig. 2.

The mental rotation test is developed in the experimental environment PEBL (Psychology Experimental Building Language). The geometries were shown in randomized orientations and are either mirrored or equal [5]. The subject had to decide about the equality in between 3 s and if he or she was too slow the test jumped to the next pair of geometries. After every decision the participant became feedback (right/wrong) to keep the motivation up over the whole time. Every subject had to do 128 pairs of figures in each stress level with a fixed number of equal pairs in a randomized order. The program collects a set of information like the reaction times and the amount of mistakes, as well as the number of trails in which the participants were either too slow or did not answer.

4 Data Analysis and Results

The following data analysis is divided in three parts which constitute useful to each other. Starting with the subjective and the objective data across the physiological data and ending with a detailed analysis of subject specific data.

In a first step all the recording were reviewed to check them for any interferences or problems concerning the eye detection. There were six subjects with either defective recording from the field camera or defective recordings from the eye cameras, which could not be cleared by the recalibration. Therefore these six (Participants number 18, 24, 27, 33, 37 and 45) datasets had to be excluded from the further examination. Before starting the examination of the remaining datasets all the variables were analyzed using the Kolmogorov-Smirnov-Test for normal distribution which can also be adopted by the central limit theorem due to the size of the sample (~50). Each variable needed for an analysis of variance (ANOVA) was tested for homogeneity of variances by the Levene-Test before.

4.1 NASA-TLX and Performance Parameters

In a first step the subjective data (NASA-TLX) as well as the performance parameters are analyzed, which are used for a manipulation check of the physiological data. The performance parameters consist of the number of mistakes in the secondary tasks (flight task) and the reaction times as well as the number of mistakes in the comparisons of the geometries (mental rotation). Like mentioned before both test parts consist of three stress levels. At first it is necessary to know if the parameters are able to reproduce the diverse task-induced demands leading to different cognitive effort.

4.1.1 Flight Task

For the flight task there were only significant differences between the NASA-TLX values, $F(2, 94) = 59.16, p < .05$. The Bonferroni post-hoc test showed that the first setting without secondary task differs significant form both conditions with secondary task. Neither NASA-TLX nor the number of mistakes in secondary task induced significant differences regarding acoustic and visual condition. The effect described by these two parameters can be related to resource theory by Wickens et al. [40], indicating that two visual tasks have a high inference and so a potential higher demand, but in this case there is often the effect of overlaying primary and secondary task. The task-induced various demands are hard to identify in the individual workload [40]. The subjective and objective data correlate significantly ($r = .40$, $p < .05$).

4.1.2 Mental Rotation

For the mental rotation test, an ANOVA showed a significant main effect between the three stress levels concerning NASA-TLX values, $F(2, 94) = 22.71$, $p < .05$. The pairwise comparison with Bonferroni revealed that NASA-TLX ratings of the most difficult level were significantly higher than of both other two situations. Also the number of mistakes in the mental rotation task differed between the stress levels, $F(21.76, 82.82) = 33.65$, $p < .05$. A Bonferroni post-hoc test indicated a significant higher number of mistakes in high demanding than in the medium demanding situation, whereas the low and the medium demanding situation differed not significantly. There are deviations between the subjective and the objective data. The correlation analysis confirmed this impression ($r = .07$, *n.s.*). The drift between NASA-TLX and number of mistakes can be attributed essentially to two possible reasons. On the one hand the usage of the NASA-TLX with untrained people can, despite the simplicity, cause distortion and on the other hand NASA-TLX may not be appropriate to map such a granular differentiation.

All in all it can be summed up that the subjective perception of mental workload operationalized by the NASA-TLX doesn't fit the objective performance parameters perfectly. It may be noted that the significant deviations in the NASA-TLX are accompanied by great jumps in the number of mistakes. The NASA-TLX scores moves around ten on the scale from 0 (low) to 20 (high) [31], indicating a medium task-induced demand in both conditions.

4.2 Physiological Data

Now the parameters tracked by the head mounted eye-tracking system, explained in the part methods, move into the focus Sect. 2. For this step of analysis the common way of regarding the mean value of the physiological parameters for every condition of both tasks is performed, before having a closer look at the individuals in the next chapter.

At first the various parameters are analyzed task-divided concerning their sensitivity to the diverse stress levels with an ANOVA, before comparing them to the subjective and performance parameters.

For the flight task there weren't any significant differences between the stress levels. As it can be seen in illustration 1.3 the certain parameters follow a remarkable and increasing trend like it was in the subjective or objective data, but these alterations lie on a very small range, so that there are no significant differences to be mentioned Fig. 3.

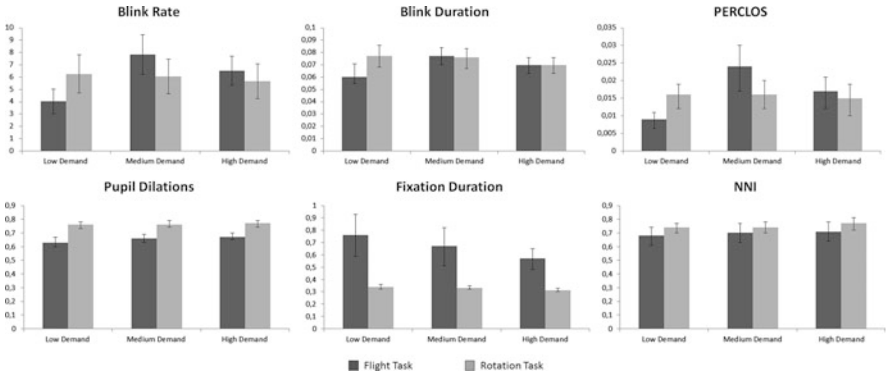


Fig. 3 Means of all physiological parameters for flight and rotation task

Like with the flight task there is none of the six parameters that indicate significant differences between the stress levels of the mental rotation task. Illustration 1.3 clarifies that by showing all six parameters Fig. 3. The majority of the parameters are stable on a specific level, so that there is no possible differentiation between the certain levels. In this task none of the parameters is neither able to detect differences between the levels of difficulty nor able to describe subjective or objective perceived workload.

After analyzing the sensitivity for the deviation in difficulty, the property of each parameter to predict the subjective sensation of mental workload as well as the objective parameters is considered. Therefore a correlation analysis is used. For the mental rotation task there are no significant correlation between the physiological measures and the subjective or objective data. The correlation coefficients are all close to zero indicating independence excepting a slightly positive relation between the number of pupil dilations and the NASA-TLX. For the flight task there are significant correlations. The NNI correlates significant with the NASA-TLX ($r = .27$, $p < .01$) as well as the number of mistakes in the secondary task ($r = .27$, $p < .01$). There is also a significant correlation between pupil dilations and NASA-TLX ($r = .23$, $p < .01$).

In summary it can be said that none of the used physiological eye-tracking measures is able to have a reliable mental workload assessment for both types of tasks in this sample of 48 subjects. In the flight task the NNI and the pupil dilations tend to fit well to the subjective data whereat the explained variance is still relatively low. There are mainly two possible reasons for the missing reliability of the physiological parameters. On the one hand a task dependence of the parameters can be postulated due to the differences between the tasks, and on the other hand the mentioned individual coping strategies may influence the reliability that much. These results motivate the following detail examination of the individual subjects concerning the coping strategies.

4.3 Detail Examination of Subjects

Based on the described theory of individual strategies for coping with high workload in the introduction (Sect. 1) and the results of the examination of the entire sample (Sect. 4.2), it is necessary to have a closer look at the individuals. Therefore the following analysis is divided into three steps. In the first part a cluster analysis was used trying to find groups of subjects with similar characteristics in terms of coping strategies. Afterwards each subject was analyzed separately over both test parts with each three stress levels concerning the relation between the physiological parameters, including the different temporal resolutions (Sect. 2), and the subjective as well as performance parameters. At least these two steps were considered together looking for dependencies and conclusions.

4.3.1 Results of the Clustering Analysis

When all physiological parameters with all possible resolutions were considered, there were 13 parameters per subject, which could be taken into account. These 13 parameters for both tasks with each three stress levels were considered for a correlation between the subjects. The result was a triangle matrix with the correlations between all subjects, which was the input for Multidimensional Scaling (MDS). The MDS calculated the Euclidian Distance based on the correlation coefficients between all subjects and plotted them into a two dimensional coordinate system. A MDS of all 42 participants indicated that there were two outliers being spatially wide separated from the remaining sample. A closer look at the questionnaires of both subjects showed that they are real outliers due to their NASA-TLX scores weren't higher than 1 in any tasks or stress levels as well as their number of mistakes, which were nearly zero. The results of another MDS carried out after ignoring these two participants indicated that there were certain groups of subjects with similar characteristics Fig. 4. To specify this visual impression a k-means analysis was accomplished. For this type of clustering algorithm a desired number of clusters had to be defined before running the analysis. The k-means analysis is then followed by an ANOVA checking the found clusters for differences. After trying all possible numbers of clusters (1–40), the best fitting solution contains eight clusters. These eight clusters (A to H) significantly differed from each other, so that this solution could be regarded as robust Fig. 4.

If both tasks (flight task & rotation task) are regarded separately, the k-means analysis prevents also eight clusters which are significantly different from each other. The main positions of the subjects in the coordinate system and their cluster allocation are consistent with the overall picture Fig. 4. Since the variations in the precise mapping are rather low, the hypothesis of task dependency of the parameters (Sect. 4.2) can be discarded.

For the last step of this cluster analysis, the naming of the axes, it is useful to consider the following individual examination of all variables, so that the naming is placed below this step.

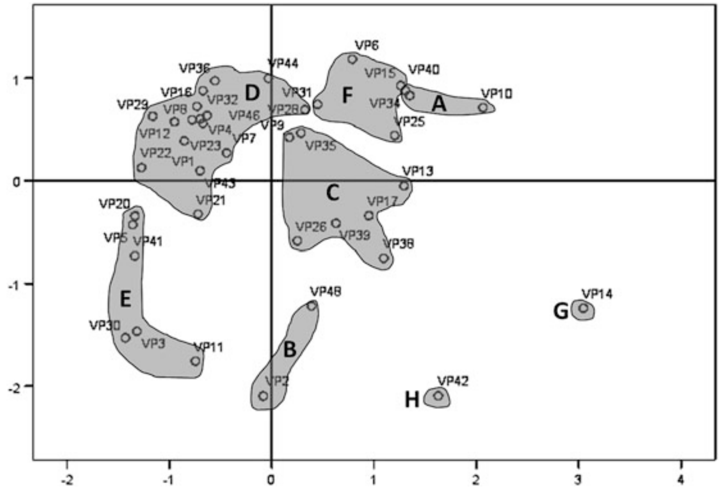


Fig. 4 Results of MDS with k-means clusters

4.3.2 Results of the Detail Examination of all Parameters

Another step further to a reliable naming of the axes concerning the extracted clusters is to have a closer look at each individual. Each of the 13 variations of physiological parameters was correlated to the number of mistakes as well as the NASA-TLX scores for every subject. Due to the independence of the task, all six conditions of both tasks were considered, so that there are six values for every parameter. A first look at the correlation coefficients showed that there are high significant relations between the number of mistakes and different parameters for each subject. Considering the clusters found by the k-means analysis, it turned out that the highly significant predictors for the number of mistakes and the NASA-TLX distinguish between the clusters, whereas within the clusters the parameters are homogeneous. For example all subjects in cluster A can be significantly described by mean of the pupil dilations, the min values of the NNI and also by the Blink Rate as well as the PERCLOS. In contrast cluster B can be described by the first values of the pupil dilations and the maximum value of the NNI. If each cluster is analyzed with the best fitting parameters, the mean of the correlations between these physiological parameters and the performance parameters for the whole sample is highly significant and shown in Table 1.

The differences in the sample indicated by the clusters can be confirmed by the detail examination of physiological parameters for each subject.

4.3.3 Development of Coordinate System

Taking into account both, the results of the cluster analysis and the differences found concerning the best predicting physiological parameters, a coordinate system can be developed, which refers to the theory of individual coping strategies Sect. 1.

Table 1 Means of correlation coefficients between the physiological data and the performance data

	Blink rate	Blink duration	PERCLOS	Pupil dilations	Fixation duration	NNI
Number of mistakes	$r = -.79$, $p < .05$	$r = -.80$, $p < .05$	$r = -.78$, $p < .05$	$r = .84$, $p < .05$	$r = -.81$, $p < .05$	$r = .79$, $p < .05$

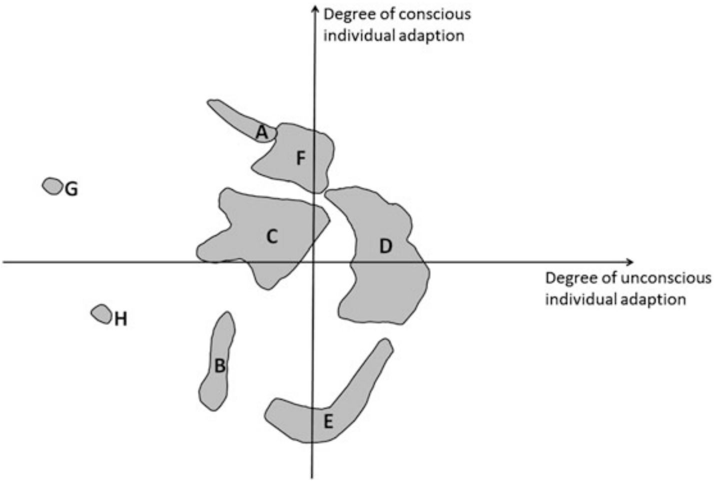


Fig. 5 Clusters after rotation of axes

After rotating the coordinate system around 45°, which is a common method in cluster analysis, an explicit assignment can be found (Fig. 5).

The X-Axis can be named with "Degree of unconscious adaption" according to the unconscious adaption or coping with high workload, whereas the Y-Axis shows the degree of conscious adaption of the individual to high task-induced demands Sect. 1. In terms of the pupil dilations it is remarkable that the best predicting resolution changes along the X-Axis. For example cluster D can be better described by the first values of the pupil dilations, whereas clusters B, G and H can be best described by the mean value. For group D the mean of the first values of the pupil dilations are significantly higher than the following values, $t\ (226) = -2.0$, $p < .05$. The indicated workload by the pupil dilations decreases over the process of one specific stress level, whereas in groups, which can be better predicted by the mean value, the pupil dilations stay stable over the time. This effect indicates that the subjects, best predicted by the first values, develop unconscious individual adopting strategies. The NNI describes the gaze behavior or strategy of individuals. For example group A can be best described by the minimum values, whereas group B can be best described by the maximum value and for group C the mean value fits best. Regarding the NNI for these three groups of individuals over the time of one

Table 2 Changing parameters concerning the defined axes

	High	Low
Degree of unconscious individual adaption (X-Axis)	Pupil dilations decrease over task time Blink Duration as a good predictor	Pupil dilations stay stable over time Blink Duration not a significant
Degree of conscious individual adaption (Y-Axis)	Changing individual gaze behavior over task time Blink Rate as a good predictor Low individual effort (NASA-TLX)	No changes in gaze behavior over task time Blink Rate as moderate predictor

stress level, it becomes clear the NNI changes over the time for the subjects with maximum or minimum values as best predictors. In case of the minimum values the curves look like bath tubs, whereas for maximum values the bath tubs are inverted and the mean values stay stable. In a further step the data stream is cut in three intervals with borders around the areas of the curves with the steepest decreases or increases. The ANOVA shows significant differences between the levels of the three time intervals for the minimum value group, $F(2, 210) = 2.84$, $p < .05$. The subjects for who the mean is the best fitting value don't show this effect. So there is a significant change in the individual gaze behavior or informational strategy over the time for those who can be best described by the minimum values of the NNI.

It can be seen that the theory of conscious und unconscious coping strategies can be confirmed by these parameters. Also the other recorded parameters change along the new defined axes, which are summarized by the following Table 2.

All in all it can be said, that depending on the individual coping strategies the sample can be divided in eight significant clusters. After a 45° axes rotation in combination with the individual examination of various variants of all physiological parameters there are significant differences between these clusters concerning the different adaption types described in Sect. 1. This adaption of physiological parameters can obtain very good task overreaching results concerning the reliability of workload assessment.

5 Discussion

This study shows the application of a portfolio of promising physiological eye-tracking measures for mental workload assessment. Therefore two tasks with each three different stress levels were chosen to analyze the parameters in different cases of use. The analysis clarifies that the individual coping strategies have a significant impact on the reliability of mental workload assessment. A clustering of the sample accompanied by an individual examination of all the introduced methods and instruments provide a substantial increase in reliability of mental workload assessment. There is for each subject at least one parameter which significantly represents the mental workload operationalized by the number of mistakes as well as the NASA-TLX. The combination is very variable due to the different clusters in the defined coordinate system (conscious and unconscious individual adaption to workload). It is possible to assess mental workload reliable and additionally independent from the task. An upcoming problem with this approach is the need of an objective measure or parameter for workload like the number of mistakes in a secondary task for the differentiation in the physiological parameters, which is needed together with the clustering for the integrated workload assessment. In most practical experimental setting this is hard to realize. A required further step in the development of such an approach for reliable mental workload assessment is the generalization for all, laboratory and field, settings. Therefore our institute is currently in the intensive development of software with which a fast calibration of any eye-tracking system is introduced. The software calculates automatically all named physiological measures in all variations and adapts them to the individual subjects coping strategies based on objective data and the cluster analysis.

Acknowledgement This study was conducted within the BMBF project AMIKA (Development of an adaptive and multi-modal interface for complex automation systems).

References

1. Abe T, Nonomura T, Komada Y, Asoaka S, Sasai T, Ueno A, Inoue Y (2011) Detecting deteriorated vigilance using percentage of eyelid closure time during behavioural maintenance of wakefulness tests. Int J Psychophysiol 82:269–274
2. Backs R, Walrath L (1992) Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. Appl Ergon 23:243–254
3. Beatty J, Lucero-Wagoner B (2000) The Pupillary System. In: Cacioppo J, Tassinary L, Berntson G (eds) Handbook of psychophysiology. University Press, Cambridge, pp 142–162
4. Benedetto S, Pedrotty M, Minin L, Baccino T, Re A, Montanari R (2011) Driver workload and eye blink duration. Transp Res F 14:199–208
5. Berteau-Pavy D, Raber J, Piper B (2011) Contributions of age, but not sex, to mental rotation performance in a community sample. PEBL Tech Rep Ser 11(2)
6. Brookhuis K, de Waard D (2010) Monitoring drivers' mental workload in driving simulators using physiological measures. Accid Anal Prev 42:898–903
7. Cardona G, Quevedo N (2014) Blinking and driving: the influence of saccades and cognitive workload. Curr Eye Res 39:239–244
8. Demberg V (2013) Pupillometry: the index of cognitive activity in a dual-task study. In: Knauff M, Pauen M, Sebanz N, Wachsmuth I (eds) Proceedings of the 35th annual conference of the cognitive science society. Austin TX, Berlin
9. Di Nocera F, Camilli F, Terenzi M (2007) A random glance on the flight deck: pilots' scanning strategies and the real-time assessment of mental workload. J Cogn Eng Decis Making 1 (3):271–285
10. Donoho D (1995) De-noising by soft-thresholding. IEEE Trans Inf Theory 41:613–627
11. Donoho D, Johnstone I (1995) Adapting to unknown smoothness via wavelet shrinkage. J Am Stat Assoc 90:1200–1224
12. Gagnon J, Durantin G, Vachon F, Causse M, Tremblay S, Dehais F (2012) Anticipating human error before it happens: towards a psychophysiological modal for online prediction of mental workload. Proceedings of the human factors and ergonomics society 2012, pp 1–11

13. Gawron V (2008) Human performance, workload, and situational awareness measures handbook. CRC Press, Boca Raton, FL
14. Gopher D, Donchin E (1986) Workload – an examination of the concept. In: Schieber F (ed) Handbook of perception and human performance. Wiley, New York, NY, pp 1–49
15. Halverson T, Estepp J, Christensen J, Monnin J (2012) Classifying workload with eye movements in a complex task. Proceedings of the human factors an ergonomics society annual meeting, 58, pp 168–172
16. Hart S, Staveland L (1988) Development of the NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock P, Meshkanti N (eds) Human mental workload. North-Holland, Amsterdam, pp 139–183
17. Kahneman D (1973) Attention and effort. Prentence-Hall, Englewood Cliffs, NJ
18. Kahneman D, Beatty J (1967) Pupillary responses in a pitch-discrimination task. Percept Psychophys 2:101–105
19. Kane M, Conway A, Miura T, Colfesh G (2007) Working memory, attention control, and the n-back task: a question of construct validity. J Exp Psychol 33:615–622
20. Loewenfeld I (1993) The Pupil: anatomy, physiology and clinical applications, 1. Iowa State University Press, Ames, IA
21. Manzey D (1998) Psychophysiologie mentaler Beanspruchung. In: Rösler F (ed) Enzyklopädie der Psychologie. Hofgrefe, Göttingen, pp 799–864
22. Marshall S (2000) Patentnumber. 6,090,051. USA
23. Marshall S (2002) The index of cognitive activity: measuring cognitive workload. IEEE Human Factors Meeting, 7, Scottsdale
24. Marshall S (2007) Indentifying cognitive state from eye metrics. Aviat Space Environ Med 78:165–175
25. Marshall S, Pleydell-Pearce C, Dickson B (2003) Integrating psychophysiological measures of cognitive workload and eye-movements to detect strategy shifts. Proceedings of the international conference on system science, 36, Hawaii
26. Martins R, Carvalho J (2015) Eye blinking as an indicator of fatigue and mental load – a systematic review. In: Arezes P et al (eds) Occupational safety and hygiene III. Taylor & Francis, London, pp 237–259
27. Matthews G, Campbell S (1998) Task-induced stress and individual differences in coping. Proceedings of the human factors and ergonomics society 42nd annual meeting
28. Misiti M, Misiti Y, Oppenheim G, Poggi J (1996) Wavelet toolbox user's guide. The MathWorks Inc., Natick
29. Parasuraman R, Rizzo M (2007) Neuroergonomics: the brain at work. University Press, Oxford
30. Robert G, Hockey J (1997) Compensatory control in the regulation of human performance under stress and high workload: a cognitive – energetical framework. Biol Psychol 45:73–93
31. Rubio S, Díaz E, Martín J, Puente J (2004) Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. Appl Psychol Int Rev 53:61–86
32. Sardy S (2000) Minimax threshold for denoising complex signals with waveshrink. IEEE Trans Signal Process 48:102
33. Schleicher R, Galley S, Briest S, Galley L (2008) Blinks and saccades as indicators of fatigue in sleepiness warners: looking tired? Ergonomics 3:2–29
34. Schubert P, Kirchner M (2014) Ellipse area calculations and their applicability in posturography. Gait Posture 39:518–522
35. Schwalm M (2009) Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotive Kontext. Doctoral dissertation, University of Saarland, Saarbrücken
36. Steinhauer S, Siegle G, Condray R, Pless M (2004) Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. Int J Psychophysiol 52:77–86
37. Straeter O (2005) Cognition and safety: an integrated approach to systems design and assessment. Ashgate, Brussels

38. Watling C, Smith S (2013) In the blink of an eye: the circadian effects on ocular and subjective indices of driver sleepiness. In: Australasian road safety research, policing & education conference, 28–30. Brisbane, QLD
39. Wickens C (2002) Multiple resources and performance prediction. Theor Issues Ergon Sci 3 (2):159–177
40. Wickens C, Hollands J, Parasuraman R, Banbury S (2012) Engineering psychology & human performance. Pearson, München
41. Wilson G (2002) An analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int J Aviat Psychol 12(1):3–18

Analysis of Dynamic Performance Data for the Assessment of Cognitive States: Results from Aviation, Assembly Tasks and Maritime Transportation

Marcus Arenius, Jürgen Klippert, Georgios Athanassiou, and Oliver Sträter

Abstract

The rapid development and application of biometric recording devices has resulted in a plethora of human performance data in various industry settings. This article will demonstrate how the data generated by these recording devices can be used for inferring cognitive states in a variety of field settings and thus applied complementary to traditional data collection methods. The authors will present results from several studies demonstrating how data from portable motion and eye tracking devices can be used to assess three main aspects: the transition between safety-relevant cognitive coping strategies by pilots in aviation, the onset and time-based development of worker mental fatigue in assembly tasks and the time-based impact of situation-induced affective states on visual attention and decision-making in maritime transportation. Furthermore the paper outlines the underlying theoretical framework (the cognitive processing loop; Sträter (2005) Cognition and safety: An integrated approach to systems design and assessment. Ashgate, Aldershot). The benefits of this kind of approach will be demonstrated by showing how work systems can be improved by providing a better fit between cognition and workplace design.

Keywords

Mental fatigue • Coping mechanisms • Affective states • Assembly tasks • Aviation • Maritime transportation

M. Arenius (✉) • G. Athanassiou • O. Sträter
FG Arbeits- und Organisationspsychologie, Universität Kassel, Kassel, Germany
e-mail: arenius@uni-kassel.de; athanassiou@uni-kassel.de; straeter@uni-kassel.de

J. Klippert
IG Metall Vorstand, Frankfurt a. M., Germany
e-mail: Juergen.Klippert@igmetall.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_14

1 Introduction

Although the optimization of physical work demands has been a priority issue in the wake of the demographic change across various industry sectors in Germany, the optimization of work for cognitive demands is still very much work-in-progress. One of the main reasons being, that the assessment of cognitive states still poses significant challenges. As they mostly provide “snapshot” representations of cognitive states, traditional subjective, self-report measurements (e.g. questionnaires, interviews) are poorly suited for addressing the fluid and dynamic nature of human cognition (questionnaires), or are vulnerable to known distortions of memory (e.g. interview data). Advances in the technology of portable performance measurement provide a promising approach for addressing these types of issues and provide a complementary view to traditional methods.

One additional question that affects the design of cognitive work and the respective supporting systems consists in the explanation and understanding of the interdependency of affective responses and information processing and its impact on observable performance. Affect and emotion are acknowledged as aspects of everyday work that influence human performance [13, 23]. Designing work systems adapted to affective responses recently has attracted the attention of some scholars as a way to improve system and equipment design beyond usability and user satisfaction [9]. However, the impact of affective states on behavioral responses in the context of dynamic monitoring processes and system control has been rather marginally investigated to the present day [7]. One reason for this may lay in lack of appropriate methods and the inherent difficulty of modeling and integrating the potential impact of affect-related factors on—cognitive—performance. Most importantly, the fringe presence of affective phenomena in human factors research may be associated with a normative view of the role of affective and emotional states in the behavior of rational actors in the working context. Affects are considered as mostly negative interferences at the cost of rational decision-making and behavior, and hence are treated as unwelcome and unacceptable behavioral by-products that should not be there at the first place.

Despite this persistent view, affective phenomena constitute integral features of everyday life and everyday work [11, 23]. As such they should not be treated in a normative but rather in a pragmatic, descriptive manner. Research should focus on the assessment and analysis of the specific ways that such phenomena drive behavior especially in safety-critical situations.

Several studies will be presented how dynamic performance data can be used to address different issues related to cognitive states.

2 The Cognitive Processing Loop as an Explanatory Model for Dynamic Human Performance Data

The cognitive processing loop (CPL) is a framework for describing and modeling human information processing and relating it to human performance [23]. The CPL consists of several main components (Fig. 1):

– Human perception through sensory organs.
– The central comparator, which aligns the sensory input to the experiences and concepts of the human by minimizing dissonance between the internal world and the external world.
– The emotional system, which strongly influences the evaluation of sensory information and cognitive processing though affective activation.
– The vegetative nervous system which regulates the bodily functions.
– The motor system, which is involved in movement execution.

As the term ‘loop’ in CPL indicates, the information processing is treated as an iterative process. For simple single-loop stimulus-response tasks, the reaction times stated in the model are applicable: 50 ms for processing sensory information, 100 ms for evaluation and 50 ms for motor execution which leads to action.

The different studies are associated with different aspects of the CPL (Fig. 1):

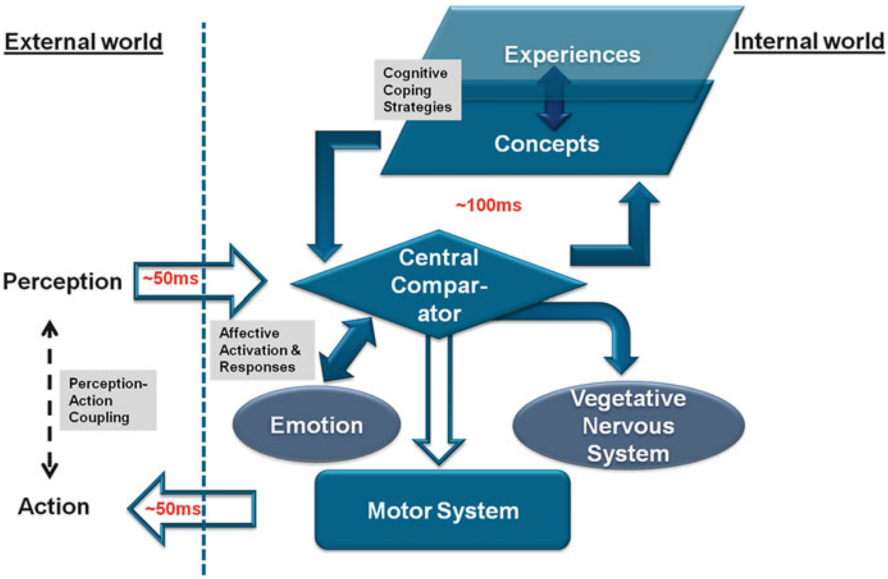


Fig. 1 The different research activities (*grey boxes*) addressing different aspects of the cognitive processing loop (CPL, [23])

– The first study from the aviation domain demonstrates how cognitive coping styles change over time given different task-demands. Via eye-tracking, the way how pilots use their concepts and experiences can be evaluated e.g. when noticing an engine failure.
– The second study taken from assembly work demonstrates how the patterns of time-based coupling of human perception, information processing and motor actions can be related to the issue of mental fatigue
– The third study from the maritime domain demonstrates how information processing is related to affective responses and how the latter in turn influence human perception and performance (as evident in eye-tracking data).

3 Analysis of Eye-Tracking Dynamics in Aviation: Cognitive Coping Strategies

In order to assess cognitive coping strategies in aviation, we have to move from traditional eye-tracking data to the analysis of eye-tracking dynamics [1]. A short summary on what is meant by cognitive coping will be given.

Human performance is always rational from the actors' point of view. In other words, people use their knowledge to pursue their goals; therefore a particular course of actions always makes sense from the practitioners' perspective at a given point in time [24, p. 16]. In a seminal paper, a cornerstone of this reasoning was defined—satisficing [21]. In essence, satisficing means that is not feasible for the human to process all of the relevant information to find the optimal solution in information-rich environments due to tight resource constraints. Instead, humans adopt a stance which is described as satisficing—they process information in such a way that solutions are produced which may not be optimal but rather "good enough" to be delivered in a timely manner. Otherwise the human would be constantly occupied with processing of information in order to find a perfect solution, an (reactive) approach that is not applicable in complex work places. As it is not possible for the human to process all relevant information, any actor essentially has to base his or her actions on an incomplete picture of the overall situation.

Therefore the human performance is always approximate in nature, as humans have "fill the gap" of their incomplete picture of the working environment by relying on heuristics and trade-offs [12, p. 93–94]. This produces variability in human performance, which often goes unnoticed, as it in the vast majority of cases leads to the desired results. However, in very rare cases performance variability may be associated with undesired outcomes, if the assumptions behind trade-offs or heuristics are violated (e.g. due to unanticipated complications or very unusual events).

Conversely, if human performance is approximate, that is, characterized by necessary short-cuts and heuristics, this should be reflected in the way human operators interact with their working environment and react to changes and disturbances.

Based on this reasoning, a study was conducted to assess the impact of an air-traffic control (ATC) information display for airborne separation on the performance of commercial pilots. Ten male pilots from a major German airline flew a safety-critical flight scenario in a full flight simulator using an ATC display in addition to the usual cockpit layout. An expert rating pilot trained by the airline scored their total performance.

The flight route chosen for the investigation is a standard scenario for the pilots and consists of a flight from Athens to Heraklion. On route, after ~60 s of flight time, the pilots receive a traffic warning of a potential head on collision with another aircraft in the ATC-display followed directly by the sudden onset of an engine failure. Thus, the pilots have to resolve the traffic conflict with the other flight on collision course and notice/react to the engine failure at the same time. In this situation, the pilots are facing a goal conflict: Should they focus on resolving the collision warning or on the engine failure? Here, the engine failure should be treated as the priority issue over the traffic warning following the principle of "fly the aircraft first" [3, 20]. The results of the study demonstrated that some pilots focused on the ATC display and thus did not notice the engine failure in time [3, 20].

The difference in coping styles between the pilots who noticed the engine failure right away/without delay and the pilots who did not were examined using NMDS representations of respective eye-tracking data: The non-metrical multidimensional scaling technique is an ordination technique (NMDS) that produces a graphical representation of a distance matrix by preserving the original distances as well as possible. Using a correlation matrix of eye-tracking data on Areas of Interest (AoI) as distance matrix, a two-dimensional representation of the correlation between the visual focus on the information sources can be calculated [22]. The closer the points are in the NMDS representation, the stronger the correlation between the gazes on the two information sources, that is, the stronger the common visual focus on them. This representation can thus reveal the division of visual attention associated with this coping strategy of prioritizing the traffic warning over the engine failure.

In order to create the representation, the eye-tracking data of a pilot that did not notice the engine failure in time has to be separated into meaningful segments that provide insight into the coping process associated with the onset of the traffic warning and the engine failure. The change in the strategy applied by the pilot is clearly visible with the onset of the collision warning as shown in the left hand side NMDS representation of Fig. 2. The autopilot, the ATC-display with the traffic warning and the central console are in joint visual focus, while the other displays zone out of visual focus.

The right hand side NMDS representation of Fig. 2 shows the visual attention after the pilot has noticed the engine failure, about 20 s after its onset. The primary flight display (PFD) and the engine and warnings display (E/WD), both relevant for handling the engine failure, are in visual focus—together with the ATC-display. Visual focus on the latter is not required if the principle "fly the aircraft first" is followed. This means, that the most safety-critical issue is a pilot focusing on resolving the traffic warning although the engine failure has been identified.

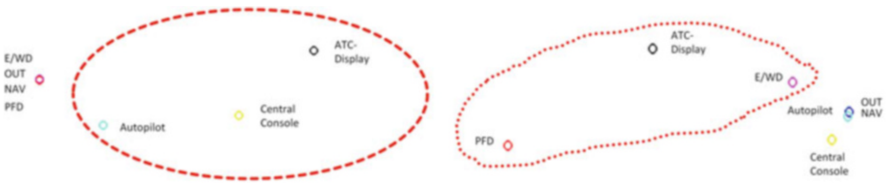


Fig. 2 Left-hand: After the traffic warning, information sources relevant for collision avoidance are focused (*red circle*). Right-hand: The pilot notices the engine failure occurring directly after the traffic warning late; however still also focuses on ATC-display (irregular red shape) [Note: the NMDS representation is dimensionless]

Using eye-tracking, this issue can be identified by the fact that visual attention is directed away from the most pressing task—the handling of the aircraft—to resolving the traffic warning.

Thus, the NMDS representation shows the different coping strategies associated with safety-critical changes (traffic warning and engine failure) in the work environment. In terms of the CPL, coping strategies are cognitive actions steering the visual attention towards those perceptual elements of the external world that fit the actual elements of the conceptual level.

4 Eye-Tracking for Assembly Tasks

Most industrial assembly tasks are characterized by simple manual tasks performed by workers. The mental effort required for performing these kind of tasks is characterized by recognition-action-coupling also termed “object related actions” (ORAs, [16]): The worker’s task is grasping objects and place them on defined positions. Performing ORAs requires spatial orientation of the whole body and the recognition of task-relevant objects. Spatial orientation and recognition of task-relevant objects in turn requires cognitive processing of visual information by the worker. Before performing an ORA, the action relevant object therefore is first fixated with the eyes and then grasped with the hand.

The mental workload associated with simple manual tasks is comparatively low (underload). Therefore, the mental fatigue of the workers usually stems from mental underload as a consequence of the repetitive characteristics of the tasks.

Commonly, two types of methods are applied in order to measure mental fatigue: Either measuring primary task performance or by subjective response (e.g. questionnaires). Measuring mental fatigue by primary task performance seems to be inadequate because operators are able to compensate for low workload by increasing their effort to maintain vigilance. Therefore, performance outcomes of simple manual tasks are no adequate measures for mental fatigue. The measurement of mental fatigue by subjective response (e.g. questionnaires) is also problematic. It may produce distorted results as performance and perception tend to deviate. An alternative is the use of physiological parameters to assess mental fatigue. This seems as a promising approach [8]. We follow the hypothesis that cognitive processing time differs depending of the prevalence of mental fatigue. We assume that the time between fixating and grasping is the time required for cognitive processing of the information for controlling the body's motor functions. We call this time span eye-hand-latency (EHL; [19]). In the course of several studies, the visio-motoric measure EHL in milliseconds (ms) was tested as an indicator for mental fatigue in simple manual tasks [14, 15]. Both studies indicate a fatigue-related pattern of the EHL over time: At the beginning of the shift the EHL is by trend shorter than at the end of the 8 h shift.

In both cases, the measurements were done for the complete working shift and we compared similar tasks: The tasks studied were conducted in u-shaped production cells with several machines served by one worker. The type of production cells under study were organized following the chaku-chaku approach. It is a concept originating from the philosophy of the Toyota production system. The Japanese term "chaku chaku" means "load load" and defines the workers task very accurately: stepping from machine to machine and loading and unloading objects manufactured automatically by the machines. Although the tasks in focus of the studies [14, 15] are very similar, there are some characteristic differences which should be pointed out. They require a different range of actions to be performed. In the first study, the work consisted of simply loading and unloading objects from the machines. If a technical disturbance occurs during work, it is strictly prohibited to take appropriate action to eliminate the interference by the operator himself. In the second study, besides loading and unloading objects, the elimination of technical disturbances was explicitly defined as part of the operator's task. The workers were specifically trained for eliminating disturbances.

For both cases a *t*-test for independent samples was conducted to compare the EHL values for the beginning and the end of an 8 h shift. Table 1 shows the results for the work system with very simple tasks.

In this case, the test confirmed an effect on α-level of 0.1 ($p = 0.09$): The EHLs at the end are longer than in the beginning. Cohen's d (0.35) and the Bravais-Pearson Correlation (Effect $r = 0.19$) confirm a slight effect [14].

The results for the enriched task with the broader range of activities are shown in Table 2. Here, the *t*-test is significant on α-level of 0.05 ($p = 0.011$). The Bravais-Pearson Correlation (Effect $r = 0.28$) confirms a slight to medium sized effect and Cohen's d (0.59) confirms a medium sized effect [15].

There is a difference in the effect size between the two tasks. This may have been due to the differences in the work content between the simple and the enriched task. The simple task leaves the working person very little freedom for action which means that there is a very low variation in task performance for the worker. The person performing the task is forced to follow the machines step by step and to load and unload the objects. The worker does not necessarily need to know anything about the production process and the technology at all. The elimination of technical disturbances is prohibited for the worker. In the case of a technical disturbance the operator is required by management order to call a member of the maintenance staff who will eliminate the interference.

Table 1 *T*-test for independent samples for the simple task [14]

Groups	Count	Mean (ms)	SD	Variance			
Beginning of shift	22	476.36	229.58	52709.95			
End of shift	27	854.81	1431.99	2050602.84	Standard error	*t*	*df*
Pooled				1157927.30	279.90	1.35	47
		df	27.63				
		p-value	**t-crit**				
One Tail		**0.09**	**1.31**				

Table 2 *T*-test for independent samples for the enriched task [15]

Groups	Count	Mean (ms)	SD	Variance			
Beginning of shift	37	569.72	303.95	92391.59			
End of shift	30	764.00	360.13	129693.79	Standard error	*t*	df
Pooled				109034.11	82.58	2.35	65
		df	56.88				
		p-value	**t-crit**				
One Tail		**0.01**	**1.67**				

The enriched task is more diverse compared to the first one. The elimination of disturbances is explicitly part of the work task and the workers had the appropriate qualification to do so. This additional task may contribute to task diversity which may be the cause of the differences in effect strength measured in the two studies.

The hypothesis is that we measure a mix of mental fatigue and fatigue like states. In order to gain better understanding of the effects, the concepts used in the study have to be differentiated further. This is the difference between states of mental fatigue and fatigue-like-states [10]. However, two studies provide an insufficient basis for verifying this conclusion and thus further studies have to be conducted. To study these combined effects, indicators are needed which allow the differentiation between mental fatigue and mental states similar to fatigue.

The eye-hand-latency has a direct correspondence to the CPL. A change in the latency can be directly related to the delay in the alignment of concepts with the external perceptual demands, hence mental fatigue can be seen as a result of fatigued elements of the CPL.

5 Affects and Cognition, Time-Bound Effects in Eye-Tracking

Affects are ubiquitous aspects of behavior as they represent constant evaluations of the actual or anticipated relationship between the human and the external world. Such evaluative processes take place on a conscious or a subliminal level and enable humans to adjust their behavior in a respective manner in order to establish an acceptable match between the interacting environmental demands and the behavior-driving goals and choose action [5, 6, 18, 23]. It is reasonable enough to consider working behavior as a subset of overall human behavior. Accepting that the above assumption is true, one can argue without much controversy that affect ubiquity is also true for work-related behavior and in-work performance. In agreement with a descriptive view on affects, the main questions that need to be addressed consist in the specific fashion in which affective and cognitive responses are related to each other; the direction of the relation and the respective behavioral adaptations with regard to distinct affective states and situational demands; and the behavioral parameters and levels on which the manifestation of affective responses can be observed and documented in a valid manner. Athanassiou [4] addresses these issues within the scope of a PhD activity conducted in the Department of Human and Organizational Engineering and Systemic Design at the University of Kassel to be published soon. The research has been driven by the assumption that cognitive and affective functions are tightly interrelated [23]. Thus they underlie similar principles of activation that drive information processing and behavioral adaptations. However, additional overtly affective load may induce distinct—observable—adaptation responses, which will be reflected in relevant behavioral parameters, such as visual attention. This fundamental research question was tested empirically in the context of simulated ship management tasks using eye-tracking measurements.

The empirical study was conducted in the full-mission ship handling simulator facility at Jade University of Applied Studies. N = 13 students of nautical studies participated in the study. Participants were assigned to two task-based scenarios, of which, one was designed to employ situation-induced affect through task-integral trigger events. The specific trigger event considered within the scope of this article was associated with acute time and space challenges for the safety of the operational process (crossing situation with another vessel in shallow and narrow waters and restricted available time for evasive measures). Affect-inducement in the affective treatment condition was operationalized through the behavior of the traffic "opponent", which was characterized by uncooperative and face-loss communication.

A mixed 2 × 5 factorial design was employed for the empirical study. The two levels of the between-group factor represented the two experimental conditions under consideration, i.e. task performance with additional affect-inducement versus "normal" task performance (control condition). The within-group factor represented the subsequent 5 min directly after the onset of the affect-inducing trigger event. Visual attention has been operationalized through measures of fixation duration and fixation dispersion towards crucial information sources on bridge as well as overall fixation shifts for the five consequent minutes immediately after the trigger onset of. The results concerning the view out/windscreen (due to the criticality of the information source for unexpected situations) and the aspect of overall fixation shift number will be briefly showcased.

The results of the mixed ANOVA (Pillai's Trace) showed a significant interaction between affect–inducement and visual attention over time (F (4, 8) = 4.581, $p = .032$, partial $\eta^2 = .696$) in terms of fixation duration rates towards the view out of the windscreen (Fig. 3). This result suggests that individuals made distinct use of the information source when faced with explicit affective load in conditions of severe time constraints and the need of fast decision-making. Visual information search and consultation were directed more intensively towards the outside world and not its representations (i.e. Radar screen) during the 3rd and 4th minutes after the trigger onset (when the situation developed in a manner that posed severe threats for vessel safety). Lower rates for the rest of the time suggest an increased ability to distribute visual attention towards available information in a less fixated manner. The behavior was inversed (and somehow with more moderate "deflections" during time interval) for individuals facing a similar situation with only the respective cognitive load, hence suggesting differences in necessary control efforts in order to adjust to the situational demands due to the affect inducement.

The same statistical procedure was followed for the analysis of potential effects of the affect-inducement on the overall fixation shifts towards all information sources available in the bridge operational environment over time (Fig. 3). The results showed a main effect of time on fixation shifts for both conditions (F (4, 8) = 3.843, $p = .050$, partial $\eta^2 = .698$) suggesting an increased visual activity in the progress of time as a consequence of the situational conditions of operating action. However, an additional nearly significant main effect of the scenario-based affective treatment on the overall fixation shifts towards information sources on bridge was obtained by the statistical analysis (F (1, 11) = 3.878, $p = .075$, partial $\eta^2 = .261$). This result suggests that affect-inducement was associated with an increased need for fixation shifts towards and between information sources on bridge right from the very onset of the trigger event. The observed difference between the conditions indicates increased efforts to maintain control and hence distinct adaptive performance of human operators that can be assigned to the specific impact of task-integral affect and to the subsequent affective responses during dynamic, time-bound task performance.

The maritime study showed the interrelation between the cognitive and the affective parts of the CPL. Cognitive dissonance, as a result of the central comparator, impacts the cognitive quality of decision making and in parallel this results into an affective reaction on the dissonance.

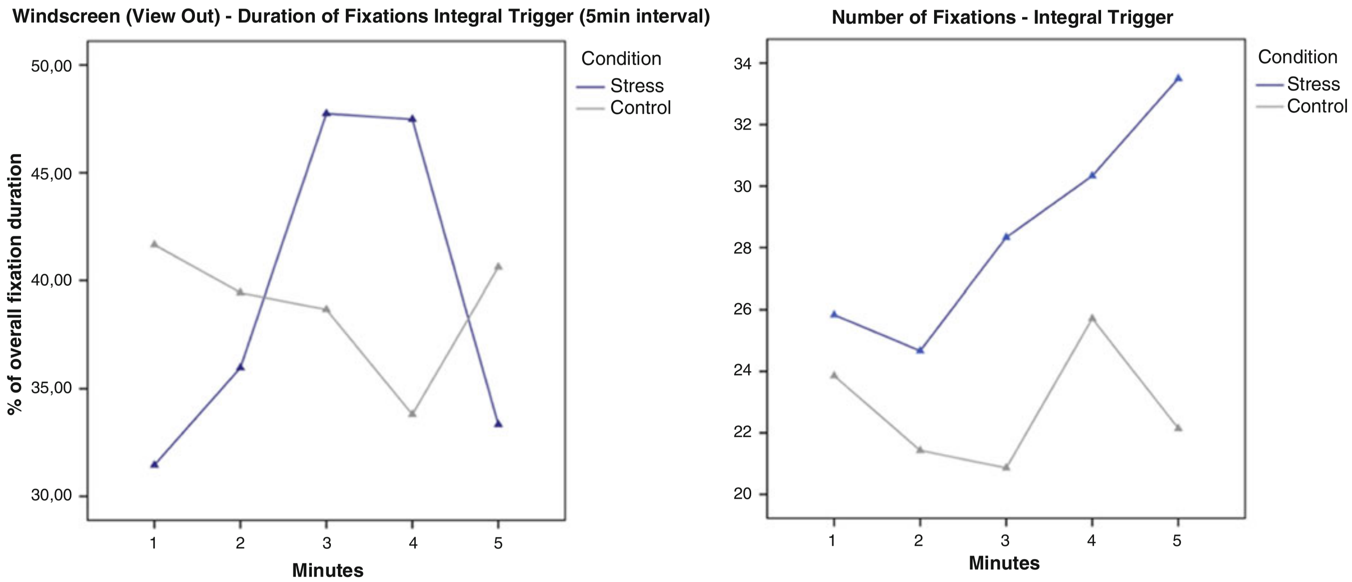


Fig. 3 Mean values for fixation duration rates towards the view out (*left*) and number of fixations towards information sources on the bridge (*right*) within the 5 min interval immediately after the onset of the affect-inducing trigger event

6 Discussion

This article has described three different research activities using dynamic performance data for inferring cognitive states based on the cognitive processing loop (CPL).

– The aviation study demonstrated how the NMDS-representation can be used to capture dynamics of eye-tracking behavior to derive cognitive coping strategies when facing adverse conditions (e.g., traffic warning/engine failure in aviation). The cognitive coping strategies are related to the way that pilots align the events in the external world (the engine failure/traffic warning) with the internal world (previous experiences and concepts).
– The analysis of eye-hand latency (EHL) for assembly tasks demonstrated how mental fatigue-related effects can be identified in assembly tasks in field settings by analyzing the eye-hand-latency over time. The EHL is defined as the time needed to couple perception and action in the CPL.
– Finally, the maritime study showed how eye-tracking performance can be linked to the reciprocal relationship of affective and cognitive states. The results allow for understanding the specific ways, in which affect-eliciting instances and the respective affective responses influence the way that human operators interact with their work environment.

The studies showed how performance date with the underlying model of the CPL can serve as a basis for better task or workplace design:

– Different coping strategies can be measured and plausible courses of action can be identified e.g. to develop appropriate qualification programs or support-systems based on this information,
– Task-design can be enriched by taking the effects of mental fatigue over time into account,
– The effects affective responses on task performance can be used to raise awareness for, to enhance the behavioral repertoire of the human element, or to guide system design efforts in a manner, which will ensure that the work system stays resilient even when facing emotionally challenging situations.

7 Outlook

As the aviation study showed, a NMDS-representation can yield satisfying results regarding the differentiation of eye-tracking dynamics before and after critical events. However, one open issue is the appropriate aggregation of the results: How should the NMDS-representations be compared between individuals? That is, what are appropriate statistical methods for aggregating and comparing the differences in the NDMS representations? Furthermore, the NMDS should ideally be applied is such a way that it enables the analyst to identify coping styles without prior knowledge of critical events; i.e., for cases in which it is unknown when systematic changes in coping styles will occur. Once this issue is solved the next question is, which level of resolution or scope (time window) should be used to calculate the NMDS representation in order to capture the changes in coping styles. Ideally, a systematic application of the NMDS should answer these questions. These issues will be addressed in the course of a current PhD-activity funded by the EU and Eurocontrol Network HALA!—Towards Higher Levels of Automation in Air Traffic Management [2].

The implementation of eye-tracking in the context of task performance in the maritime domain allowed for insights concerning the interdependency as well as the distinctiveness of affective and cognitive processes in in-work situations. However, the affect-cognition reciprocal relationship and mutual influencing are not restricted to situations of which affect is a clear and integral constituent. Moreover, they may refer to situations, on which affective responses constitute the result of carry-over effects of affect, but with a similar relevance for attentional activity, thorough information processing and risk perception [17, 23]. In addition, different situational aspects, such as expertise, will lead to distinct appraisals of the situation at hand and thus to distinct affective experiences with different attentional and action strategies and responses. Eye-tacking data can be a powerful source of information in regard to the assessment of these indirect effects of distinct affect-laden situations in the course of events and hence provide empirical validation and guidance towards an integration of affect in modeling human performance.

Finally, the link between human factors and affective psychology does not have to be restricted to dynamic data on the level of individual performance. A great importance is ascribed to the issue of teamwork and safety performance in human factors research as well as in practice. Hence, it is reasonable to guide efforts towards the exploration and derivation of relevant knowledge concerning the interrelations of affective responses and safety-crucial aspects of performance on the team level. Eye-tracking data provide only one option for such endeavors. Observational data, based on sound practice-relevant criteria and collected through domain-specific and well-constructed methods, provide another option, which allows for significant complementary information regarding the respective question of assessing affect-related effects on performance. Athanassiou [4] is addressing both issues in his current research activities.

References

1. Arenius M (2014a) Function allocation in human-automation interaction by means of dynamic performance data. In: GFA (ed) Proceedings of the 60. GFA-Conference: Gestaltung der Arbeitswelt der Zukunft. GFA Press, Dortmund
2. Arenius M (2014b) D09 – Progress Report. HALA! SESAR Research Network – Towards Higher Levels of Automation in Air Traffic Management. Eurocontrol, Brussels
3. Arenius M, Schneider M, Straeter O (2014) Evaluation of a display for airborne separation in safety-critical situations. In: 6th International Conference on Research in Air Transportation (HALA! SESAR Research Network Deliverable). Eurocontrol, Brussels

4. Athanassiou G (in press) Mariner's Adaptive Performance under Stress: Individual visual performance and team safety performance as indicators of adaptive responses to task-integral cognitive and affective workload during a complex ship management task. PhD Thesis, University of Kassel
5. Cunningham WA, Zelazo PD (2007) Attitudes and evaluations: a social cognitive neuroscience perspective. Trends Cogn Sci 11(3):97–104
6. Damasio AR (2006) Descartes' error: emotion, reason and the human brain. Vintage, London
7. Gaillard AWK (2008) Concentration, stress and performance. In: Hancock PA, Szalma JL (eds) Performance under stress. Ashgate, Aldershot, pp 59–75
8. Hancock PA, Meshkati N, Robertson MM (1985) Physiological reflections of mental workload. Aviat Space Environ Med 56:1110–1114
9. Hancock PA, Pepe AA, Murphy LL (2005) Hedonomics: the power of positive and pleasurable ergonomics. Ergon Des 13(1):8–14
10. Hacker W, Richter P (1984) Psychische Fehlbeanspruchung: psychische Ermüdung, Monotonie, Sättigung und Stress. Band 2. Springer, Berlin
11. Hollnagel E (2002) Time and time again. Theor Issues Ergon Sci 3(2):143–158
12. Hollnagel E (2009) The ETTO principle: efficiency-thoroughness trade-off: why things that go right sometimes go wrong. Ashgate Publishing, Surrey
13. Kjellén U (2000) Prevention of accidents through experience feedback. Taylor & Francis, London
14. Klippert J (2014a) Blickbewegungen als Indikator für psychische Beanspruchung? Zur Bedeutung der Wahrnehmungs-Handlungs-Kopplung am Beispiel manueller Tätigkeiten in der Industrie. Paper presented at the GFA-Frühjahrskonferenz 2014, München
15. Klippert J (2014b) Wahrnehmungs-Handlungs-Kopplung bei teilautomatisierten Montagetätigkeiten. Paper presented at the Psychologie der Arbeitssicherheit und Gesundheit, Dresden
16. Land MF, Mennie N, Rusted J (1999) The roles of vision and eye movements in the control of activities of daily living. Perception 28:1311–1328
17. Lerner JS, Keltner D (2001) Fear, anger and risk. J Pers Soc Psychol 81(1):146–159
18. Miceli M, Castelfranchi C (2015) Expectancy and emotion. Oxford University Press, Oxford
19. Pelz JB, Canosa R (2001) Oculomotor behavior and perceptual strategies in complex tasks. Vision Res 41(25–26):3587–3596, http://doi.org/10.1016/S0042-6989(01)00245-0
20. Schneider M (2014) Entwicklung eines Zusatzinformationssystems für Piloten und dessen Untersuchung hinsichtlich Auswirkungen auf das Verhalten von Piloten in kritischen Flugsituationen (Master Thesis). University of Kassel, Department of Human and Organisational Engineering & Systemic Design, Kassel
21. Simon HA (1956) Rational choice and the structure of the environment. Psychol Rev 63 (2):129–138
22. Straeter O, Arenius M (2014) Assessing the complex interaction and variations in human performance using non-metrical scaling methods. In: Kadry S, El Hami A (eds) Numerical methods for reliability and safety assessment – multiscale and multiphysics systems. Springer, Berlin, pp 579–598
23. Sträter O (2005) Cognition and safety: an integrated approach to systems design and assessment. Ashgate, Aldershot
24. Woods D, Dekker S, Cook R, Johannesen L, Sarter N (2010) Behind human error, 2nd edn. Ashgate Publishing, Surrey

Effect of Head Inclination on Neck Muscular Activity, Tracking Performance and Subjective Neck Strain: Visual and Biomechanical Conditions for Designing the Computer Workstation

Wolfgang Jaschinski, Alwin Luttmann, and Matthias Jäger

Abstract

At office computer work, musculoskeletal complaints can occur when workplace conditions are not designed according to physiological dispositions of the office worker. Regarding neck pain, the vertical position of the monitor should be adjusted in a way that the head adopts a comfortable posture. For investigating this physiological condition, this research compares several measures of performance and strain as a function of the inclination angle of the head. Head movement performance was measured when the head moves horizontally in a tracking task, while the myoelectrical activity of two types of neck muscles was recorded, i.e. of the sternocleidomastoid muscle and splenius capitis muscle. Additionally, the subjective rating of perceived musculoskeletal strain was assessed by questionnaires in the tracking task and in an office field study. All these measures represent ergonomic stress-strain functions that can be useful in physiologically based ergonomic research. Lowering the head inclination by a few degrees was able to reduce the subjectively rated musculoskeletal strain in field studies. The inclination of the head and vertical gaze direction to the monitor should agree with individual physiological dispositions.

Keywords

Neck strain • Computer work • Head inclination • Head tracking • Electromyography

W. Jaschinski (✉) • A. Luttmann • M. Jäger
Leibniz-Institut für Arbeitsforschung an der Technischen Universität Dortmund, Dortmund, Germany
e-mail: jaschinski@ifado.de; luttmann@ifado.de; mjaeger@ifado.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_15

1 Introduction

1.1 Stress-Strain Functions in Visual and Biomechanical Ergonomics

It is the aim of ergonomics to design the working conditions according to the physiological disposition of the worker. For this purpose, ergonomic research may measure physiological effects of work (referred to as objective indicators of strain), e.g. by recording the electrical activity of muscles, or by modelling the load on cervical intervertebral discs. Furthermore, the worker may rate the perceived strain or complaints by means of questionnaires (referred to as subjective indicator of strain). These indicators of strain can be measured as a function of the workload (stress) that is imposed on the organism, as suggested by the stress-strain concept of Rohmert [52]. Such a functional relation can easily be explained for the task of lifting a certain weight: when the amount of the weight increases, the muscular electrical activity will increase and the worker will report a higher degree of perceived strain. Such functions may be referred to as ergonomic stress-strain functions.

There are also other, less intuitive examples as in visual tasks at short viewing distances: the workload on the eye muscles of convergence and accommodation is—obviously—not a weight, but is given by the viewing distance in the unit 1/m, since this measure is proportional to the physiological strain of the eye muscles [25]. One can easily experience that the perceived ocular strain increases when a visual target is shifted closer to the eyes and the viewing distance is shortened.

However, given the complexity of physiological mechanisms, a single stress-strain function may not be adequate in many cases. For example, Heuer et al. [17] tested a model with two superimposed stress-strain functions to predict the favourable viewing distance at computer work: near vision strain depends on (1) the discrepancy between the current viewing distance and the individual resting position of convergence of the ocular muscles [16, 18, 24] and (2) the extent to which this discrepancy depends on gaze inclination [19, 21, 28]. Thus, gaze inclination is a relevant ergonomic parameter [48]. Moreover, visual functions in near vision are modulated by inclination of the vertical gaze direction since a more downward gaze direction facilitates the convergence and also the accommodation of the eyes [51]. Jaschinski et al. [29] described several physiological mechanisms with stress-strain functions that play a role for choosing the vertical monitor position: head inclination, eye inclination (with their effects on accommodation and convergence) and—for presbyopic computer users—the vertical zone of clear vision with general-purpose progressive lenses [30–33, 45].

It is an important aim of ergonomic research to investigate the different physiological stress-strain functions that are involved in real working tasks. In the case of computer work, both the visual and the musculoskeletal system are involved. The present report focuses on the stress-strain functions as a function of head inclination.

1.2 The Effect of Head Inclination in Physiological Studies

Different studies reported on the favourable mean head inclination which corresponded to an eye-ear line from 6 to 16 deg relative to horizontal [2, 22, 23, 42, 43].

When the head is inclined more downward, the mechanical torque of the head in the sagittal plane increases and has to be compensated by increasing activity of the neck muscles, mainly of the splenius capitis muscle. Accordingly, the perceived strain increases and 12 % of the maximum neck muscle force is exerted at a 30° neck inclination angle as shown by Chaffin and Anderson [8]. Lee et al. [34] recorded the electrical activity at the cervical region over 2 h: no increase over time was found at a 25° downward head inclination, while inclinations of 45° and 65° showed increases by about a third and two thirds, respectively; moreover, across the three levels of head inclination it was found that larger electrical activity was associated with a higher rating of the perceived strain. Straker et al. [56] calculated biomechanical model predictions of the extensor over a large range of 120°. When the head is declined from a raised to a lowered position (within a range relevant for workplaces) the C7 gravity moment and the cervical relative strain (moment demand due to gravity around the cervical vertebra C7) increases. Sommerich et al. [54] concluded that in most cases, the electrical activity recorded from the posterior upper cervical reflects the monitor location, at least between extreme locations.

There is a conflict between an upright head position to reduce musculoskeletal strain of the neck and a low head inclination to reduce oculomotor load (see Sect. 1.1). Apparently, a compromise must be found between these two concurrent aims. Sommerich et al. [55] described this conflict by a qualitative model that includes several factors of influence, including individual differences (spectacle wear in presbyopia, musculoskeletal health, sex) and workplace and task factors (type of task, time on task, task/skill match, glar, viewing distance). This model of head inclination can be understood as part of a more general model of musculoskeletal disorders at computer work as described by Wahlström [57].

For the investigation of a muscular mechanism, the researcher has to choose a certain experimental task where the muscles of interest are involved in. The above studies used different tasks in which the subject held the head stationary at certain postures or did simulated office work in the laboratory with different head inclinations. The physiological regulatory mechanisms can be investigated using tracking tasks: a continuously moving stimulus is provided and the subject has to follow this stimulus with a continuous movement response. Typical examples include movements of the hand [40], the arm [14, 49], the trunk [59, 60], or the whole body [5, 50]. The precision in tracking describes the quality of the movement performance which may be limited by neuro-muscular incapability or noise, as well as movement variability. At computer work, head movements must continuously be coordinated with eye movements when the gaze is shifted across the computer screen or between different objects at the workplace. This eye-head coordination may be impaired by progressive lenses in presbyopia [45].

In our research team, Masseida et al. [43] used a head tracking task at different head inclinations; part of the results are summarized in this chapter and extended regarding stress-strain functions.

1.3 The Effect of Head Inclination at Office Workplaces

Given the above physiological effects, the practical ergonomic question remains whether changes in head inclination affect the subjective ratings of discomfort or complaints at work, either in laboratory simulations or in office field studies. In office field studies, Ariëns et al. [4] found a trend of a positive relation between neck flexion and neck pain, while Marcus et al. [41] reported the opposite effect, i.e. a greater downward tilt was associated with a lower risk of neck/shoulder symptoms and disorders.

The head inclination that is adopted at the workplace depends on the vertical position of the monitor relative to the eyes. So the question arises, which vertical monitor position allows for a physiologically favourable head inclination.

Figure 1 illustrates the relations of the anatomical parameters of head inclination to the geometrical dimensions at the computer workstation, including the viewing distance (from the eye to the centre of the monitor), the gaze inclination angle (direction of gaze from the eye to the screen centre, relative to horizontal), the monitor inclination (inclination of the screen surface relative to vertical), and the head inclination (measured as eye-ear line relative to horizontal). The Frankfurt-line (connecting the tragus at the ear and the lower border of the orbit) is—on average—11° below the eye-ear line [47]. Downward inclinations are indicated by negative signs.

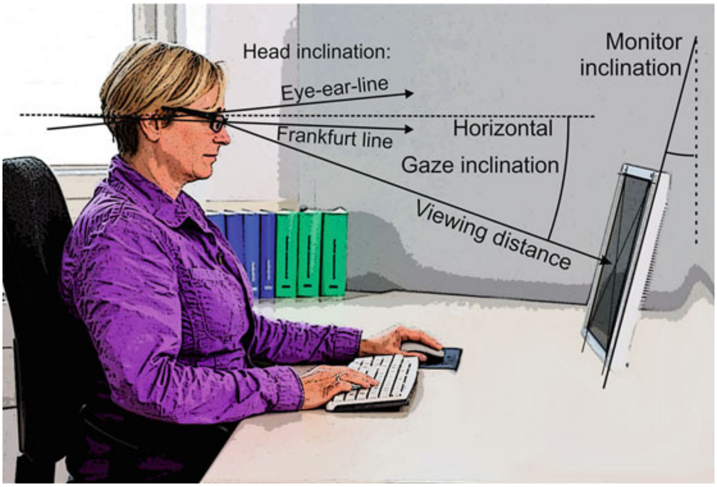


Fig. 1 Photograph at a computer workplace with included lines of reference for geometrical analyses of ergonomic workplace parameters (see text). From: Meinert et al. [46]

Some studies measured the effect of vertical monitor position on head and eye inclination. Burgess-Limerick et al. [7] showed that, e.g., a 10° change in gaze inclination induced an about 4° change in eye inclination relative to the head and an about 4.4° change in head inclination relative to the neck; the remaining change was performed by the neck relative to the trunk. From mechanical modelling, Burgess-Limerick et al. [7] suggested a favourable gaze direction of 15° below horizontal. Dellemann and Berndsen [10] varied the vertical monitor position in a laboratory task of computer work: a screen centre 10° below eye level was preferred and resulted in the minimal postural discomfort in the neck and back. The corresponding favourable head posture during work was 5° lower than when looking straight ahead along the horizontal; a 10° change in gaze inclination led to an about 5° change in head inclination; see also [9].

In the studies of Grandjean et al. [13], Jaschinski et al. [26, 27] participants adjusted both the viewing distance and vertical monitor position which resulted in a rather long viewing distance round 90 cm, probably to reduce the load on accommodation and convergence [24]. In this condition, the preferred downward gaze inclination was about −10°.

The following studies used closer viewing distances around 60 cm. Lie and Forstervold [35] reported more favourable conditions at a gaze inclination of −45° (compared to the condition of −15°) in a laboratory study with non-presbyopic users. Fostervold et al. [12] found a long-term benefit (particularly in musculoskeletal symptoms) at a gaze inclination of −30° (compared to −15°) in a 1-year field study with participants of all age groups. The physiological implications of lower gaze inclinations are reviewed by Fostervold [11].

Thus, the results of these studies partly differ considerably. There seems to be an effect of viewing distance: the closer the monitor is placed, the higher the load on accommodation and convergence of the eyes, so that lower monitor positions are favourable (see Sect. 1.1). A further influence could be the change in monitor technology over the last decades. The first studies were made around 1980 [13], when monitors with large CRT-screen could only be used in rather high positions. Today, flat screens allow for much more flexible monitor positions; notebooks are used with rather low gaze directions. Actual field studies showed preferred gaze inclination around −20° [1, 46, 58], which is lower than most earlier studies. Thus, different technologies might have induced a bias in study results.

1.4 Approach of the Present Study

It is the aim of the present research to investigate head inclination with respect to implications for the computer workplace. Therefore, the head inclination was varied as the independent variable and several dependent variables were measured in the following two studies of our research team:

– The laboratory head tracking study of Masseida et al. [43] reported the tracking precision, the electrical activity of the sternocleidomastoid muscle and the subjective rating of neck strain.

– The office field study of Weidling and Jaschinski [58] showed subjective musculoskeletal strain as function of head inclination.

The present paper refers to these studies and provides the following additional analyses of stress-strain functions regarding head inclination:

1. the electromyographical analyses are more elaborated compared to Masseida et al. [43] since additionally the activity of the splenius capitis muscle is analysed [20].
2. a synopsis across these studies covers the following dependent variables
   – the electrical activity of neck muscles that represent the physiological effects
   – the subjective rating of strain that the participants reported in questionnaires, both in the laboratory head tracking tasks and in the field study of computer work
   – the motor performance in a laboratory head tracking task, i.e. the precision of moving the head horizontally.

2 Methods

2.1 The Laboratory Head Tracking Task

The methods are fully described in Masseida et al. [43] and are shortly summarised here.

2.1.1 The Head Tracking Procedures

The head mouse system SmartNAV (Natural Point)—originally designed for computer interaction—was applied on a 44° wide back projection screen (Fig. 2). The camera of this system was installed above the screen and recorded the position of an infrared reflecting marker that was attached at a helmet worn by the participant sitting in front of the screen at a 80 cm viewing distance. The SmartNAV software analysed the head position in the video image and moved the cursor (red line on the screen) depending on head position. In this way, the participants performed horizontal head movements to keep a cursor (vertical line) on a target (rectangle) which moved in a line from 20° left to 20° right relative to the screen centre. The velocity was basically a 2 cm/s linear movement, superimposed by two sinusoidal components with frequencies of 0.1 and 0.3 Hz, each with 3.7 cm amplitude. This target mostly moved from left to right, but sometimes moved backward. This tracking task was made with an upright head posture and with head inclinations higher and lower by 10 and 20°, respectively. An upright head posture is defined as a horizontal Frankfurt line, which is the line connecting the tragus at the ear and the inferior margin of the orbit. The latter bony socket can easily be

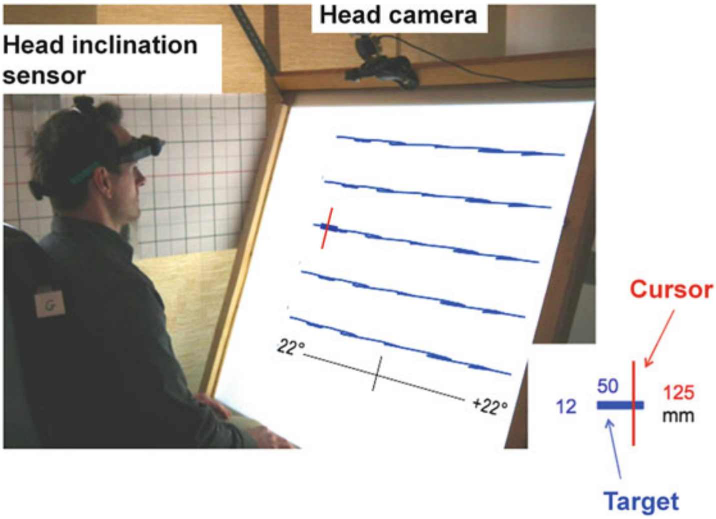


Fig. 2 Experimental setup of the head tracking task presented on a large back-projection screen. The participant is wearing a helmet with an infrared marker of the head-mouse-system. The participants task was to keep a curser (*red line*) on a horizontally moving target (*blue bar*) by moving the head horizontally. This head tracking task was performed at five different head inclination angles. At each head inclination, the target was presented at a vertical position on the screen that allowed for a comfortable gaze direction of the eyes. The trace of head movements is illustrated by the *blue lines* in this figure, but not on the back-projection screen during the experiment. Adapted from Masseida et al. [43]

identified at the scalp in anatomic studies, but is inconvenient in ergonomic studies since it cannot readily be observed by the experimenter. We generally used side-view photographs of the study participants where the eye-ear line can directly be identified geometrically. The eye-ear line connects the tragus at the ear and the outer canthus at the eye. On average, the eye-ear line is 11° higher (more positive) than the Frankfurt line [47]. Therefore, the chosen levels of head inclination in the present study were eye-ear lines of −9, 1, 11, 21, 31°, which correspond to Frankfurt lines of −20°, −10°, 0°, +10°, +20° (calculated based on the mean difference of 11°); positive and negative angles mean upward and downward head inclinations, respectively. The dependent variable was the tracking precision, conventionally expressed as the tracking error, i.e. the standard deviation of the horizontal deviation of the cursor from the target.

2.1.2 Electromyogram (EMG)

For recording the electromyogram of the sternocleidomastoid muscle, a pair of electrodes was attached below the ear along the fibre direction of this muscle which is almost vertical in an upright head position. A few centimetres more backward, another pair of electrodes was attached for recording the activity of the splenius capitis muscle, at a lateral neck position where this muscle is not covered by the trapezius muscle (Fig. 3). Two pairs of electrodes were attached on both sides of the

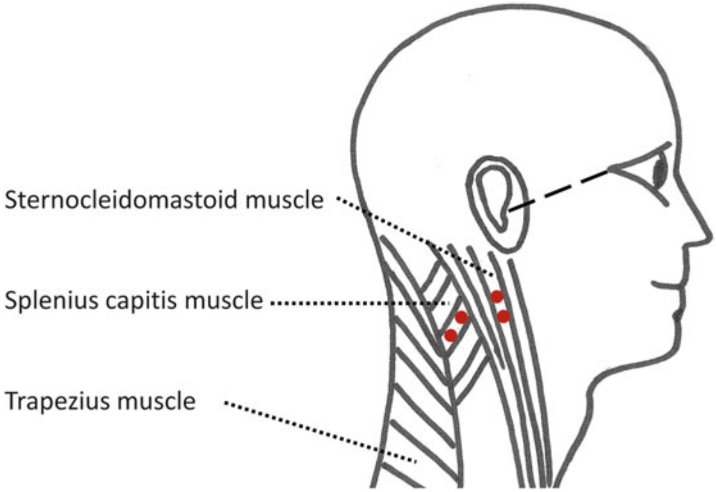


Fig. 3 Illustration of neck muscles and the corresponding positions of EMG electrodes. Note that the splenius capitis muscle is mostly covered by the trapezius muscle and can only be reached for electromyography between the sternocleidomastoid muscle and the trapezius muscle. The eye-ear line is included as anatomic reference line of head inclination

neck, so that four bipolar recordings were made with a reference electrode near C7. The EMG recorder PS 11 UD (Thumedi) was used. The myoelectrical activity was derived from the band-pass filtered raw electromyogram (10–500 Hz) by rectification and time moving averaging with a 250 ms window. For details see Hägg et al. [15].

2.2 Ergonomic Intervention Field Study with Presbyopic Employees

Weidling and Jaschinski [58] performed an ergonomic intervention field study of computer work in offices that were equipped with computer monitors with flexible support systems. The aim of this study was to lower the monitor position to better correspond to physiological properties of eye and head inclination and to the optical properties of general-purpose progressive lenses. The intervention had two steps. First, the monitor position was lowered by the participants themselves, based on ergonomic and optometric information provided in a flyer. Second, an optometric expert measured the individual physiological and optometric parameters and lowered the monitor accordingly. This lowering of the monitor position led the participants to lower their head inclination. The subjective rating of musculoskeletal strain was assessed with a 4-item questionnaire. The study comprised two subgroups wearing different types of spectacle lenses to correct for presbyopia. With respect to head inclination effects, these two groups gave basically the same results. Therefore, the mean data of the complete sample of 25 participants are reported here.

3 Results

3.1 Head Tracking Task: Precision, Subjective Strain, Muscular Activity

In the laboratory head tracking task of Masseida et al. [43], the precision of head tracking showed a nearly linear change with head inclination (Fig. 4a): the more the head was raised the smaller was the tracking error, i.e. the better was the tracking precision. Although the effect was small in amount (about 7 %) it was significant since the distribution of the 16 individual linear regression coefficients deviated from zero ($t(15) = 2.7$, $p = 0.015$, two-tailed). Moreover, the individual effects had a correlation of $r = 0.42$ ($p = 0.053$, $n = 16$) between the two experimental sessions that were made; this suggests a trend that some participants seemed to have a larger tracking effect than others.

The subjective rating of neck strain during the head tracking task showed a parabolic function (Fig. 4b) since the lowest ratings appeared at an upright head posture (eye-ear line 11° above horizontal) and the ratings increased the more the head was raised or lowered. The distribution of the 16 individual quadratic regression coefficients deviated from zero ($t(15) = 3.85$, $p = 0.0016$, two-tailed).

The activity of the two types of muscles is plotted in Fig. 4c, d. The splenius capitis muscle is known to predominantly hold the head against the gravity that exerts an increasing torque the more the head is declined. Accordingly, the electrical activity of the splenius capitis muscle was found to linearly increase when the head was lowered (Fig. 4c). The sternocleidomastoid muscle showed a parabolic function in electrical activity: the minimal activity occurred at an upright head position and the activity increased the more the head was raised or lowered (Fig. 4d). The two different response functions of these two types of muscles were replicated in two recordings that were made at the left and right side of the neck.

3.2 Computer Work in an Office Field Study

In the course of a two-step intervention procedure, Weidling and Jaschinski [58] induced lower head inclinations in two steps relative to the initial head inclination. During this intervention, the mean head inclination was significantly lowered from an initial level of 14.1° to 11.7° and 9.8° after the first and second step of the intervention, respectively ($F(2,48) = 16.267$, $p < 0.01$). The corresponding levels of the reported musculoskeletal strain were 3.1, 2.6 and 2.3 and showed a significant reduction ($F(1.252, 30.040) = 10.635$, $p < 0.01$). Both effects were significant, although the range of change in head inclination was only 4.3°. These three

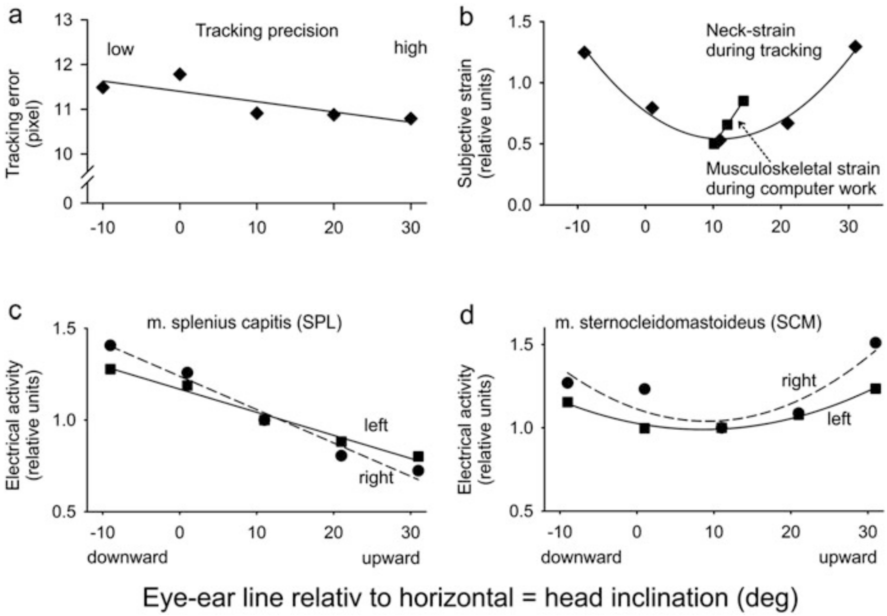


Fig. 4 Overview of the different ergonomic stress-strain functions for head inclination, given as eye-ear line relative to horizontal (11° corresponds to an upright head posture, on the average): tracking error (pixel) (**a**), subjective ratings of strain comparing the neck strain in the tracking task and the musculoskeletal strain at computer work in the office field study (**b**), electrical activity of the splenius capitis muscle (**c**) and the sternocleidomastoid muscle (**d**), both recorded from the left and right side of the neck. Data are normalized to facilitate comparisons

observations represent a stress-strain function of subjective musculoskeletal strain versus head inclination in Fig. 4b.

4 Discussion

4.1 Electrical Activity of Neck Muscles

Concerning the methodology, it can be confirmed that the left and right side of the neck gave similar functions of the electrical activity when the head inclination was varied over a range of 40°. Thus, these recordings appear valid and the observed changes agree with the general findings reviewed by Sommerich et al. [54] that changes in neck muscle electrical activity can typically be found when the head inclination is changed.

The linear increase in the activity of the splenius capitis muscle with lowered head inclination confirms the expectation that this muscle compensates the increasing effect of gravity when the head is lowered [56]. More interesting is the parabolic function depending on head inclination that was found for the sternocleidomastoid muscle which is known to perform—besides others—the horizontal rotation of the head besides the stabilization of the head against gravity.

Bexander et al. [6] investigated effects of left-right rotation of the head (that was held upright): the electrical activity of the left sternocleidomastoid muscle and also of the right splenius capitis muscle increased with right rotation, but not with left rotation. This agrees with our findings. Additionally, we found that increasing activity in the sternocleidomastoid muscle as well as in the splenius capitis muscle may be accompanied by decreasing activities in the corresponding contra lateral muscles.

4.2 Tracking Error and Electrical Muscular Activity

In many studies of motor control, the size of the tracking target is changed as independent variable in order to manipulate the required precision of a tracking task; parameters of the movement and the muscular electrical activity are measured as dependent variable. In goal-directed arm movements smaller tracking errors are associated with an increased co-activation of antagonistic muscles [14, 49]. However, a different pattern of result was found in the following studies of trunk movements. Willigenburg [60] varied the size of a tracking target at different trunk inclinations: the movement pattern was affected as expected, but trunk muscle activity was not consistently affected by the precision demand in the first study. In a follow-up study [59], increasing trunk inclination lead to larger tracking errors, but increasing muscular activity. Therefore, Willigenburg et al. [59] suggested that trunk motor control is regulated by feedback corrections and not by a co-activation strategy as in arm movements. Thus, the mechanisms of motor control may differ between muscular systems and therefore no clear hypotheses and predictions could be made beforehand in the present head inclination study that includes horizontal movements at different vertical head inclinations.

If we wish to explain the linear increase in tracking error with lowering head inclination, the sternocleidomastoid muscle seems not to play a role, since for this muscle the observed stress-strain function was not linear, but parabolic. Rather, the linear stress-strain function of the splenius capitis muscle showed that with more downward head inclination (angles with negative sign) the electrical activity of this muscle increased and the tracking error also increased. This suggests the interpretation that the tracking error increases since the tension of the splenius capitis muscle increases. This view is different from the above hypothesis for hand or arm tracking, where a large co-contraction is associated with smaller tracking errors. However, note that hands and arms are used for reaching or grasping, while the head is vertically hold more or less stable against gravity. Moreover, the splenius capitis muscle operates mainly against the weight of the head, whereas antagonistic activities of other muscles may play a minor role. This resembles the control of the inclination of the trunk, which also has to be hold against gravity. The present tracking results for head inclination and those of Willigenburg et al. [59, 60] for trunk inclination resemble in the way that larger electric activity was associated with a larger tracking error. For this situation, Willigenburg et al. [59, 60] suggested a feedback control mechanism for the trunk, as opposed to antagonistic control mechanisms that may apply to hand or the arm.

4.3 Stress-Strain Function for Subjective Strain

The subjective neck strain during tracking showed a parabolic function depending on head inclination in a way as it was expected since the lowest strain was found at an upright head posture with an increase at both higher and lower head inclinations. This subjective stress-strain function does not correspond to the linear stress-strain functions that appeared for the tracking error and for the electrical activity of the splenius capitis muscle. Rather, the stress-strain function of the sternocleidomastoid muscle showed a parabolic function and—therefore—one might suspect that the subjective neck strain could physiologically originate from the activity of the sternocleidomastoid muscle.

It is tempting to compare the results of the laboratory tracking task with those of the subjective musculoskeletal strain reported in the office field study. It is true that the conditions of testing in the laboratory and work in an office differ considerably. So the level of strain will be different, but the stress-strain functions may be compared as in Fig. 4b, after a normalisation at upright head position. Obviously, the range of head inclination is much smaller at the office computer workstations and therefore a parabolic function cannot be expected. But still, both functions agree in that raising the head starting from about 11° (eye-ear line) leads to an increase in subjective strain. Similarly, Meinert et al. [46] found a preferred head inclination of $8.0 \pm 5.5°$ (eye-ear line) after mostly non-presbyopic employees had adjusted their monitor to a comfortable position in a web-based ergonomic intervention.

Summarizing these studies it can be concluded that in the present two studies, the average favourable head inclination (eye-ear line) is near 11°, which resembles the average of previous studies that reported a range of 6–16° [2, 22, 23, 42, 43]. However, such a favourable mean eye-ear line cannot be applied to each individual user because the eye-ear line is not directly related to the anatomy muscular functions of the neck and, further, it depends on the anatomical eye and ear position of the individual head. More important is the finding of the field studies, that lowering the head inclination by only a few degrees is associated with a reduction in subjective musculoskeletal strain. This is confirmed by an optometric intervention study of Jaschinski et al. [30] where participants compared two types of glasses to correct for presbyopia at computer work: as expected, the computer-vision progressive lenses allowed for a lower head inclination ($15.4 \pm 5.2°$) than general purpose progressive lenses ($13.1 \pm 5.1°$); this lowering of head inclination led to a significant reduction in subjective musculoskeletal strain.

Therefore, in a practical emulation of the present results, the computer user should not try to precisely measure the eye-ear line. Rather, he/she may experience higher and lower head inclinations to find a favourable condition between these extreme positions as suggested by Masseida et al. [43] and Weidling and Jaschinski

[58]. Our ergonomic-optometric website consultation tool provides further information (http://ergonomic-vision.ifado.de).

4.4 General Conclusions

The present study represents a first step to investigate the work-related physiological mechanisms of head movements using an integrated approach that includes different indicators of strain, as performance in a head movement task (tracking error), physiological reaction (electrical muscle activity) and subjective rating of strain; the latter were assessed in a laboratory tracking task and an office field study. The reported general findings refer to the mean results in these samples, while data also suggested marked individual differences that would require further, more detailed analyses. The important role of individual differences was also emphasized by Straker et al. [56]. The present methodology seems to be appropriate for further research and tracking tasks may allow investigating the underlying muscular control mechanisms. The analyses of stress-strain functions appear to be a useful ergonomic research strategy to explain the relation between the physiological mechanisms, the quality of movements and the subjective strain experienced by the users of computer workstations [38, 39, 44].

Certainly, the present data do not yet allow for definite conclusions about head movement control mechanisms during computer work because of several limitations of the present studies. These could be overcome by considering the following aspects in future research. Different types of tracking tasks should be investigated and the concurrent eye movements should also be included. One could stepwise change the task from artificial, laboratory tracking tasks (as in the present study) to more realistic tasks at computer workplaces. Additional types of neck and back muscles may be investigated in order to represent the complexity of the muscular mechanisms. The time of task should be extended to cover effects of fatigue over time. The samples of subjects may not only include young subjects, but also subjects with a limited power of accommodation so that near vision glasses are required in the age of presbyopia as summarized in this volume by König and Jaschinski [33].

The present research supports the concept of combining different types of ergonomic stress-strain functions as outlined in Sect. 1.1. This may be a useful research tool for uncovering the underlying physiological mechanisms of strain. Moreover, since computer work represents a load on the eyes, neck, and the back, it is evident that visual and musculoskeletal aspects should simultaneously be included in such research [3, 36, 37, 53].

Acknowledgments The authors wish to thank the members of the present research team including Ewald Alshuth, Mirjam König, Ute Lobisch, Judith Masseida, Jens Philipp, Carola Reiffen and Patrick Weidling.

References

1. Allie P, Bartha MC, Kokot D et al (2010) A field observation of display placement requirements for presbyopic and prepresbyopic computer users. Proceedings of the human factors and ergonomics society annual meeeting 54:709–713
2. Ankrum DR, Nemeth KJ (2000) Head and neck posture at computer workstations – What's neutral? In: Society THFaE (ed) 14th Triennial congress of the international ergonomics association. Santa Monica: The Human Factors and Ergonomics Society, San Diego, California, USA, pp 565–568
3. Anshel JR (2005) Visual ergonomics handbook. CRC Taylor & Francis, Boca Ratoon, FL
4. Ariens GA, Bongers PM, Douwes M et al (2001) Are neck flexion, neck rotation, and sitting at work risk factors for neck pain? Results of a prospective cohort study. Occup Environ Med 58:200–207
5. Bardy BG, Oullier O, Bootsma RJ et al (2002) Dynamics of human postural transitions. J Exp Psychol Hum Percept Perform 28:499–514
6. Bexander CS, Mellor R, Hodges PW (2005) Effect of gaze direction on neck muscle activity during cervical rotation. Exp Brain Res 167:422–432
7. Burgess-Limerick R, Plooy A, Fraser K et al (1999) The influence of computer monitor height on head and neck posture. Int J Ind Ergon 23:171–179
8. Chaffin DB, Andersson GBJ, Martin BJ (1999) Occupational biomechanics. Wiley, New York, NY
9. Delleman NJ (2004) Head and neck. In: Delleman NJ, Haslegrave CM, Chaffin DB (eds) Working postures and movements: tools for evaluation and engineering. CRC Press, Boca Raton, FL, pp 87–105
10. Delleman NJ, Berndsen MB (2002) Touch-typing VDU operation: workstation adjustment, working posture and workers' perceptions. Ergonomics 45:514–535
11. Fostervold KI (2003) VDU work with downward gaze: the emperor's new clothes or scientifically sound? Int J Ind Ergon 31:161–167
12. Fostervold KI, Aaras A, Lie I (2006) Work with visual display units: Long-term health effects of high and downward line-of-sight in ordinary office environments. Int J Ind Ergon 36:331–343
13. Grandjean E, Hunting W, Pidermann M (1983) VDT workstation design: preferred settings and their effects. Hum Factors 25:161–175
14. Gribble PL, Mullin LI, Cothros N et al (2003) Role of cocontraction in arm movement accuracy. J Neurophysiol 89:2396–2405
15. Hägg GM, Luttmann A, Jäger M (2000) Methodologies for evaluating electromyographic field data in ergonomics. J Electromyogr Kinesiol 10:301–312
16. Heuer H (1990) Die Ruhelage der Augen und ihre Konsequenzen. In: Meinecke C, Kehrer L (eds) Bielefelder Beiträge zur Kognitionspsychologie, pp 139–191
17. Heuer H, Bruwer M, Romer T et al (1991) Preferred vertical gaze direction and observation distance. Ergonomics 34:379–392
18. Heuer H, Hollendiek G, Kröger H et al (1989) [Rest position of the eyes and its effect on viewing distance and visual fatigue in computer display work] in German. Z Exp Angew Psychol 36:538–566
19. Heuer H, Owens DA (1989) Vertical gaze direction and the resting posture of the eyes. Perception 18:363–377
20. Jäger M, Philipp J, Masseida J et al (2014) Untersuchungen zur Aktivität der Halsmuskulatur bei regulatorischen Kopfbewegungen zur Identifizierung einer günstigen Kopfneigung. Umweltmed Hygiene Arbeitsmed 19:133
21. Jainta S, Jaschinski W (2002) Fixation disparity: binocular vergence accuracy for a visual display at different positions relative to the eyes. Hum Factors 44:443–450

22. Jampel RS, Shi DX (1992) The primary position of the eyes, the resetting saccade, and the transverse visual head plane. Head movements around the cervical joints. Invest Ophthalmol Vis Sci 33:2501–2510
23. Jaschinski W (2008) [Low monitor positions when using general purpose progressive lenses at computer work] in German. Zbl Arbeitsmed 58:172–180
24. Jaschinski W (2002) The proximity-fixation-disparity curve and the preferred viewing distance at a visual display as an indicator of near vision fatigue. Optom Vis Sci 79:158–169
25. Jaschinski W, Heuer H (2004) Vision and eyes. In: Delleman NJ, Haslegrave CM, Chaffin DB (eds) Working postures and movements, Tools for evaluation and engineering. CRC Press, Boca Raton, FL, pp 73–86
26. Jaschinski W, Heuer H, Kylian H (1998) Preferred position of visual displays relative to the eyes: a field study of visual strain and individual differences. Ergonomics 41:1034–1049
27. Jaschinski W, Heuer H, Kylian H (1999) A procedure to determine the individually comfortable position of visual displays relative to the eyes. Ergonomics 42:535–549
28. Jaschinski W, Koitcheva V, Heuer H (1998) Fixation disparity, accommodation, dark vergence and dark focus during inclined gaze. Ophthalmic Physiol Opt 18:351–359
29. Jaschinski W, König M, Masseida J et al (2014) Vertical position of computer monitors – a physiological concept. In: Arbeitswissenschaft Gf, Jäger M (eds) 60. Kongress der Gesellschaft für Arbeitswissenschaft: Gestaltung der Arbeitswelt der Zukunft. GfA-Press, Dortmund, TU und Hochschule München, pp 498–500
30. Jaschinski W, König M, Mekontso T et al (2015) Comparison of progressive addition lenses for general purpose and for computer vision: an office field study. Clin Exp Optom 98:234–243
31. Jaschinski W, König M, Mekontso TM et al (2015) Computer vision syndrome in (pre-) presbyopia: a questionnaire field study on the effects of spectacle lens types. Clin Exp Optom 98:228–233
32. König M, Haensel C, Jaschinski W (2015) How to place a computer monitor: measurements of vertical zones of clear vision with presbyopia corrections. Clin Exp Optom 98:244–253
33. König M, Jaschinski W (2016) The influence of glasses on the physiology and ergonomics in presbyopic people at visual display workstations. In: Advances in ergonomic design of systems, products and processes: proceedings of the annual meeting of GfA 2015. Springer, Heidelberg, pp 239–248
34. Lee KS, Waikar AM, Aghazadeh F, Tandon S (1986) An electromyographic investigation of neck angles for microscopists. In: Proceedings of the human factors society 30th Annual meeting, Santa Monica, CA: Human Factors Society, pp 548–551
35. Lie I, Fostervold KI (1995) VDT work with different gaze inclinations. In: Grieco A, Molteni G, Picolli B, Occipinti E (eds) Work with Display Units 94. Elsevier, Amsterdam
36. Long J (2014) Forging partnerships between optometrists and ergonomists to improve visual comfort and productivity in the workplace. Work 47:365–370
37. Long J, Helland M (2012) A multidisciplinary approach to solving computer related vision problems. Ophthalmic Physiol Opt 32:429–435
38. Luttmann A, Kylian H, Schmidt K-H et al (2002) Untersuchung von Muskelbelastung und Beschwerdehäufigkeit bei Büroarbeit. Zbl Arbeitsmed 52:305–317
39. Luttmann A, Schmidt K-H, Jäger M (2010) Working conditions, muscular activity and complaints of office workers. Int J Ind Ergon 40:549–559
40. Manzey D, Lorenz TB, Heuers H et al (2000) Impairments of manual tracking performance during spaceflight: more converging evidence from a 20-day space mission. Ergonomics 43:589–609
41. Marcus M, Gerr F, Monteilh C et al (2002) A prospective study of computer users: II. Postural risk factors for musculoskeletal symptoms and disorders. Am J Ind Med 41:236–249
42. Masseida J, Perske F, Jaschinski W (2013) Conditions of physiologically favourable gaze inclination. Z Arb Wiss 67:221–231
43. Masseida J, Philip J, Wicher C et al (2013) Several indicators of the physiological favourable head inclination. Z Arb Wiss 67:207–219

44. Masuda K, Masuda T, Sadoyama T et al (1999) Changes in surface EMG parameters during static and dynamic fatiguing contractions. J Electromyogr Kinesiol 9:39–46
45. Mateo B, Porcar-Seder R, Solaz JS et al (2010) Experimental procedure for measuring and comparing head-neck-trunk posture and movements caused by different progressive addition lens designs. Ergonomics 53:904–913
46. Meinert M, König M, Jaschinski W (2013) Web-based office ergonomic intervention on work-related complaints – a field study. Ergonomics 56:1658–1668
47. Menozzi M, Von Buol A, Itschner L et al (1996) Die Eignung verschiedener cephalometrischer Bezugslinien für die Bestimmung der vertikalen Ausrichtung des Kopfes. [The suitability of various cephalometric planes for the determination of vertical alignment of the head]. Z Morphol Anthropol 81:205–215
48. Menozzi M, Von Buol A, Krueger H et al (1994) Direction of gaze and comfort: discovering the relation for the ergonomic optimization of visual tasks. Ophthalmic Physiol Opt 14:393–399
49. Meulenbroek RG, Van Galen GP, Hulstijn M et al (2005) Muscular co-contraction covaries with task load to control the flow of motion in fine motor tasks. Biol Psychol 68:331–352
50. Oullier O, Bardy BG, Stoffregen TA et al (2002) Postural coordination in looking and tracking tasks. Hum Mov Sci 21:147–167
51. Ripple PH (1952) Variation of accommodation in vertical directions of gaze. Am J Ophthalmol 35:1630–1634
52. Rohmert W (1984) Das Belastungs-Beanspruchungs-Konzept. Z Arb Wiss 38:193–200
53. Rosenfield M (2011) Computer vision syndrome: a review of ocular causes and potential treatments. Ophthalmic Physiol Opt 31:502–515
54. Sommerich CM, Joines SM, Hermans V et al (2000) Use of surface electromyography to estimate neck muscle activity. J Electromyogr Kinesiol 10:377–398
55. Sommerich CM, Joines SM, Psihogios JP (2001) Effects of computer monitor viewing angle and related factors on strain, performance, and preference outcomes. Hum Factors 43:39–55
56. Straker L, Skoss R, Burnett A et al (2009) Effect of visual display height on modelled upper and lower cervical gravitational moment, muscle capacity and relative strain. Ergonomics 52:204–221
57. Wahlström J (2005) Ergonomics, musculoskeletal disorders and computer work. Occup Med (Lond) 55:168–176
58. Weidling P, Jaschinski W (2015) The vertical monitor position for presbyopic computer users with progressive lenses: how to reach clear vision and comfortable head posture. Ergnomics. doi:10.1080/00140139.2015.1035764
59. Willigenburg NW, Kingma I, Hoozemans MJ et al (2013) Precision control of trunk movement in low back pain patients. Hum Mov Sci 32:228–239
60. Willigenburg NW, Kingma I, Van Dieen JH (2010) How is precision regulated in maintaining trunk posture? Exp Brain Res 203:39–49

The Influence of Glasses on the Physiology and Ergonomics in Presbyopic People at Visual Display Workstations: Summary of Three Studies

Mirjam König and Wolfgang Jaschinski

Abstract

The present paper provides a summary of three studies concerning the influence of glasses on the physiology and ergonomics in presbyopic computer users. The first study was a survey of 175 VDU (visual display unit) workstation users wearing glasses, age 35 and older. In the second study 23 subjects of the survey study (first study) tested a progressive addition lens (PAL) for general purpose and a PAL for computer vision. Within the third study near- and far-point curves (range of clear focus) as a function of eye inclination of the 23 people of the lens-testing study (second study) were determined. People wearing glasses for far-vision (single vision lenses) reported higher scores of eye strain, musculo-skeletal strain and head complaints with increase of the daily duration of computer work: the longer they worked at computer stations the higher the strain were scored. This effect did not appear with the other lens types. The subjects of the lens-testing study (second study) showed a head inclination 2.3° lower when wearing the PAL for computer vision. 44 % of the subjects decided in favor of the PAL for computer use. Focus range curves can show descriptively where clear vision is possible and where to best place the monitor (third study).

Keywords

Glasses • Monitor • Range of clear focus • Presbyopia • Eye inclination • Head inclination

M. König (✉) • W. Jaschinski
Leibniz-Institut für Arbeitsforschung an der Technischen Universität Dortmund, Dortmund, Germany
e-mail: koenig@ifado.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_16

1 Introduction

Presbyopic computer users need additional positive refraction for near-vision to compensate for the reduced accommodative capacity—they need glasses for clear near-vision. The use of progressive addition lenses (PAL) is widespread and well-tolerated in general [7]. However, for the use of progressive addition lenses at the VDU workstation optometric and ergonomic aspects should be taken into account.

The present paper provides a summery of three studies concerning the influence of glasses on the physiology and ergonomics in presbyopic computer users. The first study is a survey of 175 VDU workstation users wearing glasses, age of 35 and older [2]. A smaller group of participants in the survey study formed the sample group of the second study—testing two different progressive addition lenses [3]. In the third study, on the basis of measurements during the use of the two different PALs, clear focus rages (vertical zones of clear vision) were determined for these people [5].

2 First Study: Online-Questionnaire

2.1 Method

In the first study the employees of the Inland Revenue office in Dortmund, age 35 and older (mean ± SD = 52.0 ± 6.7) completed an online questionnaire. At the VDU-workstations the participants all needed—different kinds of—spectacles. The questionnaire contained questions about the glasses which are used at the workplace as well as questions about visual conditions and possible complaints. From a list of 25 items of the questionnaire, a factor analysis identified five factors of complaints regarding “Vision at the monitor”, “Ocular strain”, “Musculoskeletal strain”, “Dizziness” and “Dynamic vision” and a single item on “Headache”.

Different ergonomic conditions, e.g. the head inclination while working at the monitor, were documented by side-view photographs taken in typical working conditions from 69 of the 175 subjects who agreed to this procedure. The participants were asked to maintain a normal working posture and look at the centre of the monitor. Head inclination angle was described by the eye-ear line (from the outer canthus to the tragus of the ear) relative to horizontal. More positive angles indicate more upwardly tilted inclinations.

This study was not an intervention study. The aim was to describe the status quo to identify possible differences in computer vision syndrome based on the use of different types of lenses to correct for presbyopia as they typically apply to real working conditions.

The visual functions of the participants and the optical conditions of the glasses were not measured. Employees without glasses were not included, because the aim of this study is to compare influences and conditions of different types of lenses on the physiology and ergonomic at visual display workstations.

2.2 Results

The 175 participants in the study used five different types of glasses. 85 employees used progressive addition lenses (PAL) for general purpose, 13 participants wore PALs for computer use, only two people relied on bifocal glasses for computer use, single far-vision lenses were used by 41 people and the common single near-vision lens was used by 34 employees. The age of the glasses were up to 2 years in 55 %, up to 4 years in 25 %, and up to 3 years in 9 % of the sample; 11 % had older glasses. The following data referred to the members of the subsample (n = 114) performing primarily occupational tasks with vision at the monitor and paper documents. They did not have to speak with customers, where intermediate or far vision would be necessary. The duration of the daily computer work varied between 2 and 8 h. Wearers of single far-vision glasses (n = 25) reported increasing "Ocular strain", "Musculoskeletal strain" and "Headaches" the more working hours they had to spend at VDU workstations (Fig. 1).

The development of presbyopia differs between individuals. Therefore, employees starting 35 years and older were recruited to include the beginning stages of presbyopia. Thus, in this sample, there were also a few people who wear single far-vision glasses which do not correct presbyopia. Here the interesting effect was that these people have had more complaints when the length of time spent on computer work was increased. This effect was smaller for other lens types (Fig. 2).

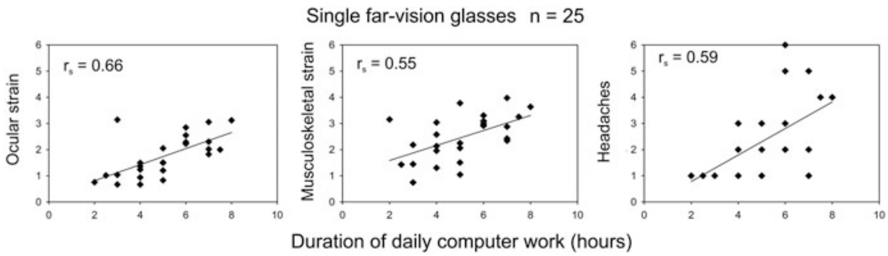


Fig. 1 Correlation between complaints and daily duration of computer work for wearers of single far-vision glasses

Fig. 2 Correlation between "Ocular strain" and daily duration of computer work for wearers of single far-vision glasses, single near-vision glasses and PALs for general purpose

The head inclination of the subjects who worked with general purpose PAL (17.4 deg ± 9.2 deg, n = 39) was, on average, about 7° higher than the head inclination of single far- or near-vision lens wearers (10.3 deg ± 6.1 deg, n = 14; 10.6 deg ± 9.4 deg, n = 10). The latter two types of single-vision lenses do not restrict head inclination and resulted in very similar head inclinations.

3 Second Study

3.1 Method

The second study was a lens-testing study. A group of subjects (n = 23; 13 females; mean ± SD of age: 55 ± 4, range: 46–61 years) of the survey study (first study) tested two types of progressive addition lenses (PAL): a PAL for general purpose with continuous clear vision between infinity and near-reading distance and a PAL for computer use with a wider horizontal zone of clear vision at the monitor and for near-vision but no clear vision into the distance. All participants were experienced PAL wearers. They were from departments where employees worked primarily at the computer in their offices and had no regular conversations with costumers or colleagues. The daily duration of work was 7.9 ± 1.6 h (mean ± SD, range from 4 to 10 h). Most employees worked 5 days per week, three participants worked 4 days and one participant worked 3 days. The mean spherical refractive error, averaged across the left and right eye, was −0.90 D ± 2.61 D (range from −7.25 D to +4.00 D) and the mean cylindrical refractive error was −0.84 D ± 0.66 D (range from 0.00 D to −2.75 D); the mean additional power for near-vision was 2.06 D ± 0.40 D (range from 1.00 D to 2.50 D). Within a period of 4 weeks there was a regular change of the two PALs every week (Table 1).

Part one: One half of the participants included in the study began with the general purpose PAL, the other half with the PAL for computer use. Thus, "glasses one" (see Table 1) was for half of the participants the general purpose PAL and for the other half the PAL for computer use. In this double blind design only the lens producer knew the mapping. The participants only got the information that they have to test two types of PALs for computer work. The questionnaire of the first study was applied with an additional question concerning the quality of "Distance vision". After 4 weeks the participants got complete information about the two

Table 1 Sequence design of the lens-testing study

	Part one: predefined wearing of glasses one and two (only at work)				Part two: free use of glasses one and two (at work and private)							
	Adaptation phase		Test phase									
Week	1	2	3	4	5	6	7	8	9	10	11	12
	Glasses one	Glasses two	Glasses one	Glasses two								

Daily survey in week three and four; weekly survey at the end of week one, two, three and four each spectacles with all their advantages and disadvantages, their principle of operation and their handling.

Part Two: The test persons were allowed to use the glasses freely according to their individual preferences, both at work and at home, for 8 weeks.

At the end of part one and at the end of part two the participants were asked which glasses they prefer (forced choice). The ergonomic setting at the workplace was adapted to the employee according to the BGI 650 at the beginning of the study. This setting should be remained constant over the "part one" with testing the two PALs. To control the ergonomic conditions, a side-view photograph was taken once at the beginning and at the end of week three and four.

3.2 Results

Already at the end of part one when the subjects have had no detailed information about the glasses they observed differences between the two lens types. The vision into the distance was estimated to be better with the general purpose PAL, whereas "Vision at the monitor" was reported to be better with the computer vision PAL (Table 2).

"Musculoskeletal strain" was significant lower (Table 2) when working with computer-vision PALs (intra-individual comparison) where head inclination was significant lower (Table 3). This finding resembles the one in Jaschinski et al. [4].

As shown in Fig. 3a after 4 weeks of quasi-experimental use of the spectacles (part one), 65 % of the participants choose the computer vision PAL (forced-choice). After the information about the optical differences between the lens

Table 2 Statistics of the differences between the two lens types; positive differences indicate a preference for computer vision PALs

	General purpose PAL versus computer vision PAL (mean difference ± SE)	Paired Wilcoxon-test (two-tailed p-values)
Vision at the monitor	0.51 ± 0.13	$W = 35.0$; $p_{cor} = 0.008$
Distance vision	-0.90 ± 0.26	$W = 169.0$; $p_{cor} = 0.021$
Musculoskeletal strain	0.27 ± 0.08	$W = 43.0$; $p_{cor} = 0.044$

Table 3 Ergonomic parameters when viewing the screen center

	General purpose PAL mean ± SD	Computer vision PAL mean ± SD	*t*-test (two-tailed)
Viewing distance (cm) from eye to screen centre	79.7 ± 10.3	78.9 ± 8.7	$t(22) = 0.91$ $p = 0.3$
Gaze inclination (degrees) to the screen centre	-18.3 ± 4.2	-16.0 ± 9.2	$t(22) = -0.71$ $p = 0.3$
Head inclination (degree); angle between eye-ear line and horizontal	15.4 ± 5.2	13.1 ± 5.1	$t(22) = 2.45$ $p = 0.023$

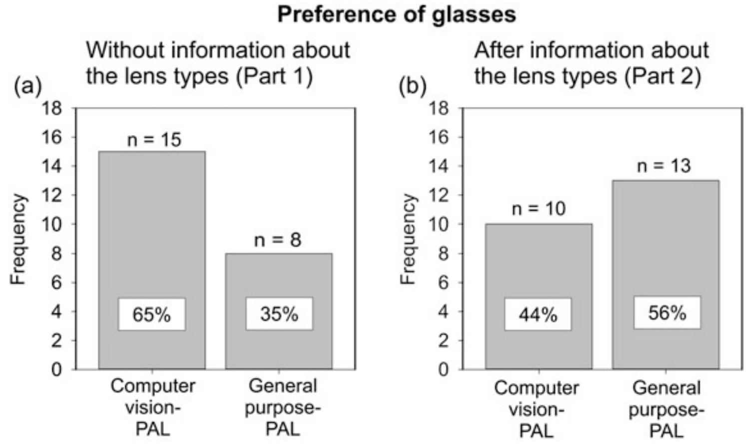


Fig. 3 Percentage of preferences of the glasses; (**a**) after part one, (**b**) after part two

types and after 8 weeks of free use of the two spectacles, 44 % chose the computer vision PAL (Fig. 3b).

While the viewing distance and the gaze inclination was very similar for the two glasses, the head inclination was significantly higher by 2.3° for general purpose PALs compared to computer vision PALs (Table 3).

4 Third Study

4.1 Method

With an "Inclined Optometer", near- and far-points of accommodation for different eye inclination from 0° to −50° with comfortable head posture were measured (Fig. 4). Thus, it was possible to determine for each participant in the second study the near- and far-point curves for the two different PALs.

The near-point was determined by shifting the visual target from a blurred near-position into the distance until clear vision was reported. The far-point was measured by moving the visual target from a blurred distant position closer towards the test person until a sharp target (circle) was seen. To ensure a large range of eye inclinations, a Binoptometer Type 1 was disassembled and the optical parts were rearranged in a new housing that can be tilted in steps of 10° in the range from 0 to −50° downwards (Fig. 4). While collecting the data, a comfortable head position of the test person was indispensable. In such a way relevant results under optimal ergonomical and physiological conditions concerning the monitor position were achieved. The "Inclined Optometer" includes a concave mirror to shift the visual target of constant angular size from a near-point (33 cm) to infinity (proposed by Reiner [9]). To test the vertical zones of clear vision for office working conditions,

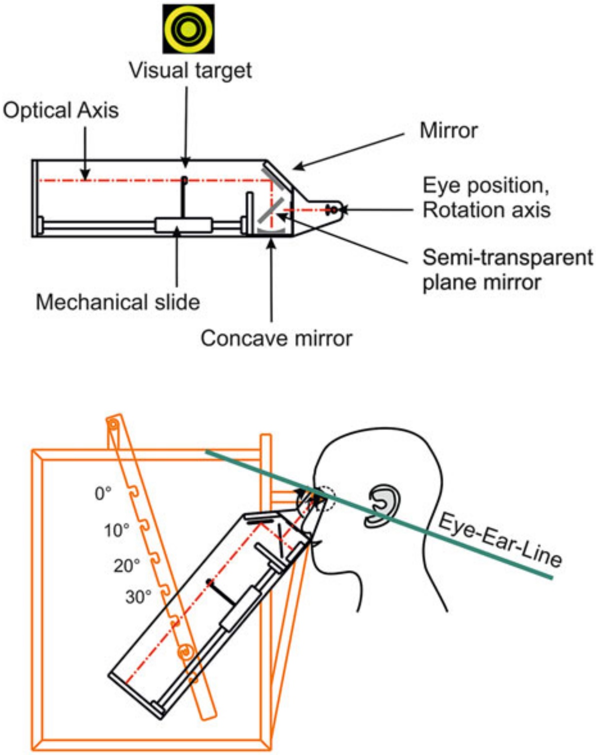


Fig. 4 Design of the “Inclined Optometer”, including the principle of the Binoptometer with a concave mirror to present a visual target of constant angular size from infinity to 33 cm when shifting a target over a range of 25 cm

the mechanical rotation axis of the “Inclined Optometer” was arranged in agreement with the centre of the eyes when the observer adopted a comfortable body and head posture at an office desk. To control for this eye position, two small cameras were installed in the tilt axis to display the position of the two eyes on control monitors. The device comprised a rest for the chin and for the forehead, which were very flexible and adjustable in a way to reach the individual and comfortable head position. The “Inclined Optometer” was used on a motor-adjustable table to easily fit different body sizes.

4.2 Results

The comfortable head inclination (angle between the eye-ear line (see Fig. 4) and the horizontal line from the eye) while measuring the near- and far-points was, on average, 15.75 ± 5.5° (range 4.4–28.7°, n = 23).

Vertical zones of clear vision (focus range curves) could be compiled and evaluated for 22 of the 23 subjects participating in the lens-testing study (second study). Near- and far-points (mean values for 22 observers) were determined and displayed in a diagram with a coordinate system showing horizontal distances

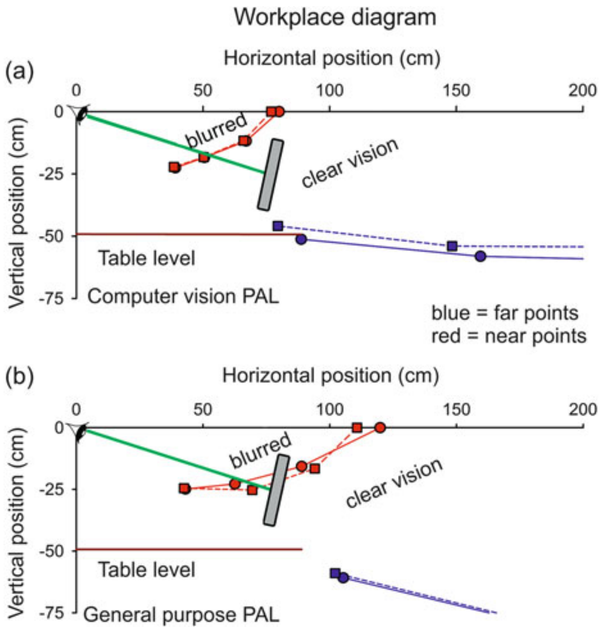


Fig. 5 The vertical zones of clear vision as an average of 22 observers wearing either a computer vision PAL (**a**) or a general purpose PAL (**b**)

(cm) at the x-axis and vertical distances (cm) at the y-axis relative to the eye position (Fig. 5).

Near- and far-points for different gaze inclinations from 0° (points on the x-axis) till −30° built the curves in Fig. 5a and b. The area between these near- and far-point curves was the vertical zone where clear vision is possible. The eye position is at the origin. Figure 5a was the result of the measurement with the computer-vision PALs, Fig. 5b the result of the measurement with the general-purpose PALs. The monitor positions as used by the participants during their daily office work were included (mean values); the green line illustrates the gaze inclination to the centre of the monitor.

At higher gaze inclinations, the near-point curve was transiently shifted to longer viewing distances. The near-points at an inclination of −30° (nearest and lowest red point in the diagrams) were similar for computer vision PAL and for general purpose PAL. But the near-points, measured with an inclination of −20, −10 and 0°, were nearer and the curve was steeper for computer vision PAL with the result that for general-purpose PALs with a comfortable head position, the upper part of the monitor was above the near-point curve and thus appeared blurred. To see the upper part of the monitor clearly with the general-purpose PAL the head must be inclined backwards.

5 Discussion

The questionnaire applied in the first study is able to assess the complaints of computer users depending on the type of lenses which are worn. Effects of the daily duration of computer work that have been reported in other studies [1, 8, 11] are replicated and specified with respect to the type of lenses. It seems to be a risk factor for complaints among people in the early stages of presbyopia to work with single far-vision glasses at the VDU workstation for longer time. "Musculoskeletal strain" is correlated with the daily duration of computer work among the users of single far-vision lenses despite the fact that these lenses do not restrict eye and head movements. Two interpretations of this finding might be applicable: (1) these participants might have transferred higher ratings for "Ocular strain" to other factors without any physiological meaning; or (2) higher "Ocular strain" due to the greater accommodation might induce greater "Musculoskeletal strain" due to a physiological interaction between ocular and musculoskeletal mechanisms [6, 10]. The latter possibility is supported by the finding that "Musculoskeletal strain" exhibit the highest inter-factor-correlation ($r = 0.56$) with "Ocular strain". Physiological studies are needed to decide between these hypotheses because correlation studies cannot provide causal explanations.

In the final forced-choice ratings, concerning the lens-testing study, approximately half of the test persons prefer either the computer-vision PAL or the general-purpose PAL. Individual factors, such as the occupational task, the costs of the glasses, individual visual habits and a possible change of glasses during work, seems to play a role for the individual decision.

Vertical zones of clear vision for different lens types are rarely measured and illustrated. Von Buol [12] measured the horizontal and vertical dimensions of the zones of clear vision for some PALs with an extensive test procedure and he shows a tube-like space of clear vision, which extends from the low and near positions relative to the eyes to a high and distant position. Such focus range curves (measured with comfortable head inclination) show descriptively where clear vision at the workplace is possible. With this information it is easy to place the monitor into the clear field of view. This information is often not given to the wearers of glasses but could be impart by occupational physicians or optometrists.

The measurements with the "Inclined Optometer" show for wearers of general-purpose PALs that the upper edge of the monitor should be, on average, at least 15 cm below the eyes and should stand at a distance of about 75 cm. Higher monitor positions are possible with special PALs for computer vision. Individual visual properties, properties of the different lens types together with the occupational task and individual habits should be considered, ideally.

To give the users of glasses information about the different lens types which can be used at the workplaces and to explain the ergonomically and physiologically relationships, we created the website http://ergonomic-vision.ifado.de/.

Furthermore, a brochure from the "Bundesanstalt für Arbeitsschutz und Arbeitsmedizin" (BAuA) with the title "Gutes Sehen im Büro. Brille und Bildschirm – perfekt aufeinander abgestimmt" can be downloaded here: http://www.baua.de/de/Publikationen/Broschueren/A93.html.

References

1. Hayes JR, Sheedy JE, Stelmack JA, Heaney CA (2007) Computer use, symptoms, and quality of life. Optom Vis Sci 84:739–745
2. Jaschinski W, König M, Mekontso T, Ohlendorf A, Welscher M (2015) Computer vision syndrome in presbyopia and beginning presbyopia: effects of spectacle lens type. Clin Exp Optom 98:228–233
3. Jaschinski W, König M, Mekontso T, Ohlendorf A, Welscher M (2015) Comparison of progressive addition lenses for general purpose and for computer vision: an office field study. Clin Exp Optom 98:234–243
4. Jaschinski W, Luttmann A, Jäger M (2016) Effect of head inclination on electrical neck muscle activity, tracking performance and subjective neck strain. In: Deml B et al (eds) Advances in eronomic design of systems, products and processes. Springer, Berlin, pp S223–S237
5. König M, Haensel C, Jaschinski W (2015) How to place a computer monitor: measurements of vertical zones of clear vision with presbyopic corrections. Clin Exp Optom 98:244–253
6. Lie I, Watten RG (1987) Oculomotor factors in the aetiology of occupational cervicobrachial diseases (OCD). Eur J Appl Physiol Occup Physiol 56:151–156
7. Menozzi M, Bergande E, Sury P (2008) Evaluation of progressive lenses by means of a field study. Z Arbeitswissensch 62:200–209
8. Portello JK, Rosenfield M, Bababekova Y, Estrada JM, Leon A (2012) Computer-related visual symptoms in office workers. Ophthalmic Physiol Opt 32:375–382
9. Reiner J (1980) Ein neues Sehtestgerät. Zbl Arbeitsmedizin 30:110–113
10. Richter HO, Zetterlund C, Lundqvist LO (2011) Eye-neck interactions triggered by visually deficient computer work. Work 39:67–78
11. Steenstra IA, Sluiter JK, Frings-Dresen MHW (2009) The eye-complaint questionnaire in a visual display unit work environment: internal consistency and test-retest reliability. Ergonomics 52:334–344
12. Von Buol A (2002) Der Einfluss von Gleitsichtbrillen auf Kopf-und Augenbewegungen, Dissertation, Zürich

Display Concepts for the Vehicle

The Comparison of an "Emissive Projection Display" and a Conventional Head-Up Display

Verena Knott, Stefan Demmelmair, and Klaus Bengler

Abstract

Driver information systems are increasingly installed in any vehicle. Such support systems include in some cases already Head-Up Displays (HUDs), which show the driver information while driving. According to Bubb et al. [Automobilergonomie (ATZ/MTZ-Fachbuch, Aufl. 2015). Springer Fachmedien Wiesbaden GmbH, Wiesbaden. 2015], the HUD has the advantage that the required accommodative capacity of the eyes is lower or even eliminated for older persons, since the virtual image is visible about 3–4 m ahead of the vehicle. In addition to conventional HUDs, the research is also engaged in contact analogue Head-Up Displays (cHUDs), where the information is represented locally correct (Israel et al., Kontaktanaloge Anzeigen für ACC – im Zielkonflikt zwischen Simulation und Ablenkung. 4. Tagung Sicherheit durch Fahrerassistenz, 15.–16. April 2010, München, TÜV SÜD, 1–7, 2010). However, these concepts cannot be used for all applications. Therefore, a new approach deals with a similar presentation of additional information. In the literature, the concept is called Emissive Projection Display (EPD) and the information is displayed directly on the plane of the windshield and not in a

With permission of Springer Science + Business Media, this chapter is a complement to the following publication, in which already parts of this chapter have been published: [20, Knott V, Demmelmair S, Bengler K (2015) Distraction and Driving Behavior by Presenting Information on an "Emissive Projection Display" Compared to a Head-Up Display. In Harris Don (Ed.) Engineering Psychology and Cognitive Ergonomics. 12th International Conference, EPCE 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2–7, 2015, Proceedings. Springer International Publishing].

V. Knott (✉) • S. Demmelmair • K. Bengler
Lehrstuhls für Ergonomie, Technische Universität München, Munich, Germany
e-mail: knott@lfe.mw.tum.de; stefan.demmelmair@gmail.com

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_17

virtual distance. A driving simulator study investigates the differences in the driver's attention and visual demand.

Keywords
Emissive Projection Display (EPD) • Head-Up Display (HUD) • Dual-task-method • Lane-Change-Task (LCT) • Distraction • Workload

1 Assistance Systems for Driver Information

Due to increasing traffic density and increasing functions in motor vehicles, the demands on the driver are steadily rising. In order to assist the driver in the task of driving, today a variety of driver assistance systems are installed in vehicles. Thus, the traffic safety can be increased and the driving comfort can be improved as well [17]. Since the possibilities for passive safety systems, i.e. systems that reduce the consequences of accidents (such as airbags) are already investigated in detail, now many car manufacturers focus on active safety systems namely driver assistance systems. These systems are used to avoid accidents by either intervening in the driving, for example via a steering torque, or warn drivers if dangers occur [17].

Assistance systems for driver information are increasingly installed in vehicles of middle and/or luxury-class. Such support systems include i.e. Head-Up Displays (HUDs), which show the driver information while driving about speed or navigation. According to Bubb et al. [2], the HUD has the advantage that the required accommodative capacity of the eyes is lower or even eliminated for older persons, since the virtual image is displayed about 3–4 m ahead of the vehicle. The Head-Up Display offers the driver the opportunity to concentrate on traffic, because the parameter "eyes off the road" is reduced or is not required for information perception compared to reading information on the instrument panel [19, 21, 26]. Head-Up Displays are becoming increasingly popular due to the short reading times [26] and reduced visual demand. In addition to conventional HUDs, research is also engaged in contact analogue Head-Up Displays (cHUDs). In this development, the information is displayed correctly on location [14]. However, these concepts cannot be used for all applications. For example, the display area of a HUD is limited to a small area above the engine hood. In addition, HUDs need a large amount of space due to complicated mirror optics, are costly and the information can only be read from the driver's position. To counteract these aspects and to support not only drivers of passenger cars but also other applications, such as trucks or industrial vehicles, a new approach focuses on a similar visualization of supplementary information. The difference to the HUD is that information is displayed directly into the plane of the windshield and the information can be projected on the entire windshield. In the literature, this display type is also known as "Emissive Projection Display" (EPD). Already in 2007, Sun and Liu [23] presented a way for implementation of such a display system (see also Sun and Cheng [22]). The replicated system of the laboratory of the Institute of Ergonomics at the Technische Universität München (TUM) has two main components: An ultraviolet laser

(Spectrum eLite, KVANT, customized) activates a fluorescent film (Media Glass™, Superimaging™) coated on a Plexiglas windshield by the laser beam, at a wavelength of 405 nm to illuminate. In future, this display technique could for example be used in poor visibility conditions, to highlight directly onto the windshield, the driving lane or animals, pedestrians and road signs [18].

But the question is which effects occur for the driver by using an Emissive Projection Display compared to a conventional HUD. As part of a study on volunteers in the static driving simulator, the Institute of Ergonomics examined the effects of such a display concept on the driver's distraction compared to a conventional HUD.

1.1 Head-Up Display (HUD)

1.1.1 Conventional Head-Up Display (HUD)

The Head-Up Display has its origin in aviation. In the late 1950s, such display systems were first installed in military aircraft and helicopters to support the pilots [21]. However, in 1977, the first HUD was ready for series production [12]. In the 1980s, General Motors installed this technology for the first time in motor vehicles [21]. A disadvantage of the former systems was that the displays allowed only visualization in black and white. In addition, the displays were not configurable [7]. An adaptation to the ambient brightness also had to be done manually [21]. In 2003, BMW brought a new type of Head-Up Display to the market, which consists of a thin-film transistor display (TFT display) with a LED light and a mirror unit. This HUD allows a color display and the presentation of different types of information [12, 21]. The typical functioning of a Head-Up Display is shown in Fig. 1.

For image formation, TFT displays are mainly used in the currently installed HUDs, which are illuminated from behind by a powerful LED light source. The light beams are then directed onto the windshield by several pivotable mirrors. Meanwhile HUDs of this version are offered by many car manufacturers worldwide [19, 26].

In Head-Up Displays, different information about the speed limit, warnings and navigation can be provided for the driver. In the lower area of the windshield a virtual image with information appears at a distance of 3–4 m [2]. Thus, the driver doesn't have to adapt his eyes from the driving scenery to an object or information which is closer. This is especially an advantage for older drivers [2]. In addition, the virtual image of the HUD is in the center of the driver's field of view and blends into the driving scene, so that the driver does not have to look away from the road to get the information. Considering that the driver no longer needs to move his head down and no accommodation is needed, the time is significantly reduced for reading the information. However, in order to avoid over-stimulation, the HUD display should not contain too much information [19, 26].

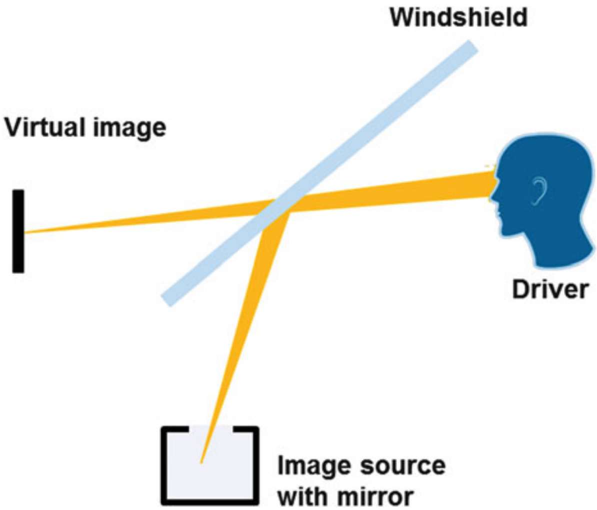


Fig. 1 Functionality of conventional Head-Up Displays (HUDs, according to Schneid [21]): Virtual image behind the windshield

1.1.2 Contact analogue Head-Up Display (cHUD)

Contact analogue Head-Up Displays (cHUDs) can be seen as a further development of conventional HUDs. A display concept is “contact analogue” if reality is supplemented with additional information—comparable to the augmented reality technology [21]. In cHUDs, different information is projected directly into the driver’s field of view and seems to merge directly with the environment. As a result, navigation arrows can be displayed as if they are stationary and lying in the street [14]. Alternatively, information presenting the distance to the vehicle ahead may be displayed [16]. Israel [13] in his work, summarizes a variety of examples to show the design of contact analogue Head-Up Displays. Also the edge of the road or pedestrians at night that were already discovered by a night vision system, can be highlighted by using a cHUD [21].

Due to the direct embedding of information in the real world, the driver is no longer forced to process information represented abstractly and can react more quickly to these [13, 16]. However, the literature also summarizes challenges and disadvantages of contact analogue Head-Up Displays. The virtual image of a contact analogue HUD has a much larger distance, compared to conventional Head-Up Displays, so that the brain of the driver perceives the image as lying in the direction of travel [16]. Accordingly, in cHUDs a larger display area and thus a larger installation space are required for implementation. Another challenge is that information must be displayed in real time and must coincide with the environment. Information, which is deferred or not correctly shown may distract the driver or lead to irritation—as well as a too large amount of information [13, 16].

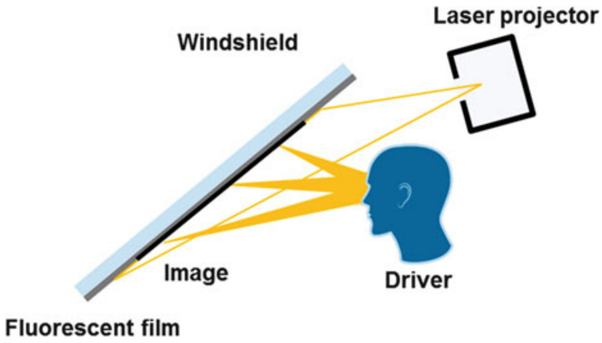


Fig. 2 Functionality of Emissive Projection Displays (EPDs, according to Sun and Liu [23] and Sun and Cheng [22]): Virtual image in the plane of the windshield

1.2 Emissive Projection Display (EPD)

The Emissive Projection Display is following a different approach. Sun and Cheng [22] describe the structure of a so-called EPD-system, which allows a projection of information directly onto the entire windshield. In this presented display technique, the windshield acts as a projection screen and a hand-sized ultraviolet laser from Sun Innovations Inc. as a projector [22]. To make the laser projection visible on the windshield, it is coated with a transparent, fluorescent film. When the laser hits this layer, it is emitted as visible light and can be perceived by humans [18]. Figure 2 schematically illustrates an Emissive Projection Display.

General Motors uses the EPD technology to project various information directly on the entire windshield. For this reason, these displays are also called full windshield Head-Up Display (fw-HUD). In difference to conventional Head-Up Displays, the information projected by laser can be seen from any position of the car, as there is an infinitely larger viewing angle [23, 24]. Fw-HUDs also offer other advantages compared to conventional Head-Up Displays. Laser projected images are for example always sharp. A further advantage is that only a small projector is required, thus much space can be saved. In addition, the projection extends across the entire windshield and is therefore much bigger than that of a conventional Head-Up Display [22, 23]. The projection screen can for example be used to highlight roadsides or signs, pedestrians and animals on the roadside in poor visibility conditions (i.e. night, fog). Moreover navigation arrows can virtually be placed directly on the road and the desired travel destination can be marked on the windshield [18, 25]. In Günthner et al. [8], an attempt was made to use this technology to support drivers of industrial trucks. However, in order to install these displays in road vehicles or industrial trucks for series production, it will still take several years.

This may be due to challenges, which still need to be solved. Disadvantages in the current implementation of Emissive Projection Displays relate in particular to laser safety. The laser projectors used with a wavelength of 405 nm are usually class 3B-laser [24]. The laser is mounted either under the car roof or in the dashboard. Direct eye contact and reflections should definitely be avoided, as a direct beam in the eye would be very harmful [23].

As the projected laser beam is fully absorbed by the fluorescent film, there is no hazard for the eyes behind the windshield [22].

1.3 Differences of HUDs and EPDs

In summary, the projections on the entire plane of the windshield using Emissive Projection Displays differ from conventional HUDs in two fundamental properties:

– **Display level:** First, using the laser projection, the information is directly presented on the windshield at a distance of about 1 m from the driver. This means that the driver's eyes must adapt to read the displayed information. In contrast, the HUD information floating above the engine hood, no or less adaption of the eyes required, which also reduces the time for reading [2, 19, 26].
– **Display size:** Secondly, information in an Emissive Projection Display is positioned in a much larger area, so that the driver has to take his eyes off the road and possibly even has to turn his head to focus and perceive it. This may result in longer fixation times and a possible hazard for the driver and other road users.

The literature review shows that no studies exist on the investigation of distraction as well as readability of information presented on a windshield by using the EPD-technology. In addition a transfer of previous findings resulting from studies investigating the driver's distraction by Head-Up Displays (HUD) is not possible for the Emissive Projection Display due to the different display levels. In order to determine whether the above mentioned properties "display level" and "display size" affect the driver's attention, basic research is required.

2 Research Questions

As part of a driving simulator study at the Institute of Ergonomics, the effects of a windshield display on the driver's attention on the road and on the driving behavior were investigated in comparison to a conventional HUD in order to get new findings. Research questions regarding the driving behavior and the usability are summarized in Knott et al. [20]. In the context of distraction the following research questions are under consideration:

2.1 Hypothesis "Response Time"

Which differences are observed regarding the driver's distraction between using the Emissive Projection Display and the HUD-technology in order to provide information for the driver?
It shall be demonstrated, there is a difference regarding the driver's distraction between the two display concepts—the simulated HUD and the EPD-technology.

To measure the distraction, a secondary task visualized on the HUD and on the EPD is used. The reaction time for performing the visual secondary task is used as a parameter for measuring the distraction. The HUD projection as well as the driving scenery in the simulation is shown on the wall screen. The alternative display concept projects the information directly on the windshield and the driver's eyes have to focus on the windshield. The disadvantage is that this affects time as well as comfort for perception of the information. So it can be determined how much time the participants need for the perception and execution of the secondary task.

2.2 Hypothesis "Workload"

Is there a connection between the types of visualization of information—Head-Up Display or Emissive Projection Display—and the test person's subjective assessments with respect to the workload in the different test runs?
A further postulate of this study is that the subjectively determined Overall Workload Index (OWI) by using the NASA-TLX questionnaire differs between the Head-Up Display and the EPD-technology. The two concepts differ with respect to the distance to the object display which affects the visual accommodation ability. But also the type of presentation is different: The simulated HUD projection is displayed as objects in the driving scenery on the wall screen in a distance of 3.65 m, whereas the laser projection glows on the fluorescent film which is affixed on the windshield in a distance of 0.7 m. It is assumed that the subjective results also show the influence on the workload.

3 Study in the Driving Simulator: Method

To answer the research questions an appropriate experimental design with repeated measurements was created by the Institute of Ergonomics (Technische Universität München, TUM) for the implementation of a dual-task method.

3.1 Experimental Setup

Figure 3 shows the experimental setup of the investigation in the static simulator of the Institute of Ergonomics (TUM) with a test participant wearing laser safety glasses. The static driving simulator-mockup consists of a driver's seat and the following components: a Plexiglas windshield coated with a fluorescent film (1, MediaGlass TM, Superimaging TM), a projection screen, a video projector, a Spectrum eLite 1.6 W laser projector (2, KVANT Laser Germany, customized) and a 7" TFT touch screen monitor for visualization of the speedometer (3). A PC, which is connected to the video projector, the steering wheel (4, Logitech G27) and appropriate pedals (5, Logitech G27), calculates the virtual driving environment with the simulation software SILAB (Version 3.0, WIVW GmbH, Germany). The

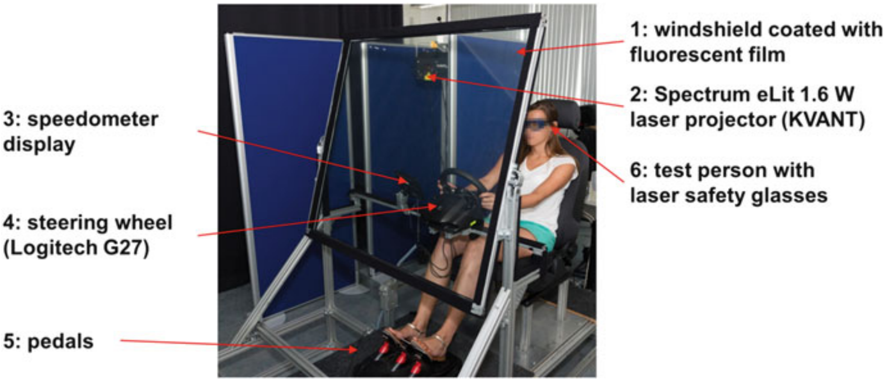


Fig. 3 Experimental setup

virtual driving environment on canvas (size 3080 × 2310 mm) is realized by a video projector, which is mounted directly above the driver's seat. The driving environment with its textures is shown on a wall screen 3.65 m away from the participant's eyes. The objects for simulation of a conventional head-up display are displayed directly in the driving simulation as a texture. The display on the windshield is implemented by the Emissive Projection Display (EPD)-technology according to Sun and Liu [23] and contains the laser projector mentioned above and a fluorescent film which is mounted on the windshield. The laser is placed laterally to the right of the driver and protected with a laser protective wall so that the subject has no direct eye contact with the laser beam. In addition, laser safety glasses belong to the experimental setup which the subject has to wear during his stay in the mockup. To simplify the EPD-display, the inclination angle of the windshield has been reduced in comparison to a conventional vehicle to 15° (to the vertical). The ultraviolet laser beam activates the fluorescent film mounted on the windshield at a wavelength of 405 nm and makes the film light up at this precise spot. The reason for perceiving the projection as a static image by the human eye is that the laser passes the given vector graphic 60 times per second. Behind the mockup is the control center. From there, the test procedure is coordinated and the data is recorded. The network with the User Data Protocol (UDP) is used for communication between the SILAB driving simulation and the KVANT laser projector.

3.2 Primary Task: Lane Change Task (LCT)

For the assessment of automotive systems, generally—as in this study—dual-task methods are used [6]. Applying the standardized driving task "Lane Change Task" (LCT) according to ISO 26022:2010 [15] and a secondary task, a statement regarding the impact of information presentation of the two display concepts to humans is possible by comparison to a reference drive (baseline run).

In this investigation, the LCT is used as the primary task to assess driving performance when different objects, which the driver has to identify, are shown either on the wall screen (HUD-concept) or directly on the windshield (EPD-concept). The LCT is a very simple and analytical test which is designed to quantitatively record the worsening of the driving performance by performing a secondary task while driving. It is designed for simulated driving environments, so it is not usable for a real road experiment. During the LCT a test participant has the task to drive along at a constant, system controlled speed of 60 km/h. The driving environment simulates a straight 3-lane highway and about a 3000 m long section of this road. Participants have the task to change the lane fast and precisely at standardized intervals. Signs on both sides of the road show the instruction which lane has to be taken and appear 40 m before passing them. A total of 18 lane change operations are carried out, because there are six different ways in which the lane change can occur [15].

3.3 Secondary Task: Landolt Rings

To keep the comparison of the two display concepts as conceptual as possible and not to restrict it to specific graphics, the following secondary task was used. In addition to the primary task (LCT) there is a visual task to detect the direction of the gap of a Landolt ring. The Landolt rings are shown in the windshield by using the laser (EPD-technology) as well as a texture embedded into the driving simulation scenery (HUD). Figure 4 shows examples of the visualization.

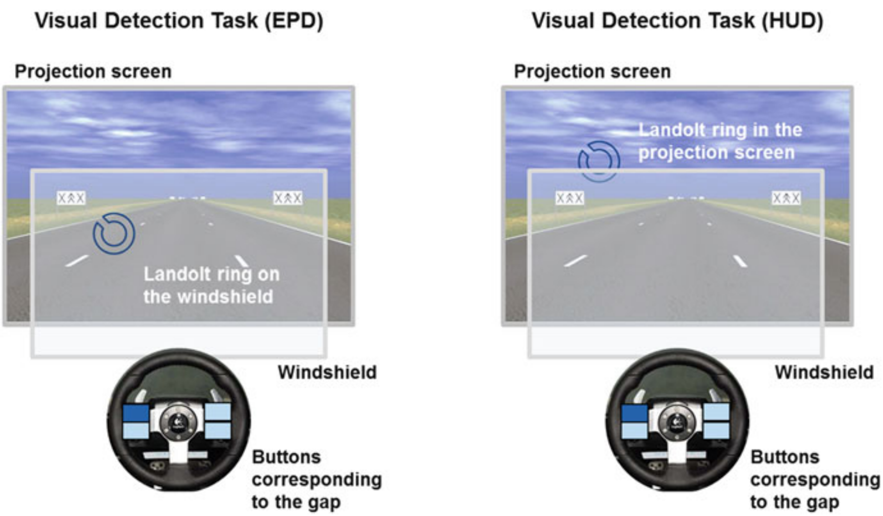


Fig. 4 Examples of the secondary task with presentation of the Landolt ring in the EPD (*left*) and in the HUD (*right*)

Fig. 5 Steering wheel with the button corresponding to the orientation of the gap of the Landolt ring

A Landolt ring is named according to the Swiss ophthalmologist Edmund Landolt and is used as a ring with a gap normally for testing the visual acuity. It has the advantage that cognitive recognition and similarity, play only a minor role compared to letters. The position of the gap varies and exact dimensions of the rings are defined in the standard DIN EN ISO 8596:2009 [4]. In the case of the study at the Institute of Ergonomics, four orientations of the gaps are used. As can be seen in Fig. 5, in each case, the gaps are rotated 45° to the vertical and horizontal. In contrast to the standard DIN EN ISO 8596:2009 [4] and due to the fact that the reflecting light of the fluorescent film is blue (hex color code: #0000CD), the Landolt rings are also displayed in blue color as well. According to the DIN EN ISO 15008:2009 [5] standard, the size of the Landolt rings is designed depending on viewing distance. However, their size has also been adapted to the capabilities of the laser. For the Emissive Projection Display a size of 0.04 m was chosen; according to this, also the size of the rings in the Head-Up Display was enlarged (HUD size: 0.14 m). The rings with the four possible gap directions are visualized in Fig. 5.

Both the gap direction and the interval between the inserted rings vary. In total, 18 Landolt rings are displayed in both the EPD and the HUD in various designs and at different positions. The position of the display area of the simulated HUD was not fixed—as in reality. The Landolt rings are distributed over the entire surface of the windshield as well as the driving scenery.

As already described in Knott et al. [20], the following procedure for the design of the reaction time for the ring detection was used: As soon as a Landolt ring is shown, the test person has to detect the correct position of the gap and push the corresponding button on the steering wheel as quickly as possible. Four buttons on the steering wheel, which are also shown in Fig. 5, are used for the detection. The time elapsed between the first appearance of the Landolt ring and the detection is recorded. This aspect makes it possible to measure the reaction time objectively which was needed to perform the visual task.

According to the "Human Processor"-Model [3] the necessary duration is defined for a valid hit. In the following the steps performing the secondary task are outlined. The relevant processors according to Card et al. [3] are parenthesized.

– Step 1: Focusing onto the randomly appearing Landolt ring (perceptual processor)
– Step 2: Recognizing the opening of the Landolt ring (perceptual processor)
– Step 3: Determination of the corresponding button (cognitive processor)
– Step 4: Pressing the button on the steering wheel (motor processor)

In the following, this information can be used to calculate the range in reaction time t according to Card et al. [3]:

$$t = 2\,\tau_P + \tau_C + \tau_M \qquad (1)$$

t: reaction time [ms]
τ_P: time frame for one perceptual process cycle [ms]
τ_C: time frame for one cognitive process cycle [ms]
τ_M: time frame for one motor processor cycle [ms]

In accordance with signal detection theory a hit is valid when the correct button is pushed between 300 ms and 3000 ms after first appearance of the Landolt ring. When no button is pushed the Landolt ring disappears after 3000 ms.

There are four possible occurrences for pushing the button which are explained below: Hit, Miss, Fail and False Alarm.

Hit: The test person has detected the correct opening of the ring and pushed the corresponding button on the steering wheel between 300 ms and 3000 ms after the ring's first appearance.

Miss: No button was pushed.

Fail: The wrong button was pushed between 300 ms and 3000 ms after the first appearance of the ring.

False Alarm: Either the test person pushed earlier than 300 ms (deception) or later than 3000 ms.

3.4 Experimental Procedure

For the implementation of the experiments in the static driving simulator of the Institute of Ergonomics, an experimental protocol with a checklist, explanations and instructions was created in advance. The protocol provides a standardized

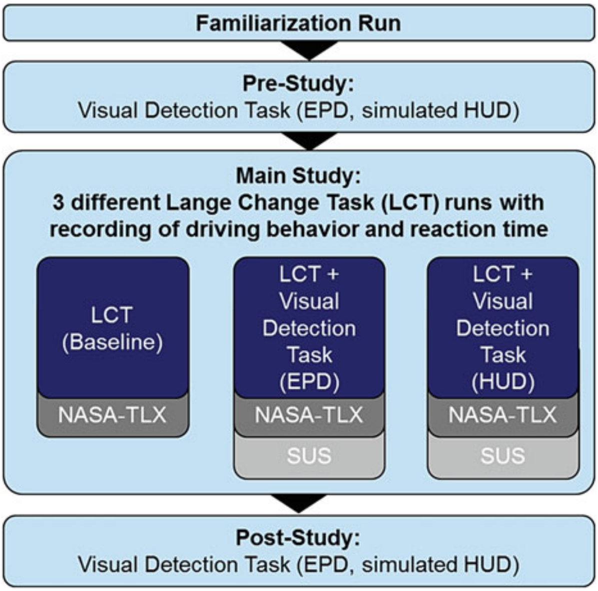


Fig. 6 Test procedure

investigation and equal conditions for each participant. The experiment consists of the following steps. In addition, the Fig. 6 illustrates the process graphically.

At the beginning of each experiment, each subject is welcomed at the Institute and informed about the study. Subsequently, the personal data of the test participant is recorded by the use of a demographic questionnaire and a consent to participate voluntarily in the study is signed. In this case, this agreement includes also taking note of the safety information for this test. Since a laser of the class 3R is used for the projection onto the windshield, risks and appropriate safety precautions were explained. A presentation with respect to laser safety and laser safety glasses, which have to be worn during the test, belong to the safety measures. Also a safety fence for the laser is available. To guarantee the anonymity of the trial participants, the consent to participate in the study with the names and signatures of participants is stored separately and does not remain with the other questionnaires.

After this, a familiarization run follows, in which the secondary tasks are displayed on the EPD in order to get a feeling for the test situation. In the following part, the so-called preliminary study begins. The subject only processes the visual secondary task without the driving task. The objects are simulated in the HUD (HUD pre) and the windshield (EPD pre) is performed. The display order (whether first HUD or EPD) must be selected for each participant at random. While the visual task is carried out in the preliminary study, a freeze image of the driving simulation is displayed so that the same conditions to the subsequent test runs prevail. The recorded and later edited data (hit rate, reaction time) serve as comparison to the following test runs later. Subsequently, the three test runs are again performed in randomized order to eliminate learning curve effects. In the first baseline trip, the subject drives the LCT-route in the simulator without visual task (BL(LCT)). The dynamic driving data (x and y coordinates of the vehicle, lane, etc.) are recorded. This is followed by two baseline runs with the added visual task, i.e. the display of Landolt rings with the corresponding detection of the gap direction. The Landolt rings are displayed both on a run on the windshield (BL(LCT) + EPD) and in the simulated HUD (BL(LCT) + HUD). Thus in the following, it can be compared how the different accommodation levels (screen distance vs. windshield distance) affect the driving performance and the distraction. The dynamic driving data of the LCT and data relating to the key operation (hit response time) of the visual task are recorded.

After each run of the main study, each test persons answers the NASA TLX (NASA Task Load Index) questionnaire to assess the subjective perceived workload in the individual test runs. In this study a German translation was used. In addition to the NASA TLX, the System Usability Score (SUS) is also used after each run with the completion of the secondary task during the main study. The SUS is a simple tool with ten items—each on a 5-point Likert-scale—to measure the usability of the two different display concepts. Also a German translation was used for the questionnaire in combination with the original text in English.

Last the post study (HUD post, EPD post) is executed also in randomized order in the same way as the preliminary study.

4 Data Collection and Statistical Data Analysis

During the experiment in the driving simulator, different data measurement for subsequent evaluation is recorded. In the following we focus on this data, which are relevant for answering the research questions (see Sect. 2).

4.1 Objective Data

For the secondary task, the reaction time between the first appearance of the Landolt ring and the button actuation has to be determined. Furthermore it is necessary to know which button was pressed. Finally there is a time interval set between 300 ms and 3000 ms after the first appearance of the sign for a valid hit. It was calculated by human reaction time (see Sect. 3.3). For the determination of the reaction time and the number of correct hits, the time stamp of the buttons pressed, as well as which one of the buttons is recorded. Furthermore the time of the appearance of the ring and its opening direction is also recorded for subsequent analysis.

The objective measurement data for driving behavior is based on the recorded data of the simulation software SILAB (WIVW GmbH, Veitshöchheim, Germany). During the experiment SILAB records the general simulation data (simulation time, time of displaying the Landolt rings), dynamic driving variables (speed, vehicle distance to the right edge of the roadway) as well as the steering wheel operation (keystrokes) as a plain text file in ASCII characters. By using the recorded time from the beginning of displaying the Landolt ring as well as the time when pressing the button, conclusions regarding the reaction times of the subjects can be drawn. The plain text file can be opened with an ordinary text editor (Crimson Editor, Notepad) and assessing programs such as Microsoft Excel or MATLAB MathWorks.

4.2 Subjective Data

The NASA Task Load Index (NASA TLX) questionnaire, which was developed by the Human Performance Group at the NASA Ames Research Center is used to get information about the workload while driving. This questionnaire is a standardized and multi-dimensional method in order to obtain the workload under different driving conditions [10, 11]. The six dimensions for the subjective assessment of stress are: Mental demand, physical demand, temporal demand, effort, performance and frustration level [11]. The test participants rate the amount of their workload on a rating scale from 0 to 100 for each of the six dimensions. Low values correspond to low stress. In order to get no distortions, the evaluation takes place immediately and without delay after the test drive. For the evaluation, it should be noted that markings, which are set in the middle of two scale marks are allocated to the higher value. It is also possible to give each dimension additional weight that produce a higher workload. In this study, all dimensions are equally weighted. To compare the subjective assessment of workload of the individual test runs, the Overall Workload Index (OWI) is used. The OWI is the result of each test drive in the unweighted version of the total of the individual ratings, divided by the number of dimensions (6) [10].

4.3 Statistics: Paired *t*-Test

Based on Bortz and Schuster [1] a paired *t*-test is used to test the defined hypotheses. The significance level, also called error probability is set to 5 % for hypotheses testing. The paired *t*-test assumes that two samples are collected dependent on each other [1]. Since in this study each subject was tested repeatedly and these measurements are compared with each other, the paired *t*-test of IBM SPSS Statistics software (version 22) was used.

5 Results

The study aims to determine the effects of different display concepts in the vehicle, regarding both the driver’s attention and the subjective assessment of workload. While for the objective data, the average response time for the processing of the visual secondary task and the driving behavior are used, the subjectively perceived workload and the usability are used for the subjective statements. For results

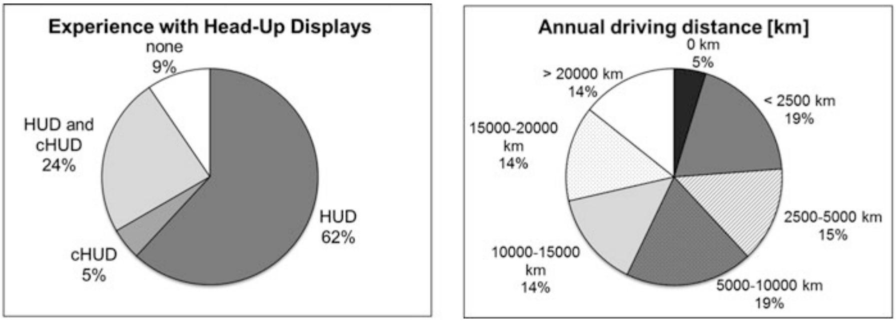


Fig. 7 Characteristics of the test sample: Experience with Head-Up Displays (*left*); Annual driving distance (*right*)

concerning the driving performance and the system usability, it is referred to Knott et al. [20].

5.1 Test Sample

The sample of this study involves 21 test participants. All of the subjects are taken from the engineering field and with a female proportion of 24 %, five women participated in the study. They together with the male participants provided the data for the analysis. With an age range of 24 years for the youngest participants and 70 years for the oldest subjects, the arithmetic average age of the subjects is 31.95 years, with a standard deviation of $SD = \pm 10.85$. It is obvious, there is also a concentration of participants with 11 people in the age group "26–30". Furthermore, 57 % of the participants reported the use of a visual aid in the form of eye-glasses or contact lenses. No participant of the study had an important vision impediment for the test procedure, but one subject had a color deficiency in form of a red-green color blindness. 95 % of the participants were familiar with driving in a driving simulator. More than 90 % of all participants had experience with a conventional Head-Up Display (HUD), a contact-analogue Head-Up Display (cHUD) or even gained experience with both display concepts. The annual driving distance among the subjects is balanced (see Fig. 7).

5.2 Hit Rate and Response Time

As explained in Sect. 3.3, the processing of the secondary task—time for selecting and pressing the appropriate button—was evaluated by using the signal detection theory (hit, miss, fail, false alarm). For reasons of comparability, only the hit reaction times of the secondary task were used for the derivative of distraction. The comparison of the hit rates in the experiment shows that the secondary task was processed almost perfectly in the test drives as well as in the pre- and post-study. Figure 8 shows the hit rates in the six different parts of the experiment, the

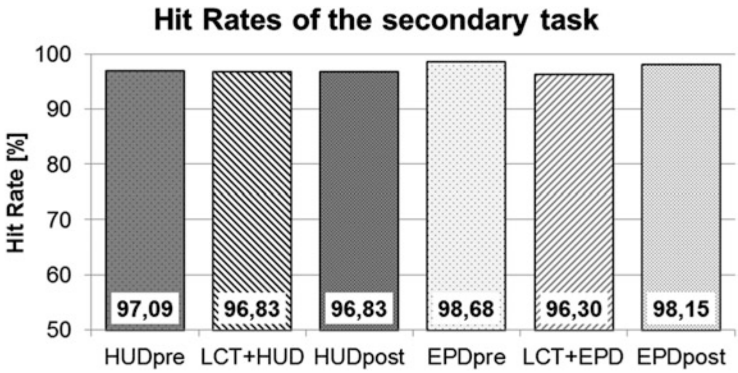


Fig. 8 Hit rates in percent in the different test drives with processing the secondary task

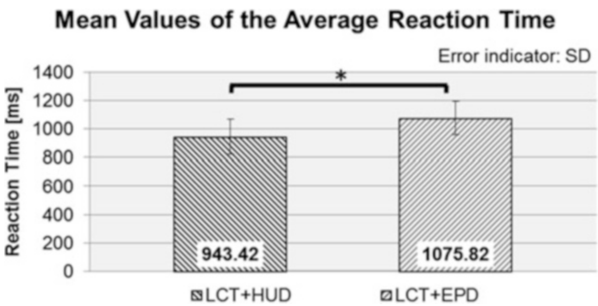


Fig. 9 Mean values of the reaction time [ms] of the test runs with visual task [20]

maximum hit rate of 100 % corresponds to 18 hits. However, it is also apparent that the processing of the secondary task is slightly worse during the use of the assistance systems in a test run, compared to the pre- and post-study.

Regarding the objectively recorded response time, the test participants needed an average reaction time of $M_{RT,\ EPD} = 1075.82$ ms (SD $= \pm 117.55$ ms) for processing of the secondary task in the Emissive Projection Display. For the condition of presenting the information in the simulated HUD, the average response time of the test persons was $M_{RT,\ HUD} = 943.42$ ms (SD $= \pm 123.16$ ms). For testing the significance with respect to the hit reaction time between the two experimental conditions—HUD and EPD—the procedure described in Sect. 4.3 was used at a significance level of $\alpha = 5$ %. The *t*-test (two sided, paired) shows there is a significant difference in the response times of the two different display concepts (LCT + HUD and LCT + EPD: t [20] $= -5.49$, $p = 0.000 < 0.05$). In summary, the reaction time for processing the Landolt ring-task on the windshield is significantly higher compared to the other condition using the HUD technology (see Fig. 9). This indicates an increased response time for the test persons in the test condition with information on the Emissive Projection Display and should thus be avoided.

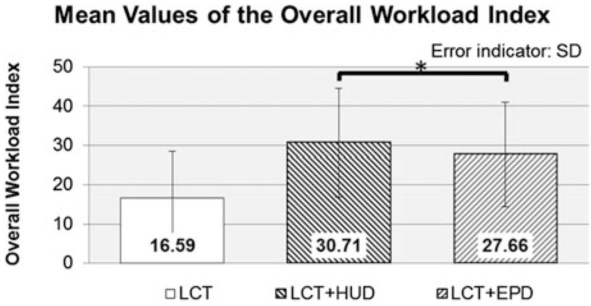


Fig. 10 Subjective ratings of the mental workload in the individual conditions [20]

5.3 Workload

With respect to the subjectively perceived workload, the Overall Workoad Index (OWI) as a measurement parameter has an average value of $M_{OWI,\ EPD} = 27.66$ ($SD = \pm 13.55$) while presenting the information in the EPD. In the simulation run (HUD) the average OWI is $M_{OWI,\ HUD} = 30.71$ ($SD = \pm 14.18$). Statistics show the value for the Overall Workload Index and consequently the subjectively assessed workload during the test run with information presentation on the simulated HUD as being significantly higher compared to the condition presenting the information using the EPD-technology (LCT + HUD and LCT + EPD: $t\ [20] = 2.10$, $p = 0.049 < 0.05$). Figure 10 also shows the comparison to the test run LCT without processing a secondary task.

6 Discussion

The fact, that the subjectively perceived workload recorded by the OWI for the simulated HUD is rated higher than for the EPD may be due to the aspect that the subjects had to concentrate more on the wall screen display. The Landolt rings were displayed in the HUD in a larger area (full screen) on one hand and on the other, the subject had to observe the entire screen carefully. The Landolt rings in the EPD were also distributed across the entire windshield, which was comparable to conventional vehicle windshields. Another possible explanation for the lower OWI-rating for the EPD is the brightness of the Landolt rings on the windshield compared to the simulated HUD, which is caused by the realization possibilities of the EPD.

At the same time, also challenges and problems arise with the implementation of an EPD. The current possibilities to implement an EPD—as in this study—need the use of a laser, which is a hazard for the eyes. The installation of this technology in vehicles is irresponsible for safety reasons. As part of this experiment, the subjects were instructed in laser safety and they confirmed this in writing. In addition, laser safety glasses were used and reflecting objects on the hands and arms of the subjects had to be removed. This will not be reviewable when using the system in the real world. Consequently and first of all, further development in the field of the displays and their realization is required.

Apart from that, the objective results provide a statement to the contrary. In relation to the significantly longer results in reaction times for detection of the gap of the Landolt ring on the Emissive Projection Display in comparison to the simulated display on the wall screen, it can be concluded, the information presentation on the EPD requires more time for recording and editing than a presentation of information in the HUD. The reason for this is the accommodative capacity of the eye. While the driving scenery and the secondary task are at the same accommodation level in the simulated HUD, the driver must first focus on the windshield when using the EPD technology. For this, dynamic adjustment time is required. As part of the doctoral thesis of Schneid [21], the influence of image distance at the time of information perception was analyzed. The results show also an extension of time for reading information by a shortening of the image distance from the eye [21]. In addition, the fact can also be added, that reading time increases with increasing age. With a concentration of participants of 11 people in the age group "26–30", the average age of the subjects collective in this study is relatively low and the ability to change between different focus levels decreases with age [9]. This is also consistent with the statement of Bubb et al. [2], that a HUD has the advantage in the required accommodative capacity of the eye and is significantly lower or even eliminated for older persons.

Acknowledgements The authors would like to thank Mr. Ilja Feldstein (Institute of Ergonomics, TUM) for his support during the experimental setup and Mr. Matthias vom Stein and Mr. Michael Förster (both Institute for Materials Handling, Material Flow, Logistics, TUM) for the implementation of the laser display. We also thank Springer Science + Business Media.

References

1. Bortz J, Schuster C (2010) Statistik für Human- und Sozialwissenschaftler (7., vollst. überarb. u. erw. Aufl). Springer, Berlin
2. Bubb H, Bengler K, Grünen RE, Vollrath M (2015) Automobilergonomie (ATZ/MTZ-Fachbuch, Aufl. 2015). Springer Fachmedien Wiesbaden GmbH, Wiesbaden. 2015
3. Card SK, Moran TP, Newell A (1983) The psychology of human-computer interaction. L. Erlbaum Associates, Hillsdale, NJ
4. DIN EN ISO, 8596 (2009). Augenoptik - Sehschärfeprüfung - Das Normsehzeichen und seine Darbietung (ISO 8596:2009); Deutsche Fassung EN ISO 8596:2009. Beuth Verlag GmbH, Berlin
5. DIN EN ISO, 15008 (2009). Straßenfahrzeuge - Ergonomische Aspekte von Fahrerinformations- und Assistenzsystemen - Anforderungen und Bewertungsmethoden der visuellen Informationsdarstellung im Fahrzeug (ISO 15008:2009); Deutsche Fassung EN ISO 15008:2009. Beuth Verlag GmbH, Berlin
6. Ecker R (2013) Der verteilte Fahrerinteraktionsraum. Doctoral Thesis, Dissertation, Ludwig-Maximilians-Universität München, München. Retrieved from: http://edoc.ub.uni-muenchen.de/15760/1/Ecker_Ronald.pdf
7. Fuchs A (Display Magazin., Hrsg) (2013) Head-up Display im Straßenverkehr. Höhere Verkehrssicherheit durch neuartige Technologie. Verfügbar unter http://display-magazin.net/thema/displays/head-up-display-im-strassenverkehr. Checked on 19.04.2015
8. Günthner WA, Bengler K, Vom Stein M, Knott V (2015) Einsatz der Aufmented-Reality-Technology zur Unterstützung des Fahrers von Flurförderzeugen. Forschungsbericht, Technische Universität München, München

9. Hahn G-A (2012) Kurzlehrbuch Augenheilkunde. Thieme, Stuttgart
10. Hard SG (NASA Ames Research Center, Ed.) (1986) NASA Task Load Index (TLX): Volume 1.0 Paper and Pencile Package. Retrieved from: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000021488.pdf
11. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (Eds.) Human mental workload. Elsevier, Amsterdam
12. Imhof T (Die Welt., Hrsg) (2012) Mit dem Tiefflug-Display über die deutsche Autobahn. Retrieved from: http://www.welt.de/motor/news/article108347262/Mit-dem-Tiefflug-Display-ueber-die-deutsche-Autobahn.html. Checked on 19.04.2015
13. Israel B (2012) Potenziale eines kontaktanalogen Head-up Displays für den Serieneinsatz. Doctoral Thesis, Dissertation, Technische Universität München, München. Retrieved from: http://mediatum.ub.tum.de/doc/1172451/1172451.pdf
14. Israel B, Seitz M, Senner B, Bubb H (2010) Kontaktanaloge Anzeigen für ACC - im Zielkonflikt zwischen Simulation und Ablenkung. 4. Tagung Sicherheit durch Fahrerassistenz, 15–16 April 2010, München, TÜV SÜD, 1–7
15. ISO, 26022 (2010) Road vehicles. Ergonomic aspects of transport information and control systems. Simulated lane change test to assess in-vehicle secondary task demand: Beuth Verlag GmbH
16. Jansen A, Israel B, Spiessl W (2012) Augmented Reality Navigation in zukünftigen Head-Up Displays. Prototypenaufbau und erste Bewertung mittles Probandenstudie. In Optische Technologien in der Fahrzeugtechnik. 5. VDI-Tagung, Karlsruhe, 24 und 25 April 2012 (VDI-Berichte, Bd. 2154, nichtredigierter Ms.-Druck, S. 205–216). VDI-Verlag, Düsseldorf
17. Kassner A, Vollrath M (2006) Akzeptanzmessung als Baustein für die Entwicklung von Fahrerassistenzsystemen. In Integrierte Sicherheit und Fahrerassistenzsysteme. Tagung 12./13. Oktober 2006, Wolfsburg (VDI-Berichte, Bd. 1960, S. 97–112). VDI-Verlag, Düsseldorf
18. Kedor T (2010) GM macht die Windschutzscheibe zum Head-Up-Display. Retrieved from: http://www.motor-talk.de/news/gm-macht-die-windschutzscheibe-zum-head-up-display-t2622751.html. Checked on 19.04.2015
19. Knoll P (2012) Anzeigen für Fahrerassistenzsysteme. In: Winner H (Ed.), Handbuch Fahrerassistenzsysteme. Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort; mit 45 Tabellen (2., korr. Aufl, S. 330–342). Vieweg + Teubner, Wiesbaden
20. Knott V, Demmelmair S, Bengler K (2015) Distraction and Driving Behavior by Presenting Information on an "Emissive Projection Display" Compared to a Head-Up Display. In Harris Don (Ed.) Engineering Psychology and Cognitive Ergonomics. 12th International Conference, EPCE 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2–7, 2015, Proceedings. ISBN: 978-3-319-20372-0. Springer International Publishing
21. Schneid M (2009) Entwicklung und Erprobung eines kontaktanalogen Head-up-Displays im Fahrzeug. Doctoral Thesis, Dissertation, Technische Universität München, München. Retrieved from: http://mediatum.ub.tum.de/doc/652901/652901.pdf
22. Sun T, Cheng B (2013) Novel emissive projection display digitizes glass windows. Inf Displ Mag 29(6):18–24
23. Sun T, Liu J (2007) Turning the Windshield into a Display: a projection system with fluorescent conversion technology allows the size and angle of an automotive display to be almost limitless. Retrieved from: http://www.photonics.com/Article.aspx?AID=28347. Checked on 15.04.2015
24. Sun Innovations (2011). Full-Windshield Head Up Display Developmental Kit. Retrieved from: http://www.sun-innovations.com/images/brochure/HUD.pdf. Checked on 15.04.2015
25. Vasilash GS (2010) GM Purposes Augmented Reality. (Automotive Design & Production, Hrsg). Retrieved from: http://www.autofieldguide.com/articles/gm-pursues-augmented-reality. Checked on 19.04.2015
26. Winner H (Hrsg) (2012) Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort; mit 45 Tabellen (2., korr. Aufl). Vieweg + Teubner, Wiesbaden

User-Centered Design of Font Size and Polarity of Assistance Systems in Production Planning and Control

Jochen Nelles, Sinem Kuz, and Christopher M. Schlick

Abstract

Due to a lack of responsiveness of unplanned incidents and deviations between planning and reality, manual overplanning in production planning systems should be reduced with the help of high resolving and consolidated data. In this context, one promising approach is a specific decision-making support of the production supervisor as a decision-maker by using action recommendations of an assistance system. With vertical human-machine interaction, the specific capabilities according to the Men-Are-Better-At/Machines-Are-Better-At (MABA/MABA) principle can be applied. In accordance with intelligent information visualization of the assistance system and a tablet-application for usage in a production hall environment, the research presented in this paper focuses on the investigation of the polarity and the font size respectively angular character height of the assistance system application regarding to task performance and mental effort. In an eye-tracking study, 15 participants were presented a search task on a touchscreen either with positive or negative polarity. In addition in each trial of the experiment the angular character height (16, 20, 24, 28 arcmin) was varied randomly. The results show that the mean search time can be improved significantly with an angular character height of 24 arcmin. Moreover, the results indicate that positive polarity leads to smaller average pupil diameter. Which is associated with a lower mental workload.

J. Nelles (✉) • S. Kuz • C.M. Schlick
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany
e-mail: j.nelles@iaw.rwth-aachen.de; c.schlick@iaw.rwth-aachen.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_18

Keywords
Human-machine interaction • Self-optimizing detailed planning system • Production planning and control • Industrie 4.0 • Automation • Assistance system • Font size • Angular character height • Polarity • Mental workload • Pupil diameter

1 Introduction

1.1 Cyber-Physical Production Systems

Technical improvement in the field of microelectronics and information and communication technology has led to the development of so-called (socio-)cyber-physical systems (CPS). In cyber-physical systems, environmental information are acquired via sensor technology and acted upon by actuators. Thus the physical and digital world are directly connected by sensors and actuators. Therefore cyber-physical systems are autonomously able to solve problems. A comprehensive implementation of cyber-physical systems in production as well as the development and use of a system based information management leads to so-called cyber-physical production systems (CPPS) and is referred to as the fourth industrial revolution (Industry 4.0). In this way, a much improved use of existing information in business environment is enabled and a new level of collaboration productivity is achieved [3, 17, 26].

Nowadays, production systems are characterized by complex material flows and a high variance of the manufacturing and assembly processes. Collectively, logistical performance indicators like leadtime, timeliness or time of delivery gain importance because they are perceived as differentiating characteristics [25].

In spite of latest software developments, manufacturing companies still suffer from a lack of responsiveness to unforeseen events like urgent customer orders, machine failures or absence due to illness. The reasons for this deficiency are a strong simplification of complex situations and limited data quality [2]. Due to this lack of responsiveness, as well as restricted up-to-dateness and maintenance of data, the production supervisor has to rely on manual overplanning. This increases the discrepancy between planning and reality. In addition, complex IT structures and deficient transparency of the control logic affect the understanding of the production supervisors. The combination of manual overplanning by the operator and non-transparent algorithms of the production planning system (e.g., Enterprise Resource Planning (ERP), Manufactoring Execution Systems (MES), Advanced Planning and Scheduling Systems (APS)), which calculate order sequences for production resources on a daily basis, leads to considerable difficulties in understanding the order sequence and process. Additionally there is no sufficient overview of the interaction between timeliness, capacity, circle time and inventory.

On the basis of all these restrictions, it is necessary to develop innovative assistance systems in order to improve comprehension of production planning and control. Currently, the cyber-physical production systems are however, technologically still in development. The implementation and wide distribution of these technologies are still low both in industrial production and in intralogistics. Thus, propositions about technology and work organization and related skills development paths cannot sufficiently clearly be met. Two polar opposite scenarios can be considered more closely. In the automation scenario, the work of the production supervisor is determined by the cyber-physical system and the autonomy of the professionals is restricted. Due to lack of knowledge and experience the understanding of the production supervisors needed in case of failure cannot be built. In the specialization scenario the production supervisors are in control of the cyber-physical system and represent the domain of industry 4.0. In this case the expertise of the production supervisors is everyday needed and in case of failure available. The necessary skills require the appropriate training and support of the production supervisors [12].

It should be noted that the complex production planning and control system cannot be controlled solely by the working person. Therefore, the production supervisor is dependent on the reliability of the planning generated by the IT systems. The methodology of the research project is based on a cyber-physical detailed planning system with an evaluation of the detailed planning results [15].

1.2 The ProSense Assistance System

One approach to improve the planning quality and transparency in production and to provide the production supervisor is the assistance of the production supervisors with the help of a decision support system. This allows—according to the Men-Are-Better-At/Machines-Are-Better-At principle—an optimal combination of the specific abilities and characteristics of humans and machines. In production planning and control, the production supervisor uses typically a variety of IT systems like ERP-, MES- and APS-systems as well as programs like MS Excel. This common kind of production control is characterized, according to the model of the four steps of automation by Parasuraman et al. [21], by manual processing of the following components: (1) information reception, (2) information analysis, (3) decision-making/operation choice, and (4) decision execution.

In this research project, a specific IT structure, a data base, and high resolving sensors are developed and implemented in two consortium partner factories as well as in a demonstration factory at the RWTH Aachen University. In this case, the self-optimizing detailed planning system supports the production supervisor on the basis of 3 of 4 steps of automation: information reception, information analysis and based on computational simulation, recommendations in terms of decision-making. In this vertical human-machine cooperation the production supervisor himself has the possibility to select the action and is responsible for operation execution (Fig. 1).

1.3 User-Centered Design of Font Size and Polarity

This paper focuses on the user-centered design of the software application for assistance and decision support of the production supervisor in production control. The research object of this study is the investigation of angular character height and

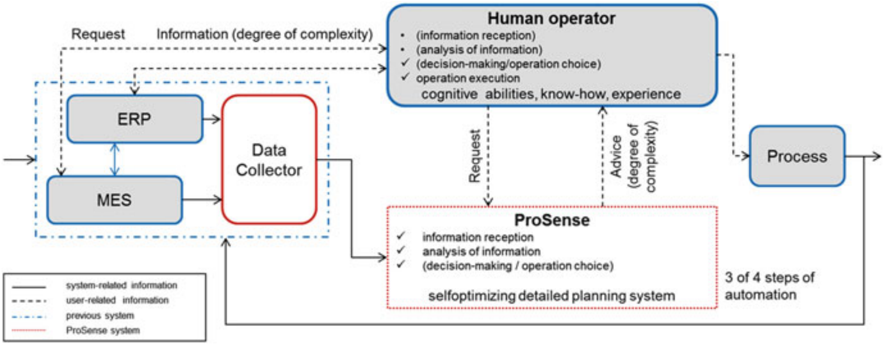


Fig. 1 Vertical human-machine cooperation and decision-making support by the assistance system (figure based on Millot et al. [21])

polarity effects of the software application in the context of an assistance system in production planning and control. In this case, there is the use of a touchscreen respectively tablet on the one hand, and the lighting environmental condition and ceiling height of a manufacturing hall on the other hand. It is important to enhance usability in this special context, because a usable and task appropriate software application can save time and increase transparency, acceptance and trust. Therefore, information visualization plays a key role in this study. In particular, the influence of polarity and angular character height on the task performance and on mental effort is in the focus of our attention as the decision-making support and assistance system is determined to be used on touchscreens in manufacturing companies.

The first aspect considered is polarity. Polarity results from negative and positive contrast. When dark letters are presented on a light background, the contrast is positive. Hence, polarity is positive. Light letters on dark background therefore refer to a negative contrast and polarity [4]. So far, studies have shown contradicting results regarding the effects of polarity. While some findings replicate earlier findings [4] that positive polarity leads to higher legibility [14, 32], other more recent findings do not show any advantage of positive or negative polarity [4]. Cushman [8] demonstrates that there was no polarity effect on reading time and comprehension, but visual fatigue was greater in negative polarity. A possible explanation for the preference of positive polarity might be its familiarity because traditional information media are presented in positive polarity (e.g., newspaper, books). Buchner and Baumgartner [4] explain that the advantage of positive polarity is caused by higher display luminance than in negative polarity. Display polarity per se does not influence legibility. In spite of the preference for positive polarity, the current draft of the decision-making support system has, similar to the design of a control desk, a design of negative polarity.

The second aspect investigated in this study is the optimal font size for the use on touchscreens. Touchscreens do not only gain importance in our everyday life, but also in business and production applications. In the context of industry 4.0, touchscreens can be used as a flexible and innovative tool [28]. Studies indicate that font sizes of 12–14 pt are optimal on an E-Book Reader (9, 7″) [19], though elder people prefer larger font sizes than younger people [9]. A study by Parhi et al. [22] shows that the optimal font size is depending on the task (serial or discrete target object) and on the style of use (one-handed or two-handed). The question of the optimal font size on touchscreens is essential within this research project as the decision-making support application is meant to be used on a tablet in the production hall in order to ensure high flexibility for the production supervisor. Possible use cases are production planning on the shop floor, e.g. in case of machines malfunction, team meetings in small manufacturing groups and production supervisor training.

In addition to the influence of font size and polarity on task performance, we can consider the effect on mental workload. One aim is the improvement of usability, as the assistance system combines a variety of information. To measure the subjective mental effort a self-report index for workload evaluation was used. It could be selected between various measures: NASA Task Load Index (TLX), Modified Cooper Harper (MCH) Scale and Subjective Workload Assessment Technique (SWAT). Both SWAT and the TLX show a high sensitivity to different demand manipulations in the flight environment and in other systems [31]. All three represent globally sensitive measures of operator workload. Again both SWAT and TLX are multidimensional and can provide some diagnostic information on the sources of workload represented by the subscales of the respected procedures. Since the SWAT has only got 3 subscales, while the TLX provides 6 of them, this one was chosen for the mental workload evaluation. Furthermore, in a system test comparison, Byers et al. [6] reported higher user acceptance ratings for the TLX procedure relative to the SWAT technique. Beside the NASA TLX measuring the subjective mental effort, there was also used a physiological measurement to capture the mental workload during the performance on the touchscreen. Inter alia this kind of measurement was selected following the advice of de Waard and Lewis-Evans [10] that you cannot measure mental workload through self-reports alone. They say multiple measures should be used and you cannot rely on the task difficulty, since it is very subjective. There are a number of different physiological measurements. Wang et al. [30] recommend three for the evaluation of mental workload: facial skin temperature, eye blinks and pupil dilation. The facial skin temperature was excluded, as we did not have the instrumentation to measure it and since researchers reported that the sensitivity may not be strong enough for practical application in complex tasks. Thomas et al. [29] write that iris recognition is a good indicator of mental workload. Following them it is the most accurate biometric recognition technology, with report results of false match rates of 1 in 200 billion. And Beatty [1] reported that the pupil diameter is a physiological measurement of the mental workload. The higher pupil diameter is associated with a higher mental workload. A number of studies [5, 27, 16] have replicated this effect. Cegarra and Chevalier [7] say that the pupil dilation technique is considered to be very sensitive in many tasks and therefore can be seen as one of the most relevant physiological measures. As it can be measured with low effort, the pupil dilation technique was used in order to investigate the mental workload.

For this study we created a search task in which participants had to find a discrete target object, in order to represent one of the common tasks in manufacturing companies. As the human decision-making support system consists inter alia of an overview of all machines in a manufacturing company, searching for target objects on a screen is one of a primarily tasks in the context of the research project. The information visualization should be adjusted to that.

2 Methods

2.1 Participants

Fifteen participants (6 females, MAge = 25.93, SD = 4.23, range = 16) took part in the study. Participants had normal or correct vision, and color vision. All participants were German native speakers and were highly educated. Fourteen of them hold the German general qualification for university entrance.

2.2 Apparatus and Materials

Visual acuity and color vision were measured using the Landolt C test and the Ishihara color test provided by Optovist Software Rodabase by Vistec AG Vision Technologies. The screen used in the experiment was the ELO Touch-Monitor ET1715L-8CWA-1-G by ELO Touch Solutions Inc. The touchscreen had an active matrix TFT LCD with LED backlighting display, a resolution of 1280 $\times$ 1024 pixel, a contrast ratio 1000:1 and dimensions of 338 $\times$ 270 mm with a dark grey frame. We used SMI-Eye-tracking glasses in order to record eye movement and pupil diameter and a presentation for calibration. Data was analyzed using SMI BeGaze 3.5 software. During the experiment, participants were fixed with a chin rest 600 mm in front of the touchscreen. The experimental room did not have any windows and was enlightened by a fluorescent lamp. The lamp was turned on 15 min before the experiment started in order to ensure a constant illumination. Illumination was measured with Multi Test Master Laserliner by Umarex GmbH & Co KG. A value of 905 lux was kept constant throughout the whole experiment.

We used the NASA Task Load Index for workload evaluation. The NASA TLX is a multidimensional assessment tool which rates the perceived workload. Participants had to judge the mental effort in six dimensions (mental demand, physical demand, temporal demand, performance, effort, and frustration). In a second part participants had to weigh each dimension. The NASA TLX was developed by the Human Performance Group at NASA's Ames Research Center [13]. Beside the NASA TLX, we used the average pupil diameter as a physiological indicator for mental workload.

The INCOBI-R questionnaire is a revised version of the paper-pencil based instrument INCOBI by Richter et al. [24]. The tool measures computer anxiety, computer knowledge (procedural and declarative) and computer related attitudes. The measurement of the latter is based on 8 attitudes scales. In our study, we used two subscales of the INCOBI-R: the PRACOWI and TECOWI, in order to measure procedural and declarative computer knowledge.

2.3 Procedure

To investigate the effects of font size and polarity on performance and mental effort a 2 × 4 mixed design was used. Polarity was used as an independent variable with two levels (positive and negative polarity). Participants were randomly divided into groups (positive and negative polarity). As a second independent variable with four levels, we used font size. The font size in points (pt) was converted in relation to the ppi of the used touchscreen in the viewing angle. This results in a in minutes of arc (′) measured viewing angle, the so-called angular character height [18] of 16, 20, 24 and 28 arcmin. According to DIN EN ISO 9241-303 [11], the angular character height should be at least 16′ and should be between 20′ and 22′ for office work. As dependent variables, we measured search time, pupil diameter of the left and right eye, NASA TLX, TECOWI and PRACOWI score.

Demographics were surveyed. Visual acuity and color vision was tested at the beginning of the experiment. Then, eye-tracking-glasses were calibrated. Participants were divided randomly into two groups: One group underwent the task with positive polarity (positive condition), the other group conducted the task with negative polarity (negative condition). Each participant was tested individually. After the presentation of the instruction on the touchscreen, participants were presented with the target object which consisted of a three-digit number. Then, a table comprising five rows and five columns was presented. All in all, the table consisted of 25 cells. Each cell was characterized by a position (see Fig. 2). In the positive condition the margins of the table were black, while the background was white. In the negative condition, margins of the table were white on black background. Each cell contains a three-digit number. Before the start of the search task, participants had to confirm via a button on the touchscreen that they have understood and recognized the target object. Participants' task was to find the target object as fast as possible. When the target object was found, participants were instructed to stop the search time by tipping on any point on the touchscreen. Search

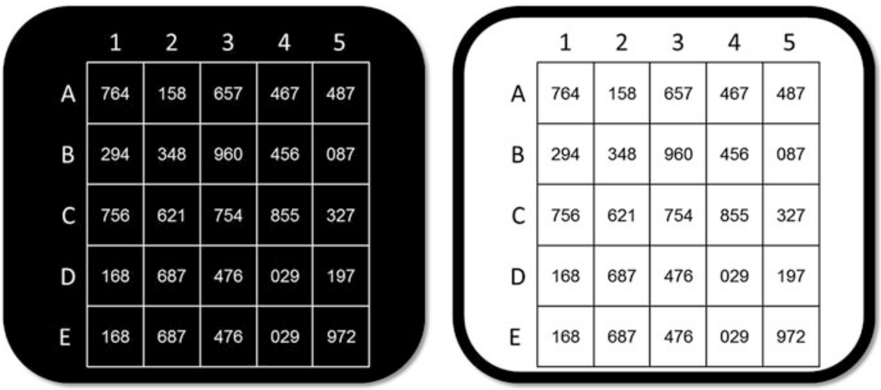


Fig. 2 Schematic illustration of the search task in negative (*left*) and positive (*right*) condition. Positions of each cell were characterized by a number and a letter (e.g. A1 for the upper left cell)

time was defined as the time between confirmation of the target object and the first tipping on the touchscreen. Participants had to say out loud the position of the target after tipping on the cell. The second tipping on the respective cell led to selecting the searched object. Angular character height of the digits varied randomly from 16, 20, 24, to 28 arcmin. Each angular character height appeared 20 times. Hence, the search task consisted of 80 trials. Right after the search task, participants had to fill out the NASA TLX questionnaire. The experiment lasted for about 20 min.

2.4 Statistical Analysis

For the purpose of the investigation of polarity and font size, we conducted a repeated-measures analysis of variance (ANOVA) using polarity (two levels: negative and positive polarity) as a between-subject factor and the font size (four levels: 16, 20, 24, 28 arcmin) as a within-subject factor. Mauchly-test of sphericity was not significant, χ^2 (5) = 0.582, p = .028, n.s. Therefore, sphericity can be assumed.

3 Results

60 % of the participants stated that they are familiar (or very familiar) with handling a tablet. 26.7 % use tablets in their everyday life, mostly for E-Mails, internet surfing and social media (respectively 100 % of the participants who use a tablet). 60 % of participants agreed to the statement that they can learn the features of software quickly. They also agreed to the statement that they have fun trying out new software. 53.3 % stated that they know a lot about software.

When conducting the repeated-measures ANOVA, we found a significant effect for the factor font size, $F(3, 39) = 4.679$, $p = .007$. We found neither a significant effect for the factor polarity, $F(1, 13) = 3.54$, $p = .082$, nor a significant interaction effect between the two factors font size and polarity. Bonferroni post-hoc tests revealed a significant effect in search time for 24 arcmin compared 20 arcmin at $p < 0.05$. The results show that the mean search time can be improved significantly with an angular character height of 24 arcmin. Results are shown in Fig. 3.

Furthermore, we analyzed mental effort using the NASA TLX score. We conducted a *t*-test using the NASA TLX score in positive and negative condition. We found no significant effect, when comparing negative and positive condition, $t(13) = 0.64$, $p = .535$, n.s. Therefore, polarity did not have an significant influence on subjective mental workload.

Moreover, we used the average pupil diameter as a second, physiological and objective measurement for mental effort. For this purpose, we excluded blinking in each eye from the analysis. Blinking was indicated by a pupil diameter of 0. Then, the average pupil diameter was calculated for each eye and each participant individually. A *t*-test for independent samples revealed a significant effect for the left eye, $t(11) = 3.23$, $p = .008$, and for the right eye $t(11) = 3.36$, $p = .006$, when comparing the average pupil diameter for each eye in the negative and positive condition. Results for the left eye are shown in Fig. 4. Results for the right eye are

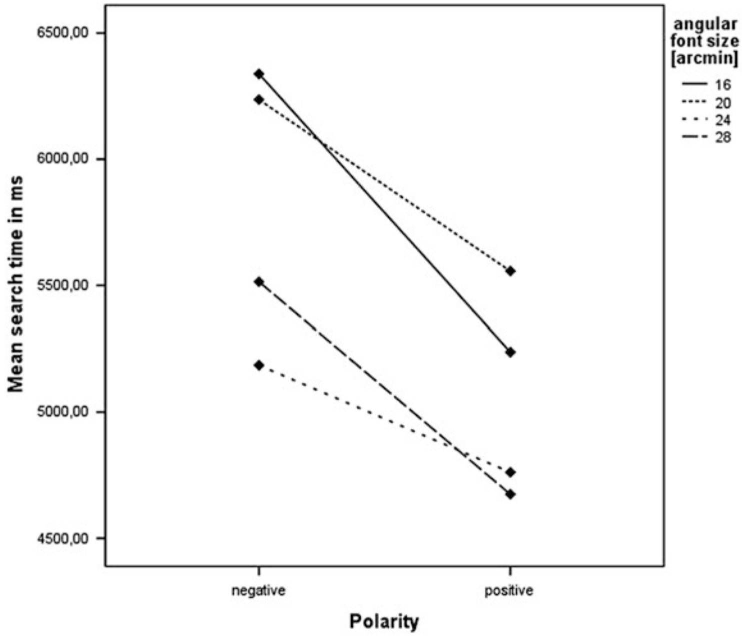


Fig. 3 Effect of font size in negative and positive polarity

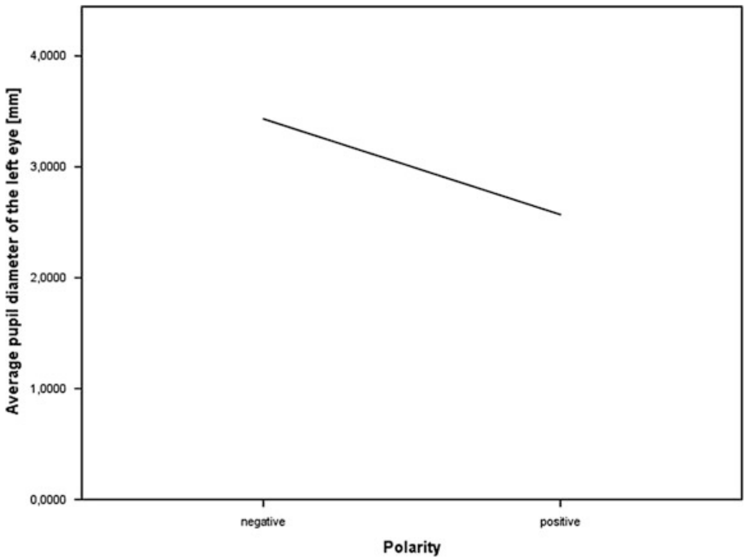


Fig. 4 Effect of average pupil diameter in the left eye in negative and positive condition. Results for the right eye are similar

similar. The results show that positive polarity leads to lesser average pupil diameter. A lower pupil diameter is associated with a lower mental workload [1].

We calculated the correlation between NASA TLX score and TEWCOWI and PRACOWI scores. This analysis did not reveal any significant correlation, $r = -0.034$, $p = .903$, n.s. (correlation between NASA TLX and TECOWI score), $r = 0.123$, $p = .658$, n.s. (correlation between NASA TLX and PRACOWI).

4 Discussion

The aim of the study was the investigation of the influence of polarity and font size on performance and mental workload in a search task. From the results, it can be concluded that font size 24 arcmin leads to significantly shorter search time than the font size 20 arcmin. Polarity did not have any significant effect on search time, though a small tendency that positive polarity might lead to shorter search time can be found. Additionally, polarity did not have any significant effects on mental effort, as there was not observed any difference in the NASA TLX score when comparing positive and negative groups. We did not find any significant correlation between computer knowledge and mental effort. Therefore, we cannot conclude that higher computer knowledge leads to less mental effort. The search task we used in the experiment can be seen critical for the correlation as it does not require any computer knowledge to solve the task. We suggest a more complex task involving several computer functions as an extension of this study. Furthermore, many researchers are of the opinion, that a single technique does not provide an appropriate indication of mental workload and suggest a combination of various measures to enhance the explanatory power [10, 30, 31]. Regarding polarity, we could find only a small tendency for positive polarity. This might be caused by methodological aspects. Our sample size consisted only of 15 participants. A larger sample size could contribute to a significant effect. Furthermore, our sample consisted of highly educated subjects. Most of them hold the German general qualification for university entrance and are quite familiar with the use of tablets. It can be useful to test workers of manufacturing companies as the software applications target group in further research.

Although we could find a significant effect in the average pupil diameter, when comparing negative and positive polarity, we cannot conclude that this is caused by polarity. Positive polarity appeared to be associated with a lower average diameter. In fact, this can be caused by several influences such as a higher display luminance or its individual variability. We suggest the use of other physiological measurement for mental workload (e.g., EEG, heart rate, skin response conductance [27]). Wang et al. [30] report that blink intervals correlate positively with increased mental workload, while blink duration seems to decrease against more intense workload. And Qiu and Helbig [23] found a significant effect of body posture on mental workload. They carried out a video based analysis and found out that the distance between head and display decreases with increasing mental effort.

Concerning the results of the angular character height, our results are in agreement with previous literature results [9]. In the context of the project, this study gives a concrete recommendation: the angular character height of 24 arcmin is optimal for the use of touchscreens in search tasks. Whether this is also adequate for other, more complex tasks is needed to be investigated. We also suggest to use positive polarity thought we did not observe any significant effect of polarity. But previous studies give us a hint, that positive polarity might be preferable due to higher display luminance and less visual fatigue than negative polarity.

Acknowledgements The presented project ProSense (production control based on cybernetic support systems and intelligent sensors) is supported by financial resources (02PJ2490) of the Federal Ministry of Education and Research (BMBF) within the frame concept "Research for the production of tomorrow" and support initiative "Intelligent networking in production—a contribution to industry 4.0". It is supervised by the research center of Karlsruhe (PTKA).

References

1. Beatty J (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol Bull 91:276
2. Brackel T (2009) Adaptive Steuerung flexibler Werkstattfertigungssysteme: Nutzung moderner Informations- und Kommunikationstechnologien zur effizienten Produktionssteuerung unter Echtzeitbedingungen. Gabler-Verlag, Wiesbaden
3. Broy M (2010) Cyber-physical systems. In: Broy M (ed) Cyber-physical systems. Innovation durch softwareintensive eingebettete Systeme. Springer, Berlin, pp 17–32
4. Buchner A, Baumgartner N (2007) Text-background polarity affects performance irrespective of amient illumination and colour contrast. Ergonomics 50(7):1036–1063
5. Buettner R (2013) Cognitive workload of humans using artificial intelligence systems: towards objective measurement applying eye-tracking technology. In: Advances in artificial intelligence. Springer, Berlin, pp 37–48
6. Byers C, Bittner C, Hill G, Zaklad L, Christ E (1988) Workload assessment of a remotely piloted vehicle (RPV) system. In: Proceedings of the human factors society 32nd annual meeting. Human Factors and Ergonomics Society, Santa Monica, CA, pp 1145–1149
7. Cegarra J, Chevalier A (2007) Theoretical and methodological considerations in the comparison of performance and physiological measure of mental workload. In: Engineering psychology and cognitive ergonomics, Don Harris. 7th International conference, EPCE 2007. Springer, Berlin
8. Cushman WH (1986) Reading from microfiche, a VDT, and the printed page: Subjective fatigue and performance. Hum Factors 28:63–73
9. Darroch I, Goodman J, Brewster S, Gray P (2005) The effect of age and font size on reading text on handheld computers. In: Human-Computer-Interaction-INTERACT. Springer, Berlin, pp 253–266
10. de Waard D, Lewis-Evans B (2014) Self-report scales alone cannot capture mental workload. Cogn Technol Work 16:303–305. doi:10.1007/s10111-014-0277-z
11. DIN EN ISO 9241-303:2011: Ergonomie der Mensch-System-Interaktion – Teil 303: Anforderungen an elektronische optische Anzeigen
12. Dworschak B (2015) Kompetenzen der Zukunft in der Industrie 4.0 – Anwendungsfälle, Szenarien, Entwicklungspfade. In: VerANTWORTung für die Arbeit der Zukunft. Bericht zum 61. Arbeitswissenschaftlichen Kongress vom 25–27 Februar 2015, Hrsg: Gesellschaft für Arbeitswissenschaft e.V. (GfA), GfA-Press, Dortmund, ISSN 978-3-936804-18-8, pp 1–5

13. Hart S, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 52:139–183
14. Hasegawa S, Omori M, Watanabe T, Matsunuma S, Miyao M (2009) Legible character size on mobile terminal screens: estimation using pinch-in/out on the iPod touch panel, human interface and the management of information. In: Information and interaction. Springer, Berlin, pp 495–402
15. Hauptvogel A (2015) Bewertung und Gestaltung von cyber-physischer Feinplanung. PhD Thesis, RWTH Aachen University
16. Iqbal S, Zheng X, Bailey B (2004) Task-evoked pupillary response to mental workload in human-computer-interaction. CHI'04 extended abstracts on Human factors in computing systems. ACM, pp 1477–1480
17. Jeske T (2015) Gestaltungspotentiale für die Produktionsarbeit 4.0, In: VerANTWORTung für die Arbeit der Zukunft. Bericht zum 61. Arbeitswissenschaftlichen Kongress vom 25–27 Februar 2015, Hrsg: Gesellschaft für Arbeitswissenschaft e.V. (GfA), GfA-Press, Dortmund, ISSN 978-3-936804-18-8, pp 1–5
18. Jochems N (2010) Altersdifferenzierte Gestaltung der Mensch-Rechner-Interaktion am Beispiel von Projektmanagementaufgaben. PhD Thesis, RWTH Aachen
19. Lin H, Wu FG, Cheng YY (2013) Legibility and visual fatigue affected by text direction, screen size and character size on color LCD e-reader. Displays 34:49–58
20. Millot P, Debernard S, Vanderhaegen F (2011) Authority and cooperation between humans and machines. In: Boy GA (Hrsg) The handbook of human-machine interaction. Ashgate, FL, pp 207–234
21. Parasuraman R, Sheridan T, Wickens C (2000) A model for types and levels of human interaction with automation. IEEE Tran Syst Man Cybern A Syst Humans 30:286–297
22. Parhi P, Karlson A, Bederson B (2006) Target size study for one-handed thumb use on small touchscreen devices. In: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services. ACM, pp 203–210
23. Qiu J, Helbig R (2012) Body posture as an indicator of workload in mental work. Hum Factors 54:626–635
24. Richter T, Naumann J, Horz H (2010) Eine revidierte Fassung des Inventars zur Computerbildung (INCOBI-R). Zeitschrift für pädagogische Psychologie 24(1):23–37
25. Schuh G, Stich B (2011) Produktion am Standort Deutschland. Aachen
26. Schuh G, Stich V, Nyhuis P, Franzkoch B, Hempel T, Hering N et al (2014) Cyber physical production control. In: Brecher C, Klocke F, Schmitt R, Schuh G (ed) Integrative Produktion. Industrie 4.0 Aachener Perspektiven. AKW Aachener Werkzeugmaschinen-Kolloquium. Aachen, 22.23.5.2014. Werkzeugmaschinenlabor WZL der RWTH Aachen. Shaker Aachen, pp 117–137
27. Schultheis H (2004) Pupillengröße und kognitive Belastung
28. Spath D (2013) Produktionsarbeit der Zukunft-Industrie 4.0. Fraunhofer Verlag
29. Thomas N, Du Y, Artavatkun T, She J (2009) Non-intrusive Personalized Mental Workload Evaluation for Exercise Intensity Measure. In: Digital Human Modeling, Vincent G. Duffy. Second International Conference, ICDHM 2009. Springer, Berlin
30. Wang L, Duffy V, Du Y (2007) A composite measure for the evaluation of mental workload. In: Digital Human Modeling. Springer, Berlin, pp 460–466
31. Wierwille W, Eggemeier T (1993) Recommendations for mental workload measurement in a test and evaluation environment. Hum Factors 35:263–281
32. Zuffi S, Brambilla C, Beretta G, Scala P (2007) Human computer interaction: legibility and contrast. 14th International conference on image analysis and processing

Measuring Work Environment Factors by Everyone Using Smartphones

User-Oriented Consideration of Applications for Measuring Work Environment Factors Noise, Climate and Lighting

Michael Spitzhirn, Torsten Merkel, and Angelika C. Bullinger

Abstract

The impact of work environment factors has to be considered in an integrated ergonomic analyzing and designing process. To capture relevant environmental exposures, measurements must be carried out. In companies, an increasing use of smartphones for initial measurements of work environment factors can be noticed. Here, the question arises whether it is possible to determine adequate data for scientific work analyses by using built-in sensors in smartphones. For that purpose, the operational readiness of smartphone applications to measure the work environment factors noise, climate and lighting is investigated. Here, the assessing of measuring accuracy and the recordable measurement quantities for different combinations of applications and smartphones is analyzed. In result, a wide variety of the measurement accuracy for the analyzed environmental factors can be observed.

Keywords

Climate measurement • Environment factors • Lighting measurement • Noise measurement • Smartphones devices

M. Spitzhirn (✉) • A.C. Bullinger
Professur Arbeitswissenschaft und Innovationsmanagement, Technische Universität Chemnitz, Chemnitz, Germany
e-mail: michael.spitzhirn@mb.tu-chemnitz.de; bullinger-hoffmann@mb.tu-chemnitz.de

T. Merkel
Institut für Produktionstechnik, Westsächsische Hochschule Zwickau, Zwickau, Germany
e-mail: torsten.merkel@fh-zwickau.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_19

1 Introduction and Motivation

For integrated ergonomic analyzing and designing of work systems and processes, work environment factors such as noise, climate, or lighting have to be covered [9]. The importance of working environment factors in Germany is shown for instance by the employment survey 2011 [7]:

– 31 % men and 19 % women work in noisy environments, in which 45 % men and 65 % women feel oppressed,
– 26 % men and 13 % women work under poor climate conditions, in which 51 % men and 64 % women feel oppressed,
– 10 % of men and 9 % of women work in poor lighting conditions, in which 53 % of men and 65 % of women feel oppressed.

Basically, work environment factors can have an effect on human health, human performance or well-being [41]. These effects might be positive such as increase of well-being or performance or negative, such as harassment, disturbance of work activity or health effects [6, 40, 42, 44, 47].

In order to avoid negative impacts, limiting values must be observed [15, 17, 47] and appropriate arrangements have to be carried out [41]. To capture relevant exposure such as noise, climate or lighting parameters, measurements must be performed [32, 41]. Depending on the purpose of measuring, different levels of accuracies according to noise measurement [16], climate measurement [18] and lighting measurement (DIN 5032-6:2006) [13] are defined. The utilized instruments have to comply with these accuracies.

In recent years, an increasing use of smartphones for the measurement and assessment of work environment factors such as noise can be noticed in companies [29, 31]. These tools are currently used by persons for indicative measurements, in assumption to obtain adequate measurement results. In conjunction with the numerous applications and advancement in smartphone development, the expectation in terms of usage scenarios of application in industrial context is steadily increasing. Here, the question arises whether it is possible to determine adequate data for scientific work analyses by using built-in sensors in smartphones. In this case, the use of expensive professional technology may be reserved for special tests and a wide range of people will be able to perform environmental investigations in more detail.

For assessing the operational readiness of applications on smartphones to measure work environment factors such as noise, climate and lighting, the Chair for Ergonomics and Innovation Management at the Technische Universität Chemnitz, and the Chair of Ergonomics at the University of Applied Sciences Zwickau conduct scientific studies. The operational readiness is going to be evaluated by assessing the measuring accuracy and the recordable measurement quantities. Hereby, the accuracy of applications is compared to professional measuring instruments. Furthermore, measurement requirements and the characteristics of the applications as well as possible sources of application error will be investigated. The main focus will be put on the operational capability of Android smartphones due their broad use in Germany [28]. Using the presented results, companies and interested individuals get a guideline for the use of smartphones for ergonomic investigation.

2 Applicability and Examination Results of Applications for Noise Measurement

2.1 Sensors, Measures, and Recommendations for Noise Measuring

A-weighted sound pressure level and the peak sound pressure levels are key factors for noise evaluation. In addition, in certain situations a frequency analysis shall also be conducted to identify special exposure hazards for example arising from low-frequency noise [11, 41]. Moreover, practical requirements are placed on the storage and documentation of measured data [29].

To capture noise-related parameters, the built-in microphones in smartphones can be used. However, these are adapted for voice communications and differ in quality depending on the model and manufacturer of microphones, as shown by test reports of various institutes [19, 25]. Figure 1 shows the usual location of usable sensors and states measurable noise parameters.

Using the built-in sensor the sound pressure level can be measured in dB, dB (C) or dB(A) as well as the frequency of the noise source. Here, the measurable variables and built-in functions per application differ significantly [23]. While measuring the sound pressure level, selected applications allow also displaying time profiles, the output of statistic figures such as minimum, maximum or the distribution of the sound pressure level within a certain time period as well as the calculation of a time-average value for measured parameters. Some programs also offer the possibility to perform a frequency measurement by frequency distribution or the dominant frequency plotting [24, 46]. The calculation of the dominant frequency is calculated by means of Fast Fourier Transformation (FFT), and it describes the frequency range with the highest amplitude [5].

Some applications also give the user information about the device handling and the performing of sound measurement. This supports inexperienced persons in dealing with the device in the right way. In addition, simple calculation to determine the noise exposure value LEX (8 h), evaluation tools for the sound pressure

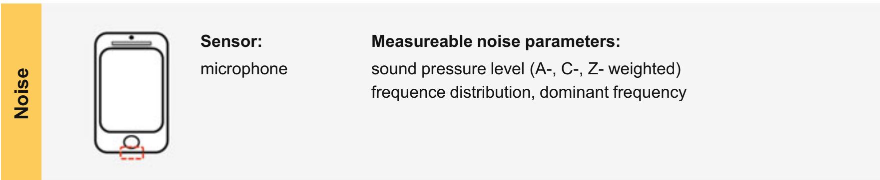


Fig. 1 Measurable noise parameters

level/LEX (8 h) in the form of verbal comments or by coloring, “noise tables for comparison of daily sounds” [3] or also instructions in association with noise are partially integrated. The measurement results can partly be shared via Twitter, Facebook or on special webpages. Furthermore, some application allows the export of measured data via screenshot or file saving which is appropriate for practical use [2].

Special applications offer on noise measurement beyond services such as testing of hearing (Application: “Job Stop scan”) based on “classical business situations such as telephone conference [. . .]” [8], or selecting of appropriate hearing protection (Application: “uvex”) based on measured noise [22]. Such applications can be useful tools for health and safety issues, but also require reliable data. Applications such as “NoiseTube” [34] or “Noisemap” [20] generates noise maps and are mainly implemented for the noise mapping based on measurements by residents of an urban district. For noise mapping in working situations, no applications are available so far, but it would be a useful tool for the documentation of noise levels in companies.

Before the measurement with a smartphone is carried out, it should be checked whether the operational readiness is limited according to the manufacturer. When selecting the appropriate application, restrictions regarding the sound pressure level and the frequency range need to be considered. However, this may only be determined in laboratory studies or to be extracted from existing studies. The provider Valtech [46] for instance has indicated that its application should be used in the range of 40–100 dB(A). Selected applications such as Application “SPLnFFT” can be used to measure up to 114 dB(A) [48]. Furthermore, the detected kind of noise has to be considered. Valtech [46] indicates that noise with dominant low frequencies can lead to sound pressure level underestimation. Unfortunately, the majority of applications give no hints for this problem, which can lead to misjudgments in use. It should also be noted that in own measurements, applications show around 85 dB(A) no changes in displayed value in spite of increasing SPL. An indication of a “refusal” of the measurement is not given to users. Under occupational health and safety aspects, this is a critical issue.

While selecting the application, it should also be examined whether a calibration function is available or not. Before the operational measurement, calibration should be done with a calibrated precision sound level meter and be repeated at least every 2 years. For that purpose, the company must have such instruments or has to purchase external support. This complicates the usage of smartphones for noise measurement. The use of applications without calibration should mainly be avoided. To perform noise measurement, relevant regulations, such as DIN EN 9612:2009, have to be considered. In addition, the smartphone microphone must be oriented towards the sound source. For that purpose the smartphone has to be turned by 180° because the microphone is usually located at the lower surface. As shown by practical tests at the University of Applied Sciences Zwickau, this can lead to significant measurement errors [37].

For noise measurements in companies, sound level meter at least of Class 2 must be used due to its sufficient accuracy [27]. The measurement tolerances, which are used to evaluate the noise measurement accuracy, are shown in Table 1.

Table 1 Permissible limits for measuring sound level meter (Germany/EU, USA)

	Limits for measuring tools for:	
	DIN EN 61672-1:2014 [16] (Germany/Europe)	ANSI S.1.4:1985 [1] (USA)
Type 1: laboratory measurement	+/− 0.8 dB	+/− 1 dB
Type 2: actual measurement	+/− 1.1 dB	+/− 2 dB

2.2 Studies on the Measurement of Noise Variables by Applications

Several authors conducted investigations to the applicability of smartphones for measuring noise.

Kardous and Shaw [29] analyzed 130 iOS and 62 Android noise applications on their applicability for scientific work purposes. For this, the applications had to meet the following measurement criteria: (1) unweighted or A-weighted sound pressure level; (2) 3 dB or 5 dB exchange rate; (3) slow or fast adaptation rate, and (4) detecting the Leq or time-weighted average (TWA). Only 10 of 130 Apps of IOS system, 4 of 62 apps for Android and no Windows app met the criteria. The investigation was conducted with 4 different IOS, 4 Android and 2 Windows devices. The sound pressure level was varied in the range of 65–95 dB(A) in 5 dB(A) stages. As a result, the iOS based applications achieved better results than Android applications due to the quality of the sensor. The measurement of A-weighted sound pressure levels revealed for the iOS application "Sound Meter" a deviation of −0.52 dB(A), which is within the DIN 61672-1:2014 [16] permissible deviations for operational measurements. Overall, three iOS applications ("Noise Hunter", "NoiSee", "Sound Meter") were within ANSI S.1.4:1985 [1] standard (valid in USA) of +/− 2 dB(A) for operational measurements. None of the four Android-based applications fulfilled the requirements for operational measurement. Significant differences were observed between the devices and within the applications [29].

Nast, Speer and Prell [38] conducted a study on the frequency-dependent measurement accuracy of iOS applications. Five applications on the smartphone iPhone 4 were compared to the reference sound meter Brühl & Kjär 2250 (Type 1) in three different sound pressure levels (0, 50 and 70 dB(A)) in the respective frequency bands (0.25; 0.5; 1; 2; 4; 8 kHz) to test the measurement accuracy according to A- and C-weighted sound pressure levels. As a result, the measurement differences between the individual applications differed significantly from each other. Only the application "Sound Meter" was able to meet all testing conditions results within 5 dB(A) for both the A- and C-weighted sound pressure level. The errors of measurement at A-weighted sound level were for this application in all test conditions within 2 dB(A), for the C-weighted sound pressure level within 3 dB(A). For other applications, combined variations for the A- and C-weighting were measured of partially more than +/− 10 dB [38].

Mahler [33] dealt with the frequency response and linearity of sound level variations (dynamics) of smartphones. Various combinations of applications and smartphones were tested. In the frequency band analysis in each case, the measurement errors were determined in intervals of thirds from 63 Hz to 16 kHz with a constant sound pressure level of 75 dB(A). For the dynamic measurements, a weighted wideband noise to VDI 3770 in the range 25–105 dB(A) in 5 dB (A) was used [33]. As a result of investigations, substantial differences in the measurement accuracy in the applications and tested smartphones could be detected. Thus, the applications differ in terms of measurable sound pressure level and frequency range. The most accurate results were obtained with the application "Noise Immission Analyzer 1.2" in iOS system, where the tested deviations for all devices were less than 2 dB(A) in the range of 63 Hz to 13 kHz and sound pressure level of 30–110 dB(A). A calibration of the equipment was not necessary. For Android-based applications, only the application "Audio Tool v.5.8" (35–95 dB(A)) achieved deviations of less than 2 dB(A) under an adjustment of −6.5 dB(A). In general, high and application-specific variations were observed in experiments [33].

2.3 Test Series for Noise Measuring Using Applications

Previous studies mainly analyzed iOS applications in terms of the linearity of level variations and frequency response. Our study presents a comparison of measurement deviation with android applications using smartphone S3 and S4 of the manufacturer Samsung (Class 1 Brühl & Kjär 2250). For that purpose, the measuring series L1 and L2 were performed amongst others:

– Test series L1—Comparison of different applications on three smartphones depending on various sound pressure levels at a frequency of 1000 Hz
– Test series L2—Comparison of an application on three different smartphones depending on SPL and frequencies

2.3.1 Experimental Design

The applications to be examined must allow at least the collection of A-weighted sound pressure level and it should have a customer ranking by at least 3 points on a 5-point scale. This approach was selected since it corresponds to the behavior of a normal user ("everyone").

Before starting the measurement, calibration of the reference sound level meter Brühl & Kjär 2250 is performed by using a sound calibrator Class 1. Calibration of the application does not take place, since only one constant additive sound level change in the systems is possible. The noise measurement is made at a distance of 2 m from the sound source in a recording height of 0.8 m. A frequency generator in combination with an amplifier is used to produce a test sine-wave. To minimize the impact of any fluctuations in sound level, a measurement recording of 10s is performed based on Kardous and Shaw [29]. In the experiment, the L1 measurement deviations are measured in gradual increase of 5 dB(A) of 40 dB(A) up to the maximum detectable sound pressure level in a sinusoidal tone of 1000 Hz. A total of 14 applications are tested for their accuracy. 5 repeated measures are done by each sound level for each application. Based on the test results of L1, the deviation for one application by sound pressure levels (40, 50, 60, 70, 80, 90, 94 dB(A)) at different frequencies (16, 32, 125, 250, 500, 1000, 4000, 8000, 16,000 Hz) is determined in the experiment L2.

2.3.2 Test Results and Discussion

Test series L1—Measurement deviation depending on various noise pressure level

A critical point in using smartphone applications is a limited measurement range. The maximum detectable sound pressure level in dB(A) varies between the applications on one smartphone as well as one application on different smartphones. Difference of more than 15 dB(A) are observed by using different applications on the same smartphone. Difference of same applications on different smartphone are smaller, but can be measured until 10 dB(A). These can be caused by the quality of the built-in sensors as well as the programming of applications. Hereby, none of the applications supply the user with an active notice when the measuring range is exceeded. This can lead to false interpretation of noise situation and may result in a risk for the health of workers.

Furthermore, there are significant measurement differences between noise applications by using the same smartphone. The average error of measurement curves in the sound pressure level range from 40 to 90 dB(A) are shown in Fig. 2 on the example of four selected applications using smartphone S3-1. In addition, the allowable measurement deviations according to DIN EN 61672-1:2014 [16] of +/− 1.1 dB(A) [red] and ANSI S.1.4:1985 [1] of +/− 2 dB(A) [orange] are presented for a sound level meter Class 2.

Two out of 14 applications have measurement deviations within a range of +/− 3 dB (A) compared with the reference sound meter by using S3-1. However, none of the applications reaches the prescribed measuring tolerance of +/− 1.1 dB (A) according to DIN EN 61672-1:2014 [16] or +/− 2 dB(A) according to ANSI 1.4:1985 [1]. Some applications have measuring deviations of more than 10 dB(A). In particular, the majority of tested applications underestimate the true sound pressure level. This is a significant problem in relation to the assessment of health risks.

The best result is shown by Application A with an absolute deviation between −0.3 dB(A) and 3 dB(A) with a standard deviation of 1.7 dB(A). The variation of

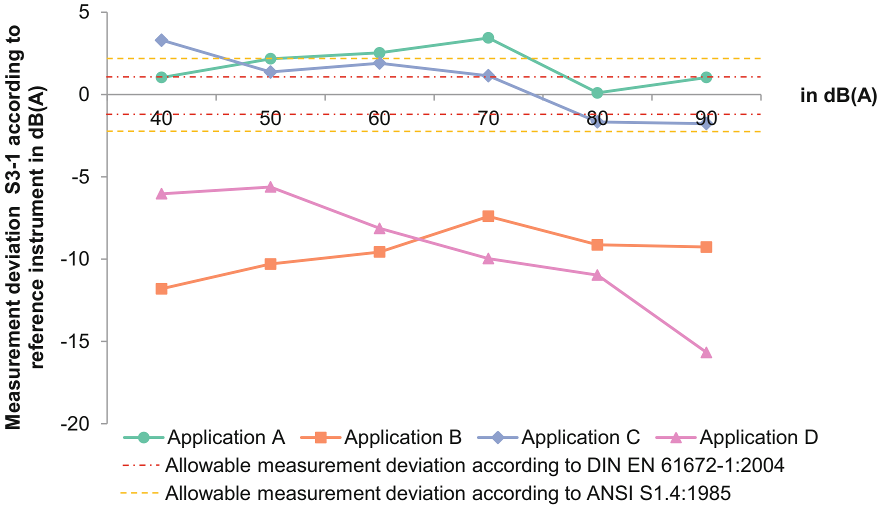


Fig. 2 Measurement deviations smartphone S3-1 according to reference sound level meter

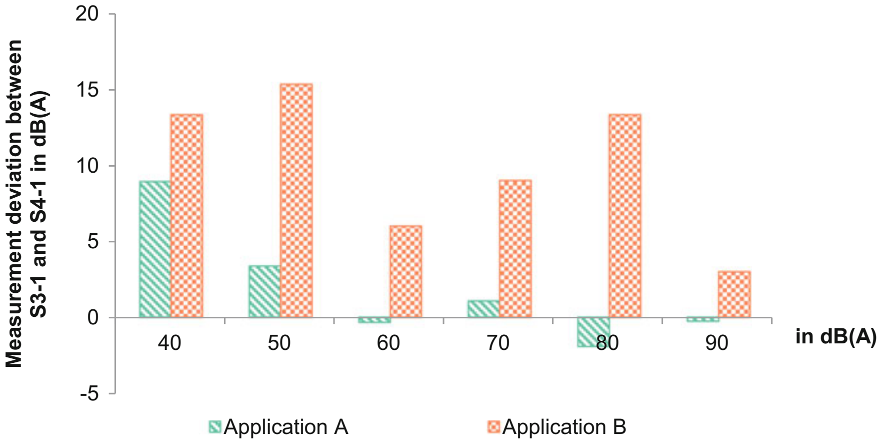


Fig. 3 Sound pressure level differences of same applications on different smartphones

the individual measured values depending on sound pressure level stage is between 0.2 dB(A) at 80 dB(A) and 2.4 dB(A) at 60 dB(A). The application B has the lowest standard deviation with 1.4 dB(A), but a high absolute measurement deviation between −12 dB(A) and −7 dB(A). By using the calibration function the absolute deviation can be reduced to reach the uncertainty range of +/− 3 dB(A). The individual measured values differ depending on the SPL level of 0.3 dB(A) at 70 dB(A) and 1.2 dB(A) at 40 dB(A).

There are also significant measurement deviations by using the same applications on different smartphones. Figure 3 shows the related measurement differences between smartphone S3-1 and S4-1.

The measurement errors of smartphone S4-1 tend to be higher compared to the smartphone S3-1. While using the Smartphone S4-1, none of the applications comply with a range of +/− 3 dB(A) to the reference sound level meter or meet the requirements for measurements in companies.

The significant deviation of the measurement results between different smartphones is contrary to a general use. Since the measurements were accomplished with only a small number of devices, no general statements should be made at this point. In experiment L2, it is going to be examined whether the results at 1000 Hz can be transferred to other frequency bands.

Test series L2 - Measurement deviation depending on various noise pressure level over different frequency bands by using application A

In test series L2 the application A will be tested. For the Smartphones S3-1 and S3-2, measurement deviations between 0.2 and 1.7 dB(A) in the range of 40–70 dB(A) at 1000 Hz are detected. However, with increasing sound pressure level, the measurement error increases up to 7 dB(A) by 90 dB(A). The consequence of this result could be that for each smartphone a laboratory study has to be carried out. Using the example of smartphone S3-2, the measurement deviations for different frequencies are shown depending on the sound pressure level in a range of 40–90 dB(A).

As a result of the measurement series, a high absolute error of the measurement and frequency-dependent variation between reference sound level meter and Smartphone S3-2 is noticed. Similar results were achieved for the other investigated smartphones. The strong variation would require a frequency-dependent calibration, which is not practicable (Fig. 4).

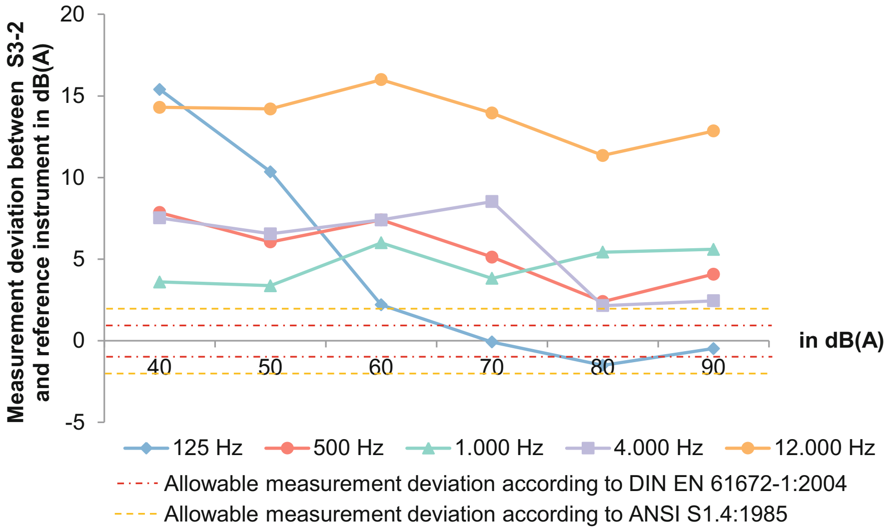


Fig. 4 Frequence-dependent deviation of smartphone S3-2 according to reference instrument

Conclusion: Noise Measurement with Smartphone Applications
The use of android applications to measure the sound pressure level cannot be recommended for "everyone" due to the significant derivation between different applications and same applications on different smartphones. None of the applications has reached the required measurement accuracy according to DIN 61672-1:2014 [16] or ANSI 1.4:1985 [1]. There are also significant deviations depending on the frequency.

On the other hand, different studies show that iOS applications have higher measurement accuracy compared with android applications. For that reason, the use should not be generally rejected but requires a careful choice of application and testing in laboratory according to its applicability on a certain smartphone. In conclusion, a general use of internal sensors of smartphones to measure noise parameters cannot be recommended for "everyone" due to high uncertainty of the measurement results quality.

3 Applicability and Examination Results of Applications for Climate Measuring

3.1 Sensors, Measures, and Recommendations for Climate Measuring

The following factors have to be considered for an ergonomic evaluation of climatic condition effects [4, 18]:

- air temperature, air humidity, air velocity,
- heat radiation from objects e.g. machines,
- Surface temperature, air quality,
- Individual-related parameters such as energetic workload and clothing.

The air temperature and air humidity, as well as the air pressure can be measured by means of internal sensors in Smartphones. For other factors such as the acquisition of air speed or the CO_2 concentration in air, professional instruments must be used. Here, different companies of professional instruments offer a crosslinking to Smartphones. External sensors can also be connected to Smartphones. Figure 5 shows the usual location of usable sensors and states measurable climate parameters.

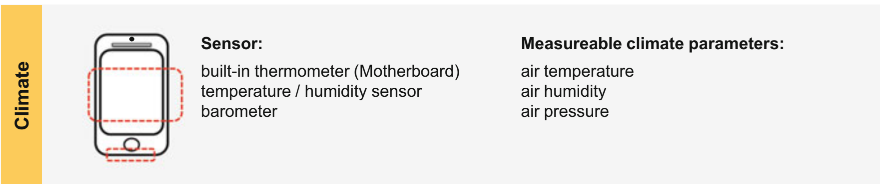


Fig. 5 Measurable climate parameters

Four approaches in applications are used to capture the air temperature and humidity:

- Obtaining climatic data from weather services [35],
- Determining the climate data using the built-in thermometer of the motherboard [21],
- Record of climatic data using special internal temperature and humidity sensors [10], or
- Coupling with an external measuring station [26].

The present study focused exclusively on the use of special internal temperature and humidity sensors and their accuracy. The measurement of air pressure is recorded by means of a barometer, but should not be investigated at this point.

The functionality of climate applications differs. Selected applications offer an additional calibration function, changeable temperature units (°C or °F), adjustable measure intervals or time-oriented diagrams as well as climate assessments. Hereby assessments are carried out by color coding or text. Storage of the measured data is sometimes available.

By performing climate measurements, it is important to consider that heavy usage, storing in closed and narrow areas (such as in pockets) or direct sunlight may cause a strong deviation from the actual air temperature and humidity [43]. According to DIN EN ISO 7726:2001 [18], climate measuring instruments have to meet minimum requirements in relation to the objective of the investigation regarding the measuring range, the measurement accuracy and the 90 % adjustment time [18]. The study is focused on climate comfort measurement. In a range from 10 to 40 °C the temperature measurements are not allowed to exceed a maximum measuring deviation of +/−0.5 °C and absolute humidity accuracy of +/− 0.15 kPa is required within a range of 0.5–3 kPa [18]. Professional instruments are not allowed to exceed a humidity deviation of +/− 2 % [45]. The adjustment time should be as short as possible and must be specified by the instrument.

3.2 Studies on the Measurement of Climate Variables by Applications

Overeem et al. [39] from Wageningen University in the Netherlands collected about 1.3 million temperature records by participation of smartphone users in eight cities worldwide in a research project. The data were accumulated from the temperature sensor on the motherboard. By using a self-developed heat transfer model, Overeem et al. [39] tried to exclude any deviations due to the different usage of smartphones from the measured data. As a result, the team could achieve an average deviation of 1.5 °C [39]. The authors also noted that without the algorithm “Smartphones, on which for example a performance-intensive 3-D game is running, are almost 20 °C warmer than the same model which lies unused in an air conditioned room” [36].

Scientific studies on the investigation of special internally built temperature and humidity sensor could not be found.

3.3 Test Series for Climate Measuring Using Applications

To test the accuracy of special built-in temperature and humidity sensors (see approach 3 in Sect. 3.1), the deviation between two smartphones and a reference measuring tool is analyzed. Here, only test series K1—climate measurement by means of internal sensors in constant conditions - is presented. Further results can be found in Spitzhirn et al. [43].

3.3.1 Experimental Design

In order to measure the air temperature and humidity, two Smartphones of the type Samsung S4 are examined. The climate measurement Testo 435 is used as a reference tool (temperature measurement accuracy: +/– 0.3 °C; humidity: +/– 2 %). Only those applications are chosen for the investigation which use built-in S4 sensors and have a customer ranking by at least 3 points on a 5-point scale. Before the measurement is started, all applications on the smartphone have to be turned off. After that, the smartphone must lie on the table for at least 30 min to ensure that precious use cannot affect the measurement results. As Merkel [37] notes, some sensors need to cool down for more than 1 h.

3.3.2 Test Results and Discussion

Figure 6 presents the measurement deviation of temperature and humidity for 12 applications (application A to L) by using Smartphone S4-1 in constant climatic conditions (T = 23 °C, RH = 35 %) to t = 0 s and after t = 60 s.

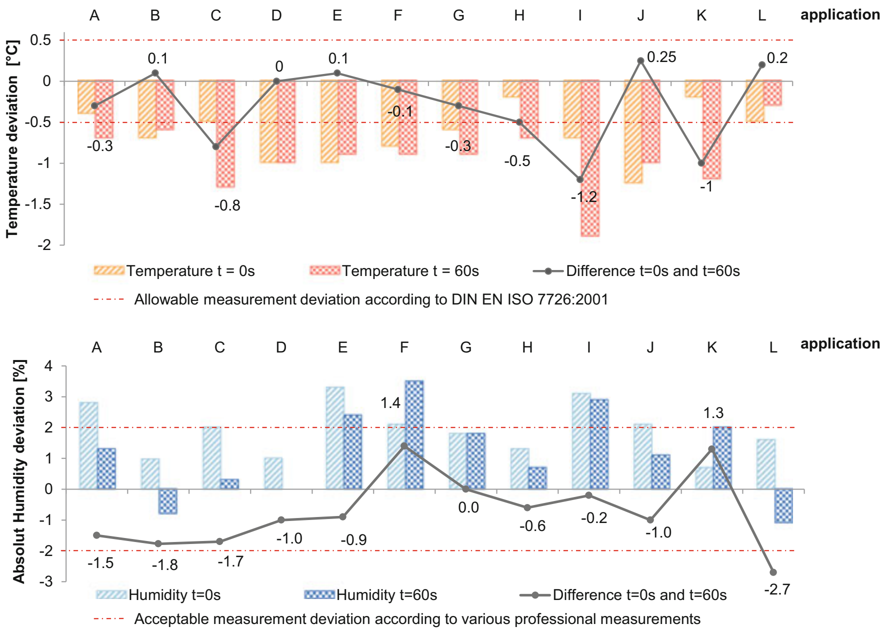


Fig. 6 Temperature and humidity measurement deviation of samsung S4-1 according to the reference tool.png

Five applications meet the requirements of DIN EN ISO 7726:2001 [18] for temperature measurement in t = 0 s. In general, the measured temperature of application is lower compared to the reference instrument. After a measurement time of 60 s, some applications increase or decrease the absolute deviation. The absolute deviation is between −0.2 °C and −1.2 °C at t = 0 s and increases to −0.4 °C and −1.8 °C until t = 60 s. The majority of applications show higher air humidity than the reference instrument, which, however only vary up to 4 %. Most of the application achieve the requirement of +/− 2 % after t = 60 s.

The two tested smartphones S4-1 and S4-2 show temperature deviations between 0.2 and 0.7 °C and humidity deviations of 2 % by using different applications.

Conclusion: Climate Measurement with Smartphone Applications
Climate applications enable relatively accurate temperature and humidity measurements under constant climate conditions. In contrast to noise measurements, no large differences between two smartphones could be observed. However, for adequate climate measurements the user has to prepare the measurement carefully. Heavy use and a locked storage of the smartphone can lead to strong warming, which results in big differences compared to the actual temperature. In addition, the adjustment time also has to be considered especially when climate conditions are changing [43]. For that reason, an adhoc measurement is not possible in all cases. This limits the applicability for "everyone".

4 Applicability and Examination Results of Applications for Lighting Measuring

4.1 Sensors, Measures, and Recommendations for Lighting Measuring

For ergonomic evaluations, the following lighting quality parameters must be observed: [14, 41]

– Illumination (lighting level), light density distribution,
– Limitation of direct reflection and glare by reflection,
– Body rendering (shading) and avoiding from disturbing shadows,
– Light color, color rendering and flicker.

Currently, only the nominal illumination (lighting level) can be measured by using the ambient light sensor. Figure 7 shows the usual location of the sensors and the detectable measuring variables.

The range of functions varies depending on the application. Some applications offer additional functions such as the theoretical calculation of the luminous flux and the distance from the light source. Furthermore, statistical information such as

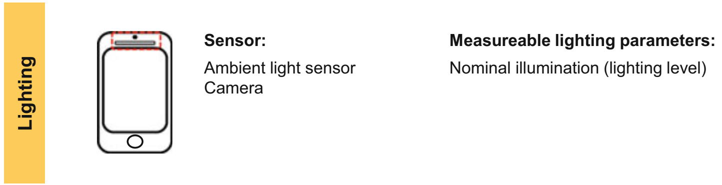


Fig. 7 Measurable lighting parameters

minimum or maximum values is retrieved. Some applications also provide a calibration function (multiplier), storage function or time-oriented diagrams.

Before the use, the position of the sensor must be checked. The user also has to hold the device in the right direction.

For indicative measurements in companies, at least class C requirements with a measurement tolerance of up to 20 % have to be met and a calibration of the device has to be carried out every 2 years [13, 12]. In addition, the v (lambda) adjustment may have a measurement error of up to 9 % [12]. Only the maximum measuring error as a function of the measured light level will be investigated in this study.

4.2 Studies on the Measurement of Lighting Variables by Applications

Klein et al. [30] verified the operational capability of smartphones to the lighting level measurement on the basis of the application “AndroSensor”. Hereby, the inverse square law of the distance was used to demonstrate that smartphones are at least as well suited as traditional methods with photodiode for the quantitative study of light intensity. In addition, the built-in ambient light sensor can also be used to capture the directional characteristics of surfaces [30].

4.3 Test Series for Lighting Measuring Using Applications

In this study, the accuracy of various applications in combination with smartphones for the detection of lighting level will be investigated by lighting level measurements at the workplace under uncalibrated conditions.

4.3.1 Experiment Design

Two smartphones of type Samsung S4 are used to measure lighting level. The precision photometer type 1105 and PocketLux LMT are used as reference instruments. Only applications with a customer ranking of at least 3 on a 5-point scale are selected for the investigation. 14 applications are going to be investigated.

4.3.2 Test Results and Discussion

The measurement results on the example of application L-B by using the smartphone S4-1 as well as the classification of lighting level according to DIN 5032-7:1985 [12] as a function of the total measurement error is shown in Fig. 8.

12 out of 14 applications have the same measurement deviation as application L-B, however, two applications differ significantly from the other measured values. However, most of the applications meet the requirements of Class C for indicated measurements within a lighting level range of 100–1000 lux. Due to the installed sensors, a measurement below 30 lux is not possible. With increasing light intensity, an increase of measurement deviation of smartphone S4-1 is detectable. The measurement deviation is 48 % or 720 lux at 1500 lux. Furthermore, there is a smartphone related measurement deviation. Smartphone S4-2 has steadily increasing measurement deviation according to S4-1 and the reference system, starting from 7 lux at 50 lux to 60 lux at 800 lux. Samsung S4-2 also meets the requirements for indicated measurements according to DIN 5032-7:1985 [12].

Conclusion: Lighting Measurement with Smartphone Applications

Applications for light measurement only allow the measurement of lighting level. In the experiment only Samsung devices were used. For those devices measurement deviation within 20 % could be reported which meets the criteria for indicative measurement in the range of 100–1000 lux. However, measurement deviations between the two tested smartphone were obtained. For that reason, a calibration of the smartphones is recommended. For an ergonomic evaluation of working places, additional parameters such as light density have to be considered. For that purpose, separate tools have to be used.

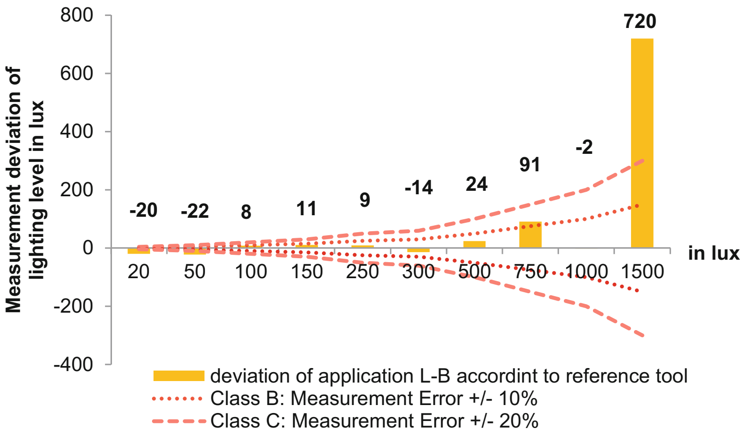


Fig. 8 Deviation of lighting level between application L-B and reference tool using smartphone S4-1

5 Summary and Outlook

The purpose of this study is the review of the applicability of smartphones by using built-in sensors for measuring the work environment factors noise, climate and lighting. First of all, the user of a smartphone application has to consider different aspects to get "adequate" results. It means user cannot only put the device on the table and start the measurement. The user has to consider the sensors position within the hardware and functionality of the different applications. Additionally, knowledge about ergonomic work environment factors and about measurement procedures is also needed to carry out measurements and to evaluate the results properly.

As a result, it becomes clear that the measurement of work environment factor using Smartphones is associated with a high uncertainty of the acquired data. To obtain reproducible results, the described instructions for selecting and performing measurement such as handling of smartphones must be observed. Especially, the fact that the use of same applications on different smartphones or different applications on the same smartphone can result in measurement deviations reduces the general applicability due to causing uncertainty and the demand to check the devices before usage.

The measurement of noise using internal smartphone sensors cannot generally be recommended for everyone. Although individual combination of smartphones and application, as own studies or investigations by Mahler [33] and other authors show, reach accuracy within +/− 2 dB(A), significant variations in measurements of other applications with more than 10 dB(A) can be observed. Hereby, iOS applications mainly have higher measurement accuracy as compared to android applications. This may be caused by a different quality of built-in sensors. Due to this observation, a testing of the combination smartphone—application should be carried out with a precision sound level meter before the measurement starts. Unfortunately, this limits the applicability strongly, but is necessary to avoid false measurement results and evaluations.

The results of the measurement series for climate and lighting parameters are more positive. Smartphones with special built-in temperature and humidity sensors can be used for indicative measurements especially in constant climatic conditions. One advantage, compared to noise measurements, is that no large differences between two smartphones could be observed. However, there are differences between applications which have to be tested before the use of an application. Furthermore, the user has to consider warming factors such as heavy use and locked storage of smartphone which can result in big differences compared to the actual temperature. This limits the general usage for everyone slightly.

Light measurements can be done for indicative measurements in the range from 100 to 1000 lux with the majority of the tested applications. In addition, most of the applications have a calibration function which can be used to adjust the smartphone according to a reference tool. However, the measured parameters are limited which restricts the usage of smartphone to the measurement of lighting level. For more comprehensive measurements, professional tools have to be used.

In future, more faithful and comprehensive work environment measurement can be achieved by receiving improvements in the following field. First of all, the used sensor quality for measurement with smartphones has to be improved. Here, companies with a reputation in work environment measurement should develop appropriate external sensors for the intended use. Furthermore, the applications should be more user-friendly. That means that the application contains adequate guidelines for using this program for environment measurements. A user could be guided when starting the application for the first time. Explanations and hints to first steps such as calibration and measurement procedure would help to get more accurate data.

Moreover, applications should be developed that combine various functions of different application such as measurement, statistic evaluations, recommendations, storage and documentation and so on. In addition, noise, climate, lighting and other ergonomic issues could be collected and be evaluated by one system, which would enable an integrated ergonomic evaluation. Furthermore, the visualization of the data using for example maps, which visualized the current situation, would support the user to identify a critical ergonomic situation more quickly.

Regardless of the technical development and future trends, the accompanying analyses with respect to the currently usable technology are necessary to generate recommendations in order to avoid misuses and misinterpretations of work environment measured data by using smartphones.

References

1. American National Standards Institute (1983). ANSI S1.4-1983 (r2006/ansi s1.4a-1985 (r2006). American national standard specification for sound level meter. New York, NY: American National Standards Institute.
2. appgefahren.de (2010) Gratis-App WideNoise misst den Geräuschpegel. Online unter: http://www.appgefahren.de/gratis-app-widenoise-misst-den-gerauschpegel-3042.html
3. Ärzte Zeitung (2012) Apps im Gesundheitswesen - Lärm messen mittels App. Springer Medizin. Online unter: http://www.aerztezeitung.de/praxis_wirtschaft/w_specials/gesundheitsapps2011/article/800860/laerm-messen-mittels-app.html
4. BAuA (2004) Ratgeber zur Ermittlung gefährdungsbezogener Arbeitsschutzmaßnahmen im Betrieb – Handbuch für Arbeitsschutzfachleute, 4.aktualiserte Auflage, Dortmund/Berlin
5. BSB Mobile Solutions (2014) deciBel. Online unter: https://play.google.com/store/apps/details?id=bz.bsb.decibel&feature=search_result
6. Bullinger HJ (1994) Ergonomie – Produkt- und Arbeitsplatzgestaltung. Springer Fachmedien, Wiesbaden
7. Bundesministerium für Arbeit und Soziales [BMAS], Bundesanstalt für Arbeitsschutz und Arbeitsmedizin [BAuA] (2013) Sicherheit und Gesundheit bei der Arbeit 2011. Unfallverhütungsbericht Arbeit. BAuA. Dortmund/Berlin/Dresden
8. Bundesverband der Hörgeräte-Industrie (2015) Lärm. Online unter: http://www.ear-fidelity.de/das-gehoer/laerm/
9. Caffier G, Steinberg U, Liebers F (1999) Praxisorientiertes Methodeninventar zur Belastungs- und Beanspruchungsbeurteilung im Zusammenhang mit arbeitsbedingten Muskel-Skelett-Erkrankungen. Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin – Forschung – Fb 850. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Dortmund/Berlin http://www.baua.de/cae/servlet/contentblob/673906/publicationFile/46978/Fb850.pdf

10. CHIP Digital GmbH (2013) Galaxy S4: Die wichtigsten Sensoren des Handys. Online unter: http://www.chip.de/news/Galaxy-S4-Die-wichtigsten-Sensoren-des-Handys_61493678.html
11. DIN 45680:2013. Messung und Beurteilung tieffrequenter Geräuschimmissionen. Entwurf
12. DIN 5032-7:1985. Lichtmessung. Klasseneinteilung von Beleuchtungsstärke- und Leuchtdichtemeßgeräten
13. DIN 5035-6:2006. Beleuchtung mit künstlichen Licht – Teil 6: Messung und Bewertung
14. DIN 5035-7:2004. Beleuchtung mit künstlichem Licht – Teil 7: Beleuchtung von Räumen mit Bildschirmarbeitsplätzen
15. DIN EN 12464-1:2011-08. Licht und Beleuchtung - Beleuchtung von Arbeitsstätten - Teil 1: Arbeitsstätten in Innenräumen
16. DIN EN 61672-1:2014. Elektroakustik - Schallpegelmesser - Teil 1: Anforderungen
17. DIN EN ISO 11690-1:1996. Akustik – Richtlinie für die Gestaltung lärmarmer maschinenbestückter Arbeitsstätten – Teil 1: Allgemeine Grundlagen
18. DIN EN ISO 7726:2001. Umgebungsklima – Instrumente zur Messung physikalischer Größen
19. Focus Online (2013) Diktiergerät gegen Handy – Wie gut sind Tonaufnahmen mit dem Smartphone?. Online unter: http://www.focus.de/digital/handy/tid-33811/diktiergeraet-gegen-handy-wie-gut-sind-tonaufnahmen-mit-dem-smartphone_aid_1105550.html
20. Frank C (2013) 200 Nutzer messen 150000 Werte. Forschung: Mit der von TU-Informatikern entwickelten Smartphone-App "Noisemap" werden Lärmdaten ermittelt. Darmstädter Echo. Online unter: https://www.wiso-net.de:443/document/DECH__2013121647510
21. GIGA (2013) Wie funktionieren eigentlich: Thermometer-Apps. Online unter: http://www.giga.de/apps/google-play-store/specials/wie-funktionieren-eigentlich-thermometer-apps/
22. GIT Sicherheit (2011) App für mobile Lärmmessung und schnelle Gehörschutzauswahl. Online unter: http://www.git-sicherheit.de/produkte/management/app-fuer-mobile-laermmessung-und-schnelle-gehoerschutzauswahl
23. Google (2015a) Google play-Suche "Lärmmessung". Online unter: https://play.google.com/store/search?q=l%C3%A4rmmessung&c=apps&docType=1&sp=CAFiDgoMbMOkcm1tZXNzdW5negIYAIoBAggB%3AS%3AANO1ljIdPoI&hl=de
24. Google (2015b) Google play – Noise Exposure. Online unter: https://play.google.com/store/apps/details?id=se.av.buller&hl=de
25. Google (2015c) Google-Web-Suche "Vergleich interne Mikrofone Qualität Smartphone". Online unter: https://www.google.de/search?q=Vergleich+interne+Mikrofone+Qualit%C3%A4t+Smartphone&start=10&cad=h
26. Ibb-wetter.de (2015) ibb-wetter für Smartphone. Online unter: http://ibb-wetter.de/wetterdaten/smartphone
27. IG Metall Vorstand [IGM] (2009) Lärmprävention in Büro und Produktion. Lärmprävention im Betrieb. Arbeitshilfe 24. http://library.fes.de/pdf-files/netzquelle/igm/laerm.pdf
28. Kantar Worldpanel (2015) Smartphone OS sales market share. Online unter: http://www.kantarworldpanel.com/global/smartphone-os-market-share/
29. Kardous CA, Shaw PB (2013) Evaluation of smartphone sound measurement applications. Division of applied research and technology engineering and physical hazards branch. EPHB Report 349-12a
30. Klein P, Kuhn J, Wilhelm T, Lück S (2014) Beleuchtungsstärken mit AndroSensor messen. Physik in unserer Zeit 45(4):201–202
31. Kohlrausch RC (2010) Bei Lärm zum Handy greifen – und messen. Hannoversche Allgemeine. Online unter: http://www.haz.de/Ratgeber/Bauen-Wohnen/Uebersicht/Bei-Laerm-zum-Handy-greifen-und-messen
32. LärmVibrationsArbSchV §3 (2007) Verordnung zum Schutz der Beschäftigten vor Gefährdungen durch Lärm und Vibrationen http://www.gesetze-im-internet.de/bundesrecht/l_rmvibrationsarbschv/gesamt.pdf
33. Mahler N (2015) Messtechnische Untersuchungen im Projekt "Smartphone als Schallpegelmesser". Untersuchungsbericht Nr. 5'214'001'633. Bundesamt für Umwelt BAFU. EMPA

34. Maisonneuve N, Stevens M, Nissen ME, Hanappe P, Steels L (2009) Citizen noise pollution monitoring. In: Proceedings of the 10th Annual International Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government, Digital Government Society of North America, pp 96–103
35. Mattgey R, Von Brockedorff S (2012). Was taugen Internet-Wetterdienste. Online unter: http://www.chip.de/artikel/Die-besten-Wetter-Apps-und-Online-Wetterdienste-im-Test_55968168.html
36. mbe (2013) Temperaturmessung: Wettervorhersage mit Smartphone-Batterien. Online unter: http://www.spiegel.de/wissenschaft/technik/temperaturmessung-smartphone-batterien-sollen-wetter-vorhersagen-a-916625.html
37. Merkel T (2014) Smartphones und Tablets im Alltag der ergonomischen Arbeitsplatzbewertung Hat die spezifische Messtechnik ausgedient?. Industrial Engineering. Fachzeitschrift des REFA-Verbandes 02/2014; REFA-Bundesverband e.V., Darmstadt; ISSN 1866-2269, pp 20–24
38. Nast DR, Speer WS, Le Prell CG (2014) Sound level measurements using smartphone “apps”: usefull or inaccurate? Noise Heath 16(72):251–256
39. Overeem A, Robinson JCR, Leijnse H, Steeneveld GJ, Horn BKP, Uijlenhoet R (2013) Crowdsourcing urban air temperatures from smartphone battery temperatures. Geophys Res Lett 40:4081–4085
40. Schierz C, Krüger H (2002) Vor- und Nachteile intensiver Beleuchtung am Büroarbeitsplatz. Zeitschrift für Arbeitswissenschaft, Sonderausgabe Ergonomie im Büro 56(4):269–274
41. Schlick CM, Bruder R, Luczak H (2010) Arbeitswissenschaft, 3rd edn. Springer, Berlin
42. Seppanen O, Fisk W, Faulkner D (2003) Cost benefit analysis of the night-time ventilative cooling in office building. Conference: Healthy Buildings 2003, Singapur. Online unter: http://www.osti.gov/scitech/biblio/813396
43. Spitzhirn M, Merkel T, Bullinger AC (2015) Untersuchungen der Einsetzbarkeit von Smartphone-Applikationen zur Messung von Arbeitsumweltfaktoren am Beispiel Klima. 5. Symposium Produktionstechnik. Heft 6. ISSN: 1863-1916.
44. Szalma JL, Hancock PA (2011) Noise effects on human performance: a meta-analytic synthesis. Psychol Bull 137(4):682–707
45. Testo AG (2015) Praxisratgeber: Behaglichkeitsmessung am Arbeitsplatz. Testo. Lenzkirch. Online unter: https://www.testo.de/behaglichkeit/praxisratgeber.jsp
46. Valtech (2015) How to protect people from the dangers of noise. Online unter: http://sg.valtech.com/work/case-studies/how-protect-people-dangers-noise
47. VDI 2058 Blatt 2:1988-06. Beurteilung von Lärm hinsichtlich Gehörgefährdung
48. Weber D (2013) 30 iPad Sound Measurement Apps Reviewed. Safety Awakenings. Online unter: http://www.safetyawakenings.com/safety-app-of-the-week-42/

Integration of Eyetracking-Data into Virtual 3D-Development Tools

Description and Potential of a Procedural Principle

Kai Bürkle and Martin Schmauder

Abstract
Concerning vision from mobile equipment machinery there is a lack of methods that sufficiently take dynamic criteria into account. In consequence, inspections regarding direct and indirect vision have so far been conducted with static methods, and, in general, carried out within a late phase of the product development process, using real prototypes. This article describes a method to analyse vision by focussing on the integration of dynamic user and process characteristics in early development phases. Therefore, body and eye movements of operators and associated machine processes were recorded during virtual work task performances using an interactive simulator. By transferring these data to a virtual machine model and digital representatives of the operators, work case scenarios can holistically be copied into a 3D environment. Aggregated dynamics of operators and machines provide a broad range of analysing possibilities, including the evaluation of semi-transparent visual interfaces.

Keywords
Mobile equipment machinery • Operator's vision • Virtual prototype • Interactive simulator • User orientation • Process orientation • Transfer of eyetracking data • Semi transparent visual interfaces • Superimposed visual information • Head-up display • Head-mounted display

K. Bürkle (✉)
Bombardier Transportation, Berlin, Germany
e-mail: kai.buerkle@rail.bombardier.com

M. Schmauder
Institut für Technische Logistik und Arbeitssysteme, Technische Universität Dresden, Dresden, Germany
e-mail: martin.schmauder@tu-dresden.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_20

1 Motivation and Objectives

About 80–90 % of relevant information required to perform driving and working tasks are obtained through visual perception [6, 9, 10]. Based on this fact, accidents with mobile machines involved often can be inferred from deficient visual conditions. For several years, Clemens has been recording and publishing exceeding tragic accidental occurrences which often result in heavy, permanent injury or even fatal consequences for involved humans [5]. Results of analyses concerning accidents in the earth-moving sector, carried out by the Employers' Liability Insurance Association, reinforce today's urgent needs to take constructive measures against deficient visibility [8].

Tasks performed with mobile machines, i.e. moving of specific machine dependant kinematics, takes centre stage within a high dynamic environment. As a result, characteristics of process-related body and eye movements clearly differ from solely driving task motivated movements such as "car driving". Furthermore, the relationship between task driven body and eye movement strategies, individual anthropometric properties and vision related matters is abundantly clear. However, especially these user and process oriented aspects seem not to be adequately taken into account within the well-established standard ISO 5006 [7], used of late by manufacturers to prove compliance with legal requirements. ISO 5006 describes a testing method based on shadows projected on the ground, arising by the reason of light being masked by machine components. For this testing method light sources are positioned at assumed locations of an operator's eyes by using a substitution model as specified in ISO 5006. Visual obstructions due to opaque parts of the machine appear as core shadows. These are analysed in terms of quantity, spatial location and size, referring to a standardised distance on the so-called "visibility test circle" around the tested machine. Then, the determined criteria are evaluated by comparing these to the requirements listed in the standard. However, these standardized requirements only underlie average anthropometric body measures of worldwide male operators. Hence, their validity for operators with deviating body measures seems to be questionable. In addition, the substitution model described in ISO 5006 only takes limited head and upper limb movements into account. As a consequence, possible eye positions can only be found on an imaginary horizontal circular disc. At last, the dynamics of the machine and its components is strongly abstracted and replaced by a simplified geometric issue as a mere static problem.

Figure 1 opposes shadows as formed for a standard compliant arrangement (a) to a setting with exemplarily deviating positions of light sources and alignments of machine components and attachments (b). There are remarkable differences of the shadow sizes. Locations of the shadows even change affected sections on the visibility test circle. Additionally, the modality for documenting test results by sketching in markings on the circle (c) reveals a problem for tracing shadows back to a specific opaque machine component or attachment causing the masking. In case of testing a machine non-compliant to the standard, this way of documentation turns out to be inconvenient for design and constructing engineers in retrospect.

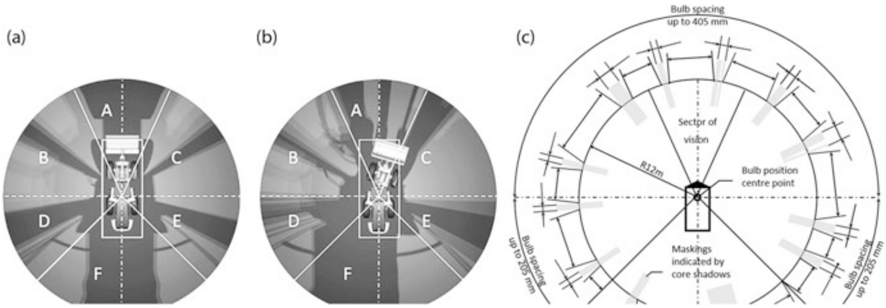


Fig. 1 Shadows on the visibility test circle for a wheel loader with an ISO 5006:2006 compliant arrangement (**a**), shadows of the same wheel loader with deviating positions of light sources and alignment of the machine and it's equipment respectively (**b**), detail of an exemplary documentation of test results on the visibility test circle according to the standard ISO 5006 (**c**)

The requirements for developing the "DYNASICHT" method[1] were to holistically integrate dynamic factors of vision for virtual prototypes. These factors were expected to appear as process motivated influences on operators, machines and environmental facets, varying on specific work tasks. This implied to firstly develop an adequate approach to analysing and defining user- and process-oriented aspects. Particular attention was payed to the variability of body and eye movements for anthropometrically differentiated operators, to dynamic aspects of focussed earth-moving machinery, to the characteristics of work task assignments and especially to the concatenation of all these criteria. It soon became apparent that the idea of manually handling all of the collected data was not efficiently applicable. Hence, collation and processing of data had to meet special demands in order to provide computeraided and automated evaluation possibilities within the field of digital ergonomics.

2 Method

The approached method development basically consists of three sequential phases.

Phase 1

The first phase deals with acquiring data by analysing different work task assignments carried out with real and virtual machines. In order to design work tasks in a virtual environment, experimental field studies were performed. After the transfer of acquired information to a virtual 3D-environment (Fig. 2), test persons can operate a machine model of a heavy wheel loader to work on provided representative work tasks. The hexapod moving system of the interactive earth moving machinery simulator of the Technische Universität Dresden (TUD) ensures a realistic experience, providing a highly immersive setting. This simulator is

[1] With funds from the German Research Foundation (DFG).

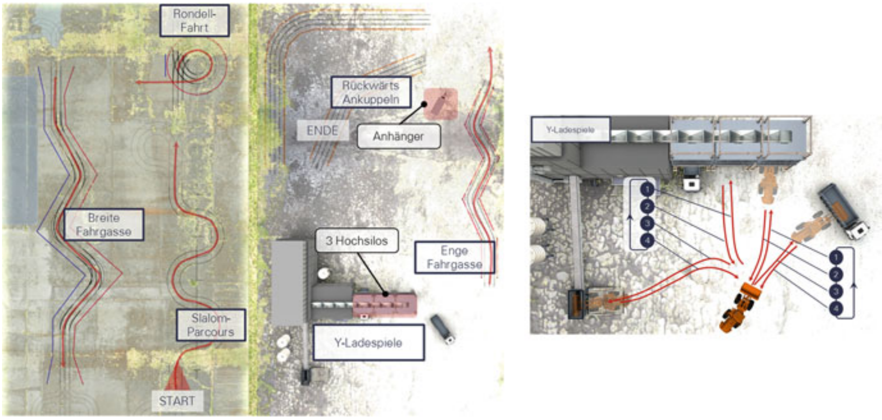


Fig. 2 Virtual work tasks in a 3D-environment [3]

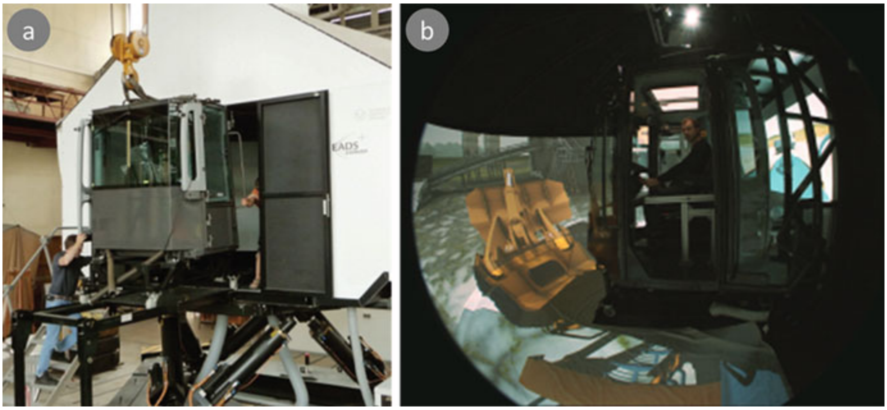


Fig. 3 Installation of a wheel loader cabin into the interactive simulator of the TUD (**a**) and view to the inside with the multiprojection system in operation (**b**)

capable of accommodating whole cabins to provide authentic real-world operating interfaces. Figure 3 shows the interactive simulator being prepared for an experimental setup by mounting a cabin (a) and a view to the inside of the simulator showing virtually operated components and attachments of a wheel loader model (b).

Individual anthropometric data of each of the participating test persons were measured. The experimental setup not only delivers data regarding process related dynamics of a virtual wheel loader and it's components but also serves to acquire corresponding movements of the operators' bodies and eyes. Body movements are recorded with assistance of the Institute for Occupational Safety and Health of the

Fig. 4 Holistic virtual copy of work tasks, including body and eye movements of test persons attached to their individual digital representatives, combined with the movements of the wheel loader as virtually operated by the test persons in the interactive simulation [4]

German Social Accident Insurance, using the CUELA-System.[2] Eye movements are acquired by using the binocular eyetracking system "Eyelink II" provided by SR Research.

Phase 2

Within the second phase, all of the acquired data sets are unified within the modelling and rendering tool 3ds Max by Autodesk. The integration of several data sets, each of which are originally supposed to be analysed separately into only one modelling tool, requires an applicable strategy to deal with extensively big data files. One of the successful approaches is to highly abstract virtual work task scenarios using simple primitives and trying to only manipulate small bits of data sets for transferring trial purposes of short representative sequences. Data processing of complete data set-driven body and eye movements of the digital human model and the machine dynamics finally require abundance of patience. The intermediary results from phase 2 are highly realistic integral virtual copies of work task sequences absolved by the test persons as shown in Fig. 4.

Phase 3

The virtual copy of work case scenarios obtained within phase 2 provides the basis to further develop an automatically running procedure for analysing vision related aspects. The final result benefits from featuring a quantitative functionality which is applicable to any 3D-objects within the virtual 3D-environment. Thus there is no need to be tied down to any specific areas of interests and to place optical markers prior to conducting experiments. Additionally, the granularity of the quantifying function can be adapted—as needs arise—by simply manipulating the sizes and shapes of affected grids of objects. Additional utilities help to differentiate between lines of vision as a function of their kind of collisions encountered on their way (details see Sect. 3).

The following chapters further describe details of the transfer procedure and analysing functions regarding eyetracking data within virtual 3D-modelling tools as

[2] Computerunterstützte Erfassung und Langzeit-Analyse von Belastungen d. Muskel-Skelett-Systems.

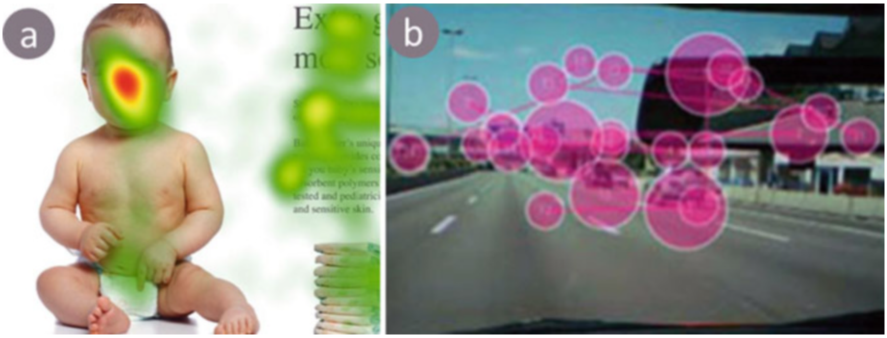


Fig. 5 Examples for established ways to represent eyetracking results. Heat-map visualization [2] (**a**), gaze-trail representation [11] (**b**)

developed within phase 2 and 3. The availability of dynamic lines of vision based on eyetracking data does not only represent an innovation in the context of vision related analyses with virtual prototypes but also enables to take effects of advanced semi-transparent visual interfaces such as Head-up displays (HUDs) into account.

2.1 Procedural Principle for Transferring 2-Dimensional Eyetracking Coordinates to 3D-Design and Modelling Tools

Figure 5 shows established possibilities to represent results from eyetracking analyses. A scene-camera delivers an approximate view from the perspective of the eyetracked subject. The locations of eye fixations are highlighted and displayed within 2D-videos as so-called "overlay-videos".

By regarding technical parameters of applied hardware components and formats, such as specifications of the objective lens of the scene-camera or the resolution of the overlay-video sequences, it is possible to visualise eyetracking data within 3D-environments.

In case of the Eyelink II system, the scene-camera enables a field of view with a horizontal angle of $\alpha = 95°$. The processed overlay-video has a resolution of $B = 720$ pixels at width and 480 pixels at height.[3] The distance A to an imaginary projection plane (Fig. 6a) which matches the aspect ratio of an image section recorded by the scene-camera (Fig. 6b) and which also is oriented parallel to the frontal plane of the subject's face can be calculated with the trigonometric relation as per Eq. (1).

[3] For precise results, further details have to be considered, e.g. regarding pixel aspect ratios in NTSC based video sequences or varying distortion effects caused by objective lens parameters.

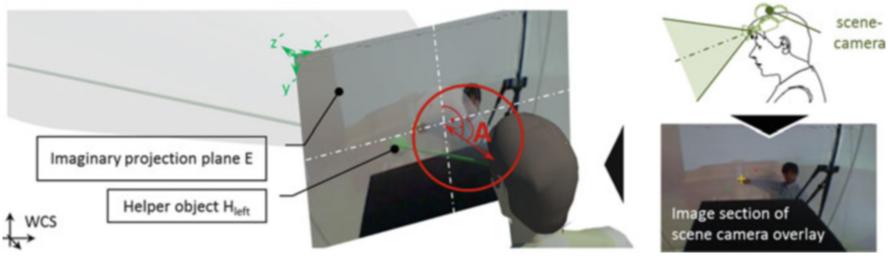


Fig. 6 Geometric relation between highlighted eyetracking coordinates in the overlay-video and elements of the hierarchical structure for the virtual construct as required to import eyetracking data (exemplified illustration for the left eye)

$$A = \left[tan \left(\frac{180^{\circ} - \alpha}{2} \right) \right] \cdot \frac{B}{2} \tag{1}$$

By configuring an appropriate hierarchical structure using software specific functions, the x- and y-coordinates contained in eyetracking data sets can then be assigned to virtual helper objects (e.g. H_{left} in Fig. 6a as an example for the left eye). It is important, that plane E is hierarchically linked to the spatial position and orientation of the eyetracked test person's head. Additionally, the helper objects have to be configured in such a manner that these inherit transformations of plane E. This way, imported eyetracking coordinates refer to the local coordinate system of plane E. Importing eyetracking data works by applying manipulated 3DS max specific xaf-animation files to the helper objects. By shifting the points of their origin to the upper left corner of plane E, helper objects move to the correct locations which correspond to the highlighted areas for visual fixations in the overlay-video sequence. In order to visualize eye movements with lines of vision, additional objects, preferably thin cylindrical primitives with their starting points equalling the spatial eye positions and their roll-axis being oriented towards the appropriate helper objects, can be applied.

2.2 Determination and Virtual Visualization of a Viewing Distance Between an Eyetracked Test Person and Fixated Targets

In case of tracking the movements of both eyes, another parameter, the viewing distance between an eyetracked test person and fixated objects can approximately be determined by including the effect of vergence eye movements. The approximation is due to the fact that lines of vision only theoretically intersect. Effectively, lines of vision for the left and right eyes, animated by eyetracking data, hardly ever have coordinates in common since outlying of square. Hence, the centre of the shortest link between both lines serves as a substitute for an intersection point. Its spatial position can be determined by solving a linear system of equations (2),

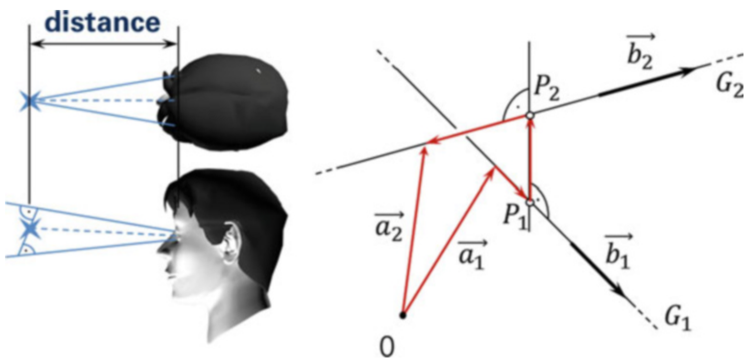


Fig. 7 Determination of the centre point of the shortest linkage between two lines of vision substituting the viewing distance target

deriving from the closed lineament as demonstrated in Fig. 7. In order to dynamically realise the determination of the distance with algorithms mapped on transformation controllers in 3ds Max, the solutions of the linear system had to be expressed parametrically with assignable variables. This task was carried out by computing solutions with wolfram alpha [12].

$$\vec{a_1} + r_0 \cdot \vec{b_1} + t_0 \cdot \left(\vec{b_1} \times \vec{b_2}\right) - s_0 \cdot \vec{b_2} - \vec{a_2} = 0 \qquad (2)$$

3 Results

After the transfer of eye movements to the 3D-environment, a scripted program can be executed to flag intersections of lines of vision with any user-defined virtual objects, based on keyframes. The script includes a functionality to quantify and to distinguish between different types of generated flags. These types vary in the sequential order both lines of vision for the left and right eye collide with objects, which are distinct from each other in terms of assigned materials. The distinction includes attributes such as

– no obstructions,
– encountered intersection(s) with transparent objects,
– indirect vision via mirrors (the lines of vision are deflected correctly on their further course)
– at least one opaque object obstructing the view of the operator.

In order to exemplarily provide a reference for an area of vision, a virtual cylinder can be generated around the machine model. It is configured to move along with the wheel loader model by inheriting the translational and rotatory transformation of the cabin and serves as an abstract element to allocate the distribution of the viewers' fixations in terms of attributes and quantities (Fig. 8a). The grid of the area of vision can be structured as needed to provide

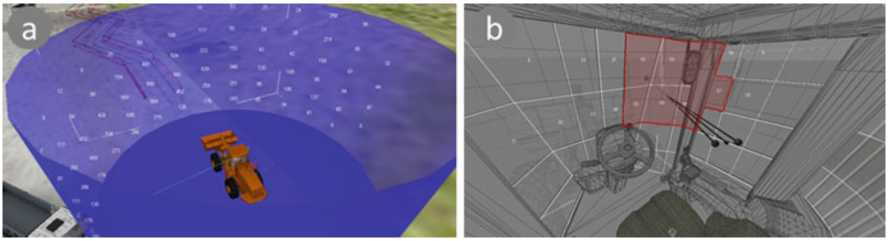


Fig. 8 Quantified intersections of lines, distinguished between the different types of collided objects. Area of vision referenced to a cylinder (**a**). Segmented windscreen with process-oriented distribution of flags, highlighting intersected grid elements by the lines of vision (**b**)

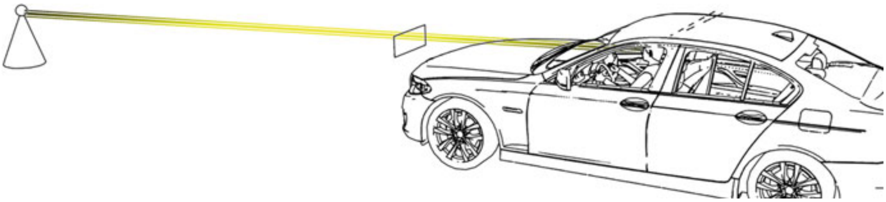


Fig. 9 Static testing setup with a Head-up display and an object in alignment to superimposed information

Fig. 10 Differentiation of fixated objects by an eyetracked test person (**a**), if these are aligned one behind the other (**b**) within an animated scenario in the 3D-modelling environment 3DS max (**c**)

sections dependent upon a desired degree of detail. For each of the grid sections the numbers of intersections are counted as for selected attributes.

Figure 8b demonstrates the same procedural principle applied to an alternative virtual element—the front shield of the cabin. The highlighted area within this example shows the grid sections of the windscreen most often intersected by the lines of vision for right turns with the wheel loader.

The determination of a viewing distance using eyetracking data sets was pilot tested with an experimental setup as shown in Fig. 9, using a Head-up Display of a car in a static testing setup.

The acquired data are transferred to the virtual environment. A script then automatically highlights collisions of components with a helper object that is bound to the orientation of the central line of vision but placed within calculated distances.

The results demonstrate, that the distances, determined by applying the procedure as described above, help to distinguish between different distances of viewing targets in the case should a virtual line of vision simultaneously intersect with a real-world object in alignment with another (traffic cone and superimposed visual information of the HUD, Fig. 10). This scenario also applies to other hardware applications which superimpose visual information such as augmented reality applications with Head-mounted Displays (HMDs).

4 Discussion and Outlook

By synchronising virtual operators' perspective views with sequences of the overlay-video provided by the eyetracking analysing software, "original" (overlay) and "virtual" eye movements can be compared with each other. The comparison reveals deviations of virtually intersected objects from focussed objects on the overlay-video within definable limits. The deviations primarily depend on the quality of each of the data sets but also originate from a complex interdependency of several error sources, e.g.

- inaccuracies in the way of presenting the virtual environment by a distorted projection with a multiple projection system on a double curved surface,
- errors occurring within adjusting human models to individual body measurements,
- deviations due to measuring and transferring errors of body movements,
- systematic biases due to the hardware and calibrating process of the eyetracking system.

Bearing these in mind, a revisal using the original 2D-overlay-video is essential should the identification of a single fixated object be the primary intension of using the virtualised eyetracking data in the 3D environment. In the event of dynamically analysing tendencies, the described procedure remains applicable yet, because deviations as stated above do not diminish the proceeds of a possibility to virtually analyse a human's view based on the frequency of occurrence for segmented areas of vision.

Dependant on the distribution of intersections, different levels of importance can be allocated to specific grid structures by assigning different weighting values. Such a parameterising of areas of vision by objectively acquired weighting factors delivers valuable information for a process-oriented engineering design. The combination of virtual lines of vision with the procedure to determine a viewing distance further enables to integrate modern devices featuring semi-transparent visual interfaces (HUDs or HMDs) into analysing methods.

In the context of determining viewing distances based on vergence eye movements, almost parallel arrangements of the axes of vision for increasing

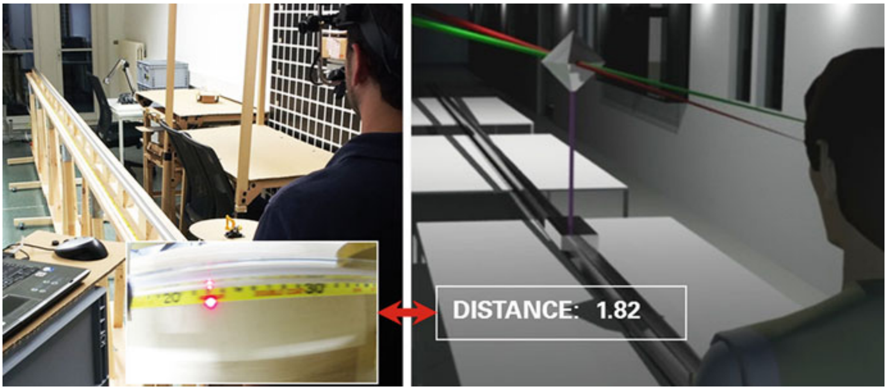


Fig. 11 Experimental setup to determine critical ranges and associated tolerances for computed distances to objects that are moved along the viewing direction, based on eyetracking data

distances are a well-known effect. However, literature research ended without results for assigning maximum distances to dynamically backwards and forwards moving objects, causing "smooth-persuit-like" vergence eye movements. For this reason, an experimental setup was developed (Fig. 11).

The setup basically consists of a linear rail with an electric motorised carriage with a mounted viewing target to be fixated by the test person. A camera, located on the carriage as well, records a spot on a distance scale highlighted with a laser pointer device, providing information about the "real" distance of the viewing target to the eyetracked person. An integrated reed sensor triggers a flashlight that is recognized by both, the camera on the carriage and the scene-camera of the eyetracking system. This optical signal serves as a time stamp signal to synchronise the real distance with the calculated distance computed within the virtual 3D-environment.

First tests with this setup show that the experimental design not only allows to empirically determining critical distances and adequate tolerances applicable to calculated distances, but also seems to be capable to provide findings concerning influences of distinct eye dominances on vergence eye movements.

Changes of the bonus scheme of the German Social Accident Insurance Institution for the raw materials and chemical industry (BG RCI) in 2014 redefine costs eligible for funding. Accordingly, only retrofitted rear view camera systems for prior machines benefit from refunds by the BG RCI whereas rear view systems are supposed to be a standard instrumentation for new machines [1].

Furthermore, the placing of additional components that could further limit the field of view of earth moving machines like diesel exhaust-gas after-treatment systems, are currently discussed as these are mandatory in order to keeping compliant to tightened exhaust emission laws [13].

These are aspects which indicate the growing demand for modern kinds of human machine interfaces for assistance systems in mobile machines. In the first instance, the increasing pertinence of having eyetracking data available in 3D design and development tools help to embed these into process-oriented concepts with virtual prototypes.

References

1. Berufsgenossenschaft Rohstoffe und chemische Industrie (2014) Das Prämiensystem der Branche Baustoffe Steine Erden 2014. http://www.bgrci.de/fileadmin/BGRCI/Downloads/DL_Praevention/Praemiensystem/RZ_Finale_Praemiensystem_Broschuere_2014_low.pdf. Accessed 11.11.14
2. Breeze J (2009) You look where they look – People look at other people's faces. http://usableworld.com.au/2009/03/16/you-look-where-they-look/. Accessed 03.06.14
3. Bürkle K (o.J.) TU Dresden: Methode zur nutzergruppenabhängigen, dynamischen Sichtuntersuchung bei mobilen Arbeitsmaschinen anhand virtueller Prototypen. Entwurf zur Dissertation
4. Bürkle K, Hoske P (2013) Method to determine the operator's view with virtual prototypes regarding dynamic process- and user-characteristics. Fachbeitrag und Tagungsunterlagen zur Automotive Seating Reloaded 2013. we.CONECT Global Leaders GmbH (Veranstalter), Berlin
5. Clemens R (2014) Arbeitsgemeinschaft Netzwerk Gesundheit und Qualifikation: Unfälle mit Baumaschinen. http://www.ral-bau.eu/unfälle-mit-bm-2014/. Accessed 15.09.14
6. Grandjean E (1979) Physiologische Arbeitsgestaltung. Ott Verlag, Thun
7. ISO/TC 127 Sc 2, ISO Technical Committee Safety, ergonomics and general requirements (2006) Erdbaumaschinen - Sichtfeld - Testverfahren und Anforderungskriterien. ISO 5006:2006
8. Leisering H (2013) Aktuelle Entwicklungen im Straßen- und Tiefbau. Braunschweig: Verein Deutscher Revisions-Ingenieure e.V. (VDRI) B. d. Bauwirtschaft (Hrsg.) 33–36
9. Schlick CM, Bruder R, Luczak H (2010) Arbeitswissenschaft. Springer, Berlin
10. Schweigert M (2003) TU München: Fahrerblickverhalten und Nebenaufgaben. Dissertation. http://mediatum.ub.tum.de/doc/601886/601886.pdf. Accessed 05.01.14
11. Tobii Technology GmbH (2011) Eye Tracking brings clarity to 'driving without awareness'. http://www.tobii.com/eye-tracking-research/global/library/customer-cases/cognitive-psychology/attitudes/. Accessed 01.08.14
12. Wolfram Alpha LLC - A Wolfram Research Company (2014) WolframAlpha computational knowledge engine. http://www.wolframalpha.com/. Accessed 07.02.2014
13. Directive 97/68/EC of the European Parliament and of the Council of 16 December 1997 on the approximation of the laws of the Memmber States relating to measures against the emission of gaseous and particulate pollutants from internal engines to be installed in non-road mobile machinery

Experimental Comparison of Sidestick Steering Configurations for an Innovative Electric Two-wheel Vehicle

Benjamin Strenge, Stefan Sieburg, and Ludger Schmidt

Abstract
In the E2V project we participated in the development of an innovative electric two-wheel vehicle for closed, non-urban spaces like parklands. In order to enable usage by handicapped and persons without a driving license, the vehicle was designed to be steered with a sidestick which was placed centrally between the two passenger seats. The active sidestick allowed for haptic feedback, e.g. warning the driver when the road is left. Using an OpenDS-based driving simulator, we experimentally compared four different sidestick configurations to identify a secure, efficient and comfortable steering concept. A longitudinally isometric configuration turned out as most popular, but the actual differences were insignificant.

Keywords
Human-machine interaction • Ergonomics • Usability • Side sticks • Assistance • OpenDS • Simulator

1 Project Background and Goals

The main objective of project E2V was the development of an innovative compact, versatile and lightweight electric two-wheel vehicle. The project was funded by the German Federal Ministry of Education and Research (BMBF) from 2011 to 2014. In total, five chairs from the Mechanical Engineering department and the Electrical Engineering and Computer Science department of the University of Kassel as well as six partners from industry participated in the project. The E2V *(Electrical*

B. Strenge • S. Sieburg • L. Schmidt (✉)
Human-Machine Systems Engineering, University of Kassel, Kassel, Germany
e-mail: b.strenge@uni-kassel.de; L.Schmidt@uni-kassel.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_21

Explorer Vehicle) was specifically designed for deployment in closed spaces that prohibit the use of conventional road vehicles, e.g. due to space constraints, tourism, or environmental protection. Examples of such spaces are town centers, traffic-free residential areas, parks and cultural heritage sites, whose infrastructure must remain unchanged. In these settings the inherent advantages of electromobility can be leveraged effectively without suffering from disadvantages like limited reach. The intended test area for the project was the *Bergpark Wilhelmshöhe* near the city of Kassel. This mountain landscape park was inscribed as a UNESCO World Heritage Site on June 23rd 2013. The park is physically demanding to explore by foot, encompassing an area of 2.4 km^2 and a height difference of 240 m between its most popular sights, castle *Schloss Wilhelmshöhe* and the Hercules statue [3]. In this context, the E2V project could be regarded as a provision to ensure the park's accessibility in the long term. Our task within the project was to conceptualize, design and evaluate the vehicle's human-machine interfaces in order to ensure usability and user acceptance [1]. This included the touristic information and navigation systems [9], the interior and the vehicle control elements. The overall goal of the study presented in this article was to identify a secure, efficient and comfortable steering concept for the E2V.

2 General Vehicle and Sidestick Properties

The E2V had only one axis with two wheels. Cornering was therefore not realized via modification of the wheel angle. Instead, different rotational speeds were applied to each wheel. This principle enables turning on the spot as known from tanks and other track vehicles. Small additional support wheels were used in parking position. The maximum speed of the E2V was limited to 10 km/h due to security and performance reasons.

As of today, road vehicles for private usage typically still use pedals, gear lever and a steering wheel for steering. These control elements historically resulted from mechanical coupling rather than from ergonomic considerations [8] and aimed at coping with high positioning forces which are not required anymore [12]. The E2V was instead designed to be controlled one-handedly with an active sidestick in order to facilitate its operation for disabled persons, e.g. paraplegics, and users who do not possess a driver's license. The sidestick was placed centrally between the two passenger seats. Left-handed drivers could accordingly occupy the right seat and use their preferred hand to gain equal comfort and control. Sidesticks have several advantages over the traditional vehicle steering concept. The removal of the steering wheel reduces the risk of injury and improves the driver's view field [10]. Furthermore, the dimensionality of the input device matches the dimensionality of the driving task [7]: The vehicle's longitudinal dynamics are controlled via just one sidestick axis instead of with three to four different control elements (gas pedal, brake pedal, gear lever and possibly clutch pedal). This also avoids the counterintuitive act of pressing a (brake) pedal forward to make the vehicle stop [cf. 11].

Because control signals are transmitted electronically between sidestick and vehicle, there is no mechanical coupling anymore. To be able to nonetheless produce haptic feedback and compare different input concepts we chose an active sidestick (Stirling Dynamics Next Generation Interceptor). This sidestick could recognize not only its position but also any forces applied to it. Conversely, the stick could be prompted to change its inertia and attenuation, and create artificial force barriers (so-called soft stops and hard stops) or arbitrary other forces. By these means it is for example possible to give the driver haptic feedback when near obstacles are approached or the road is left. This enabled the simulation of the assistance system described in Sect. 3.5.

3 Sidestick Configurations

Based on the E2V's specific properties and literature research regarding reputed advantages and disadvantages of comparable steering concepts, we defined different potentially feasible configurations for the active sidestick.

As remarked e.g. by Schieben et al. [8], the absence of mechanical coupling initially leads to a reduced information flow between driver and vehicle. An approach to compensate for this is position reflective feedback: An artificial haptic feedback system recognizes the system status, e.g. about the current wheel angle and speed, and determines the sidestick position accordingly to inform the driver, which is analogous to the human experience of direct interaction with a physical object [5, 8]. However, as mentioned above, the E2V as a two-wheel vehicle did not use different wheel angles. Furthermore, the E2V's acceleration was high in relation to its maximum speed of only 10 km/h. The longitudinal dynamics could therefore be regarded as a quasi-binary dimension (either standing or driving with maximum speed). An actual position reflective feedback was not reasonably applicable under these conditions. However, we seized the basic idea of aligning the sidestick's behavior with the vehicle's response characteristics (see configurations C and D below).

All four compared configurations are subsequently described in detail, as well as a shared assistance system that was used concertedly with all these configurations. Figure 1 serves as an illustration of the most important characteristics of each

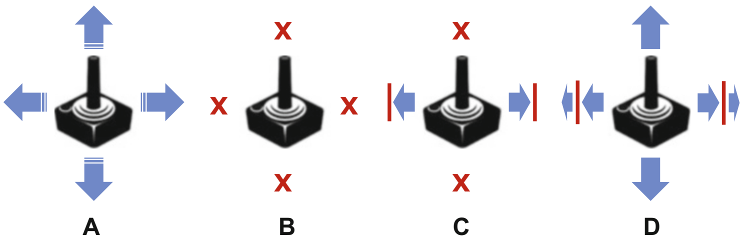


Fig. 1 Illustration of compared sidestick configurations

configuration. Technically, the respective sidestick parameters were set via the software *NGI Communication Setup* by Stirling Dynamics. As the E2V did not have a reverse, for all configurations held that pushing the stick forward caused acceleration while pulling backward or releasing in central position meant slowing down to standstill.

3.1 Configuration A

This configuration allowed moving the stick in both axes while a constant spring force pushed the stick towards its central position. Inertia and attenuation were low.[1] There were no hard stops or soft stops. Therefore the stick could be deflected to the physical maximum of 25° at both axes. The vehicle's acceleration and steering were determined proportionally to the stick's deflection and not modified by any other influences. The resulting behavior was similar to that of a common computer joystick. Because of the low inertia, the physical strength required for usage was low.

3.2 Configuration B

Steering concept B was isometric on both axes, i.e. the stick could not be moved visibly. Technically this was realized by creating hard stops around the central position. To determine the acceleration and steering, the longitudinal and lateral forces applied to the stick by the driver were used instead of the stick's deflection. This approach is supposed to lead to a more comfortable arm position, because the friction-like effects between armrest and forearm are reduced [2]. The simulated vehicle was programmed to respond proportionately to the forces applied to the stick up to a maximum of 30 N. Stronger forces had no further effect.

3.3 Configuration C

This concept was isometric on the longitudinal axis only. As in configuration B, applied forces (on the longitudinal axis) up to 30 N were taken into account. The constant angular relationships of the upper extremities are supposed to allow for a more precise perception of the stick's deflection in lateral direction [2]. Inertia and attenuation of the lateral axis were tuned to approximately correspond to the simulated vehicle's response characteristics.[2] The adjusted inertia led to a

[1] NGI software parameter values: Inertia laterally = 0.005, damping laterally = 0.05, inertia longitudinally = 0.005, damping longitudinally = 0.55. The higher damping value on the longitudinal axis prevented self-oscillation of the stick in central position.

[2] NGI software parameter values: Inertia laterally = 0.7, damping laterally = 0.7.

comparatively high degree of "stiffness" of the sidestick. Because of this, artificial force barriers were used on the lateral axis restricting lateral deflection of the stick to a maximum of 15° in order to spare the driver from unnecessary physical strain and to avoid fatigue. Due to technical limitations fixed hard stops had to be used for this purpose while overridable soft stops (like in configuration D) might have been theoretically preferable.

3.4 Configuration D

Like variant A, this configuration allowed moving the stick on both axes. However, the approach from concept C to set inertia and attenuation to roughly match the vehicle's behavior was applied to both axes here.[3] This lead to a stiffer steering compared with concept A. Soft stops were used at 15° on both axes to prevent the driver from unnecessary oversteering. The soft stops could be overridden by applying forces greater than 30 N. This was a relatively low value because the force barrier was intended to have the character of a recommendation rather than a restriction.

3.5 Assistance System

All of the above configurations (A–D) were complemented by the same assistance system: Whenever the vehicle left the road, the sidestick vibrated like driving on a "virtual gravel bed" [cf. 4] in order to warn the driver. This feature should ensure that the E2V was used on accessible roads only, contributing to the safety of passengers and the park landscape. The vibration was sinusoidal at a frequency of 15 Hz and amplitude of 20 N. It was applied on the lateral axis to indicate that the driver must respond mainly on this axis.

4 Experimental Design

This section describes how the four different sidestick configurations were evaluated experimentally.

4.1 Simulator Hardware

The frame of the E2V driving simulator (Fig. 2) was constructed of components from test productions for the actual vehicle prototype to enhance the immersion.

[3] NGI software parameter values: Inertia laterally = 0.7, damping laterally = 0.7, inertia longitudinally = 0.5, damping longitudinally = 0.5.

Fig. 2 E2V simulator in preparation

Fig. 3 E2V simulator interior and sidestick

This included the metallic car body and several plastic cover parts. As in the actual vehicle, the driver's field of view is partially occluded by the left front screen strut. Armrest and footrest frames were made of aluminium profiles. Two touchscreens were attached to the armrest as an interface to a touristic information and navigation system, but unused (i.e. turned off) during the test sessions in this study. The Next Generation Interceptor sidestick from Stirling Dynamics Ltd. was also embedded into the armrest (Fig. 3). The trigger and buttons on the sidestick had no function.

The sidestick's position (in longitudinal direction) and the height of the armrest could be adjusted to meet different body measurements. Deviating from the actual E2V design for practical reasons, there was only one driver's seat in the simulator (on the left side), which was taken into account in the experimental design.

Above the vehicle frame an Epson EH-TW8100 projector with a resolution of 1920×1080 was installed to project the driving simulation on the wall in front of the driver. Below the frame three loudspeakers were placed and aimed at different directions to create a diffuse sound field, playing back an artificial electric engine noise on acceleration, ambient sounds (chirping), and a scream sound on collision with human visitor models.

4.2 Simulation Software and Modeling

The simulation was based on OpenDS, an open source driving simulator software created for scientific purposes and distributed under the GNU GPL. This allowed us to make arbitrary changes to vehicle parameters and the general software behavior where required. Most notably the E2V's limited maximum speed, absence of reverse, the ability to turn on the spot and other relevant characteristics were implemented. Further adjustments had to be made to the OpenDS trigger system to simulate the assistance system and log performance data.

The simulated environment was a representation of the *Bergpark Wilhelmshöhe* based on freely available as well as self-made models and data of the park (Fig. 4). The test track ran partially parallel to the hill and partially downhill with variable breadth between two famous sights, from the Hercules cascades to castle *Löwenburg*. Except for one T intersection, the road course contained only obtuse

Fig. 4 Screenshot of the simulated park environment

angles. The environment off the road consisted mostly of trees and some landmarks. At road forks and branches, 3D objects of road-blocking fences striped in red and white have been placed in order to guide the driver to the intended way. Furthermore, a total of 37 static 3D objects representing individuals and small groups of human visitors were placed on and near the test track. Their positions were determined such that frequent deviation from the racing line was necessary to avoid collisions. A few of the visitor models were placed on the wayside where they did not constitute obstacles, but increased the level of immersion. The placement of all models, including human visitors, remained unchanged for all test sessions and trials.

4.3 Operationalization in Terms of Usability Research

ISO 9241-11:1998 defines usability as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Here the products were the different sidestick configurations embedded in the overall vehicle system. The goal was specified as driving from one position to another as fast as possible without causing any damage to people or the environment. Efficiency was therefore measured by the driving time related to the effectiveness at getting from the starting point to the finish, staying on the road, and avoiding collisions. Satisfaction is usually measured by asking users for some kind of subjective assessment, in this case regarding the sidestick configurations and the assistance system. The context of use comprised the simulated vehicle within the simulated ambience of Bergpark Wilhelmshöhe including human visitors and was devised to approximate evidently relevant aspects of the actual vehicle's context of use.

4.4 Participants and Test Procedure

We used a within-subjects design, i.e. each participant consecutively used each of the four sidestick configurations. A total of $N = 24$ participants were recruited (21–58 years old, $\varnothing = 30.1$ years, 17 males, 7 females; Fig. 5), which enabled us to use each of the $4! = 24$ permutations of the four configurations exactly once. This equals complete counterbalancing to avoid systematic bias due to order effects like learning [cf. 6]. Participants were naïve with respect to the assignment of configurations. They were also not given any information regarding the configurations' specific properties to mitigate the Pygmalion effect.

At the beginning of a test session the participant was asked to take a seat in the E2V driving simulator and close the frame. The participant was then introduced to the simulation environment and hardware features. The task given to the participants demanded that the predefined test track be finished as quickly as possible while avoiding any collisions with visitors and staying on the road at all times. Then the first trial was started. After a short intro track, logging started on the

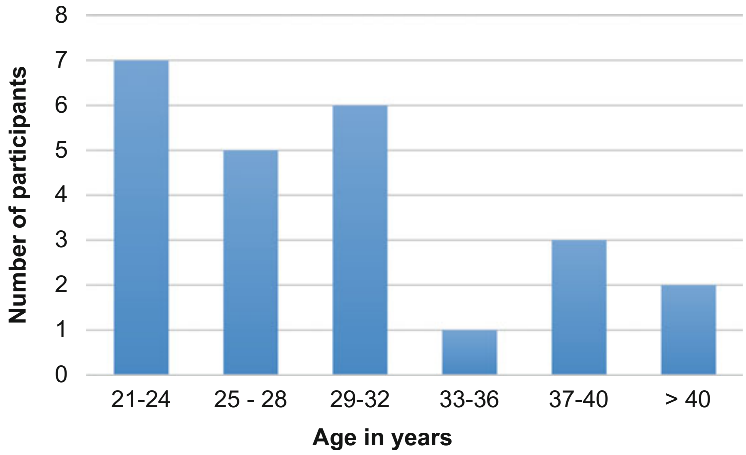


Fig. 5 Participants' age distribution

road at the level of the Neptune basin (a landmark below the Hercules cascades) which was visible from the starting position. On collision with a visitor an audio file of a neuterly sounding scream was played. Additionally, the sidestick vibrated with an amplitude of 20 N at 15 Hz in longitudinal direction to indicate the crash by haptic feedback. A single faultless run required circa 4:50 min as determined by one of the investigators. When the finish was reached, this was signaled by a window popup. The number of collisions, times the road was left, and driving time from start to finish were recorded. The vehicle position in the simulated environment was then reset to the start and another sidestick configuration was loaded.

After trials two, three and four, the participant was asked to give a rank order based on their subjective preferences regarding the steering configurations they had used until then. This was realized by ordering cards, each representing a specific configuration, on a table in the simulation lab. The previous rank order could be changed at will after each consequent test run. The final rank order after all four test runs was retained for analysis by the investigator.

At the end, a questionnaire was used to retrieve demographic data (age, sex), handedness, possession of a driving license, and a subjective assessment of the simulated assistance system. The latter aimed at both the perceived strength and usefulness of the haptic feedback. The strength was hereby specified on a bipolar scale with five levels from "much too weak" through "just right" to "much too strong". Usefulness was measured on a unipolar scale with four levels from "not helpful at all" to "very helpful".[4]

[4] All level descriptions are translated from German here.

5 Results and Observations

The results of the objective measurements are shown in Table 1. When the road was left for only less than 2 s, this was not counted here, because the assistance system warned the user accordingly and the driver was obviously able to almost instantly return on the road (cf. Fig. 6).

Boxplots of driving times and number of collisions are shown in Figs. 7 and 8, respectively. As determined by two-tailed paired t-tests at $\alpha = 5\,\%$, no statistically significant differences existed between any two configurations regarding the average driving time, number of collisions, or number of times the road was left.

Two out of the 24 participants were left-handed (according to their own statement) and had to use the sidestick with their disfavored hand; nevertheless, their performances were comparable to those of the right-handed participants. This also holds for the four participants without a driving license and the seven females when compared with the respective complementary groups. As one would expect from the small subgroup sizes, there were no significant differences.

The results of our participants' subjective ratings are illustrated in Fig. 9. Configuration B was rated worst (4th) most often while configuration D was most

Table 1 Objective results wrt. driving time, number of collisions and times the road was left

	Minimum	Maximum	Mean	SD
Driving time A (s)	286.4	371	300.8	18.6
Driving time B (s)	290.4	407.1	306.2	26.8
Driving time C (s)	290.4	357.2	303.5	17.6
Driving time D (s)	289.1	522.1	310.6	49.3
Number of collisions A	0	4	0.92	1.1
Number of collisions B	0	5	1.13	1.26
Number of collisions C	0	4	0.71	1
Number of collisions D	0	3	0.75	1.15
Times road left $\geq$2 s A	0	7	1.04	2.05
Times road left $\geq$2 s B	0	13	1.5	3.11
Times road left $\geq$2 s C	0	6	1.08	1.79
Times road left $\geq$2 s D	0	6	1.13	1.8

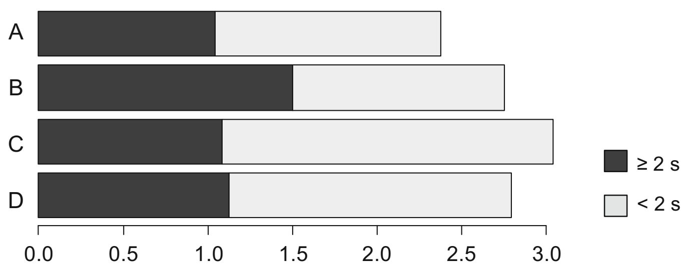


Fig. 6 Avg. number of times the road was left

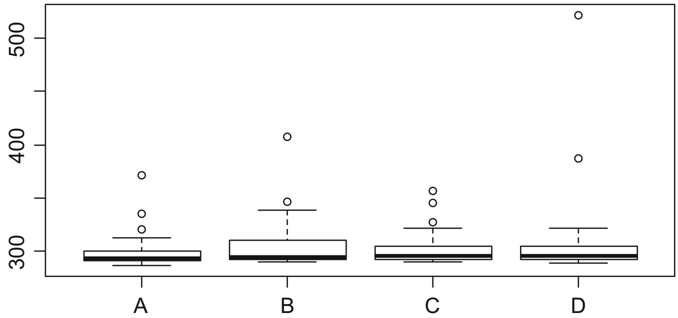


Fig. 7 Driving time (s)

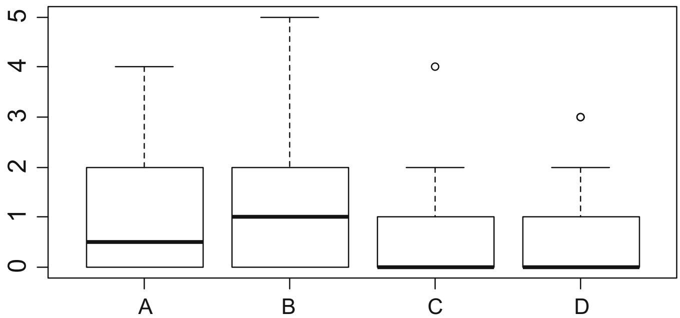


Fig. 8 Number of collisions

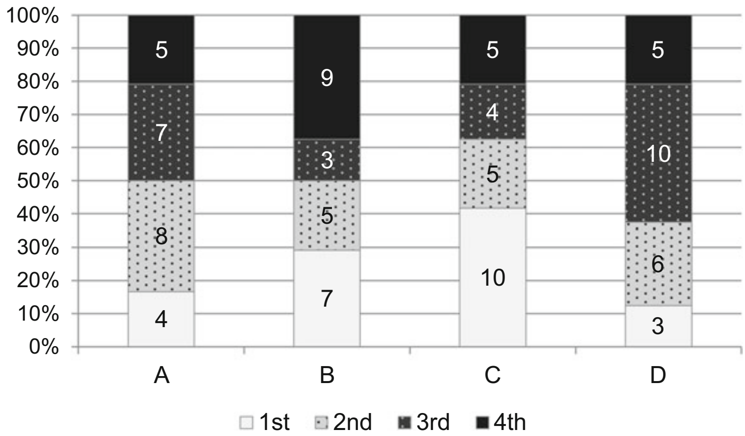


Fig. 9 Participants' subjective ratings of sidestick configurations

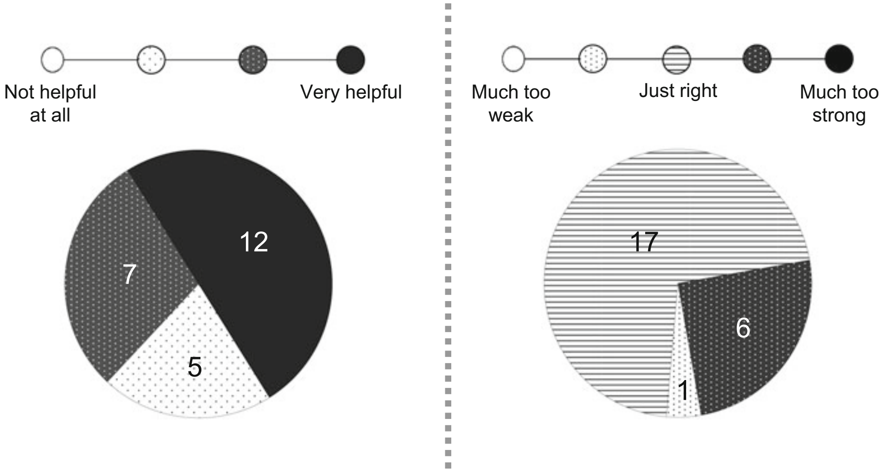


Fig. 10 Participants' assessment of the assistance system's helpfulness (*left*) and intensity (*right*)

often rated worst or second-worst (3rd + 4th). According to incidental statements by some participants, their bad rating of configuration D was due to its distinct stiffness. Configuration C was rated best (1st) as well best or second-best (1st + 2nd) most often. However, χ^2 tests showed that neither the distributions of the ranks on each configuration nor the distributions of the configurations on each rank deviated significantly from uniform distribution ($\alpha = 5$ %). Therefore, as requesting rank orders can be considered a forced choice and assuming random tiebreaking on the part of participants, it cannot be ruled out that these results have emerged by chance.

The assistance system was judged fairly helpful (29.2 %) or helpful (50 %) by the majority of participants (cf. left side of Fig. 10). Nobody rated the system as not helpful at all.[5] The feedback intensity was rated just right (70.8 %) or slightly too strong (25 %) at the abovementioned values (cf. right side of Fig. 10). The most negative ratings of much too strong and much too weak were not given by anyone.

6 Discussion and Conclusion

A notable limitation of the study is foremost the range of considered configurations. This was aimed at the most promising discriminable variants for the context of use and enabling a practicable experimental design, but a myriad of other flavors were theoretically possible. Additionally, for practical reasons the adjustment of inertia and attenuation values was conducted merely by trial and error instead of by

[5] All participants caused warnings by the assistance system in at least one trial, enabling them to make an informed statement.

systematic derivation from theory or measurements. It may very well be that meticulously fine-tuning these values would lead to more definite results. Furthermore, while the E2V was primarily intended to be used by the special group of elderly and disabled, no such constraints were made in the selection of participants. Due to the broad range of disabilities and increased deviation in the performance of elderly (compared with younger persons), an exclusive participation of such persons under the constraint of a strictly limited sample size was presumed to lead to a tremendous amount of variability in the data and therefore unusable results. Attention should also be paid to the fact that our task was to identify an appropriate sidestick steering concept for the specific properties of the E2V, a novel electric two-wheel vehicle. Obviously, the results should not be generalized to other vehicle types with levity.

As shown in Fig. 8, when using configuration C the road was left most often in total, but comparatively seldom for more than 2 s. Slightly simplified this says that upon leaving the road, the driver was able to return on it quickly in most cases. A possible reason for this may be that the assistance system's informative vibration was applied to the sidestick's lateral axis. This was also the axis on which the drivers had to countersteer and at the same time the only non-isometric (i.e. movable) axis of this configuration. This might suggest that sidestick configuration C works particularly well in combination with this type of informative assistance system. For actual usage the assistance system would have to be adjusted to already initiate the warning when the vehicle is still on the road, but anticipated to leaving it shortly, in order to prevent accidents.

While no statistically significant differences were identified between the sidestick configurations with the sample size of $N = 24$, the trends in the objective data as well as the subjective ratings provide some support for related works from Eckstein [2] and Schieben et al. [8]: The most popular configuration C had an isometric longitudinal axis and the lateral axis was adjusted to approximately correspond to the simulated vehicle's response characteristics. Interestingly, no participants made negative remarks regarding the stiffness of configuration C's lateral axis, although inertia and attenuation were objectively just as high as those of complained-about configuration D. This may suggest that the isometric longitudinal axis of configuration C led to a more comfortable lateral steering experience as postulated by Eckstein [2]. Participants also tended to cause the least collisions when using this configuration. Overall these observations may hint at a minor advantage of this type of configuration over other sidestick steering approaches. However, stronger evidence would be necessary for a definite conclusion.

References

1. Braun R, Schmidt L (2013) Benutzerorientierte Entwicklung eines für spezifische Nutzergruppen adaptierbaren, einachsigen, teilautonomen Elektrofahrzeugs. In: Verband der Elektrotechnik Elektronik Informationstechnik e. V (Hrsg.) Lebensqualität im Wandel von Demografie und Technik: 6. Deutscher AAL-Kongress (Berlin 2013). VDE-Verlag, Berlin, pp. 321–325

2. Eckstein L (2000) Sidesticks im Kraftfahrzeug - ein alternatives Bedienkonzept oder Spielerei? In: Bubb H (Hrsg.). Ergonomie und Verkehrssicherheit. (München 2000). GfA Herbstkonferenz. GfA, München, pp. 65–96
3. Hoss, S. (Mitarb.) (2013) Park Wilhelmshöhe—größter Bergpark Europas. 1. Aufl. Regensburg: Schnell + Steiner, (Parkbroschüren MHK Bd. 1)
4. Kelsch J, Flemisch FO, Schieben A, Schindler J (2006) Links oder rechts, schneller oder langsamer? Grundlegende Fragestellungen beim Cognitive Systems Engineering von hochautomatisierter Fahrzeugführung. In: Grandt M, Bauch A (Hrsg.) Cognitive Systems Engineering in der Fahrzeug- und Prozessführung. 48. Fachausschusssitzung Anthropotechnik. DGLR-Bericht, pp. 227–240
5. Kienle M, Damböck D, Kelsch J, Flemisch F, Bengler K (2009) Towards an H-mode for highly automated vehicles: Driving with side sticks. In: Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications AutomotiveUI '09. ACM, New York, NY, pp. 19–23
6. Martin DW (2008) Doing psychology experiments, 7th edn. Wadsworth, Belmont, CA
7. Penka A (2001) Vergleichende Untersuchung zu Fahrerassistenzsystemen mit unterschiedlichen aktiven Bedienelementen. Utz, München
8. Schieben A, Damböck D, Kelsch J, Rausch H, Flemisch F (2008) Haptisches Feedback im Spektrum von Fahrerassistenz und Automation. In: Proceedings of 3. Tagung Aktive Sicherheit durch Fahrerassistenz, Garching
9. Strenge B, Schmidt, L (2015) Touching problems—Formative Evaluationen eines touristischen Informationssystems für ein neuartiges Elektrofahrzeug. In: Gesellschaft für Arbeitswissenschaft e. V. (Hrsg.) VerANTWORTung für die Arbeit der Zukunft: 61. Kongress der Gesellschaft für Arbeitswissenschaft (Karlsruhe 2015). GfA-Press, Dortmund, pp. 1–6 (C.1.13)
10. Weber B, Panzirsch M, Preusche C, Deml B (2011) Der aktive Sidestick als Bedienelement im Elektromobil. In: Gesellschaft für Arbeitswissenschaft e. V. (Hrsg.) Mensch, Technik, Organisation—Vernetzung im Produktentstehungs- und -herstellungsprozess: 57. Kongress der Gesellschaft für Arbeitswissenschaft (Chemnitz 2011). GfA-Press, Dortmund, pp. 473–476
11. Weber B, Panzirsch M, Montoya-Zamarron S, Preusche C, Deml B (2011) Evaluation zweier Bedienkonzepte zur Steuerung eines Fahrzeugs mit drei Bewegungsfreiheitsgraden. In: Grandt M, Schmerwitz S. (Hrsg.) Ergonomie im interdisziplinären Gestaltungsprozess: 53. Sitzung des Fachausschusses Anthropotechnik der Deutschen Gesellschaft für Luft- und Raumfahrt. DGLR, Bonn, pp. 175–186
12. Winner H, Hakuli S, Bruder R, Konigorski U, Schiele B (2006) Conduct-by-Wire-ein neues Paradigma für die Weiterentwicklung der Fahrerassistenz. In: Freundeskreis Mess- und Regelungstechnik Karlsruhe e.V. (Hrsg.) 4. Workshop Fahrerassistenzsysteme (Löwenstein 2006). Fmrt, Karlsruhe, pp. 112–125

Development and Evaluation of an Input Method Using Virtual Hand Models for Pointing to Spatial Objects in a Stereoscopic Desktop Environment with a Fitts' Pointing Task

Ronald Meyer, Jennifer Bützler, Jeronimo Dzaack, and Christopher M. Schlick

Abstract

Virtual Reality for desktop or table mounted display environments can be a strong tool in supporting users in sighting spatial information and give a better understanding in their spatial relationships. Visualization of spatial data incorporates spatial interaction and manipulation with six degrees of freedom for data processing which is part of current research objectives. Modern tracking devices for desktop applications allow capturing a model of the human hand with three translational and three rotatory degrees of freedom which can be used as an input system allowing the control over a virtual hand model in a stereoscopic desktop system. Interaction with complex data in rich environments using virtual hand models, such as three-dimensional geographic information systems with integrated spatial data, requires visualization with a low degree of occluding virtual entities in the virtual scene but a high degree of recognition. An experimental task using a Fitts' pointing task is conducted to investigate different minimalistic modeling approaches of virtual hand models using the Leap Motion Controller in a stereoscopic desktop environment. Movement times are analyzed for pointing movements in three different depth layers. Results indicate that a kinematic visualization of a virtual hand model produces significantly better movement times in pointing over a minimalistic point cloud model.

Keywords

Virtual hand model • Fitts' pointing task • Natural user interaction • Freehand interaction • Desktop virtual reality

R. Meyer (✉) • J. Bützler • C.M. Schlick
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany
e-mail: r.meyer@iaw.rwth-aachen.de; c.schlick@iaw.rwth-aachen.de

J. Dzaack
Atlas Elektronik GmbH, Sebaldsbrücker Heerstr. 235, 28309 Bremen, Germany
e-mail: jeronimo.dzaack@atlas-elektronik.com

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_22

1 Introduction

Spatial data visualizations in virtual reality become more relevant with more accurate input methods that are capable of providing at least three rotatory and three translational degrees of freedom (DOF). Displaying data in virtual reality is a strong tool to immerse users into scenarios where presence, i.e. the feeling of “being there”, can be used to raise a user’s situational awareness [5]. Stereoscopic display system applications allow virtual reality for desktop use. A stereoscopic display supports a user’s cognition in spatial relations when spatial data is displayed by featuring additional depth perception. Stereoscopically visualized datasets of magnet resonance imagery for example can strongly support the interpretation of live spatial data recorded during magnet resonance guided interventions such as giving overviews of organ positions before surgical operations. Brunner et al. [7] used a prototypical setting where a commercial stereoscopic display was used to visualize data of a magnet resonance image scanner to track the orientation of a catheter during the insertion procedure. Ferrari et al. [8] describe an experimental setting where a head mounted mixed reality system is used to support physicians to see preoperative radiological exams in a natural, three-dimensional way. For this, the system creates augmented live image data by blending previously taken volumetric radiological images with real patient live images taken by a camera. Kockro et al. [11] compared a stereoscopic display system as a navigational tool for surgery planning of patient data which was previously recorded through magnet resonance imagery to a classical standard display system. The Results show that surgeons clearly prefer the stereoscopic visualization setting which included a virtual tool rack for interaction with the data. The depth perception improved the understanding of manipulations and helped planning the surgical strategy while reducing the need of guessing work (e.g. guessing the adjustment of human organs) during surgery. This approach promises better performance of surgeries.

Baier and Zimmer [2] conducted experiments using stereoscopic displays where both experts and non-experts were compared in their performance in estimating altitudes of airplanes in an air traffic control environment. An effect for a better performance of non-experts in using the stereoscopic environment was measured and the number of false alarms was reduced when using the stereoscopic system. Baier and Zimmer [2] emphasize the benefits of the system when the air traffic controllers process information on different channels—e.g. when incoming radio messages have to be processed together with airspace information. These findings indicate an improved cognitive processing of spatial data of user’s by visualizing the data in its correct spatial relation.

Spatial interaction with such data requires precise and robust methods for three-dimensional navigation and manipulation. The current market provides several end user products such as 3D mouses, data hand gloves or the recently released hand tracking device Leap Motion, which operates with visual tracking and therefore without the need to wear gloves with inertia or gravity sensors. These devices allow control of a virtual cursor in the virtual world providing at least three translational devices of freedom or letting the user control a distantly controlled virtual hand.

1.1 Stereoscopic Visualizations and Geographic Information Systems

Stereoscopic visualizations of terrain and map data becomes relevant when the requirement of blending three-dimensional environmental data into spatial map data which can rise situational awareness in e.g. emergency situation scenarios. Mostly, geographic information software (GIS) is used to provide situational overview and awareness. Zlatanova et al. [20] describe a variety of applications that emerge with the usage of three-dimensional GIS that are not solely created for visualization and better understanding of spatial relationships but also for manipulation and creation of data in three dimensions as well as a tool to navigate through the large spatial models in real time. Corresponding systems are used as desktop work environments and are applied for e.g. landscape and infrastructure planning, resource exploitation planning or all other tasks where query and reasoning or measurement tasks are required [9]. The applications of three-dimensional geographical information systems become more diverse with linking live data feeds of sensors to a corresponding system. Especially in the maritime area, where different spatial sensors, subsurface or underwater e.g. echo sounder sensors, or radio sensors such as radar, are applied. A mixed virtual environment of three-dimensional real time data can be a benefit to support a user with additional contextual awareness and improved situational overview. Scenarios for the application of a three-dimensional GIS provided with real-time data for maritime use case scenarios have been developed by Meyer et al. [13]. The three main foci are:

1. Harbor maintenance: Protection and monitoring of harbor areas is generally performed from a central control room, where all sensor data is collected and evaluated. Spatial data of echo sounders or autonomous underwater vehicles can be sighted in a proper spatial ratio.
2. Offshore wind park maintenance: Protection and maintenance of wind parks is usually accomplished by regular bathymetric scans and echo soundings of the underwater cables and concrete bases of the wind mills.
3. Bathymetric resource retrieval: Three-dimensional data sets recorded by echo sounders can be used to generate live images of the sea bed which can be integrated into a spatial mapping environment. Resource deposits can be tracked, and marked for later exploitation. A three-dimensional model of the sea bed surrounded by the deposit provides information about accessibility on a possible exploitation.

The advantages of stereoscopically visualizing three-dimensional data in these scenarios are apparent, since the data can directly be visualized in a mapping environment with its realistic scale providing the user a short cognitive distance to the system in relation to its surroundings.

1.2 Stereoscopic Desktop Environments

Current stereoscopic end user display technology which is applicable in desktop scenarios is equipped with active or passive glasses [17]. Active glasses, working with shutter technology and a high refresh rate of the display, provide full resolution in stereoscopic display mode and active occlusion per eye with a refresh rate of 60 Hz and synchronization to the refresh rate of the display, which has to operate at a full refresh rate of 120 Hz. Drawbacks of active shutter technology are the requirement of batteries, which have to be replaced or recharged after a period of use, and additional technology for LC shutters, which both increase the weight of the glasses.

Passive glasses, equipped with polarized light filters, work without batteries and do not need additional technology but have a loss of light intensity caused by the light filtering process. Another drawback is the loss of half the spatial resolution as every horizontal or vertical pixel row or column is equipped with a light filter which results in an occluded pixel row or column. Both systems suffer from the conflict of the eyes' convergence relating to the virtual object of interest to the contradictive accommodation on the actual display surface which results in visual discomfort after a period of usage [4].

1.3 Virtual Hand Models and Hand Input Devices

The human hand is the primary tool for performing most interaction tasks in the real world. Accurate tracking of a user's hand offers capabilities to give the user the ability to control a hand model in virtual space and give a natural experience of interaction with virtual entities. Therefore, virtual hand models are a promising technique to be used as direct pointing input for virtual environments. Bowman [5] marks direct input methods with six degrees of freedom such as hand tracking with control over a virtual cursor or hand model as an interaction method which provides short learning times with a trade-off versus desktop devices concerning precision comfort. The Leap Motion Controller is a robust device for tracking a user's hand in three dimensions with sub-millimeter accuracy [19]; comfort analysis and application for precise tasks are part of current research and is part of the investigation presented in this work. The Leap Motion Controller framework V2 facilitates tracking a user's hand including interpolated positions of the finger joints which allows the conversion of hand models using different visual styles and also supports hand gesture recognition [1].

Fundamental research on the usage of hand models in virtual environments as a pointing technique is predominantly conducted by using data gloves as input device [5]. Besides hand postures and hand gestures the three-dimensional virtual representation of the human hand in virtual systems has become a common interaction technique as a pointing device. Data gloves provide a large number of DOF which allows precise tracking of hand and finger positions, however, the user does have to wear gloves which might not fit well or make the user feel uncomfortable. In addition, data gloves might limit the user in her freedom to interact with a system when using classical devices like mouse and keyboard in parallel, which disqualifies data glove input as a tracking device for mixed work places in which two- and three-dimensional input have to be used likewise. The Leap Motion

Controller uses visual tracking as an input method with a stereo infrared camera and thereby enables mixed usage of classical two-dimensional input devices like mouses as the user's hands can quickly alternate between the three-dimensional interaction volume of the visual sensor and other devices for two-dimensional input without the need of changing or putting on any wearable. The Leap Motion Controller is a visual sensor and can only be operated without the hands being occluded from the sensor optics which must be regarded for a mixed application setting.

The application of a virtual hand model as input method is generally a conceptual approach for settings where the users' hands are completely blocked from their eyes, e.g. in full-immersive virtual reality settings with head mounted displays [5]. Teather and Stuerzlinger [15] showed that using a virtual hand has also benefit for stereoscopic environments: The distantly-controlled virtual hand model helps to avoid a visual conflict when using a stereoscopic display device which occurs with accommodation of the users' eyes on the display plane with the angle of divergence between the eyes does not fit to the currently perceived image [14]. The intrusion of the users' real hand with a sequent focus on it immediately breaks the stereoscopic effect and causes discomfort. Bruder et al. [6] confine the usage of distantly controlled hand models in a stereoscopic display table environment with users selecting objects in the stereoscopic environment. They used a Fitts' pointing experiment to compare mid-air selection using a distantly controlled hand model and direct selection with the users hand in a stereoscopic viewing volume. The participants were quicker in performing the selection task when using the direct manipulation technique with using their own hand but with a significantly higher error rate. Hence, the results suggest the usage of a distantly-controlled hand model for virtual scenes having a high information density or that are highly cluttered. The properties of a distantly controlled hand model should be chosen with a preferably low rate of occlusion, i.e. the visual properties of the virtual hand model should be sufficient for the users' to be clearly recognized but downgraded to a visibility state, where no virtual entities are occluded while it is operated in the virtual world. A reduction of the visual information of the virtual hand to a kinematic model, i.e. where hand motion is visualized through interconnecting lines between finger joints in different visual strength, or even a point cloud model, where only the finger joints and hand root are shown as moving points, can reduce the visual complexity to a grade where information of importance being displayed in a virtual system is not affected by occlusion. This downgrade of visual information can give additional benefit to virtual scenarios containing high information density in a virtual working space. As the reduction of visual information of the virtual hand model impedes the visual cognition of the model, an experimental design is developed which empirically investigates hand models with different grades of visual complexity concerning their pointing performance in a stereoscopic desktop environment. Our hypothesis is that a minimalistic visualization of a hand model using a point cloud or a kinematic model with interconnected lines is sufficient to enable a user to interact with virtual entities in a virtual system and cause low distraction from the user's real hand. Since the user's real hand is moved in the stereoscopic display volume to control the virtual hand model we investigate the differently modeled virtual hands on different depth levels on that the user enters the stereoscopic volume with her real hand. This interaction method is evaluated for a later integration into a virtual 3D GIS system featuring a combined view of a 3D terrain visualization and 3D data view.

2 Experimental Study

An empirical study was conducted to investigate the pointing performance by using the differently visualized hand models in a stereoscopic desktop environment. Different visualizations of virtual hand models were implemented with three different complexity states in their visual representation using the Leap Motion Controller. The pointing movement for performance measurement based on a Fitts' pointing task. Collected data was analyzed based on motion time.

2.1 Design

We used a repeated measures full-factorial design for the experiment. We defined three differently visualized hand models as independent variables in three levels (M_1, M_2, M_3, cf. Fig. 1). The hand models are based on a skeletal representation which was retrieved through real time sensor data from the Leap Motion framework. Length values of inter-joint connections were automatically adapted to the measured length of the user's hand by the automatic estimations using algorithms of the Leap Motion framework. Finger joints and hand root were designed as spherical objects with either interconnected lines or without and differing in their size. The index finger tip was reactive on collisions to active target objects in the virtual scene. Movement time (MT) of the index' fingertip was set as dependent variable. Hand movement of the users' real hand was directly conversed on the virtual hand including all fingers and finger joints.

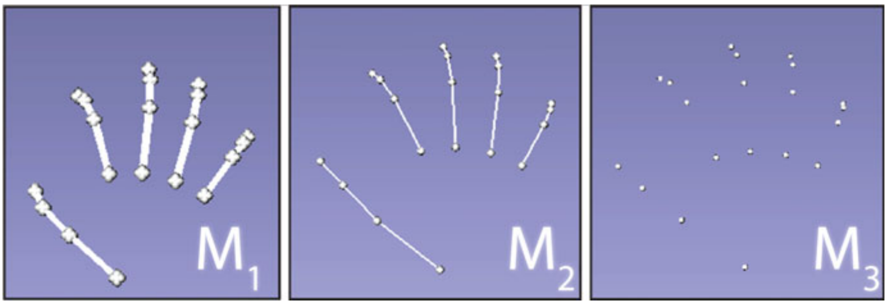


Fig. 1 Three factor hand models M1, M2, M3

The target objects were designed as spherical targets of different sizes and located at trivariate positions in their spatial alignment. We used the Shannon formulation of Fitts' Law [12] to determine an index of difficulty (ID) for the three-dimensional objects which is given by:

$$MT = a + b{*}ID \text{ where } ID = log_2\left(\frac{D}{W} + 1\right) \quad (1)$$

Factors a and b are empirically given constants that relate to a specific pointing technique [15]. The index of difficulty (ID) is measured in bits and is the crucial part of Fitts' model to determine pointing performance. Figure 2 shows the circumferential alignment of the target objects which were located on three different depth levels. The depth levels positions were with an in-between distance of approximately 100 mm each with the first depth layer lying 100 mm before the Leap Motion Device center position, the second depth layer lying directly on the center position of the Leap Motion Device and the third depth layer lying 100 mm behind the center of the Leap Motion Device as shown in Fig. 3. The pointing movement had to be conducted from an initial sphere, which was located in the center of the virtual scene with a real world position directly lying on the center position of the Leap Motion Device. All target objects were located in the upper half of Cartesian space and had a negative parallax concerning their virtual depth position.

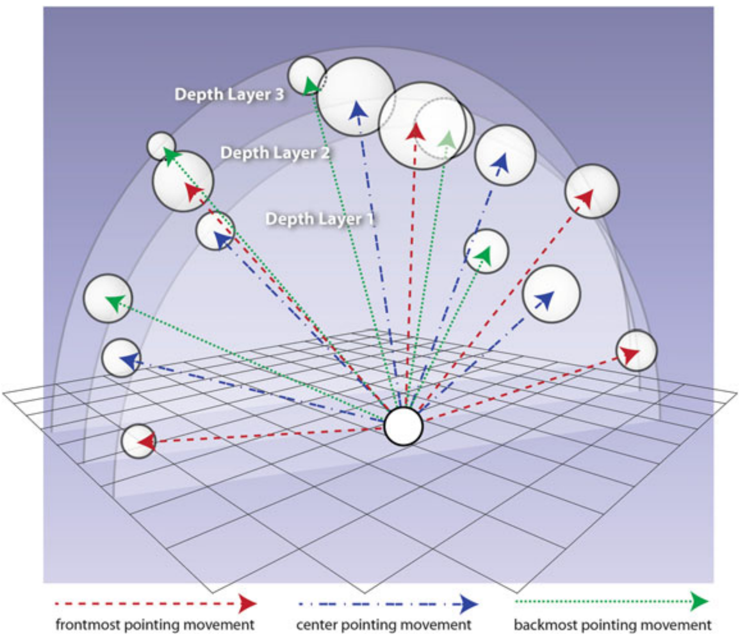


Fig. 2 Target objects on different depth layers in the virtual scene

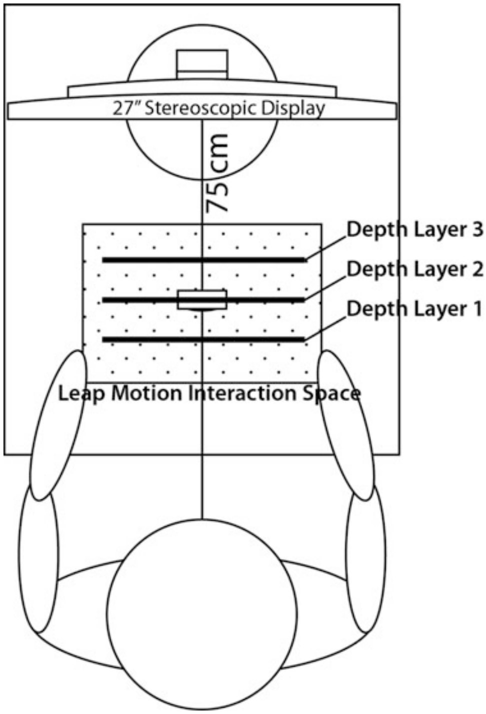


Fig. 3 Top-view of experimental setting with approximate depth positions of virtual target objects

Table 1 shows each factor *W* (target width) for every target sphere (cf. Fig. 2) and their specific indices of difficulty (*ID*) and their positions on the specific depth planes. The approach of using sphered targets over cylindrical targets in a Fitts' pointing task was validated by Teather and Stuerzlinger [15] since spheres are the more natural 3D extension of 2D circles.

The participants started the pointing movement with initially having their hand on a pre-defined position on the table outside the interaction volume of the Leap Motion Controller. As soon as the experimental task started they had to move their hand into the interaction volume and use the virtual hand model to move their virtual fingertip from the initial sphere to the target sphere. A time measurement started as soon as the initial sphere was reached and stopped when the index fingertip touched the target sphere. Spheres to hit were marked with a bright green color; inactive spheres were marked with a translucent grey color. Only the initial sphere and one target sphere were visible at the same time.

2.2 Participants

A total number of n = 14 participants took part in the experimental study, ten male and four female aged from 22 to 41 years (M = 27.3; SD = 5.8). They were all

Table 1 Target sizes of the 15 ID's regarding their actual size and their perceived size by the participant

ID (bits)	3.0	3.2	3.3	3.3	3.4	3.4	3.4	3.4	3.6	3.6	3.6	3.7	3.8	4.0	4.1
Size W (mm)	20.5	17.7	15.3	15.9	14.3	13.4	12.8	13.7	10.6	10.3	10.3	9.0	8.4	6.8	5.9
Size W (arcminutes)	94.2	81.3	69.9	72.8	65.6	61.4	58.5	62.8	48.5	47.1	47.1	41.4	38.5	31.4	27.1
Depth Layer	3	2	1	2	3	3	1	1	3	1	2	3	2	1	2

non-paid volunteers with having the ability of stereoscopic vision and a valid visual acuity. One person had to be removed from the sample for having strabismus which turned out during the experiment and leading to difficulties in usage of the stereoscopic passive glasses in the main experiment which caused a high error rate in pointing at the virtual targets.

2.3 Apparatus

The main experimental task was conducted using a 27″ stereoscopic display with integrated polarized light filters on horizontal pixel lines. The participants wore appropriate passive glasses to separate the perceived image per eye channel. Participants who wore correcting glasses were provided with a clip of passive filters which was mounted onto the participants' glasses. The viewing distance was adjusted to 750 mm before conducting the main experimental task. The Leap Motion Controller was placed centered between display and participant as shown in Fig. 3. The experimental software ran on an upper class laptop with an nVidia K3000M graphics card, which guaranteed a fluid 3D graphics performance of the software. An ASUS VG27AH was used as a display device which had to be adjusted in its height for every individual participant to guarantee a perpendicular view on the display.

The implementation of the software for the experimental carrier was conducted using VSG Open Inventor 9.4 .NET and a Visual Studio .NET 4.5 programming environment. Standard spherical objects were used from the Open Inventor framework as target objects as well as the standard collision model provided by the framework. There were no additional virtual objects in the scene except a blue gradient used as background.

3 Procedure

The participants started the experiment with a short survey concerning their demographic data and their experience in using free hand input methods in virtual environments and a measurement of their visual acuity. Then, participants were asked to do a paper-based spatial cognition test which incorporated spatial rotations of cubes featuring different figures which were pictured on their sides, which the participants had to assign to given templates. Afterwards, the participants conducted a motoric test on their right hand performance with the SCHUHFRIED Vienna Test System series for fine motor skills. The pre-testing phase was concluded with a simulator sickness questionnaire (SSQ) [10] to record the participants' well-being before the conduction of the main experiment which included wearing the passive stereoscopic glasses. The participants were accommodated to the stereoscopic environment in a two minutes experimental phase by using one of the hand models before conduction of the main pointing

Table 2 Borg rate of perceived exertion

Rating	Perceived exertion
6	No exertion
7	Extremely light
8	
9	Very light
10	
11	Light
12	
13	Somewhat hard
14	
15	Hard
16	
17	Very hard
18	
19	Extremely hard
20	Maximal exertion

task. The hand models were used in permutated order with filling an SSQ and a Borg RPE [3] (cf. Table 2) after execution of each set of pointing movements.

Pointing tasks to each target object were performed three times per object and movement time was recorded, aggregated for each ID and then grouped per depth layer and analyzed with ANOVA with a significance level set to $\alpha = 0.05$.

4 Results

The spatial cognition test and the test for fine motor skills using the Vienna test system was used to validate the sample. Collected measuring data from the main experiment was cleared within-participant from outliers based on movement time using a box plot: $MT < Q1 - 1.5 \times IQR$ and $MT > Q3 - 1.5 \times IQR$ were marked as outliers.

4.1 Descriptive Analysis

Figure 4 shows regression on means of movement time MT in milliseconds on each of the factors hand model for each specific depth level. Their means and standard deviations are stated in Table 3. Hand model factor M_2 shows a generally better performance for all three depth layers.

Subjects performing well in the spatial sense test showed shorter movement times for the Fitts' pointing task. The Simulator Sickness Questionnaire did not indicate changes in visual discomfort or a change of being during the experiment. Results of the Borg RPE indicate a slight increase of the exertion measurement scale with measurement after each task repetition R with using a different hand

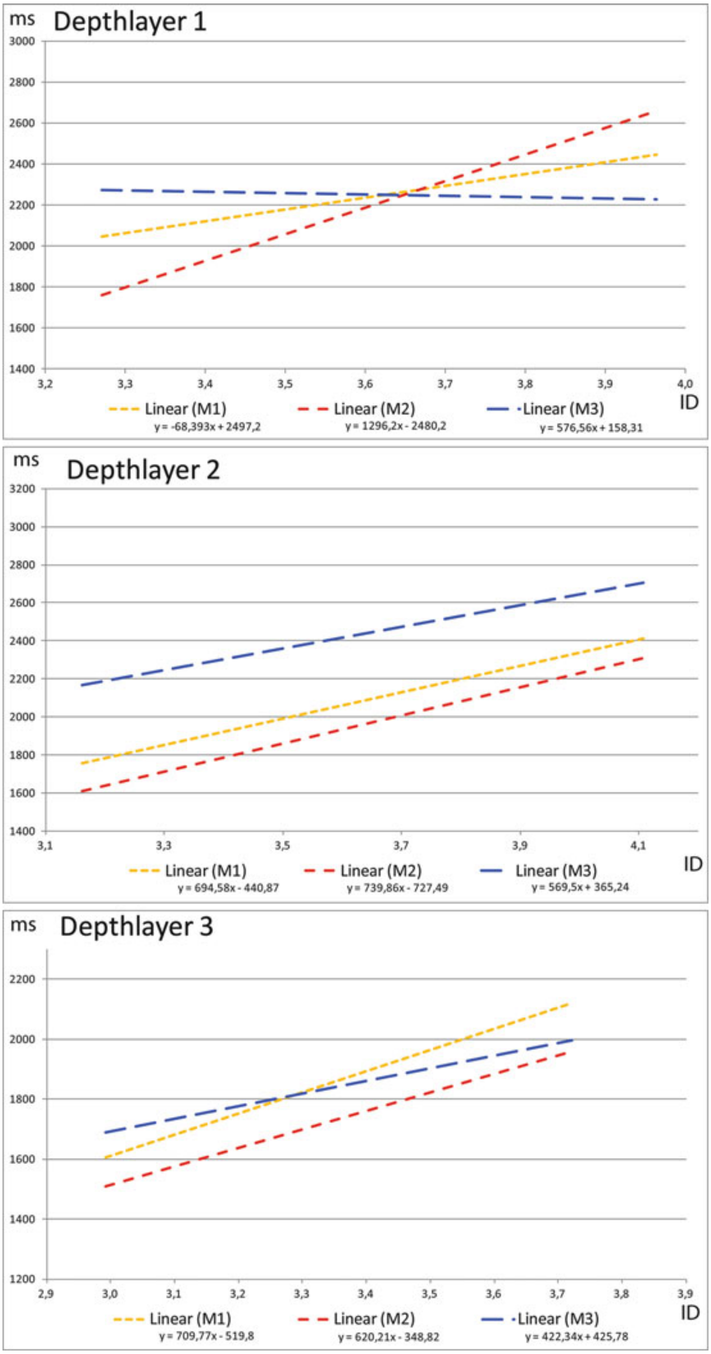


Fig. 4 Regression on movement times for each factor hand model on each specific depth level

Table 3 Means and standard deviations for each factor hand model per depth layer

Depth level	M_1 MT (ms) Mean	M_1 MT (ms) SD	M_2 MT (ms) Mean	M_2 MT (ms) SD	M_3 MT (ms) Mean	M_3 MT (ms) SD
1	2190.59	169.235	2088.60	365.80	2256.07	257.73
2	2056.87	307.46	1933.10	327.65	2413.19	434.00
3	1897.49	224.20	1763.50	227.96	1864.17	295.37

Table 4 Three levels of Borg RPE measurement after repetition R of each factor hand model in execution order

Mean R_1	SD R_1	Mean R_2	SD R_2	Mean R_3	SD R_3
10.38	2.1	11.23	2.0	11.7	1.7

model (hand models were used in permutated order) (cf. Table 4). Execution time per repetition R was approximately 7 min.

4.2 Analysis of Variance

A Kolmogorov-Smirnov test was conducted which verified the sample data was normally distributed. We analyzed the three factor hand models individually per depth layer to analyze each performance of the models concerning to the intrusion of the users' real hand into the stereoscopic display volume. We found a significant effect on movement time (MT) in pointing movements for the factor hand models on depth level two ($F = 10.608$, $p = 0.001$), containing target objects on the center depth level.

Analysis on depth layer one, or the front-most depth layer, indicated a violation of Mauchly's test of sphericity, why a Greenhouse-Geisser correction was used to correct degrees of freedom. We found no significant effect for depth level one ($F = 2.669$, $p = 0.116$) on movement time (MT) in pointing movements for the factor hand models, containing target objects on the front-most depth level and no significant effect for depth level three ($F = 1.236$, $p = 0.309$) on movement time (MT) in pointing movements for the factor hand models, containing target objects on the back-most depth level.

A post-hoc *t*-test indicated significantly better movement times for hand model factor M_1 ($M = 2062.98$, $SD = 236.96$) over M_3 ($M = 2330.06$, $SD = 288.5$) with $p = 0.19$ ($T = -2.695$) and M_2 ($M = 1934.46$, $SD = 288.83$) over M_3 with $p = 0.001$ ($T = -4.659$) for the center depth level pointing movements.

We measured an effect in the Borg Rate of Perceived Exertion data (cf. Table 4, Table 2), on which a Greenhouse-Geisser correction was used for violation of Mauchly's test of sphericity ($F = 10.957$, $p = 0.003$).

5 Discussion

Descriptive analysis of the mean values for the three depth levels and pointing movement times indicate a better performance on all factor hand models with increasing depth into the stereoscopic viewing volume. Analysis of variance showed that the effect on the factor hand models only exists for the center depth level, where a significantly better pointing performance was measured on hand model factor M_2 over factor M_1 and M_3. Performance on hand model factor M_3 was significantly inferior, which can be tracked to the difficult visibility of hand model M_3, which has heavily reduced visual characteristics that are probably not sufficient to let the user accommodate to the kinematic of the hand model during the control procedure. The interconnection of the point cloud elements with lines generated a much more fluent and more reliable usage of pointing movements in the virtual environment as the kinematic model M_2 with interconnected lines significantly outperformed the other hand models implemented in this study. When participants were using hand model factor M_3 we observed their hand trajectories differed slightly from the usage of the other two factor hand models: Participants could not properly estimate the depth position of the virtual hand in the stereoscopic viewing volume and had to correct the depth position of the virtual hand in order to hit the target objects with their fingertip. The missing effect for depth layer one and three can potentially be ascribed to the restrictions that are given by the implemented technology of the sensor, which depends on a full orthogonal view of the Leap Motion Controller's stereo camera on the users hand which is a fragile condition when the user moved her hand into the stereoscopic viewing volume (away from the user's body into display direction), as the user's wrist covered the tracking sensor partially, or the stereo camera did not fully cover the user's hand, when pointing movements on the front-most depth layer had to be conducted. The Borg RPE showed a significant increase in its mean value with the increase not being on an exceeding level. The received exertion mean increase from repetition R_1 to repetition R_2 range was 0.85, from repetition R_2 to repetition R_3 was 0.47 (cf. Table 2, Table 4). The exertion rate can be categorized between "very light" and "light" between first repetitions of the gestural task and slightly merged into "somewhat hard" in reference to the scale classification of the Borg scale. Though, participants had interruptions in the pointing task by putting down their hand down after each pointing movement, a subjective increase of exertion is measurable. The average execution time of the pointing task at 7 min.

6 Conclusion

We conducted an experiment with differently visualized virtual hand models that were distantly controlled in a stereoscopic desktop environment with the Leap Motion Controller. The conducted experiment is used as a pre-study for assessment of a pointing technique to be used in a stereoscopically visualized three-dimensional geographical information system environment. The assessment indicates that a kinematic model, with visualized hand and finger joints with interconnected lines, works significantly better over a point cloud model. The results help to find a hand model which features the optimal tradeoff between the lowest amount of occlusion in the virtual world and the highest characteristics of recognition. Pointing performance was measured for three factor hand models and on three different depth layers. Since all depth layers contained different target sizes a comparison between the depth layers could not be respected for evaluation. Therefore, we analyzed the factors hand model in the specific depth levels each. Strong effects were only measured for the center depth level which suggests that tracking the depth axis using the Leap Motion Controller requires additional adaption for a stereoscopic desktop setting, where a preferably deep intrusion into the stereoscopic viewing volume is required. Raised exertion data indicates an expected increase of physical effort in using an exceeded use of pointing movements. These first results recommend to use gestural devices as a supportive or additional input device to classical input but has to be investigated in a further study incorporating a measurement of muscular exertion using more precise measuring devices.

References

1. Apostolellis P, Bortz B, Peng M, Polys N, Hoegh A (2014) Poster: exploring the integrality and separability of the Leap Motion Controller for direct manipulation 3D interaction. In: 3D User Interfaces (3DUI), 2014 I.E. symposium on. IEEE, pp 153–154
2. Baier A, Zimmer A (2014) Advantages of stereoscopic 3D displays for conflict judgment performance in air-traffic-control: when and how to use 3D. In: Stephanidis C (ed) HCI International 2014 – Posters' Extended Abstracts, vol 435. Springer International Publishing, pp 609–614
3. Borg G, Maibaum S, Braun M, Jagomast KK (2001) Borg's perceived exertion and pain scales. Deutsche Zeitschrift Für Sportmedizin 52(9)
4. Bracco F, Chiorri C, Glowinski D, Hosseini Nia B, Vercelli G (2013) Investigating visual discomfort with 3D displays: the stereoscopic discomfort scale. In: CHI'13 extended abstracts on human factors in computing systems. ACM, New York, NY, pp 307–312
5. Bowman DA (2005) 3D user interfaces: theory and practice. Addison-Wesley, Boston, MA
6. Bruder G, Steinicke F, Stürzlinger W (2013) Effects of visual conflicts on 3D selection task performance in stereoscopic display environments. In: Proceedings of IEEE symposium on 3D User Interfaces (3DUI). IEEE Press, pp. 115–118
7. Brunner A, Groebner J, Umathum R, Maier F, Semmler W, Bock M (2014) An MR-compatible stereoscopic in-room 3D display for MR-guided interventions. Magn Reson Mater Phys Biol Med 27(4):277–282
8. Ferrari V, Megali G, Troia E, Pietrabissa A, Mosca F (2009) A 3-D mixed-reality system for stereoscopic visualization of medical dataset. Biomed Eng IEEE Trans 56(11):2627–2633
9. Goodchild MF (2008) The use cases of digital earth. Int J Digital Earth 1(1):31–42
10. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 3(3):203–220
11. Kockro RA, Reisch R, Serra L, Goh LC, Lee E, Stadie AT (2013) Image-guided neurosurgery with 3-dimensional multimodal imaging data on a stereoscopic monitor. Neurosurgery 72: A78–A88

12. MacKenzie IS (1992) Fitts' law as a research and design tool in human-computer interaction. Hum Comput Interact 7(1):91–139
13. Meyer R, Bützler J, Dzaack J, Schlick CM (2014) Development of interaction concepts for touchless human-computer interaction with geographic information systems. In: Human-computer interaction. Advanced interaction modalities and techniques. Springer, pp 589–599
14. Schlick C, Luczak H, Bruder R (2010) Arbeitswissenschaft. Springer DE
15. Teather RJ, Stuerzlinger W (2011) Pointing at 3D targets in a stereo head-tracked virtual environment. In: 3D User Interfaces (3DUI), 2011 I.E. Symposium on. IEEE, pp 87–94
16. Tukey JW (1977) Exploratory data analysis. Addison-Wesley, Reading, MA
17. Urey H, Chellappan KV, Erden E, Surman P (2011) State of the art in stereoscopic and autostereoscopic displays. Proc IEEE 99(4):540–555
18. Vetter S, Bützler J, Jochems N, Schlick CM (2011) Fitts' law in bivariate pointing on large touch screens: age-differentiated analysis of motion angle effects on movement times and error rates. In: Universal access in human-computer interaction. Users diversity. Springer, pp 620–628
19. Weichert F, Bachmann D, Rudak B, Fisseler D (2013) Analysis of the accuracy and robustness of the leap motion controller. Sensors 13(5):6380–6393
20. Zlatanova S, Rahman A, Pilouk M (2002) 3D GIS: current status and perspectives. Int Arch Photogramm Remote Sens Spat Inf Sci 34(4):66–71

Effects of Pressure and Vibration Stimuli on the Usability of Human-Machine Systems

Studies on Spatial Recognition and Response Times Using a Variable Tactile Belt

Matti Schwalk and Thomas Maier

Abstract

Users' information processing during human-machine interaction can be supported by redundant or additional tactile information, especially in situations of visual information overload or impaired visual perception. The present work is meant to describe basic thoughts on tactile semantics, the development of a variable tactile belt with 8 actuator units as well as first results of comparative studies of pressure and vibration stimuli, separately and in addition to visual perception. The examinations were conducted with 25 subjects. Spatial recognition rate and response time were recorded for pressure, vibration and visual stimuli presentation as well as all combinations of these. For purely tactile perception (pressure and vibration) reduced recognition rates occur at the sides of the waist, decreasing to approximately 80 %. Response times for discrete presentation of information via pressure or vibration were longer than for visual presentation but shorter response times of the visual presentation could almost be achieved purely tactile by the combination of pressure and vibration. The significantly shortest response times were detected for the combinations of visual + pressure + vibration and visual + pressure, revealing the potential of pressure stimuli for human-machine interaction. According to the 8 presented directions around the body, the shortest response times for tactile stimuli were detected for the actuator unit on the back. Furthermore, the article deals with subjects' estimations about their performances, distribution of input errors as well as different stimuli durations. The findings confirm expected advantages of visual-tactile information presentation in human-machine systems. Visual perception leads to high accuracy while additional tactile stimuli reduce response times. Especially pressure stimuli seem to have slight advantages compared to

M. Schwalk (✉) • T. Maier
Institute for Engineering Design and Industrial Design, Research and Teaching Department Industrial Design Engineering, University of Stuttgart, Stuttgart, Germany
e-mail: matti.schwalk@iktd.uni-stuttgart.de; thomas.maier@iktd.uni-stuttgart.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_23

vibration stimuli in terms of response times. The combination of pressure and vibration stimuli even leads to response times similar to visual presentation. As a conclusion, the potential of pressure stimuli could be shown to enhance future design of visual-tactile interfaces. In the long term, guidelines for tactile semantics are supposed to be derived.

Keywords
Human-machine interaction • Tactile semantics • Tactile belt display • Pressure stimuli • Vibration stimuli

1 Introduction

The present work is meant to describe basic ideas about tactile semantics, systematic test bench development and first findings regarding recognition rates and response times. Especially the comparison of pressure and vibration stimuli is focused on. These different types of tactile stimuli are presented separately as well as in addition to visual perception by means of a variable tactile belt with 8 actuator units and a touch screen.

1.1 Motivation and Significance

Various studies show that cognitive resources can be exploited optimally in different situations of human-machine interaction by using the still often rarely occupied channel of tactile perception in addition to the visual channel [1]. Compared to purely visual perception, shorter response times and lower error rates can be achieved [cf. 2, 3]. This is why different types of tactile displays are currently under development or already in use. The majority of these devices is limited to vibration stimuli. Besides, there are also some approaches investigating pressure stimuli. But no known device makes it possible to apply pressure and vibration stimuli at the exact same spots on the human body to conduct comparative studies of these different types of tactile stimulus presentation. However, due to the permanently increasing number of functions within various technical systems, in our opinion this kind of fundamental studies is absolutely mandatory for situational and intuitive encoding of information in the long term. Multimodal interface design should exploit the full potential of human perception to increase usability and reduce users' workload whenever possible.

To start comparative studies of the effects of pressure and vibration stimuli (as well as their combination) on the usability of human-machine systems, a special test bench was developed at IKTD (Institute for Engineering Design and Industrial Design, Research and Teaching Department Industrial Design Engineering, University of Stuttgart, Germany). The test bench features 8 actuator units placed around the waist of the subject (cf. Fig. 2). Each unit can apply pressure as well as vibration at the respective same spots on the human body. In the following, some further basic thoughts regarding cognitive ergonomics are described.

1.2 Basic Thoughts

To illustrate the context of the present work, it is referred to the basic scheme of human-machine interaction according to Maier and Schmid [4]. This is a holistic approach to consider human perception, cognition and execution for interface optimization. In the following, this basic scheme is extended with further contents of cognitive ergonomics, merging several recognized models within a single visualization. These include the model of human information processing according to Wickens and Hollands [5], the theory of situation awareness by Endsley [6], general aspects of cognitive psychology according to Wentura and Frings [7] as well as the definition of usability according to standard DIN EN ISO 9241-11 [8]. The mentioned models and contents are combined and linked with visual-tactile interface design. The visualization developed from this shows the entire circle of human-machine interaction, starting with information encoding and presentation by the multimodal interface (see Fig. 1).

Every situation of human-machine interaction is determined by the current task as well as the physical and social environment. Within this situation the user is concentrating on specific elements and actions depending on his or her focus and capacity of attention. Human sensory processing and perception systems are filtering the huge amount of information arriving at the receptors. Most important functions of the interface are drawing attention to important elements at the right time and supporting correct information filtering. Only this provides the basis for optimal cognition, respectively human information processing. The user has to perceive and comprehend the current situation before he or she can generate a projection of the near future of the system (situation awareness). This projection

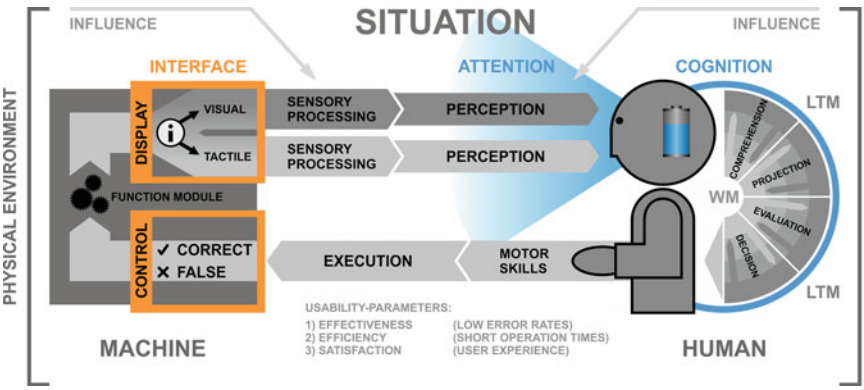


Fig. 1 Extended basic scheme of human-machine interaction for cognitive ergonomics of visual-tactile display systems

leads to an evaluation of possible actions by interaction of working memory (WM) and long term memory (LTM). The more intuitive or familiar specific human-machine interactions are, the faster the cognition can proceed. After evaluation a decision is made and by activating motor skills the user starts to perform an execution which can be correct or false, in the easiest case. Effectiveness, efficiency and satisfaction of human-machine interaction depend on the usability and cognitive ergonomic design of the interface. In support of this, tactile information encoding is used to complement visual interfaces by redundant or additional presentation of tactile stimuli. Especially in situations of visual information overload or impaired visual perception, visual-tactile display systems can enhance usability and thus safety.

Involving and investigating pressure stimuli besides vibration as a further possibility of tactile information encoding can be compared to using speech information as well as signal tones in terms of audible feedback. It is a matter of providing and using as much different opportunities of information encoding as possible. The main goal for interaction design should always be to achieve the best possible use of the user's physiological and cognitive resources. The more reliable options of information encoding, the greater are the opportunities to create optimal usability and constant situation awareness for the user within human-machine systems. However, for most systems it can be assumed, that the visual perception channel is the most important one. Tactile information is supposed to be used in support of a visual display most of the time. This is why the present work includes visual, pressure and vibration stimuli as well as all combinations of these.

In the long term it is planned to derive concrete guidelines of tactile semantics. Depending on the situation, environment, type of information and task specific forms of encoding are supposed to be recommended. Situational adaptive parameters can be stimuli combination (visual, pressure, vibration), body location of stimuli, intensity, duration, pattern etc. In the following, comparative fundamental studies on recognition rates and response times regarding pressure and vibration stimuli (also in combination with visual stimuli) are focused on, starting with state of the art research regarding tactile displays and systematic test bench development.

2 State of the Art: Tactile Displays

Tactile displays can be used for various applications. According to Van Erp and Van Veen [9] the four most important categories are safety (releasing visual workload, presenting information in a natural and intuitive way), assistance (decreased workload and faster responses), fun (silent and private communication with the machine), and efficiency (e.g. optimizing fuel consumption in driving scenarios) of human-machine interaction.

So far tactile displays are in use in the automotive industry in terms of vibrating actuators in driving seats [10, 11], steering wheels [11, 12] or seatbelts [13]. Furthermore, there are solutions under development to transmit information via force feedback of the pedals [14, 15]. In the field of medical engineering vibrating surgical instruments are used for navigation [16] as well as arm cuffs displaying vital signs of a patient [17]. Furthermore, there are tactile displays to navigate people [18] as well as for pilots [19, 20]. In general, tactile displays are often used in situations where the visual perception channel is impaired or overloaded.

2.1 Tactile Belts

In the field of tactile belts there are several works of Cholewiak, who is, for example, concerned with vibrotactile patterns [21–23] and the difference between younger and older subjects in terms of tactile perception [24]. McGrath & Rupert revealed advantages through a vibrotactile belt supporting subjects to keep their balance [20]. These effects occurred with pilots as well as in the medical environment for patients suffering from impaired balance.

Pielot et al. [25] and Heuten et al. [26] describe a presentation method with a vibrotactile belt that allows displaying directions between tactors by interpolating their intensity. They let participants determine directions while navigating along tactile waypoints in a virtual environment and compared this method with the prevalent one. They found that interpolated direction presentation significantly improved the accuracy of perceived directions. The studies revealed that users could perceive the interpolated directions with a median deviation of 15° from the presented direction. But discrete direction presentation proved to be better suited for waypoint navigation and was found easier to process.

Tsukada and Yasumrua [27] integrated GPS technology into their tactile belt, also presenting directional information. First findings showed that vibration on the back was perceived to be weaker than vibrations on the abdomen. For practical use it is supposed that four vibrators would be good enough, while eight vibrators are not necessarily needed. Reliable studies have still to be carried out to validate these estimations.

None of the found investigations on tactile belts includes pressure stimuli. But this form of tactile information encoding is used in some developments of skin stretch devices, which are described in the following.

2.2 Pressure Stimuli

Bark et al. [28] developed a portable device, strapped to the forearm that provides position and velocity information by imparting rotational skin stretch. The produced motions and sensations are comparable to those obtained by placing two fingers against the skin and rotating them together. The authors state that with brief training, users are able to use skin stretch to gain a sense of position and motion of their limbs when such feedback is otherwise missing or impaired. Applications for this device include motion training for sports or physical therapy and proprioceptive feedback for enhanced control of prosthetics as well as feedback during virtual reality and gaming. Findings indicate that skin stretch feedback may be best suited for tasks in which an afferent/efferent command loop exists so that users receive skin stretch feedback in response to motor commands. Using the same device, Wheeler et al. [29] found that the results for controlling a virtual prosthetic arm indicate the effectiveness of rotational skin stretch for proprioceptive feedback in myoelectric prostheses, particularly when vision is otherwise occupied.

Guinan et al. [30] designed and built a skin stretch feedback device capable of communicating directional cues in 5° of freedom (2 translational and 3 rotational DOFs). 4 translational and 4 rotational motions of the hand can be presented to a user by means of 2 independently controlled 2-DOF skin stretch feedback displays placed back-to-back. Additionally, 2 rotational motions of the hand can be communicated through spiral motions of the 2 skin stretch feedback displays. Thereby, the user's index finger and thumb grasp the device at the location of the moving contactor (tactor) of each of the 2-DOF skin stretch displays. Studies showed that users were able to achieve more than 98 % accuracy in recognizing cues in all 5° of freedom under 2 tactile cue modes: sustained and pulsed skin stretch cues. Subjects responded more quickly to sustained tactile cues than to pulsing tactile cues. Their responses to out-of-plane (roll and yaw) rotation cues were faster than their responses to translational direction cues. Also users' ability to match a given target angle of rotation with their wrist rotation in 1° of freedom was investigated. Thereby, users could match the target angle with an average of 7.39° of error with open-loop skin stretch feedback. The addition of proportional rate-based skin stretch feedback allows users to match the given target angle within ±2°.

Gwilliam et al. [31] prototyped and used a force feedback joystick with an integrated skin stretch display to conduct a directional matching experiment to evaluate the perception of force and skin stretch cues, individually and combined. The handle of the joystick provided placement of the user's thumb tip in the center of the skin stretch display which is located at the top of the joystick's handle. The results suggest that force and skin stretch cues are additive and integrable, indicating that skin stretch cues can be used to supplement or replace low magnitude force cues. Thereby, skin stretch cues result in higher precision, but lower accuracy than force cues. Thus, integration of both cue types results in improved precision, while retaining high accuracy. The authors also state that skin stretch cue magnitude can be increased without the risk of creating instabilities. This is why this kind of tactile feedback is supposed to be a viable candidate to augment or replace force feedback in safety-critical situations.

Skin stretch feedback as a form of haptic stimulus presentation, which is described above, surely is a combination of pressure and tangential touch, while in the present work pure, vertically applied pressure stimuli are meant to be compared with vibration stimuli. For preliminary studies some tactile displays operating with this kind of pressure stimuli were already developed and built up at IKTD. These devices can be attached to the chest respectively forearm. Diverse studies revealed benefits of supporting tactile pressure stimuli during visual searching tasks on a screen, using 4 lifting magnets placed on the user's chest. Different patterns and timing of stimuli were investigated to display one of 4 sectors on the screen in each case, giving a hint where the requested object can be found. Searching times were shorter for all configurations of visual-tactile encoding compared to purely visual presentation [2]. The tactile display attached to the forearm (also using 4 lifting magnets) was used in a memorizing task, revealing the potential of enhancing working memory capacity by visual-tactile information encoding compared to purely visual presentation of different codes consisting of random sequences of the 4 directions up, down, left and right [3].

2.3 Recommendations and Summary

In the following, some approved values for frequency and display duration of vibrotactile stimuli are cited. Jones and Sarter [32] identified optimal frequencies for the perception of tactor vibrations to be between 150 and 300 Hz which is also matching with the recommendations of Van Erp [33].

Regarding stimulus duration, Kaaresoja and Linjama [34] showed that participants prefer vibrotactile alerts between 50 and 200 ms while the torso is supposed to be the most suitable body part. Based on the findings of White [35], Barber et al. [36] set the duration of vibrotactile stimulation to 250 ms at a sinusoid frequency of 230 Hz. Intervals between stimuli were defined to be 100 ms. These parameters allowed participants to accurately perceive and distinguish individual tactors during studies of tacton (tactile icons) learning and performance.

As a conclusion it can be noted that most of the results and findings regarding tactile displays are based on pure vibration stimuli. On the other hand there are also some studies which are focusing on pressure stimuli, although most of them are using skin stretch feedback which does not create pure pressure stimuli but a combination of pressure and tangential touch on the human skin. But in particular no known device offers the possibility to apply pressure and vibration stimuli at the exact same spot on the human body to conduct comparative studies of these different types of tactile stimulus presentation. The present work is meant to address exactly this gap of information.

3 Test Bench Development

The comprehensive literature research provided the basis for concept creation and evaluation, optimized final draft and realization of the variable tactile belt. The following description of the test bench development is based on [37].

3.1 Concept Creation

Besides literature research, the concept creation was conducted on the basis of a generated function structure and requirement specification. Thereby, several solutions for all sub-functions were generated. By systematic and reasonable combination of these solutions six holistic concepts were created, covering as many different body parts as possible: upper arm cuff, belt, rucksack, forearm cuff, chair and armrest. A systematic evaluation based on the requirements was conducted according to German Engineer Association (VDI) guideline 2225 [38]. As a result the highest technical value (88 %) was calculated for the tactile belt which was consequently followed up and detailed in the further development process. This concept provides several key benefits, e.g. relevancy to practice and the possible mobility of a later close-to-production version. Furthermore, the circumferential arrangement of actuator units around the user's waist is supposed to support intuitive recognition of the display during navigation tasks. Also, the variable tactile belt can be equally used by left and right handers as well as persons with different height and corpulence because all actuator units can be positioned freely by belt clips. Finally, the weight of the actuators can be carried comfortably on the belt compared to other body parts.

In terms of human perception thresholds especially the two-point discrimination has to be taken into account for the development. This parameter describes the smallest distance of two nearby objects touching the skin simultaneously in which they can still be perceived as two distinct points. The simultaneous two-point discrimination depends on the body part and varies between 1 and 70 mm. Particular values for waist and back can hardly be found in literature. The successive two-point discrimination (stimuli presented one after the other) can be up to four times smaller [1]. Furthermore, pointed tips lead to greater perceived stimulus intensity than larger surfaces, according to Grunwald and Beyer [39]. At sensitive body parts the smallest perceivable depth of penetration on the skin is 0.01 mm. For vibration stimuli at a frequency of 200 Hz an amplitude of 0.1 μm is enough to trigger a sensation [1].

3.2 Final Draft and Realization

For the final draft of the variable tactile belt, 8 equal actuator units were designed, which can be positioned freely around the user's waist by belt clips and an adjustable strap (see Fig. 2). The number of 8 actuators is recommended by

Fig. 2 Mounted actuator units as 3D-CAD model (*left*), single unit with cover (*middle*) and described components (*right*)

Cholewiak et al. for vibrotactile stimuli [40]. It is also the highest possible number of stimuli positions around the waist which can be realized without falling below the simultaneous two-point discrimination. For verification the waist circumference of the 5th percentile female (between 18 and 25 years of age) is taken into account, which is 625 mm according to body size tables of standard DIN 33402-2 [41], leading to a minimum distance of 78 mm between stimuli positions. This value is still higher than the maximum two-point discrimination of 70 mm.

Each actuator unit features a lifting magnet with linear motion and return spring to apply pressure stimuli by moving out for 3 mm (at a force of about 3 N) from the starting position which is defined by a distance piece. The diameter of the round plastic tip (touch element) transferring the stimulus to the human body is also 3 mm. The touch elements are detachable and can be replaced by other geometries. Additionally, a vibrating actuator is integrated within the lifting axis which makes it possible to apply pressure and vibration successively or even simultaneously at the exact same spot. For vibration, motors with 8 mm diameter and 3.4 mm length were used and operated with 3 V input voltage leading to 185 Hz vibration frequency and 0.7 G amplitude (normalized gravitational acceleration). Stimulus duration was set to 250 ms for both, pressure and vibration stimuli. All actuators were actuated by a PC workstation using LabVIEW software. The actuator units are shown in Fig. 2. The weight of each of the 8 actuator units is 300 g.

4 Experimental Studies

The examinations were conducted with 25 subjects ($N = 25$, 8 female, 17 male) between 23 and 39 years of age (mean = 27.6 years, $SD = 4.4$ years). Test conditions, given task and results are described in the following.

4.1 Test Conditions and Task

For the current study all 8 actuator units were evenly arranged around the belt with equal distances to each other. Thereby, it was made sure that the units were placed exactly at the front, back, sides and the four positions in between according to the individual constitution of each subject. In addition to the tactile display a graphic touch interface was programmed showing the 8 actuator units visually in the same arrangement. At the beginning the subjects had the opportunity to test the feeling of all actuators by activating them one after another to get used to the pressure and vibration stimuli. Doing so it was also made sure that every actuator unit had proper contact to the subject and all stimuli could be perceived. After that the test was started. Thereby, one direction display at a time was presented and the subject's goal was always to react as quickly as possible and select the perceived direction out of the 8 possible directions on the touch screen. Then the next direction display would follow. By using a touch pen, a comparable operation by all subjects was ensured, independent from their individual skin conductance, finger size and

Fig. 3 Experimental setup with subject

operation force. The direction display could be visual (via lighting up one of the virtual units on the touch screen), by pressure, by vibration or any combination of these (7 types of stimulus presentation in total). By means of a programmed random generator the presentation sequence was different for each subject and thus randomized. In total every possible form of display (7 types of presentation multiplied by 8 directions) was presented three times per subject to minimize falsification by outliers (quality control by standard deviation). The subject's input errors and response times were recorded. To ensure comparability of response times, the subjects had to press a button (starting point) on a separate numpad in defined distance of 25 cm to the touch screen for a randomized period of time between 1 and 3 s before the next direction display followed. Furthermore, the subjects wore hearing protectors to avoid influence by background noises of the actuators or environment. Figure 3 shows the experimental setup and one of the subjects making an input.

4.2 Results

The data collected by Rheinländer [42] were evaluated for all types respectively combinations of stimulus presentation and can be seen in Fig. 4. By means of seven radar graphs spatial recognition rates are visualized as a percentage. Therefore, every input of every subject was analyzed in terms of correctness by recording all input errors. In addition, the response time of each stimulus presentation was measured (see further below).

As expected the results in Fig. 4 show recognition rates of 100 % for all cases which include the visual perception channel. For purely tactile perception reduced

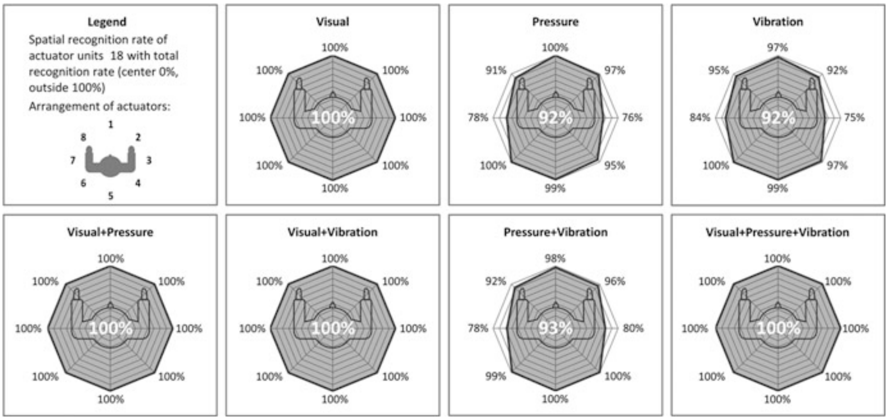


Fig. 4 Spatial recognition rate according to stimulus presentation

recognition rates occur at the sides of the body, decreasing to approximately 80 %. But in the front and back areas the recognition rate is almost 100 % also for purely tactile perception. These results correspond with the findings of Cholewiak et al. [40] and Brill and Scerra [43]. In terms of general spatial recognition no significant difference between pressure and vibration stimuli could be detected. But some interesting facts occurred regarding the input errors. The distribution of input errors for directions “left” (actuator unit 7) and “right” (actuator unit 3) can be seen in Fig. 5 for pressure stimuli. The figure shows all inputs made for the displayed directions “left” and “right” including correct inputs and directions respectively actuator units which were falsely selected instead of directions 7 and 3.

While actuators 1 (front) and 5 (back) were not selected falsely at all, the input errors in Fig. 5 show an even distribution between the front (actuators 8 and 2) and back (actuators 6 and 4) of the subject regarding pressure stimuli. In contrast, for vibration stimuli there are considerably more input errors occurring at the actuators at the back (6 and 4) than in the front which can be seen in Fig. 6.

For the combination of pressure and vibration stimuli the distribution of input errors becomes more even again between front and back, showing the greater influence of pressure stimuli (see Fig. 7).

In addition to the recognition rate, the achieved response times were recorded and evaluated. Figure 8 shows the results according to the seven different types respectively combinations of stimulus presentation (averages with standard deviation).

A two-way repeated measures ANOVA (analysis of variance) revealed an effect regarding the type of stimulus presentation, $F\ (2.54,\ 71.06) = 43.78$, $p < 0.001$ (Greenhouse-Geisser corrected). On the basis of a subsequently conducted post-hoc analysis by pairwise comparison with Bonferroni correction the following conclusions can be drawn: response time for single presentation of information via pressure or vibration is longer than for visual presentation ($p < 0.05$ and

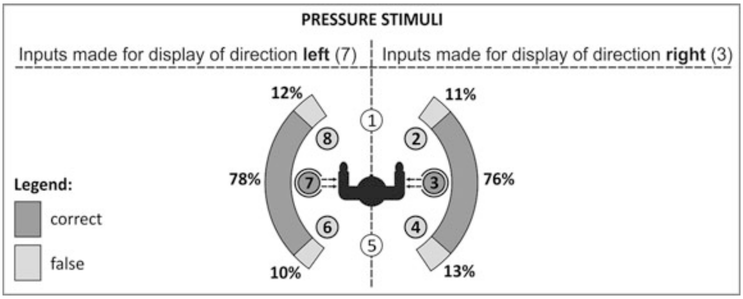


Fig. 5 Distribution of inputs: pressure stimuli

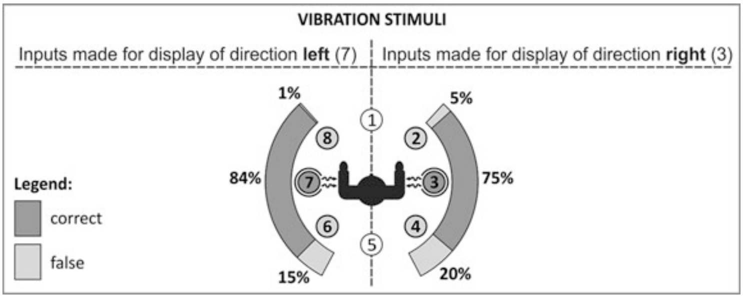


Fig. 6 Distribution of inputs: vibration stimuli

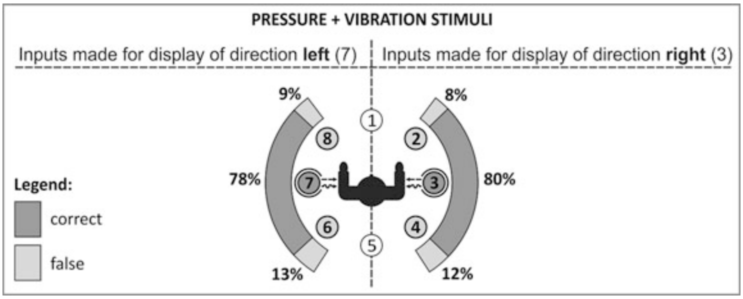


Fig. 7 Distribution of inputs: pressure + vibration stimuli

$p < 0.01$). But the shorter response times of the visual presentation can almost be achieved purely tactile by the combination of pressure and vibration, which means that there is no significant difference to the visual presentation any more. Compared to this, even shorter response times were detected for multimodal presentation of information combining the visual stimulus with pressure, vibration or pressure + vibration ($p < 0.001$, $p < 0.01$ and $p < 0.001$). Furthermore, the combination of visual + pressure + vibration stimuli leads to the shortest response times. Thereby,

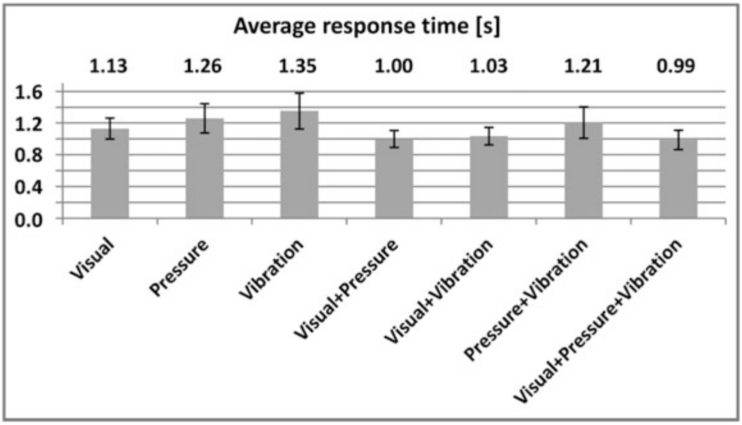


Fig. 8 Response time according to stimulus presentation

significant differences occur in comparison to all other types of information presentation ($p < 0.001$), except the combination of visual + pressure stimuli.

In addition to the recorded response times, the estimation of the subjects, which type of display they thought to be the fastest, was enquired afterwards. The order due to number of mentions (in brackets) was as follows: visual + vibration (8), visual + pressure (6), visual (5), visual + pressure + vibration (5), pressure (1), vibration (0). Thus, the subject's estimations of bimodal perception (visual + single tactile stimulus type) correspond with the recorded short response times in these cases. But it has to be stated that the combination of all three stimuli (visual + pressure + vibration) seems to be underestimated by the subjects based on their self-assessment. So the resulting advantages of this combination of stimuli are measurable but partially unconscious.

Besides the comparison of response times due to different types of stimulus presentation, Fig. 9 shows the comparison according to the 8 directions around the body (averages with standard deviation).

The results in Fig. 9 show quite balanced response times for different directions. The conducted two-way ANOVA also indicated an effect regarding the direction, $F\ (4.23,\ 115.97) = 5.77$, $p < 0.001$ (Greenhouse-Geisser corrected). A post-hoc analysis as described above revealed significant shorter response times for actuator unit 5 at the back in comparison with all other units, except unit 2 ($p < 0.05$ compared with units 1, 3, 6 and $p < 0.01$ compared with units 4, 7, 8). The quick responses for actuator unit 5 occurred particularly for purely tactile stimuli presentation by pressure, vibration and pressure + vibration. For purely or additional visual presentation, response times were almost equal for all directions. The quicker responses for purely tactile presentation may be explained by the fact that actuator unit 5 is normally the only one getting close to a bone (spine) which is probably enhancing the stimulus. Also many nerves are located in this area.

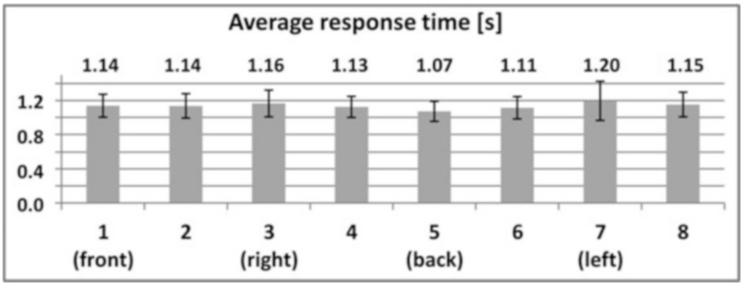


Fig. 9 Response time according to direction

It should also be mentioned that the response times in general were independent from the stimulus duration. A pretest showed no significant differences between varied display durations of 250, 500, 750 and 1000 ms. Thus, it was confirmed that the subjects react to the beginning of each stimulus, no matter which duration was selected. Especially in terms of pressure stimuli mainly position changes of the lifting magnets (activation and deactivation) can be perceived because of the adaption of responsible mechanoreceptors in the skin.

5 Discussion and Conclusion

Experimental studies were conducted to compare visual, pressure and vibration stimuli as well as all combinations of these during a task of spatial recognition using a variable tactile belt. The detected effects lead to the conclusion that combined visual-tactile information encoding can support the usability of human-machine systems in terms of achieving low error rates and short response times. Thereby, the visual display ensured response accuracy avoiding any input errors, while additional tactile stimuli led to shorter response times. Furthermore, the relevance of pressure stimuli in combination with visual or vibration stimuli became clear. The developed variable tactile belt offers diverse possibilities to advance research in this area. Special feature of the belt is the opportunity to apply pressure and vibration at the exact same spot of the human body directly after another and even simultaneously. Only this allows comparative studies of pressure and vibration stimuli under the same conditions. Due to the modular structure with belt clips and adjustable strap this test bench can be worn by persons with different waist circumferences and generally allows various arrangements of 1–8 actuator units. Further investigations in the field of tactile semantics (situational information encoding via pressure and vibration stimuli) are planned. In the long term universal recommendations for cognitive ergonomic design of visual-tactile display systems are supposed to be derived. Thus, especially in critical situations the operator can be supported to develop adequate situation awareness continuously. Practical applications are conceivable for rescue forces (navigation), doctors (patient information) or machine operators (status messages). A further advantage of tactile pressure stimuli is the possibility of discreet information presentation, depending on the actuator design. For example, the new Apple Watch features a kind of pressure feedback instead of a vibration motor, which is always generating unwanted noise during operation. The "taptic engine" of the watch silently produces the sensation of being tapped on the wrist to indicate an incoming call or other functions [44]. This kind of silent tactile information encoding can also be transferred to diverse situations of human-machine interaction in the future, as soon as respective fundamental studies on pressure stimuli have been conducted.

References

1. Schlick C (2010) Arbeitswissenschaft, 3rd edn. Springer, Berlin
2. Petrov A, Pfeffer S, Maier T (2011) Visuelle Präsentation und taktile Repräsentation – redundante, substitutive oder komplementäre Informationsdarbietung. In: 9. Berliner Werkstatt Mensch-Maschine-Systeme. Berlin
3. Schwalk M, Maier T (2014) Multimodale HMI – Untersuchungen zur Erweiterung der Arbeitsgedächtniskapazität durch visuell-taktile Anzeiger. In: Krzywinski J, Linke M, Wölfel C, Kranke G (eds) Entwerfen Entwickeln Erleben – Beiträge zum Technischen Design, Dresden, pp 233–242
4. Maier T, Schmid M (2014) IDeEn Kompakt – Manuscript of lecture Industrial Design Engineering. University of Stuttgart, Institute for Engineering Design and Industrial Design – Research and Teaching Department Industrial Design Engineering. Germany
5. Wickens CD, Hollands JG (1999) Engineering psychology and human performance, 3rd edn. Prentice-Hall, Upper Saddle River, NJ, 07458, pages 10 ff
6. Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. In: Human factors. The journal of the human factors and ergonomics society 37, pages 32 ff
7. Wentura D, Frings C (2013) Kognitive Psychologie. Springer Fachmedien, Wiesbaden
8. DIN EN ISO 9241-11 (1999) Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability
9. Van Erp JBF, Van Veen H (2001) Vibro-tactile information presentation in automobiles. In: Proceedings of Eurohaptics, pp 99–104
10. General Motors (2013) Sierra Safety Alert Seat Aids Driver Awareness (published April 30, 2013). Website: http://media.gmc.com/media/us/en/gmc/vehicles/sierra_hd/2014.detail.html/content/Pages/news/us/en/2013/Apr/0430-sierra.html. Accessed 24 Jul 2014
11. Gulde D (2014) Nervig oder hilfreich? Spurhalteassistenten im Test (published May 23, 2014). Motor Presse Stuttgart GmbH & Co. KG. Website: http://www.auto-motor-und-sport.de/testbericht/spurhalteassistenten-im-test-nervig-oder-hilfreich-8337582.html. Accessed 24 Jul 2014
12. Pudenz K (2014) Ein vibrierendes und pulsierendes Lenkrad: Forschung an haptischen Navigationssystemen (published May 7, 2014). Springer Fachmedien Wiesbaden GmbH. Website: http://www.springerprofessional.de/einvibrierendes-und-pulsierendes-lenkrad-forschung-an-haptischennavigationssystemen-15881/3952694.html. Accessed 24 Jul 2014
13. Riener A, Ferscha A, Frech P, Hackl M, Kaltenberger M (2010) Subliminal vibro-tactile based notification of CO2 economy while driving. Institute for Pervasive Computing, Johannes Kepler University, Linz, Austria
14. Gill B (2014) What's the difference between Pre-Collision and Advanced Pre-Collision? Are they the same system? Lexus, a Division of Toyota Motor Sales, U.S.A., Inc. Website: https://secure.drivers.lexus.com/lexusdrivers/magazine/articles/Vehicle-Insider/Ask-Lexus-Collision. Accessed 2 Nov 2014

15. Continental AG (2010) Leichtes Pulsieren im Gaspedal besser als Warnleuchten oder Töne (published February 23, 2010). Website: https://www.conti-online.com/www/presseportal_com_de/themen/pressemitteilungen/3_automotive_group/chassis_safety/press_releases/pr_2010_02_23_affp_hmi_de.html. Accessed 24 Jul 2014
16. Brell M (2009) Eine vibrotaktile Mensch-Maschine-Schnittstelle für chirurgische Applikationen. Ph.D. dissertation, Department of Computing Science, University of Oldenburg. Germany
17. Ng JYC, Man JCF (2004) Vibro-monitor: a vibrotactile display for physiological data monitoring. In: Proceedings of the human interface technologies conference. University of British Columbia
18. DLR Nachrichten (2012) Navigation durch Vibration: VibroTac unterstützt sehbehinderte und blinde Menschen (published September 25, 2012). Website: http://www.dlr.de/dlr/desktopdefault.aspx/tabid-10081/151_read-5318/year-ll/#gallery/7675. Accessed 20 Feb 2014
19. Self BP, Van Erp JBF, Eriksson L, Elliott LR (2008) Human Factors issues of tactile displays for military environments. In: Van Erp JBF, Self BP (eds) Tactile displays for orientation, navigation, and communication in air, sea, and land environments. NATO RTO Technical Report: TR-HFM-122
20. McGrath B, Rupert A (2014) Tactile displays: from the cockpit to the clinic. In: Ahram T, Karwowski W, Marek T (eds) Proceedings of the 5th international conference on applied human factors and ergonomics AHFE 2014. Kraków, Poland, pp 939–947
21. Cholewiak RW (1979) Spatial factors in the perceived intensity of vibrotactile patterns. Sens Processes 3(2):141–156
22. Cholewiak RW, Collins AA (1997) Individual differences in the vibrotactile perception of a “simple” pattern set. Percept Psychophys 59(6):850–866
23. Cholewiak RW, Collins AA (2000) The generation of vibrotactile patterns on a linear array: influences of body site, space, and time. Percept Psychophys 62(6):1220–1235
24. Cholewiak RW (1994) A comparison of vibrotactile pattern perception by young and old observers. Report No. NMAB-478: Currency Features for Visually Impaired People: National Research Council, National Materials Advisory Board report to the Bureau of Engraving and Printing, Department of Treasury, National Academy Press
25. Pielot M, Henze N, Heuten W, Boll S (2008) Evaluation of continuous direction encoding with tactile belts. HAID’08: Haptic and Audio Interaction Design
26. Heuten W, Henze N, Boll S, Pielot M (2008) Tactile Wayfinder: a non-visual support system for Wayfinding. In: Proceedings of NordiCHI, pp 172–181
27. Tsukada K, Yasumura M (2004) Active belt: belt-type wearable tactile display for directional navigation. UbiComp2004. Springer LNCS3205, pp 384–399
28. Bark K, Wheeler J, Lee G, Savall J, Cutkosky MR (2009) A Wearable Skin Stretch Device for Haptic Feedback. World Haptics, Salt Lake City, UT
29. Wheeler J, Bark K, Savall J, Cutkosky MR (2010) Investigation of Rotational Skin Stretch for Proprioceptive Feedback with Application to Myoelectric Systems. IEEE Trans Neural Syst Rehabil Eng 18(1):58–66
30. Guinan AL, Hornbaker NC, Montandon MN, Doxon AJ, Provancher WR (2013) Back-to-back skin stretch feedback for communicating five degree-of-freedom direction cues. In: Proceedings of the 2013 World haptics conference. Daejeon, Korea
31. Gwilliam LT, Doxon AJ, Provancher WR (2013) Haptic matching of directional force and skin stretch feedback cues. In: Proceedings of the 2013 World haptics conference. Daejeon, Korea
32. Jones L, Sarter N (2008) Tactile displays: Guidance for their design and application. Hum Factors 50:90–111
33. Van Erp JBF (2002) Guidelines for the use of vibro-tactile displays in human computer interaction. Proceedings of Eurohaptics 01/2002
34. Kaaresoja T, Linjama J (2005) Perception of short tactile pulses generated by a vibration motor in a mobile phone. In: Proceedings of the first joint eurohaptics conference and symposium on

haptic interfaces for virtual environment and teleoperator systems. IEEE, Washington, DC, pp 471–472
35. White T. (2010) Suitable body locations and vibrotactile cueing types for dismounted soldiers. Tech. Rep. ARL-TR-5186. Aberdeen Proving Grounds, MD: U.S. Army Research Laboratory
36. Barber DJ, Reinerman-Jones LE, Matthews G (2014) Toward a Tactile Language for Human-Robot Interaction: Two Studies of Tacton Learning and Performance. The Journal of the Human Factors and Ergonomics Society, Human Factors
37. Schwalk M, Sperl M, Maier T (2015) Abbildung von Druck- und Vibrationsreizen auf dem menschlichen Körper durch einen variablen taktilen Hüftgurt. In: VerANTWORTung für die Arbeit der Zukunft, 61. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft. Karlsruhe
38. German Engineer Association (VDI) guideline 2225 (1998) Design engineering methodics – Engineering design at optimum cost – Valuation of costs
39. Grunwald M, Beyer L (2001) Der Bewegte Sinn – Grundlagen und Anwendungen zur haptischen Wahrnehmung. Birkhäuser-Verlag, Basel
40. Cholewiak RW, Brill JC, Schwab A (2004) Vibrotactile localization on the abdomen: Effects of place and space. Percept Psychophys 66:970–987
41. DIN 33402-2 (2005) Ergonomics – Human body dimensions – Part 2: Values. Deutsches Institut für Normung e.V. Beuth, Berlin, Germany
42. Rheinländer J (2014) Untersuchungen zur taktilen Informationskodierung mittels Druck- und Vibrationsreizen bei der Mensch-Maschine-Interaktion (extract from student research project, unpublished). University of Stuttgart, Institute for Engineering Design and Industrial Design (IKTD). Germany
43. Brill JC, Scerra VE (2014) Effectiveness of vibrotactile and spatial audio directional cues for USAF Pararescue Jumpers (PJs). In: Ahram T, Karwowski W, Marek T (eds) Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014. Kraków, Poland, pp 958–965
44. Apple Inc. (2015) Product details of apple watch. Website: https://www.apple.com/de/watch/technology/, accessed May 4, 2015

Force-Matching. A New Approach to Determine Action Forces for the Ergonomic Evaluation in Automotive Industry

Benjamin Franzke and Mario Walther

Abstract

The collection and analysis of applied forces is an essential part of the ergonomic evaluation of workplaces. In industrial companies, direct measurements and estimates are currently considered to be accepted methods. The present study identifies alternatives to the accepted procedures for the determination of action-forces and analysed one of these in detail: force-matching.

Force-matching describes the immediate subsequent simulation of an exertion on a measuring device, without actually assembling a component. In this way, cost and time savings in comparison to the direct measurement can be achieved. Furthermore, the method increases accuracy compared to estimations. To validate the applicability, a study involving 38 subjects was carried out in order to achieve a systematic comparison between force-matching and direct measurement.

In the course of the evaluation, influence factors and key premises for the utilisation of the new method in the automotive industry are identified. The results indicate that force-matching has great potential for improving quantity and quality in the determination of applied forces in ongoing production.

Keywords

Force-matching • Force measurement • Action forces • EAWS • Ergonomics

B. Franzke • M. Walther (✉)
Volkswagen AG, Wolfsburg, Germany
e-mail: benjamin.franzke@volkswagen.de; mario.walther@volkswagen.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_24

1 Introduction

The ergonomic design of work places is progressively gaining importance as part of the entrepreneurial effort in industrial corporations of the twenty-first century. An accepted method for workplace evaluation in terms of ergonomics is the 'Ergonomic Assessment Worksheet'—EAWS [8].

The EAWS contains four sections: working postures, action forces, manual materials handling and repetitive loads. The evaluation of action forces is based on the input variables duration, number and level of applied forces. Therefore, a standardized approach is essential for the determination of action forces. Currently, direct measurements and estimations are accepted methods in industrial corporations.

During a direct measurement, the value of the force is recorded by a measurement device. This happens while the assembly of the parts is being carried out. By doing so, the results are relatively accurate. However, the experiments require a complex set-up and expensive devices.

Due to technical reasons or extreme efforts, a direct measurement is not always possible. In such a case a practitioner estimates the forces based on professional experience. A subjective estimation involves a high potential for failure in evaluating work places. This causes a need for the simplification of actions force determination as well as for the development of new solutions for difficult or not determinable applied forces.

The present paper describes an alternative to those accepted approaches for the determination of action-forces: the force-matching. Aiming for a methodology of procedures for action force determination, force-matching will be validated and systematically compared to the established ones.

2 Study Methodology

2.1 Force-Matching

The basic concept of force-matching is the reapplication of an applied force on a measuring device [11]. A similar method was introduced with the aim of comparing weights [5] before it was used to determine grip forces with a dynamometer. The method proved to be accurate and reproducible [11]. According to studies using a screwdriver and a lifting task [4] as well as a reapplication of a given force value on a dynamometer [1] the aforementioned conclusion was substantiated. In most of those studies, a high variance of the skill in force-matching in between the subjects has been observed.

► Force-matching is the subsequent simulation of an applied force using a measuring device (and the same contact conditions) without actually assembling a component.

Premises for the execution:

– Equivalent contact conditions (grip, tactile feeling, contact surface and area, body posture)
– Force-matching is to be carried out immediately after the application of force (no intermediate stimuli)
– Precedent opportunity for the subject to practice the applied force and force-matching
– Exact instruction of the subject about aim and content of the experiment

Force-matching offers a potential for significant savings in terms of money and time compared to the direct measurement. Additionally, it potentially increases accuracy compared to the estimation. Currently, there are no studies for force-matching of applied forces in the automobile industries or comparable fields of application. The present study generates information for accuracy and applicability of force-matching compared to direct measurements and estimations.

2.2 Subjects

The present study was carried out with 38 subjects, of which 29 were male and 9 were female. All of them were acquired in the automobile industry. The occupation of the subjects ranged from industrial engineering to planning and production.

2.3 Test Objects

The test objects were chosen based on a prior study, which aimed at classifying as well as determining the number of applied forces in assembly processes in the automobile [9]. Overall five components were used: two clips, two plugs and one clamp (Fig. 1). These chosen components are very frequently in car assembly. They are also easy enough to assemble and reliable to measure. A thumb-contact-grip was chosen, since this is the most widely used grip for assembly operations in the automobile industry [9]. In order to investigate force-dependent tendencies, all components were assembled with two different hole sizes. Those were consistent with the specific tolerances of the components.

The objects in combination with the certain hole sizes generate the following force ranges:

Clamp: 40 N to 60 N
Clip: 15 N to 40 N
Plug: 60 N to 100 N

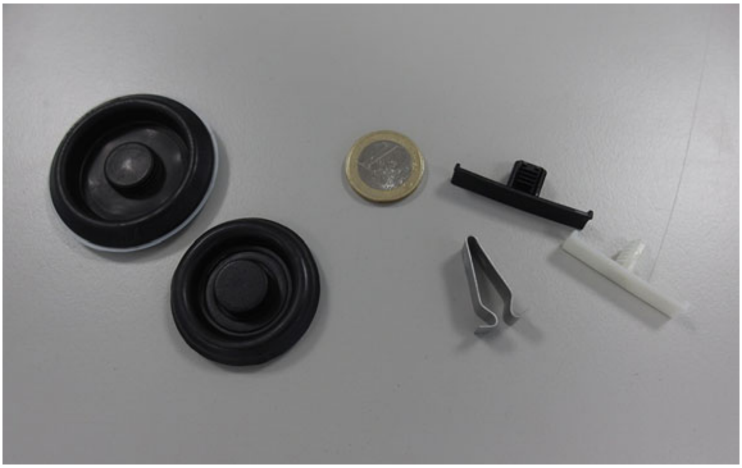


Fig. 1 Test objects

2.4 Measuring Devices and Set-Up

The measuring device was chosen to be the mobile force plate 9260AA6 made by Kistler with a 1000 Hz measuring frequency, a dimension of 50×60 cm and a maximum measurable force of 5000 N. A perforated plate was used as joint partner and placed on the force plate as seen in Fig. 2.

Due to the set-up and the procedure described in the following chapter, three issues could not be considered for the present study:

- Multidimensional forces
- Sequence related effects
- Training related effect in the subjects

2.5 Procedure

At the beginning of the experiment, the subject's size, sex and skill in terms of execution of force measurements were documented. The aim, the content and the procedure of the trial were explained to the practitioners. Afterwards, the maximum force of every individual was determined [7] to allow a comparison between estimation and force-matching. At this point the actual measurements were performed. There are ten different applied forces, resulting from the combination of five components and two hole sizes (Table 1). The sequence of those applied forces was randomly determined by a dice in order to even training and exhaustion effects in the statistical analysis.

The subjects had a 5-min individual training phase per applied force. Subsequently the measurement was started. During the actual measurement, the

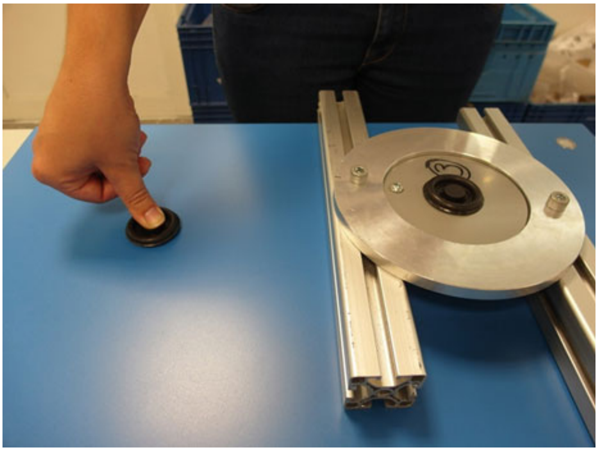


Fig. 2 Measuring set-up (*Left*: Plug for force-matching, *right*: fixture with perforated plate for the determination of the applied force)

Table 1 Overview of the applied forces divided by component and hole size

	Applied force									
	1	2	3	4	5	6	7	8	9	10
Component	Clamp		Plug 1		Plug 2		Clip 1		Clip 2	
Hole size	Small	Big	Small	Big	Small	Big	Small	Big	Small	Big

component was joined with the perforated plate and the applied force was immediately simulated with the same contact conditions, both on the force plate. Information about the actual values was hidden from the subjects in order to reduce the influence on the exertion of forces. There were ten repetitions for each applied force in order to achieve statistical reliability. After those repetitions, the participant was asked to estimate the level of the applied force on a Borg scale [3], compared to the individual maximum force measured earlier. For an optimal interpretation of the results later on, special assembly-characteristics were written down descriptively.

2.6 Analysis

There were about 4000 measurements with 400 estimations resulting from the trials. Using an experience-based algorithm [10] the relevant measuring points in the measuring curves were determined automatically. Afterwards, the location of those points was controlled and if necessary corrected manually according to Fig. 3.

In order to identify significant differences between direct measurement, force-matching and estimation, the statistical analysis was done by using SPSS and an ANOVA with repeated measurements. The significance level was set to 5 %. Afterwards a Bonferroni post-hoc test was conducted in order to describe the extent of the differences. For the direct measurement, the utilised variables were insertion force, maximum force, matched force and estimation force. In most cases the

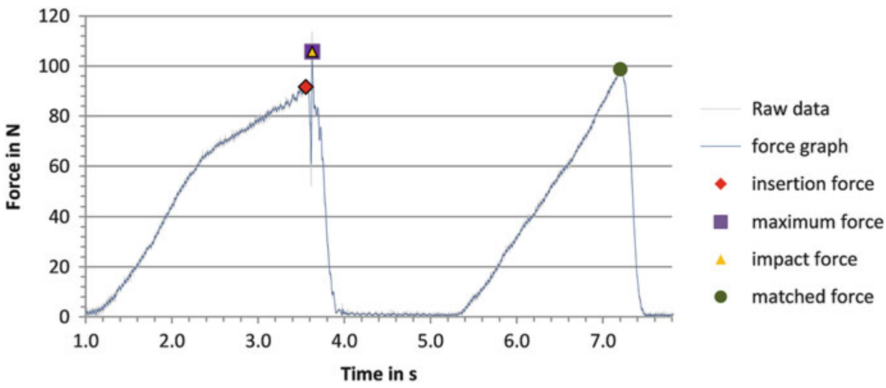


Fig. 3 *Left curve*: characteristic force-time graph using the example of a plug assembly, *Right curve*: force-time graph of the immediate force-matching of the plug assembly

maximum force was equal to the impact force. However, the maximum force is easier to determine and has the more significant influence on the matched force. Therefore, the maximum force was used for the analysis in the following results.

3 Results

3.1 General Results

As a result from the ANOVA, the three approaches: direct measurement, force-matching and estimation showed significant differences. The Bonferroni post-hoc test shows that this difference varies with the reference point taken from the direct measurement. Generally, there is a significant difference between the insertion force and the matched force, whereas in most cases there is no significant difference between maximum-force and matched force. This means that the subjects tend to match the maximum force rather than the insertion force.

If the insertion force of the direct measurement is the given reference point for the matched force, the difference between direct measurement and force-matching varies nominal between 12 N and 29 N, which is equivalent to percentages of 15–95 %. Given the maximum-force is the relevant reference point, the difference is rather 7–16 N or 10–25 % respectively.

The estimation force is, as a rule, on the same level as the matched force but shows the highest standard distribution of 42 %. The Insertion force (11 %) and the matched force (26 %) have much lower standard distributions. Those results are visualised in Fig. 4.

Examining the percentage deviations between matched force and insertion force shows that the subjects are most able to match forces for the plug. Thereafter, the clamps and finally the clips follow. For all components, the average difference is closer to the maximum-force (Table 2).

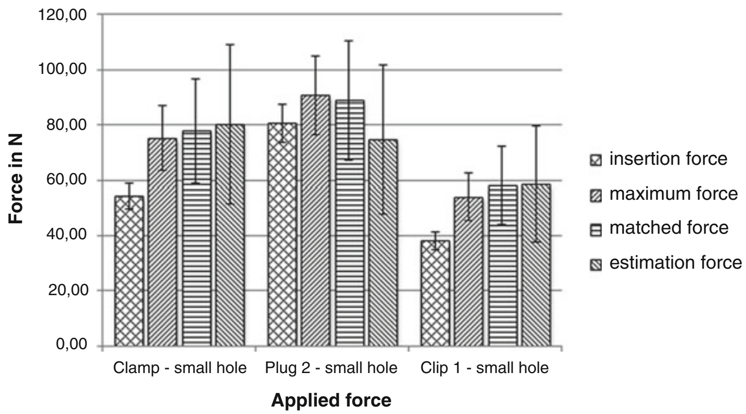


Fig. 4 Results of the applied forces 1, 5 and 7 for all subjects

Table 2 Component and applied force related overview of the percentage differences

Applied force	1 (%)	5 (%)	7 (%)	Average all components
Difference MatF–InsF	43.3	15.4	52.7	45.1
Difference MatF–MaxF	14.5	10.1	15.9	16.5

MatF Matched Force, *InsF* Insertion Force, *MaxF* Maximum Force

There was no correlation found between the ability to do well in force-matching and the sex or the experience of the participant with force measurements. Therefore, there will be no further consideration of these parameters.

3.2 Force-Level Related Results

In this section, the focus will be on the force-level related effects. The results can be examined in Table 3, which is sorted along the insertion force column. The nominal differences vary between 12 N and 29 N, in which 80 % of the numbers fall in between 13 N and 17 N. Hence, the nominal differences are relatively constant. The percentage differences lead to another conclusion. As a rule, smaller insertion forces lead to higher percentage differences.

For all the presented components, the difference for the smaller hole was lower than the one resulting from the bigger hole. Therefore, the subjects consistently performed better for higher forces on the same component.

3.3 Individual-Related Results

The comparison of the individual results shows high variation in the ability of the participants in matching the insertion force. The average deviation from the

Table 3 Overview for force-level related tendencies and all subjects

Applied force	InsF	MatF–InsF	MatF–InsF (%)
3	96.29 N	14.47 N	15.0
5	80.57 N	12.61 N	15.7
4	63.94 N	15.28 N	23.9
6	60.23 N	13.37 N	22.2
1	54.17 N	23.73 N	43.8
2	44.57 N	29.25 N	65.6
7	38.05 N	20.06 N	52.7
9	27.72 N	17.69 N	63.8
8	26.64 N	15.81 N	59.3
10	15.28 N	14.40 N	94.3

MatF Matched Force, *InsF* Insertion Force

insertion force varies between 4 N and 40 N. Therefore, a specific selection of the subjects can significantly lower the deviance of the matched force from the insertion force. For the present study the maximum of the average difference decreases step by step to 21 N for the Top 20, 14 N for the Top 15, 11 N for the Top 10 and to 8 N considering only the Top 5. Since it significantly improves accuracy whilst maintaining a sufficient sample size, the Top 10 has been chosen for the following explanations.

For the Top 10 the Bonferroni post-hoc test is generally not significant for the deviation of insertion force and matched force, but neither for the deviation between maximum-force and insertion force. The decision to only account for the Top 10 lowers the interindividual standard deviation of the insertion force by 2–8 % and the one of the matched force by 10 % down to 14 %. The change in the differences and standard deviation can be seen in Fig. 5.

When considering only the Top 10 regarding a high congruity of insertion force and matched force, the average deviation between those expressed in percentage decreases from 45.1 % to 22.9 % (Table 4). The reduction is relatively identical for all components and applied forces. When considering only the participants with the lowest average difference between insertion force and matched force, the deviation is even lower with 7.1 %. Figure 6 shows the results of the best performing subject (number 7) and therewith the potential of the method force-matching.

4 Discussion

The results from the previous chapter demonstrate that the method force-matching is technically applicable for the determination of applied forces. There is a tendency within all subjects to apply a higher force when matching forces than actually necessary to fit the part. This is in line with the results of a previous study by Bao and Silverstein [1], but is in opposition to a study carried out by Casey et al. [4]. However, it is worth mentioning that these authors utilized different applied forces. In a case where the maximum force is the point of reference, the

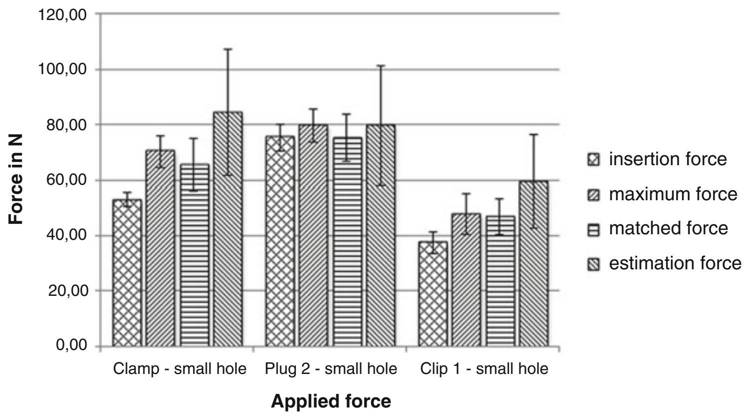


Fig. 5 Results of the applied forces 1, 5 and 7 for the Top 10

Table 4 Overview of the percentage differences of matched force and insertion force for all subjects, the Top 10 and the Top 1

Applied force	1 (%)	5 (%)	7 (%)	Average all components
All subjects	43.3	15.4	52.7	45.1
Top 10	23.3	7.1	25.0	22.9
Top 1	0.9	0.9	21.2	7.1

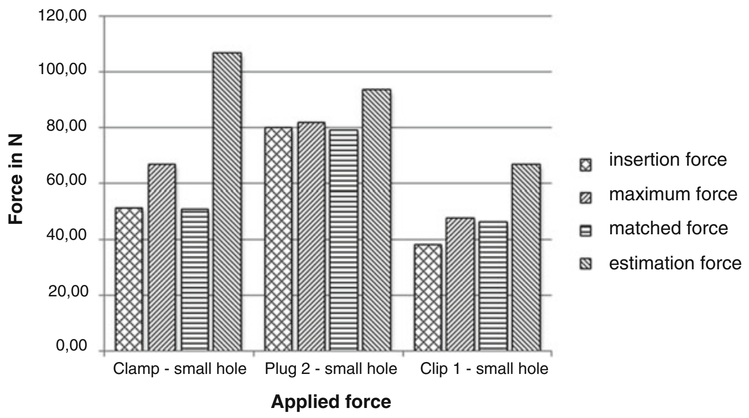


Fig. 6 Results of the applied forces 1, 5 and 7 for subject 7

force-matching offers good precision compared to the direct measurement. Compared to estimations, the method shows higher accuracy and lower variance. Force-matching is therefore the preferred method when it comes to the decision between the previously described. Figure 7 illustrates the position of force-matching as a

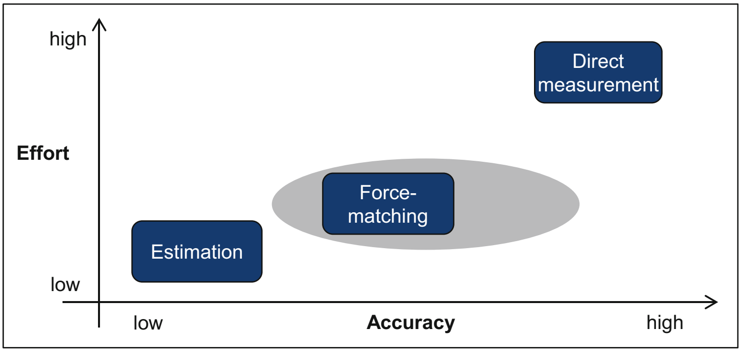


Fig. 7 Positioning of the methods along the dimensions effort and accuracy

compromise in regards to the trade-off between effort and accuracy, and therefore between the methods of direct measurement and estimation of applied forces.

In regards to the methodology of the study, a number of concerns ought to be addressed, beginning with the essential instruction of participants. Since a good instruction may improve the performance of subjects significantly, the question is justified, whether the information for the subjects was sufficient enough and if the method can be furthermore improved with a better instruction.

Furthermore, there was no subjective incentive as such to perform well in terms of matching the insertion force. The performance is therefore based solely on the intrinsic motivation of the subjects and thus could be a key factor for the large variance. This aspect also has impact on the subsequent utilisation of the method. If the force-matching is performed by employees from production, the incentive could be to apply higher forces in order to achieve better working conditions. On the other hand, planners could have an incentive to apply lower forces in order to obtain a better evaluation of the workplace. Motivation was not included as a factor in past studies, but had verbal encouragement and the subsequent increase in motivation had a positive influence on the maximum force of volunteers in past studies [6].

The experimental design of this study cannot fully evaluate the consistency of the measured values over time. In order to do so, a second series of experiments would be necessary. Thus, the reliability of the results is limited. However, the reproducibility of force-matching experiments using a dynamometer has been classified as high [11].

The validity of obtained results may be limited due to the fact that there is a lack of a ‘gold standard’ for the determination of forces as a reference point. Furthermore, there is a standard deviation of the joining force between the subjects of 11 %. The spread between the subjects in all methods could also be partly caused by exhaustion effects. In the evaluation of workplaces, an additional risk comes from a misjudgement of forces that are considered too low. However, force-matching consistently shows higher values across all components in relation to the insertion force. By using candidate-pools, the deviation of 45 % can be reduced to 23 %. In this way a result suitable for the evaluation can be generated without precise knowledge of the true value.

In terms of behavioural ergonomics, the question of what might be the reason for the strong variance of skills between subjects is especially interesting to consider. This has also been observed in similar experiments [2, 4, 5]. Some potential reasons include individual differences in understanding the instructions or the fundamental ability to reproduce a given force. In this context, the preselection of relevant volunteers and the creation of a candidate-pool would make sense. For this purpose, a large number of individuals have to be tested one-time in order to generate a sufficiently large pool. Only members of the candidate-pool may then use the method of force-matching. The potential of the method is clearly visible when looking at the experimental results of the best performing subject. In the study at hand, it remains unsolved, whether force-matching appears to be a capability (inherent capacity) or a skill (gained through exercise). Supporting the notion that force-matching is a capability is the fact that it is not related to the experience with force measurements.

5 Conclusion

The present study aimed to develop and validate a new method for the determination of applied forces in the automobile industry. With the definition of force-matching as a valid method and its subsequent investigation experimental trials, this objective has been achieved.

In summary, the use of force-matching as a method for the determination of applied forces in automobile production is possible. The results turn out to be close to the ones of direct measurements considering only a preselected pool of participants. Force-matching should therefore be the preferred method over the estimation of forces and has potential for an increased quantity and quality in the determination of applied forces in the ongoing production.

The results obtained were provided by untrained subjects that had no prior knowledge of force matching, and therefore represents only a starting point and a promising foundation for future studies. Further research questions should be investigated by using the results of this study as a basis. Pertaining to the applicability of force-matching of a method, the role of complex components and further gripping types are interesting examples to serve as questions for future endeavours.

Acknowledgements Special thanks to Dr. Sonja Schmicker, Mr. Ulrich Brennecke and Mr. Stefan Wassmann from the Institute of Ergonomics, Manufacturing Systems and Automation (IAF) at the Otto-von-Guericke University Magdeburg for the constructive cooperation.

References

1. Bao S, Silverstein B (2005) Estimation of hand force in ergonomic job evaluations. Ergonomics 48(3):288–301
2. Bao S, Spielholz P, Howard N, Silverstein B (2006) Quantifying repetitive hand activity for epidemiological research on musculoskeletal disorders – part I: individual expo-sure assessment. Ergonomics 49(4):361–380
3. Borg GA (1982) Psychophysical bases of perceived exertion. Med Sci Sports Exerc 14 (5):377–381
4. Casey JS, McGorry RW, Dempsey PG (2002) Getting a Grip on Grip Force Estimates. In: Professional Safety 47(10), p. 18
5. Gandevia SC, Kilbreath SL (1990) Accuracy of weight estimation for weights lifted by proximal and distal muscles of the human upper limb. J Physiol 423:299–310
6. Jung K, You H, Kwon O (2006) Evaluation of the FSA Hand Force Measurement System. Proc Hum Factors Ergon Soc Annu Meet 50(4):2577–2581
7. Kroemer K (1977) Die Messung der Muskelstärke des Menschen. Wirtschaftsverlag NW, Bremerhaven
8. Schaub K, Caragnano G, Britzke B, Bruder R (2013) The European Assembly Worksheet. Theor Issues Ergon Sci 14(6):616–639
9. Walther M (2015). Entwicklung und Evaluierung eines systematischen Vorgehens zur Erfassung von Aktionskräften in der Automobilproduktion. Dissertation, Technische Universität Chemnitz, Professur Arbeitswissenschaft (Unpublished)
10. Walther M, Kunze T, Toledo Munoz B, Büttner B, Kaiser A, Bullinger A (2013) Kraftmessungen an Schnappverbindungen in der Automobilproduktion. Chancen durch Arbeits-, Produkt- und Systemgestaltung – Zukunftsfähigkeit für Produktions- und Dienstleistungsunternehmen. GfA-Press, Dortmund, pp 237-240
11. Wiktorin C, Selin K, Ekenvall L, Kilbom A, Alfredsson L (1996) Evaluation of perceived and self-reported manual forces exerted in occupational materials handling. Appl Ergon 27:231–239

Part III

Design of Processes

Boundary Management as a Crucial Success Factor for Flexible-Mobile Work, Demonstrated in the Case of Home Office

Boundary-Types, Boundary-Management and Boundary-Tactics Used in Home Office

Leila Gisin, Hartmut Schulze, and Barbara Degenhardt

Abstract

The compatibility of work life and family life is a key factor in today's modern work arrangements, particularly with regard to the increase of transition to flexible working hours and places. The herein presented research study considers the question, whether the boundary-types proposed by the boundary-theory [cf. Ashforth et al. Acad Manag Rev 25(3):472–491, 2000; Nippert-Eng, Home and work: negotiating boundaries through everyday life. University of Chicago Press, Chicago 1996]. can be ascertained within flexible-mobile working, notably upon working in home office. Furthermore which difficulties the distinct boundary-types have with these new work styles and which boundary-tactics are promising in regard to those difficulties. Based on a mixed method approach consisting of an online survey (N = 395) and subsequent semi standardised interviews (N = 9) the boundary-types segmentor and integrator could be confirmed. In addition an intermediate mix type was found and empirically proven. Moreover, there is evidence that this newly described mix type has the most difficulties with his boundary-management within flexible-mobile work styles. Based on these findings the article concludes in guidance suitable for each boundary-type in favour of working in home office.

Keywords

Boundary theory • Boundary-management • Boundary-types • Boundary-work-tactics • flexible-mobile work • Home office

L. Gisin (✉) • H. Schulze • B. Degenhardt
Institute for Research and Development of Collaborative Processes, School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland
e-mail: leila.gisin@fhnw.ch; hartmut.schulze@fhnw.ch; barbara.degenhardt@fhnw.ch

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_25

1 Introduction and Central Questions

The present-day changes related to the so called knowledge society is accompanied by an increasing flexibilization of work hours and places, as knowledge work is neither temporally nor locally tied in principle [cf. 10, 19]. This so called "autonomization of work" [9] poses great challenges to employees as well as organizations. On individual level the boundaries between private and work life, which have been arisen by the industrial revolution and culturally handed down over generations, are dissolving. Significant areas of tension are constituted by means of permanent availability [2], working during ones spare time as well as the relationship between work and family in general [1]. In worst case, these areas of tension can lead to an excessive increase of sleep and health problems, thoughts circling constantly around work and therefor leading to a lasting sense of overstraining, negligence of the social environment, and conflicts in relationships with partner/family [cf. 8, 15]. Individuals which work flexible in regard of time and space are facing the challenges of boundary management in terms of having to re-examine the relationship between work and private time and adjust it suitably.

Within this context Kreiner et al. [12] identified several so called "Boundary-Work-Tactics" based on the "Boundary Theory" presented by Ashforth et al. [4] as well as the "Boundary-Types" sketched out by Nippert-Eng [14]. These "Boundary-Work-Tactics" include all strategies which help individuals to draw, defend and maintain boundaries in daily routine. Moreover, whichever "Boundary Work Tactic" is applied in which situation depends in no small part on the individual "Boundary-Type". According to the individual positioning on the role segmentation-role integration-continuum (see Fig. 1), differing "Boundary-Types" can be distinguished. They vary from "stark segmentors", which separate maximally between work and private life and are known for setting rather impermeable boundaries, to "stark integrators" which in contrary integrate work and private life up to a maximum and are known for setting permeable boundaries. Between these two extremes at least one "Mix Type" is assumed, but has not been further explored in previous research.

Previous research on flexible-mobile work and the therewith related home office work only touched upon the issue of "Boundary-Types", "Boundary-Management" and "Boundary-Work-Tactics". Notably there hasn't been any in-depth exploration

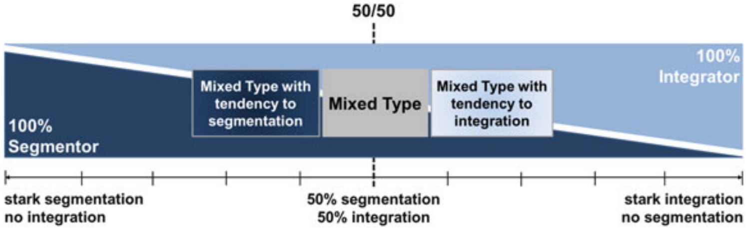


Fig. 1 Division of the role segmentation-role integration-continuum into boundary-types

in this regard; neither concerning the examination of potential boundary-types [cf. 12, 14] within the population of home office workers, nor in view of the individual challenges posed by this new work style, and the thereupon matching strategies in use to establish personal boundary congruence. Thereby boundary congruence signifies a good fit between individual needs for boundaries and the according boundary needs of the corresponding environment, and constitutes a crucial requirement for professional success and a healthy work-life balance within flexible-mobile work arrangements, in particular of the home office situation. The latter distinguishing itself specifically through blurring boundaries between job and private life. As already indicated above, previous research has focused on the archetypal poles of "stark segmentors" and "stark integrators". The mixed types lying in-between these poles are only mentioned in passing, eventhough they represent a not neglecting amount of individuals. The scientific relevance of the here presented study is located in the investigation of this research gap.

Basing on the portrayed challenges of flexible-mobile knowledge work and building on the studies of Ashforth et al. [4] and Kreiner et al. [12], following three central questions have been researched empirically in this study:

Central question 1: Can differing boundary-types be distinguished and characterized within flexible-mobile Work?

Central question 2: What impact does flexible-mobile Work have on in central question 1 identified boundary-type, with regard to health and performance? Which strains can be scientifically proven with special regard to the assumed Mix Type?

Central question 3: Which Boundary-Work-Tactics are used by which boundary-type within flexible-mobile working, notably while working in Home Office?

2 Theoretical Principles and Disambiguation

In order to embed the study and its resultant findings scientifically, the associated theoretical frame of reference is explained and key terms clarified hereafter.

2.1 Flexible-Mobile Work Incl. Home Office as a New Style of Working

With "flexible-mobile work" Schulze et al. [18] denote a new way of working. This new work style is characterized by being independent from place and time, respectively contains work which is yielded in varying places and at different times, and clearly contrasts with the so far common understanding of work being produced at a specific place (e.g. main office) and at certain periods of time (e.g. 8 am to 5 pm); furthermore "flexible-mobile work" is related to an employment relationship and the existence of a workplace at the parent company. Hence this work style focuses mainly on paid employment. The characterizing features of flexible-mobile work are, that a significant amount of working hours are performed outside the premises of the parent company at other places such as, for example at home, visiting customers or partners, when traveling by train, at the coffee shop or a co-working space. Working within the premises of the parent company is explicitly included. Schulze et al. [18] introduce this concept despite certain overlaps with the concept of telework [16], since emphasising the use of technology like in the case of telework, no longer suffice for differentiation between flexible-mobile forms of work and mere desk work (when on-site at ones company) in consequence of its implicitness nowadays. In the authors' view, the nature of flexible working methods consists of its space-time flexibility, which is expressed more adequately with the term "flexible-mobile work". The present study builds upon this comprehension and in doing so focuses on the aspect of working from home as a central place where flexible-mobile work is rendered. According to a representative survey of the Swiss working population on the issue of flexible-mobile work and working at home [20] 53 % of the 4.3 million Swiss salaried employees could in fact be working mobile due to their work tasks, but it is just under onefourth (23 %) which do so at least several times a month. The potentials on both ends (employee wise and company wise) have not been exploited to date. The situation is similar in respect of Germany, where arrangements of working at home are actually decreasing. As per Brenke [5] only 8 % of Germans working population worked partially at home in 2012. In addition to the reluctance of companies and organisations to establish flexible-mobile work systematically, the partial repatriation of workplace into home surroundings (private life) also plays a vital role in these latest developments. Ultimately it's all about cultural changes, which pose new requirements in view of boundary management of flexible-mobile workers.

2.2 Boundary Theory, Boundary-Types, Boundary-Management and Boundary-Work-Tactics

The boundary theory [4] focuses on the way in which individuals create, maintain or change boundaries, in order to simplify and classify the world around them. As a general rule, these boundaries determine the extent and scope of application of life domains such as work life and private life. The quality of these boundaries can be located on a continuum from "thin/weak", which makes them permeable and open for the integration of other life domains, to "thick/strong", which renders them impermeable and in doing so lead to the segmentation of life domains [14]. Hence segmentation and integration represent two opposing poles and lead to certain mind-sets, which differ in their overall approach to work-life-balance [14].

Nippert-Eng [14] records, that all social boundaries drawn, be that mentally or physically, are social constructs which originate from humanly classification processes. These classification processes being the essential element of culture in turn. The biggest controversies and predicaments of our times deal with boundaries and how we draw them, for they implicate when and how we interact with each other [14]. Viewed in this context, the process of socialization is nothing else than a lifelong process of learning, applying, rejecting and defying of classificatory boundaries and schemes. The boundary between work life and private life thereby being one of the most omnipresent and commonplace demarcation, where the individual is constantly called upon to draw, adjust, defend or if need be abolish boundaries with help of so-called boundary-work-tactics [cf. 12, 14]. This constant editing of cultural boundaries is thereupon named boundary management and includes all strategies, principles and practices we apply to create, maintain and modify cultural categories [14]. It is by means of this individual boundary management that the extent of personal need for integration and/or segmentation with regard to the work and private life domains is determined. In fact it establishes congruencies and distinctions between these two life domains. None the less it should be noted, that although the degree of individual need for delimitation is strongly related to personal preferences and cultural nurture, it isn't steady over time. This fact is not least related to the ever-changing environmental conditions to which individuals have to adapt, what in turn can cause changes in individual preferences with regard to delimitation needs. However, these adjustment performances ought to cause just small changes in the personal allocation on the role segmentation-role integration-continuum [3].

It is through a qualitative study conducted with laboratory employees that Nippert-Eng [14] discovered differences in the manner of how individuals create congruency between the work domain and private domain. While so-called "segmentors" (setting strong and impermeable boundaries) e.g. carry separate agendas and key rings for their work and private life, and normally keep these two domains separated from each other, so-called "integrators" (setting weak and permeable boundaries) hold on to one agenda and one key-ring for both domains and mix these living environments as required. Whilst segmentation is reflected in a fundamental separation of the domains work life and private life (e.g. in thoughts, worries and physical markers), integration represents the merging of both domains, up to the degree where no distinction is feasible between work and private anymore [cf. 12, 14].

In the model of "Work-Home Boundary Work" developed by Kreiner et al. [12] it becomes apparent that these individual need for segmentation resp. integration encounter the then again individual segmentation/integration needs of the social environment (e.g. family members, superiors, where applicable employees, profession, organizational culture). The more these segmentation/integration needs differ from those of the social environment, the more boundary incongruences can be found between personal needs and the needs of the corresponding stakeholder group. As a result, these boundary incongruences lead to so-called "Work-Home-Conflicts" (also referred to as Work-Family Conflicts in literature) caused by boundary violations due to border crossings or the setting of as unnecessary perceived boundaries. According to Kreiner et al. [12] work-home-conflicts can be overcome with the help of suited "Boundary-Work-Tactics" for the purpose of an ideal person/environment fit and hence boundary congruence. A simplified

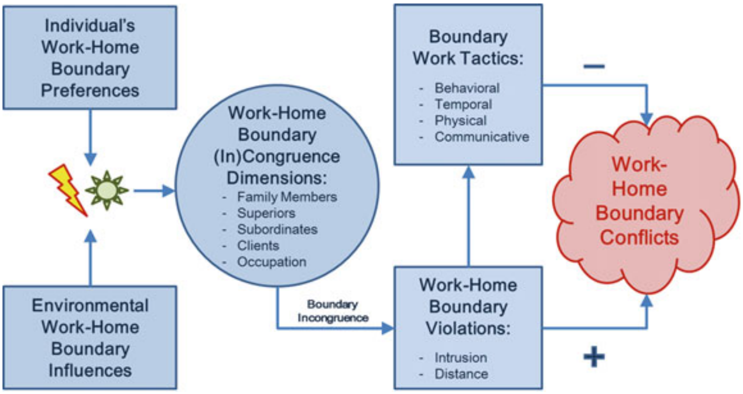


Fig. 2 Simplified "Work-Home Boundary Work" model, in accordance to Kreiner et al. [12, p. 711]

"Work-Home Boundary Work" model (in accordance to Kreiner et al. [12]) is illustrated in Fig. 2.

Kreiner et al. [12] have elaborated several boundary-work-tactics by means of two qualitative studies conducted with parish priests and have subdivided them into four categories labelled "behavioural", "temporal", "physical" and "communicative". Kreiner et al. [12] state, that they specifically chose to work with parish priests, as the nature of their occupation is particularly demanding in view of being aware of one's boundary needs and reacting upon them with the help of an appropriate boundary management. Such extremes lend themselves suitable for the construction of new theories, as they may clarify and illustrate the underlying dynamics more clearly thanks to the very quality of the extremes themselves.

This classification system of boundary-work-tactics was constructed by Kreiner et al. [12] based on detailed analysis and evaluation of the data collected in above mentioned studies, and subsequent comparative and contrasting discussion on the findings and within the research team; thereby including evidence found in previous research. Kreiner et al. [12] complete their report with the message, that the introduced reference framework resp. the typology of boundary managers shouldn't be over interpreted, as in reality conceptual overlapping can be observed, and described tactics aren't standalones; rather on the contrary they're often combined at random and in regard of the specific boundary need at hand. In an extreme case it is quite possible that the transposing of a personal boundary need involves combining integrative and segmentative boundary-work-tactics in each other. As a concrete example Kreiner et al. [12] quote a parish priest, who on one hand has his occupational mobile phone on him during holidays (integrative, behavioural tactic), for the reason of wanting to be accessible for his parishioners in case of an emergency, but let's his wife take the call, who determines if there is indeed an emergency at hand (segmentative, behavioural tactic).

Kreiner et al. [12] essentially state that (a) the proper use of boundary-work-tactics results in boundary congruency thus reducing boundary violations and

(b) the use of multiple boundary-work-tactics (within and between the categories as stated above) has a synergistic effect, which delivers additional active game facing the challenges of managing the boundaries between work life and private life.

3 Methodical Approach

The present study builds on a mixed-method scheme, combining an online questionnaire survey with a subsequent qualitative interview study. Further the quantitative and qualitative findings are triangulated per central question. This approach appeared particularly reasonable for the explorative central questions posed, as the boundary-types and their correlations with variables such as occupational strains and health could best be demonstrated through quantitative data, whereas substantiating found interdependencies is best exemplified with deeper analysis of semi-structured interviews with representatives of the quantitatively encountered boundary-types. These proceedings were successfully achieved and are described in detail hereafter.

Quantitative methods and data: Following the official "Swiss Home Office Day 2013", a widely scattered online survey has been conducted through the Institute for Research and Development of Collaborative Processes FHNW, in collaboration with the Swiss Home Office Day Consortium. A total of $N = 562$ individuals have taken part in this survey, whereof $n = 395$ experienced home office users answered the included items regarding boundary-typ and boundary-work-tactics in use. Roughly 49 % of these 395 experienced home office users are female, 51 % male. Further the majority of 85 % are between 26 and 45 years old ($M = 42.62$, $SD = 10.74$) and salaried employees (89 %). The collected data were statistically analysed with the help of the SPSS Statistics 21 and Microsoft excel software with regard to the central questions posed (also see Sects. 1 and 3.1–3.3).

Qualitative methods and data: Based on the quantitative findings from the home office day 2013 survey, the central questions 1 through 3 were further explored by means of $N = 9$ semi-structured interviews with representatives of all three statistically verified boundary-types [cf. 6].[1] For the recruitment of the interview candidates a modified "theoretical sampling" [7] has been contrived. In this specific case, three groups of resembling types (segmentors, mix type and integrators) were determined, which vary in themselves in regard to the potentially relevant factors gender and age. Thus the sample consists of five female and four male interviewees, involving an age-range from 26 to 55 years ($M = 39.89$, $SD = 9.28$). Except for one person, all interviewees are salaried employees. Due to the small case number, additional potentially relevant factors could not be included (like for instance

[1] For an in-depth understanding of the methodical approach and presented results the authors refer to the master thesis "Boundary-Typen, Boundary Management und Boundary Taktiken im Home Office" [6].

regularity of home office use or hierarchical position within the organization could have been).

Data Triangulation: In conclusion and in terms of a mixed-method approach, the quantitative and qualitative findings have been triangulated [cf. 6] in order to enhance the validity of later presented research results. Moreover systematic errors can be reduced by triangulating varying data sets.

3.1 The Distinction of Boundary-Types in Case of Working in Home Office

In a first step, in order to scientifically prove the existence of distinguishable boundary-types within flexible-mobile work, the role segmentation-role integration-continuum was divided as follows into the presumed boundary-types segmentor, mix type and integrator (also see Fig. 1 and Table 1): 100 % segmentation needs to 70 % segmentation needs (30 % integration needs) = boundary-type segmentor (N = 243); 60 % segmentation needs (40 % integration needs) to 40 % segmentation needs (60 % integration needs) = mixed boundary type; 30 % segmentation needs (70 % integration needs) to 100 % integration needs = boundary-type integrator.

As a second step, these in this manner created boundary-types haven been compared with the items on working at night or on weekends, by means of pivot tables with Chi-square tests. According to theory, there should be significant differences between segmentors and integrators. Based on Nippert-Eng's work [14] it was assumed, that segmentors tend to adhere to traditional working hours and days (8 am–5 pm, Monday through Friday), whereas integrators more likely work scattered throughout the day and week (including weekends), since the life domains work and private aren't separate spheres in their eyes. To obtain information on which subgroups significantly differ from each other, pairwise comparisons based on the exact test of Fisher and Yates were performed.

To be able to characterize these quantitatively identified boundary-types, the 9 semi-structured interviews with N = 3 segmentors, N = 2 mixed types, and N = 4 integrators have been coded and analysed with regard to specific characteristics, thereby using the qualitative content analyses method by Mayring [13]. The interview guide was developed following the model of Kreiner et al. [12] and covering central questions 1 through 3. The assignment of boundary-type was also determined by means of a prior self-allocation of the interviewees on the role segmentation-role integration-continuum (see Fig. 1 and Table 1).

3.2 Health and Productivity of Boundary-Types When Working in Home Office

To establish what probable effects flexible-mobile work can have on the documented boundary-types, varying health, performance, and satisfaction

Table 1 Division of the role segmentation-role integration-continuum into "segmentors", "mix type" and "integrators"

Segmentor	Mix Type	Integrator
100 % segmentation 90 % segmentation, 10 % integration 80 % segmentation, 20 % integration 70 % segmentation, 30 % integration	60 % segmentation, 40 % integration 50 % segmentation, 50 % integration 40 % segmentation, 60 % integration	30 % segmentation, 70 % integration 20 % segmentation, 80 % integration 10 % segmentation, 90 % integration 100 % integration

parameters have been compared with regard to their central tendencies using the Kruskal-Wallis test (H-Test). Where significant differences were found, pairwise comparisons were conducted on top; in order to find out which subgroups (boundary-types) are involved.

The transcribed data collected during the qualitative interviews were coded and analysed with regard to central question 2, thereby again using the qualitative content analyses method by Mayring [13], but this time around focusing on type-specific difficulties with flexible-mobile work.

3.3 Type-Specifically Boundary-Tactics

In order to find out which boundary-types apply what kind of boundary-work-tactics to create personal boundary congruency, the open-ended answers regarding boundary-work-tactics in the home office day survey 2013 have been assigned per type to the category system of boundary-work-tactics structured by Kreiner et al. [12]. Where advisable, the category system has been expanded in-vivo with additional subcategories. Finally the tactics were counted out and frequencies were compared between the distinct boundary-types.

The transcribed data collected in the interviews were coded with regard to central question 3, and analogously assigned per type to the category system of boundary-work-tactics [12] and finally counted out and compared between types.

4 Results

In this section the results of previously described research are presented ordered by central question 1 through 3.

4.1 The Distinction of Boundary-Types in Case of Working in Home Office

As shown in Table 2, significant differences were found between segmentors and integrators with respect to working at nights and on weekends (Working Saturdays,

Table 2 Exact test by Fisher and Yates for pairwise comparison of boundary-types with regard to working at night and on weekends

Variable	Segmentor vs. integrator		Segmentor vs. mix type		Integrator vs. mix type	
	p	Cramer's V	p	Cramer's V	p	Cramer's V
Working Saturdays	.000***	.27	.118	.09	.009	.23
Working Sundays	.001***	.21	.001**	.19	.601	.05
Working between 8:00 p. m. and 9:59 p.m.	.000***	.27	.077	.10	.024	.19
Working between 10:00 p.m. and 05:59 a.m.	.000***	.25	.012	.14	.067	.06

[Note. *p ≤ .004; **p ≤ .001]

p ≤ .001, V = .27; working Sundays, p ≤ .001, V = .21; working between 10:00 pm and 05:59 am, p ≤ .001, V = .27; working between 8:00 and 9:59 pm, p ≤ .001, V = .25).

While segmentors (as supposed) tend to adhere to traditional working hours (between 8 am and 8 pm, Monday through Friday), integrators work scattered over the course of the day and week. The effect sizes indicate small to medium-sized effects. The results concerning the mix type in between draw a less clear picture on their behaviour in regard to working hours. Nevertheless as expected, he is situated between segmentors and integrators in terms of frequencies.

The analysis of the interview data revealed, that the statistically found differences lead back to the archetypical peculiarities of segmentors and integrators, as already outlined by previous literature on the topic of boundary theory and boundary management. Whilst segmentors try to keep work life and private life separate according to possibility, prefer regulated working hours and need a lot of structure and clearly defined time periods for recovery at any rate, integrators tend to interweave work life and private life as time opportunely as possible, therefore experience regulated working hours as obstructive, and are in need of flexible and permeable structures, where they can fit in recovery periods wherever they make sense within their actual work day or work week.

Further the interview data enables to provide a first tangible characterization of the mix type, which hasn't been specified in previous literature. In the present study it becomes apparent, that mixed types have great difficulties assigning themselves distinctly to one of the poles (segmentation/integration). Ultimately, the decision of mixed types over what is handled in an integrative manner (perceived as important) and what in a segmented manner (perceived as unimportant) is steered by their

Table 3 Comparison of the individual characteristics of boundary-types

Criterion	Boundary-Type Segmentor	Mix Type	Integrator
Self-Placement boundary-type	Has no difficulties to place himself/herself on the role segmentation-role integration-continuum	Clear allocation on the role segmentation-role integration-continuum is rather difficult	Has no difficulties to place himself/herself on the role segmentation-role integration-continuum
Attitude towards working in home office	Home Office (HO) only if situational advantages outweigh	Ambiguous	HO facilitates flexibility, which is much appreciated
Working energy while working in home office	Needs rituals to create the right working energy in home office	Creates working energy when the task at hand is perceived as important	Has no troubles creating the right working energy, no matter where and when
General attitude towards place of work	Prefers to work in MO (main office) whenever possible. Even if overtime is implied	Tends to prefer working in MO, but doesn't always behaves that way	Works wherever it makes sense for the tasks at hand
General attitude towards work hours	Prefers working normal office hours (+/− 8 am–5 pm)	Wants and tries to maintain normal office hours, but doesn't always succeeds	Doesn't need regular working hours. Works whenever it makes sense
General handling of technologies	Keeps technologies separate, whenever possible. No automatic synchronisation	Handles technologies quite disparately (sometimes integrating, at other times separating them)	Whenever possible technologies are integrated. Constant synchronisation
Handling of structures	Needs a lot of structure and defined periods of recreation and relaxation	Knows that he/she needs structure, but doesn't always applies them	Doesn't needs a lot of structure. Structures are obstructive and tiresome
Handling of recovery time	Goes offline on weekends and during holidays. Is reachable by text message for emergencies. Going offline as a condition for recreation and relaxation	Are reachable for (self-perceived) important project/tasks at all times. For everything else going offline is more relaxing	During times of recreation decreased reachability. Going completely offline is not possible, would lead to decrease of perceived recreation
Embodiment of ideal leadership	Needs a segmentor as superior, or an integrator with a good understanding for segmentors	A segmenting superior is more favourable for this boundary-type, in terms of a protective function	Has to be managed as independently as possible. Needs maximal temporal freedom

underlying personal prioritization scheme. Between all detected boundary-types it is the mix type which has the most difficulties to apprehend and act upon his personal boundary needs, as following interview quote (translated from Swiss German) exemplary demonstrates:

> . . . Well, what I notice is, that as a result of being a mix type, I realize, that when I try to set boundaries I'm not taken seriously . . . I'm not able to set the necessary boundaries I need . . . (Mix Type 1)

These quantitative and qualitative findings are taken as prove and confirmation, that distinguishable Boundary-Types can be found in regard to flexible-mobile working. The existence of segmentors and integrators is validated and the mix type documented for the first time. Table 3 provides a summary of the individual characteristics of boundary-types, as substantiated on basis of the interview data collected.

The triangulation of the quantitative and qualitative results (with regard to central question 1) imply, that the boundary-types segmentor, mix type and integrator can be identified within flexible-mobile work arrangements and that they differ in view of significantly relevant work characteristics. The experienced home office users from the home office day survey 2013, which have been identified as integrators due to their self assignment on the role segmentation-role integration-continuum, work significantly more often at marginal times and on Weekends, than the identified segmentors do. The interview results reinforce these findings on the existence of Boundary-Types by being able to provide evidence and explanatory approaches for various differences in regard to relevant work characteristics.

4.2 Boundary-Type-Specifically Loads in Case of Working in Home Office

According to the statistical analysis of health, performance and satisfaction parameters of experienced home office users the proven boundary-types only differ significantly in their perceived efficiency within flexible-mobile work arrangements (see Table 4). The pairwise comparison shows, that it's the mixed type which feels significantly less efficient while working in home office, than segmentors ($\chi^2(2) = 34.66$, $p \leq .05$, $r = .22$) and integrators ($\chi^2(2) = 60.03$, $p \leq .01$, $r = .18$) do. The effect sizes of these differences are small to medium (cf. Table 4).

The interview data analysis with regard to central question 2 underpins the quantitative findings. The mix types' difficulties with flexible-mobile work arrangements start with having difficulties observing their own boundary needs and respond to them in a reasonable manner. This for instance and among other things leads to not being able to stringently enforce chosen boundary-work-tactics. Regarding the environment of mixed boundary types it gets evident, that stakeholders in turn also have great difficulties classifying the alternating work behaviour of mix types in view of segmenting and integrating between work life

Table 4 Pairwise comparison of the variable "improved job performance"

Variable	Segmentor vs. integrator			Segmentor vs. mix type			Integrator vs. mix type		
	χ	Adj. p	r	χ	Adj. p	r	χ	Adj. p	r
Improved job performance	−1.61	.651	–	2.72	.040*	.22	−3.36	.005**	.18

[Note. N = 395–399; *p ≤ .05, **p ≤ .01, ***p ≤ .001 (two-tailed)]

and private life. The consequence is that mixed types' segmentation efforts often aren't taken seriously by the immediate environment.

Although segmentors (lack of boundary mechanisms, rituals and appropriate integrational boundary-work-tactics while working at home) as well as integrators (lack of appropriate segmentative boundary-work-tactics while working at home) have initial difficulties with flexible-mobile work arrangements, they're both quite capable in facing them with appropriate boundary-work-tactics and hence in short fully benefit from the various advantages associated with flexible-mobile work.

The following table (Table 5) presents an overview of each boundary types' specific difficulties with regard to flexible-mobile work arrangements, in particular with regard to working at home.

Contrasting (triangulation) the quantitative with the qualitative findings in regard to central question 2 it gets evident, that the disparate boundary-types have distinguishable difficulties with flexible-mobile work arrangements resp. with working from home. It gets thereupon clear that it is particularly the mixed types which have the most difficulties with this new way of working. Individuals who assign themselves to the mixed type feel less productive while working at home, than segmentors or integrators do. When considering the figures of the remaining examined areas (work-life-balance, ERI-quotient, health, sleeping quality, motivation, structuring problems and satisfaction with working from home), which haven't turned out significant, it gets apparent that mixed types have invariably the lowest figures, followed by segmentors, who marginally contrast with integrators, which have the best values in regard to all queried health, performance and satisfaction parameters. These quantitative findings are then again substantiated by the interviews. Here, too, it gets obvious that it's in particular the mix type, who has the greatest difficulties with flexible-mobile work arrangements.

4.3 Application of Type-Specifically Boundary-Tactics

It can be stated in general, that spread over all three identified boundary-types all by Kreiner et al. [12] suggested categories and subcategories of boundary-work-tactics could be accounted for. Within the category of behavioural based tactics two additional subcategories could be opened up for flexible-mobile workers [cf. 6].

The content-analytically evaluation of the quantitative and qualitative data collected in regard to boundary-work-tactics in use yielded, that segmentors

Table 5 Boundary-type-specifically loads within mobile/flexible work arrangements

Criterion	Boundary-type		
	Segmentor	Mix type	Integrator
Personal difficulties	The mobile/flexible work style requires lots of demarcation efforts and rituals for the role transitions Working in home office requires adaptive performance	Has troubles identifying his/her boundary management needs and acting appropriately Has troubles enforcing chosen boundary-tactics	Has troubles keeping an overview of accomplished work hours Is very challenged to create and utilize scopes for development
Difficulties with the social environment	Has troubles dealing with heavily integrative organisational cultures and individuals Needs appreciation for the selective exemptions made for the sake of good will	The social environment has difficulties classifying his/her work style Is influenced by the social environment due to his/her ambiguous attitude. Segmentation efforts are not taken seriously by the social environment	Has troubles dealing with heavily segmentative organisational cultures and individuals Collaborating time synchronously is difficult, as integrators don't stick to the standard working hours

apply above all segmentative boundary-work-tactics (88 %) and integrators by the majority engage in integrative boundary-work-tactics (61 %) in order to establish their personal boundary congruency between individual needs and those of the environment. Regarding the mix type, the proportion of segmentative to integrative boundary-work-tactic in use lies between segmentors and integrators, whereas they indisputably utilize more segmentative tactics (see Fig. 3).

Even for the situational application of boundary-work-tactics differences can be stated between boundary-types. In case of segmentors results imply, that they predominantly use segmentative boundary-work-tactics in order to ensure the strong demarcation and distinction of the life domains work and private. Integrational strategies are introduced as exceptions, for the reason of wanting to generate a better fit between individual boundary needs and those of the immediate environment (boundary congruency), and in doing so maintaining their desired demarcations over time. Further segmentors apply integrational strategies when working at home, where they've realized, that a partial integration of private elements into their work life can have its benefits for their work life balance. In contrast to this, integrators tend to stick to integrational boundary-work-tactics, whereas they also use a considerable amount of segmentative boundary-work-tactics. These segmentative boundary-work-tactics serve integrators as enablers for the creation of recreation spots they need to boost their energy levels. For the constant integration of work life and private life definitely also has its downsides. Continuingly, integrators explicitly use segmentative boundary-work-tactics when working at home, where both domains (work/private) strongly overlap, and a minimal structuring gets necessary for the otherwise very flexible and boundary-shy integrator. Mix types on the other hand use segmentative and integrative

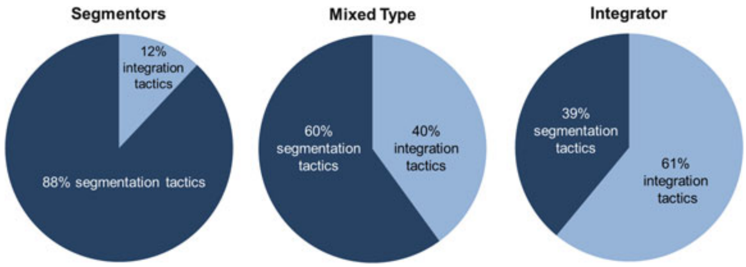


Fig. 3 Overall distribution of tactics in use according to type and sort (survey and interviews)

boundary-work-tactics to an equal degree, whereby segmentative boundary-work-tactics slightly outweigh. On closer inspection it gets moreover obvious, that mixed boundary types indeed try applying boundary-work-tactics to translate their boundary needs, but don't pursue them stringently. Whilst segmentors and integrators are comparatively successful in the application of their individual and situational appropriate boundary-work-tactics, mixed boundary types have difficulties identifying and applying them.

With central question 3 as well as with the previous two central questions, the quantitative results converge with the qualitative ones. Regarding this, the triangulation of these findings substantiates the fact that segmentors tend to use more segmentative boundary-work-tactics and integrators more integrative boundary-work-tactics within flexible-mobile work arrangements. Mixed boundary types show a slight tendency towards using more segmentative than integrative boundary-work-tactics in both data-sets. Complementary explanations for the selective integrational efforts of segmentors and selective segmentational efforts of integrators are delivered through interview data. In regard to mixed boundary types the interviews obviously point out, that they not only have trouble with the enforcement of boundary-work-tactics, but also are the boundary-type which name the fewest boundary-work-tactics and hardly elaborate them.

5 Discussion, Recommendation and Prospect

In this last section previously presented results are discussed, summarised and reflected (Sect. 1.5.1), further based on discussed results, recommendations are derived for flexible-mobile workers as well as for the organization of heterogeneous teams in flexible-mobile work arrangements (Sect. 1.5.2), and finally an outlook for future research is pointed out (Sect. 1.5.3).

5.1 Summary and Methodological Reflection

In this study a mixed method approach has been chosen to answer the question, whether the boundary-types suggested by boundary theory [4] can be determined in case of flexible-mobile work, especially in the case of working at home. Thereon the goals were to characterize identified boundary-types in context of flexible-mobile work and name their type-specific difficulties with regard to boundary management in flexible-mobile work arrangements. Furthermore, research was conducted into which of Kreiner et al. [12] suggested boundary-work-tactics are used by whom (boundary-type) in order to create boundary congruency, and if at best additional categories and subcategories of promising boundary-work-tactics could be identified and labelled.

In regard to boundary-types previous research has primarily focused on the opposites of segmentors and integrators [cf. 4, 12, 14]. The intervening mix type although mentioned, is not described in detail. By means of research on central question 1 and 2 (see Sect. 1.1) the boundary-types segmentor and integrator could be confirmed and the mixed type accounted for first-time. Distinct differences were found in view of their boundary management while working in flexible-mobile work arrangements. It is primarily the mixed boundary type which has the most difficulties in understanding his demand situation concerning boundary management and lacks handy boundary-work-tactics to reduce the therefore experienced boundary incongruences.

Research around central question 3, which focused on boundary-work-tactics and how they're used by the varying boundary-types in flexible-mobile work settings, clarified that all boundary-types involve the boundary-work-tactics suggested by Kreiner et al. [12] in their daily boundary management, but differ in terms of distribution and situational application of them. In case of the segmentors it was shown by numbers that they employ mainly segmentative boundary-work-tactics to guarantee the delimitation between work life and private life, for which they strive for. The fact that segmentors most frequently engage in boundary-work-tactics could be connected to the fact that their need for segmentation forces them to use a lot of strategies for role transitions and the partially rather artificial partitioning of live domains [cf. 4, 14]. Integrational boundary-work-tactics are only introduced in exceptional cases, but always with the goal to maintain the demarcation between work and private over time. In addition segmentors apply integrational boundary-work-strategies when working at home, because they recognized the beneficial effects of flexible-mobile work on their work life balance. In contrast to segmentors, integrators use integrative boundary-work-tactics on the majority, assuming this proportion is even bigger considering that the mere waiver to install a boundary by using segmentative boundary-work tactics could be counted as an integrational behaviour. If integrators use segmentative boundary-work-tactics it is in order to create recreational islands for energy boosts, or giving minimal structure to the times they work at home, where the overlapping of work and private life are too great, even for an integrator. As we have lived in a rather over structured world and have been given a lot of structure by our employers in former times, integrator are more familiar with dismantling them than building them up.

Mixed boundary types in turn help themselves with both segmentative as well as integrative boundary-work-tactics, whereby segmentative boundary-work-tactics are slightly more in use. By closer inspection it however gets evident, that although mixed boundary types try to manage their boundaries with boundary-work-tactics, they're not always able to keep track of them. Whilst segmentors and integrators are rather successful boundary managers within the new age of flexible-mobile work, it is the mixed boundary types which struggle the most with their personal work-life-balance. They seem to make priorities dependent on whether an issue is handled like an integrator or a segmentor, rather than having a clear preference for one of both aptitudes. It is conspicuous that mixed boundary types feel significantly less productive in flexible-mobile-work arrangements, don't elaborate on boundary-work-tactics too much and have difficulties naming them when asked to do so. If they use boundary-work-tactics, the results aren't always satisfying and they wish for appropriate strategies which could properly help them to manage their boundary needs and thus improve perceived work life balance.

As critical remark it has to be admitted, that in order to build boundary-types responders and interviewees had to self-assess their need for integration/segmentation and therefore their position on the role segmentation-role integration-continuum. As a further criticism it has to be stated, that the sample of the interview study is rather tight with only 9 interviewees. But as a sampling of boundary-types was successfully realized by means of 9 interviews and their enriching character to the broadly spread home office day 2013 survey was able to supply a solid quality of statement. Further research has to be conducted to corroborate quantitatively findings, such as studying the forms and prevalence of mixed boundary types for example. Continuingly it should be considered, if in place of simple recommendations for action, a type specified boundary type training would be more reasonable in supporting and qualifying the herein verified distinct boundary types in regard to their daily boundary management in flexible-mobile work arrangements. A first boundary type independent attempt of boundary management training has shown that boundary management can be taught [17].

5.2 Recommendation for Action and Organization Based on Presented Results

For home office users which correspond to the boundary-type of segmentors, it is recommended based on present findings that they plan home office days fixed and recurrently, in order to meet the segmentors need for rituals and structure. As working at home gets an inherent part within the working week, segmentors are able to get accustomed to it and plan/structure their weeks accordingly. Further it appears helpful, if segmentors set daily goals for their home office days, which they can consult in the evening, by means of a productivity check. As segmentors strongly link their home domain with recovery and leisure time, which then can lead to the assumption that productive work isn't possible in such surroundings. Another very important recommendation for segmentors would be, that they acquire suitable rituals for the role-transitions (private to work/work to private) when working from home. So far these role-transitions have been realized and designed by rituals such as getting dressed in occupational clothing and leaving the house thus moving to the workplace provided by the employer. These transformational processes are omitted when working at home, what can lead to the unpleasant sensation, that the appropriate work mood and energy can't be warranted like in the case of working at the main office. The more such predefined rituals are repeated and institutionalized, the easier role transitions are realized in situations of having to work at home [cf. 4]. Finally segmentors are advised to adopt personally suited integrational boundary-work-tactics when working at home, in order to maxing out the advantages of home office work. For a stubborn replication of a workday at the main office, which is purely work-related, could have negative effects on the psychological wellbeing and thus work life balance of home working segmentors.

For those home office users which correspond to the boundary-type of integrator, it is recommended based on results that they try to control their effective work hours in order to ensure that required recovery periods are incorporated where needed. Thereby it is important for integrators to prior reflect on what is experienced as straining work in sense of energy consuming, and which activities are perceived as recreational, as they have reenergizing effects and contribute to recreational needs; this independent from context (work or private), as both domains are interlaced with integrators and both have straining and energizing components, and integrators need to find their balance in this mix. Further it is important for integrators to have a maximum of time flexibility in order to being able to integrate the life domains work and private as opportune as possible. The more structures are dictated to integrators, the less flexible and effective they feel towards work issues. Instead of setting up fixed home office days, integrators should be allowed to make use of their home office as suited. Home office should be applied wherever perceived as sensible in view of the work day resp. work week at hand. As last recommendation for integrators it is advised, that they adopt personally suited segmentative boundary-work-tactics for the home office situation. The through industrialization created culture of segmenting between work life and private life has thus far forced integrators to focus on the abolishment of structures, rather than building and enforcing them. The, in sight of work, much unstructured home office situation forces integrators to segment by setting up boundaries. For the otherwise very flexible boundary-type integrator it is therefore vital to give themselves suitable structures for flexible-mobile work, in particular for the home office situation.

Concerning mixed boundary types, which have been substantiated by this study for the very first time, deepened investigation have to be conducted in order to obtain an extensive understanding for their underlying motivational mechanisms. The results at hand still allow first cautious recommendations for the case of flexible-mobile Work. As a first and most important recommendation it is advised that mixed boundary typed individuals perform a personal need assessment, in order to gain clarity about which issues and areas of life are preferentially treated in a segmentative manner, and which issues and areas of life are mainly handled in an integrative way. Building on that, mixed boundary types are advised to develop a corresponding set of rules (in sense of a policy) which they rend transparent towards their environment. It is expected that stakeholders such as teammates gain a better understanding for the functionality of this individual mixed type and therefor have a better idea about the approachability during spare time. Furthermore mixed boundary types are advised to stick to segmentative behaviour if in doubt about the personal needs at hand. As mixed boundary types have a tendency to experience more stress with the integrational behaviour mode than they have when segmenting. The results indicate that mixed boundary types wish for catchy segmentative boundary-work-tactics, in order to re-establish high productivity performances while working from home. What seemingly works for segmentors and integrators concerning boundary-work-tactics misses its desired effects with mixed boundary types.

For a successful and preferably smooth collaboration of these three very distinct boundary-types organizations are advised to set fixed work days for collaboration and team synchronization. The remaining workdays can then be designed according to the individual and unequal needs of each boundary-type. Furthermore especially the interview data point to the fact, that mutual respect and an understanding for other boundary-types' needs in regard to boundary management can be supported and fortified through communicational efforts. For this matter heterogeneous teams (meaning teams consisting of different boundary-types) should deliberately initiate such exchanges in order to avoid conflicts which arise out of their diversity. In conclusion organizations are advised to educate and enable their flexible-mobile knowledge workers for the challenges of this new way of working through boundary management trainings. As related research [17] has demonstrated, boundary management is learnable and can be supported due to specific interventions. It remains open, whether such trainings should be geared boundary type wise or not.

5.3 Conclusion and Prospect

The results of presented research show, that the boundary theory [cf. 4, 14] is able to build a better understanding for potentials and risks of flexible-mobile working arrangements, especially for the situation of working at home, and helps in the derivation and development of suited intervention strategies. It is thereupon the identification and characterization of the herein confirmed mixed boundary type which constitutes the main and most important research contribution.

It seems that the role segmentation-role integration-continuum is a valid possibility to distinguish boundary-types. Further research in this direction is recommended, as there's a good chance that there exist better and more accurate ways of tackling the self-contained boundary-types on the role segmentation-role integration-continuum. Especially the mixed boundary type delivers prospects for additional research. The questions arises, whether an extension from three to for example five boundary-types could produce further valuable insights, as there is evidence in literature, that a further subdivision is feasible [11].

References

1. Allen TD, Johnson RC, Kiburz KM, Shockley KM (2013) Work-family conflict and flexible work arrangements: deconstructing flexibility. Pers Psychol 66(2):345–376
2. Arlinghaus A, Nachreiner F (2014) Health effects of supplemental work from home in the European Union. Chronobiol Int 31(00):1100–1107
3. Ammons SK (2013) Work-family boundary strategies: Stability and alignment between preferred and enacted boundaries. J Vocat Behav 82:49–58
4. Ashforth BE, Kreiner GE, Fugate M (2000) All in a day's work: boundaries and micro role transitions. Acad Manage Rev 25(3):472–491
5. Brenke K (2014) Heimarbeit: immer weniger Menschen in Deutschland gehen ihrem Beruf von zu Hause aus nach. DIW-Wochenbericht 8:131–140
6. Gisin L (2014) Boundary-Typen, Boundary-Management und Boundary-Taktiken im Home Office. Untersuchung des Einflusses der "Boundary Theorie" auf das mobile Telearbeiten, insbesondere der Arbeit im Home Office. Master Thesis, S. 178. Olten: Hochschule für Angewandte Psychologie FHNW
7. Glaser B, Strauss A (1967) The discovery of grounded theory. Aldine, Chicago
8. Grebner S, Berlowitz I, Alvarado V, Cassina M (2010) Stress bei Schweizer Erwerbstätigen. Zusammenhänge zwischen Arbeitsbedingungen, Personenmerkmalen, Befinden und Gesundheit. Staatssekretariat für Wirtschaft SECO, Bern. S. 135
9. Horx M (2014) Das Megatrend-Prinzip. Wie die Welt von morgen entsteht. Pantheon Verlag, München
10. Karmaker U, Apte U (2007) Operations management in the information economy: information product, processes and chains. J Oper Manag 25(2):438–453
11. Kossek EE, Ruderman MN, Braddy PW, Hannum KM (2012) Work-nonwork boundary management profiles: a person-centered approach. J Vocat Behav 81:112–128
12. Kreiner GE, Hollensbe EC, Sheep ML (2009) Balancing borders and bridges: negotiating the work-home interface via boundary work tactics. Acad Manage J 52(4):704–730
13. Mayring P (2010) Qualitative Inhaltsanalyse. In: Mey G, Mruck K (eds) Handbuch Qualitative Forschung in der Psychologie. VS Verlag für Sozialwissenschaften, Wiesbaden
14. Nippert-Eng CE (1996) Home and work: negotiating boundaries through everyday life. University of Chicago Press, Chicago
15. Paridon H, Heise O (2012) Ständige Erreichbarkeit: Wie belastet sind wir? Ursachen und Folgen ständiger Erreichbarkeit. Deutsche Gesetzliche Unfallversicherung e.V, Berlin
16. Rey L (2002) Mobile Arbeit: Zwischen Hochgefühl und Untergangsstimmung. In: Ray L (Hrsg) Mobile Arbeit in der Schweiz. vdf Hochschulverlag AG, Zürich
17. Rexroth M, Sonntag K (2014) Abgrenzung der Lebensbereiche lernen - Effekte eines Boundary-Management Trainings. GfA, München
18. Schulze H, Weichbrodt J, Meissner J (2015) Gestaltung mobil-flexibler Arbeit. Gastherausgabe. Zeitschrift für Wirtschaftspsychologie, 1(17), Pabst Science Publishers
19. Vartiainen M, Hyrkkänen U (2010) Changing requirements and mental workload factors in mobile multi-locational work. N Technol Work Employ 25(2):117–135
20. Weichbrodt J (2014) SwissFlexWork 2014. Repräsentative Befragung der Schweizer Erwerbstätigen zur Verbreitung von mobiler Arbeit und Home Office. Bericht, 14 S. Olten: Hochschule für Angewandte Psychologie FHNW

Interdisciplinary Collaboration

How to Foster the Dialogue Across Disciplinary Borders?

Simone Brandstädter and Karlheinz Sonntag

Abstract

Collaborations of multiple disciplines often pose a challenge. Team members differ in their academic and professional backgrounds, leading to communicational, cultural or methodical challenges when working together. Which competences are necessary for successful interdisciplinary collaboration? In three studies, we identified crucial competences, built a model of interdisciplinary competency and showed significant relationships with various criteria. We then compared competence levels of current interdisciplinary team members with expert ratings of required values, showing that demands are largely met in regard to team and work competences, whereas leadership, topic and special competences should be fostered specifically. The competence model depicts where developmental requirements are and can be used as framework to devise recommended actions.

Keywords

Interdisciplinary competency • Competence model • Diversity in work groups

1 Background

The complexity of our current world demands collaborations of multiple disciplines and professions in numerous varying fields [1]. To address questions on sustainability, innovation or technological progress, the integration of perspectives

S. Brandstädter (✉) • Kh. Sonntag
Work and Organizational Psychology, Psychological Institute, University of Heidelberg, Heidelberg, Germany
e-mail: simone.brandstaedter@psychologie.uni-heidelberg.de; karlheinz.sonntag@psychologie.uni-heidelberg.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_26

and skills of multiple disciplines is required. Teams of professionals with different expertise sharing their approaches lead to a more profound understanding of the topic and can create ideas a single discipline would not be able to generate [8]. This is true for nearly all fields and organizational areas such as research and development.

A systematic integration of ideas of different disciplines and the development of a new joint approach is called interdisciplinarity. The National Academy of Science defines interdisciplinarity as

> a mode (. . .) that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline [23, p. 26].

According to a model by Bronstein [6, 7], interdisciplinary collaboration can be characterized by five factors. Interdependence refers to the mutual reliance and dependence when interacting and achieving project goals. Newly created professional activities are shown in collaborative acts, programs and structures. Flexibility contains role blurring and alteration, i.e. the question of who fills the leading role depending on project status. Collective ownership of goals is related to the responsibility of all members to achieve the common goal. Reflection on the process includes thinking and talking about joint work [6, 7].

Interdisciplinary teams face specific problems and challenges. Although team diversity is associated with some positive outcomes such as higher problem solving ability or higher innovation [14], it also leads to negative effects in group cohesion and team performance, such as conflicts and high financial input [16]. Indeed, interdisciplinary collaborations are often not as successful as expected [27]. Members of one discipline share a set of specific assumptions, theories and methods that differs widely between subjects and acts as a barrier for collaboration [21]. Known problem areas can roughly be sorted into five categories [9]: difficulties in communication, methodological problems, troubles to mutually define the project goal, prejudices and wrong expectations and group dynamic problems.

For example, use of specific expressions or technical terms can lead to communication problems. The same words can mean something different in various disciplines. Furthermore, one term often stands for an entire concept with specific assumptions, a history and additional information that is not obvious to other people [13]. Most of the time, experts are not fully aware of the amount of background knowledge they implicitly use, and therefore communication can be difficult and some facts and beliefs hard to explain.

The varying focus due to diverging interests and thinking styles is one of the most central reasons for conflicts when multiple disciplines have to work together. Insufficient joint definitions of project goals often lead to obstructions or abortions terminations [17]. Therefore, especially the starting phase, i.e. clarification of project goals and procedures, needs a lot of time and (financial) resources [10].

This all is intensified by the fact that members of one discipline tend to judge the value and utility of other disciplines' contributions by the centric world view of their own profession [25]. The own approach is seen as the best which leads to various conflicts when negotiating project goals or solution strategies. Referring to the concept of ethnocentrism, Percukonis, Doyle and Bliss [24] use the term profession-centrism for this common occurring attitude.

In this context, some possible influencing factors on interdisciplinary collaborations are discussed in the literature as well as fostering or hindering conditions [1]. For example, proximity, high amounts of formal and informal shared time as well as equality of the involved disciplines are seen as helpful [3, 10]. Personal contact, mutual sympathy and role clarity can further foster the success of interdisciplinary projects. Lack of communication and exchange, unequally distributed competences and diverging disciplinary cultures pose a threat for collaboration [3, 10, 19].

Bronstein [7] identified clarity of professional roles, structural characteristics, such as administrative support, history of interdisciplinary collaboration, i.e. existence of positive or negative prior experiences and personal characteristics, such as traits and attitudes, as influencing factors for successful interdisciplinary collaboration.

Team member characteristics are repeatedly mentioned to be a critical factor [1]. However, no systematic research has been conducted so far on the question of which competences are necessary to successfully work in an interdisciplinary team. Competences are defined as “measurable pattern of knowledge, skills, motivation, interests, abilities, behaviours and other characteristics a person needs to succeed in his/her tasks” [31]. Repko [26] discusses, among others, perspective taking, the ability to reflect and appreciation of diversity as crucial. Nevertheless, empirical results are missing.

The aim of the present study is to investigate which competences are necessary for successful interdisciplinary collaborations and to clarify where current team members show deficits in “interdisciplinary competency”, as well as to deduce recommendations for supporting actions to facilitate the dialogue between the disciplines.

2 Methods

The first step was to create a model for “interdisciplinary competency”. We used three studies to determine which competences are crucial for interdisciplinary collaboration, to verify those competences via expert rating and to validate the model via a survey with interdisciplinary employees [cf. 5]. In a second step, we then compared the required values known from the expert survey with the actual competence levels of interdisciplinary team members. In the following, the studies are explained in detail.

2.1 Study 1: Qualitative Pilot Study

First, we used a qualitative pilot study to investigate which competences and behaviours are perceived as crucial in interdisciplinary collaborations. We used expert workshops and semi-structured interviews with N = 33 interdisciplinarily working persons (9 female; age: 24–35 M = 28.1). Experience in interdisciplinary projects ranged from 0.5 to 5 years (M = 2.2). Participants had backgrounds in a variety of different disciplines (e.g. medicine, computer science, psychology, engineering, business or communication studies) and were asked about typical problems in their interdisciplinary collaborations and critical incidents, i.e. situations and behaviours that were very supportive or obstructing in the collaboration process [11]. They also were directly asked about necessary competences and were supposed to describe an ideal interdisciplinary co-worker. The interviews were transcribed and analysed via qualitative content analyses [22] using Software MAXQDA 10. In total, 255 statements about critical behaviours and necessary competences were collected. In a competence modelling workshop three psychological experts sorted and merged those statements and derived 40 competences. Those were further specified with a definition and behaviour examples for high and low competence levels (for an example see Fig. 2 in Appendix A).

2.2 Study 2: Expert Survey

To verify the importance of the derived competences, an expert survey was conducted online. Out of N = 64 persons who finished the questionnaire, N = 60 met the criteria of expert status, i.e. being principal investigator of an interdisciplinary project with at least 1 year experience (12 female; age: 30–67 years, M = 47.4; interdisciplinary experience: 1–42 years, M = 15.1). The sample consists of a variety of different disciplines: N = 28 sciences (e.g. physics, biology, chemistry), N = 16 humanities and social sciences (e.g. philosophy, German studies, psychology), N = 3 law, N = 5 engineering (e.g. building engineer, computer science), N = 7 human medicine. Competences with definition and behaviour descriptions were presented and the experts rated every competence regarding the importance for interdisciplinary collaboration (scale from 1: “unimportant” to 5: “very important”) and the future trend of the importance (-1: “decreasing”, 0: “constant”, 1: “increasing”). This should ensure that only (also prospectively) relevant competences are included in the model. They also were asked to rate the required competence level (scale from 1: “very low” to 7: “very high”), determining the ideal values for an interdisciplinary team member.

2.3 Study 3: Employee Survey

To validate the competence model and to compare the competence profile of interdisciplinary team members with required competence levels, we conducted another online questionnaire. N = 210 team members of interdisciplinary research projects participated. One participant had to be excluded, hence the sample consists of N = 209 (104 female; age: 25–62 years, M = 31.7; variety of different disciplines: natural sciences, humanities, law, business, engineering, medicine). Many of the participants worked in interdisciplinary fields such as oceanography, biomedical informatics, didactics of physics or land use planning with interdisciplinary experience of M = 3.1 years (range .2–30 years).

Besides general demographics (age, gender, discipline) we asked about prior interdisciplinary experience, particularly the duration in years, the number of interdisciplinary projects and a rating of how interdisciplinary own collaborations are perceived (scale: 0 %: "strictly disciplinary – no contact to another discipline" to 100 %:"strictly interdisciplinary – daily contact and all project goals and steps depend on another discipline"). The 40 competences were presented and participants had to rate their own current competence levels on all competences (scale from 1: "very low" to 7: "very high"). As further criteria we asked for work engagement (short version of Utrecht Work Engagement Scale; Schaufeli et al. [28]) as well as job satisfaction by single-item-measure (according to Wanous et al. [32]). We also asked for self-report on one's own and the supervisor's performance rating [29]. All items were answered on a 7-point Likert scale.

3 Results

First, findings regarding crucial competences for interdisciplinary collaboration and relationships of the competence model with further criteria are reported. Then, we present the results of the comparison between the required and actual competence levels of interdisciplinary team members. These indicate where there are needs for further development and show on which areas supportive actions should focus.

3.1 Competence Model for Interdisciplinary Collaboration

The qualitative pilot study resulted in 40 different competences: N = 5 professional competences (e.g. knowledge about interdisciplinary collaborations, knowledge of the strength of different disciplines), N = 11 methodological competences (e.g. project management skills, discourse and moderation skills), N = 11 personal competences (e.g. openness, adaptability) and N = 13 social competences (e.g. capacity for teamwork, communicative skills).

All 40 competences were confirmed in the expert rating. The consensus among raters was excellent with inter-rater reliability of ICC = .991 ($F(119, 7497) = 138.2$; $p < .001$; Cronbach's $\alpha = .993$. All identified competences were rated as important to very important (range: 3.3–4.5; M = 3.9, sd = .49) with constant to increasing predicted importance in the future (range: .07–0.5; M = 0.3, sd = .24). Nevertheless, the competence of "prior experience in interdisciplinary collaborations" was excluded, as it was seen more as a condition than a trainable competence. Therefore, 39 competences were included in the final model.

To validate the competence model, we analysed the factor structure of the competence model and relationships with various criteria based on the employee sample [cf. 5]. A parallel analysis [15] and principal component analysis [2] revealed five different principal components. *Topic competences* refer to cognitive abilities that enable one to work on an interdisciplinary topic, such as capacity for abstraction or critical thinking. *Special competences* contain expert competences and attitudes one needs in an interdisciplinary setting, such as knowledge about methods of different disciplines or commitment. *Leadership competences* include abilities to effectively moderate or negotiate with others. *Team competences* are such skills and abilities one needs to successfully cooperate with other people such as appreciation or friendliness. *Work competences* refer to characteristics that are required in daily work, such as expert competence or achievement motivation. Reliability of the overall competency is excellent with a Cronbach's α of .91. All five subscales have acceptable to good reliabilities: topic competences show a Cronbach's α of .80, special competences Cronbach's $\alpha = .76$, leadership competences Cronbach's $\alpha = .79$, team competences Cronbach's $\alpha = .76$ and work competences Cronbach's $\alpha = .70$. The resulting competence model with all 39 competences is depicted in Fig. 1.

We calculated the mean of the overall competence level (i.e. interdisciplinary competency) and the means of all subscales separately and looked at the relationships with criteria (for all means and correlations see Tab. 1 and 2 in Appendix B). Interdisciplinary competency does correlate with age ($r = .15$, $p < .05$) as expected, due to the fact that older people should have longer working experience and therefore should have gained more expertise. When controlling for professional experience, competences no longer correlate with age ($r = .07$, n.s.). Interdisciplinary competency does not correlate with gender ($r = .04$, n.s.), however, when look at the subscales, team competences do ($r = .18$, $p < .01$), which could reflect the fact that women show better skills for example on inclusiveness or interpersonal relations [35].

Interdisciplinary competency does not correlate with general professional experience ($r = .08$, n.s.), however it does correlate with experience in interdisciplinary collaboration ($r = .16$, $p < .05$), indicating that the model depicts a specific

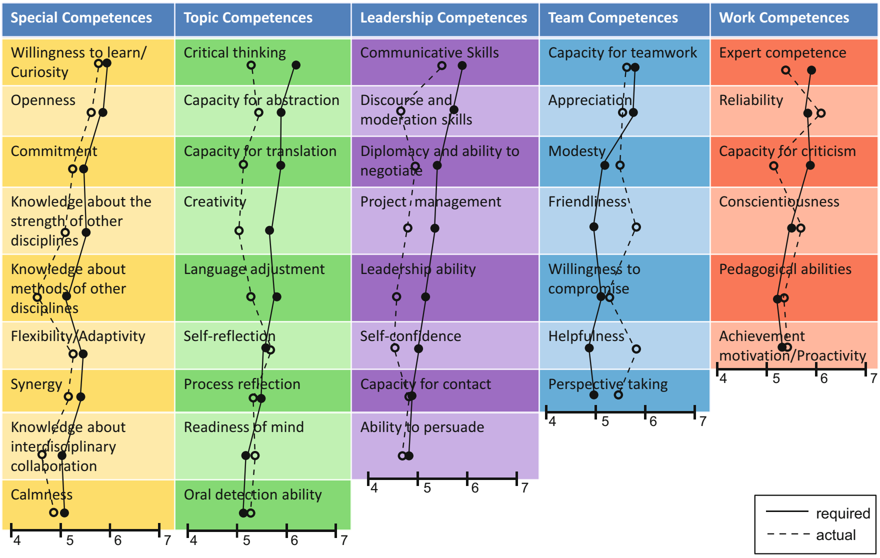


Fig. 1 Competence model for interdisciplinary collaboration, sorted by descending importance. The required competence levels (*points*) and the actual mean competence level for interdisciplinary team members (n = 209; *circles*) are indicated here. Only the upper half of the scale is depicted (4–7) in order to better differentiate

competence pattern for interdisciplinary collaboration. Looking at the subscales, especially the two more specific competence domains, i.e. topic ($r = .15$, $p < .05$) and special competences ($r = .20$, $p < .01$), correlate significantly with interdisciplinary experience. This indicates that there are some specific cognitive skills and a range of professional knowledge that develops when working interdisciplinarily. The same pattern is revealed with a growing number of interdisciplinary projects (overall $r = .14$, $p < .05$, topic $r = .15$, $p < .05$ and special competences $r = .15$, $p < .05$). The perceived interdisciplinarity of own projects correlates with overall competence ($r = .18$, $p < .01$) as well as special competences ($r = .27$, $p < .01$) and leadership competences ($r = .17$, $p < .05$). It seems that the more interdisciplinary the teamwork is, the more one needs professional expertise and communication as well as diplomacy skills.

Furthermore, interdisciplinary competency ($r = .51$, $p < .001$) and all subscales ($r = .24$ to $r = .47$, $p < .001$) correlate with engagement. Self-report of job satisfaction and performance correlates with overall competency (both $r = .30$, $p < .001$) as well as with the subscales ($r = .17$ to $r = .32$, $p < .01$ resp. $r = .26$ to $r = .27$, $p < .001$) with the exception of team competences ($r = .10$ and $r = .07$, n.s.).

3.2 Comparison of Required and Actual Competence Levels

The expert survey revealed that the required levels on all competences were rated as high medium (values 4–5) to outstanding (values 6–7) on average (range: 4.8–6.2, $M_{required} = 5.5$, $sd = .80$). Out of 39 competences, 34 revealed values above 5 which indicate the demand for outstanding competence levels. The required levels are depicted in Fig. 1.

The employee sample showed similarly high actual values (range: 4.6–6.1; $M_{actual} = 5.3$, $sd = .53$). Therefore, the interdisciplinary team members in question rated their own competence levels as at least high medium to outstanding. 29 out of 39 competences reached levels above 5. The actual values of the employees are depicted in Fig. 1 as well.

When comparing the required and actual competence levels, the results differ in the five components. First of all, only a marginally significant difference in overall interdisciplinary competency is shown ($M_{required} = 5.5$, $sd = .80$; $M_{actual} = 5.3$, $sd = .53$; $t(74.5) = 1.69$, $p < .10$). Highly significant differences are revealed regarding topic competences ($M_{required} = 5.7$, $sd = .76$; $M_{actual} = 5.3$, $sd = .68$; $t(276) = 3.49$, $p < .01$) and leadership competences ($M_{required} = 5.3$, $sd = .99$; $M_{actual} = 4.8$, $sd = .77$; $t(80.8) = 3.31$, $p < .01$), as the actual competence levels stay behind the required ones. The same is true for special competences ($M_{required} = 5.4$, $sd = .80$; $M_{actual} = 5.2$, $sd = .69$; $t(267) = 2.51$, $p < .05$). The largest deviations become apparent for discourse and moderation skills ($\Delta = 1.05$), project management ($\Delta = .60$), leadership ability ($\Delta = .56$), critical thinking ($\Delta = .91$), capacity to translate ($\Delta = .75$) and creativity ($\Delta = .64$).

No differences are to be found for work competences ($M_{required} = 5.6$, $sd = .95$; $M_{actual} = 5.5$, $sd = .66$; $t(75.9) = .77$, $p = .443$). Nevertheless, capacity for criticism ($\Delta = .91$) shows a relatively large deviation from the ideal level. In contrast, project members' actual competence levels in team competences even exceed the required ones ($M_{required} = 5.3$, $sd = 1.05$; $M_{actual} = 5.6$, $sd = .69$; $t(72.9) = -2.49$, $p < .05$). All means and standard deviations are provided in Appendix B (Table 2).

4 Discussion

The goal of the current study was to investigate which competences have to be promoted for interdisciplinary collaboration. We built a model for interdisciplinary competency in three studies. The final model consists of 39 competences in five domains. Most of these competences are discussed in previous literature, which speaks for the content validity of the model. For example, Klein [18, 19] discusses synergy, flexibility and willingness to learn, Epstein [10] team skills, language adoption and reliability and Repko [26] communication skills, creativity and readiness of mind. As professional competency can be developed in an appropriate context, people working for a longer time period in interdisciplinary teams should show higher competence levels [30]. This is exactly what the correlation analysis shows. The fact that interdisciplinary competency solely correlates with interdisciplinary and not general professional experience indicates that the competence model not only contains vocational competences per se but a specific pattern for interdisciplinary collaboration. This clearly indicates the validity of the model. The correlations with job satisfaction and performance reported here give a first hint regarding criteria validity. Interdisciplinary competency and all subscales showed considerable relationships with engagement, which is known to play a crucial mediating role for many organizational outcomes [e.g. 34]. With the exception of team competences, all other domains were also directly related to satisfaction and performance. In sum, this all suggests the predictive validity of the model. Nevertheless, we further analyzed the competence model as well as additional criteria using structural equation models for investigating criteria validity [5].

Furthermore, results are in accordance with literature about challenges in interdisciplinary collaborations [9]. The presented competences play a crucial role for overcoming specific problems such as misunderstandings, incorrectly defined project goals, faulty mutual expectations or profession-centrism.

A comparison of required and actual competence levels of interdisciplinary team members shows that leadership competences should be fostered. Discourse and moderation skills, project management or leadership ability are especially important. Those competences seem to be critical as members of different disciplines have to find agreements for diverging ideas and goals in interdisciplinary projects. Negotiations are particularly challenging because opinions, preferred working processes and thinking styles are varied [12, 33].

Furthermore, deviations regarding topic and special competence became apparent. Cognitive abilities can foster communication for example by language adoption skills or the capacity to translate own concepts to other professional fields. Critical thinking is seen as the most important competence with highest required level. This is the ability to critically reflect on and question information and facts. This includes asking the right questions and analytically dealing with reasons and conclusions [20]. This ability can be used not only for others' but also for own approaches and thereby helps to get into a balanced dialogue with other disciplines.

At present, knowledge about other disciplines' strength and methods is partly missing in interdisciplinary collaborations, leading to wrong expectations about possible contributions to the project goals [9]. Awareness of own disciplinary dependencies in thinking style can help to overcome profession-centrism and bring the collaboration to an equal level in order to effectively work together.

As team competences and work competences are mostly demonstrated by project members, it seems that most collaboration and team conflicts refer to specific interdisciplinary challenges, for example when different disciplinary cultures clash. However, capacity for criticism should be fostered within team members as results show a lack of this competence.

In order to further foster interdisciplinary collaboration, it could be useful to develop a training for interdisciplinary competency that teaches professional skills and approaches such as knowledge about difficulties and challenges within interdisciplinary collaborations and methodological competences. Exercises for practicing language adoption for communication with non-experts or training of problem solving strategies for more synergy would be useful.

Regarding this study, one possible limitation is the sample. Participants were recruited in a scientific context, i.e. interdisciplinary research projects. This may explain the relatively high required and actual competence levels. As the majority of the sample is from an academic context, the tendency of showing rather outstanding competence levels is explainable. As interdisciplinary collaboration is a daily task especially in research and development, interdisciplinary competency is of particular interest for those areas. At present, many scientific projects are already interdisciplinary and most funding organizations especially promote research jointly conducted by different disciplines [4].

Nevertheless, in increasingly more working areas interdisciplinary projects will be relevant in the future. Technical development or digitalization, for example, leads to working environments in which team members of different fields and expertise have to collaborate. Therefore, there will be an increasing demand for employees able to demonstrate interdisciplinary competency. Hence, it is reasonable to repeat the survey in other areas, such as various industries or businesses, in order to further determine whether there are differences regarding required competence levels within those fields. This would enable the development of tailored supportive measures.

As previously mentioned, there is a need to further validate the competence model. By examining a greater sample, it would be possible to confirm the factor structure and to investigate predictive validity. Furthermore, the development of a questionnaire for interdisciplinary competency as well as appropriate training programs will be the next steps on the path to further foster the interdisciplinary dialogue.

Acknowledgement We like to thank the German Research Foundation for funding this research within the research training group 1126 “Intelligent Surgery” and Alexandra Kessler for student assistance.

Appendix A (Fig. 2)

Appendix B (Tables 1 and 2)

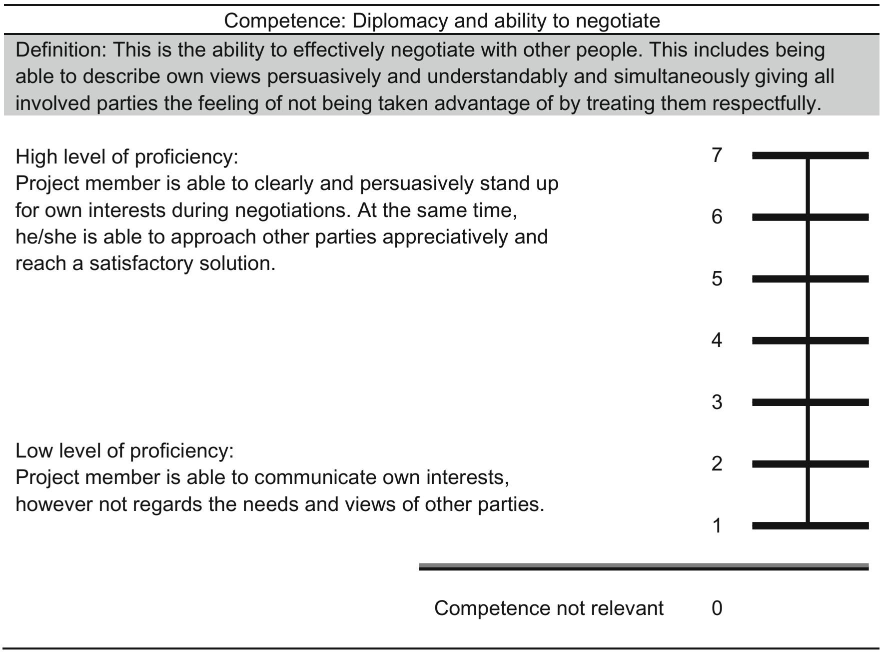


Fig. 2 Example of competence definition with behaviour description

Table 1 Means, standard deviations, internal consistency and Pearson correlations of Study 2 (employee survey)

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Interdisciplinary competency	(.91)														
2	Leadership competences	.76**	(.80)													
3	Team competences	.66**	.32**	(.76)												
4	Special competences	.82**	.49**	.48**	(.76)											
5	Topic competences	.80**	.50**	.41**	.57**	(.80)										
6	Work competences	.73**	.52**	.39**	.49**	.50**	(.70)									
7	Gender[a]	.04	.10	.18**	−.01	−.10	−.02	–								
8	Age	.15*	.07	.05	.13*	.15*	.16*	−.11	–							
9	Professional experience	.08	−.03	.07	.07	.10	.11	−.10	.82**	–						
10	interdisciplinary experience	.16*	.10	.04	.20**	.15*	.09	−.08	.56**	.54**	–					
11	Number of projects	.14*	.07	.02	.15*	.15*	.13	−.20**	.50**	.52**	.50**	–				
12	Interdisciplinarity of projects	.18**	.17*	.05	.27**	.14	.02	.06	.13	.06	.22**	.15*	–			
13	Engagement	.51**	.42**	.24**	.46**	.35**	.47**	.04	.13	.07	.09	.18*	.16*	(.93)		
14	Job satisfaction	.30**	.28**	.10	.32**	.17*	.24**	.03	−.04	−.07	−.07	.01	.16*	.55**	–	
15	Performance	.30**	.26**	.07	.27*	.26**	.27**	−.06	−.06	−.03	−.02	−.04	.05	.31**	.63**	(.66)
	M	5.27	4.84	5.63	5.18	5.31	5.52	1.50	31.7	4.86	3.06	2.54	60.0	4.95	5.32	5.07
	SD	.53	.77	.69	.69	.68	.66	.50	5.85	4.39	2.79	3.22	1.93	1.02	1.04	.82

**$p < .01$, *$p < .05$, [a]Gender (1 = male, 2 = female). Internal consistency (Cronbach's α) are presented in parenthesis in the diagonal

Table 2 Means and standard deviations of required (Study 2) and actual (Study 3) competence levels

	Required		Actual	
	M	*SD*	*M*	*SD*
Special competences				
Willingness to learn/curiosity	5.97	1.07	5.81	1.07
Openness	5.88	1.22	5.68	1.06
Commitment	5.53	1.32	5.28	1.26
Knowledge about the strength of other disciplines	5.57	1.24	5.15	1.13
Knowledge about methods of other disciplines	5.08	1.21	4.67	1.25
Flexibility/Adaptivity	5.48	1.35	5.30	1.16
Synergy	5.35	1.30	5.17	1.02
Knowledge about interdisciplinary collaboration	5.02	1.37	4.62	1.22
Calmness	5.07	1.51	4.92	1.39
Topic competences				
Critical thinking	6.23	0.89	5.32	1.12
Capacity for abstraction	5.89	1.11	5.40	1.05
Capacity for translation	5.87	1.14	5.11	1.10
Creativity	5.68	1.17	5.04	1.25
Language adjustment	5.80	1.34	5.31	1.16
Self-reflection	5.63	1.21	5.66	1.04
Process reflection	5.58	1.31	5.32	0.99
Readiness of mind	5.18	1.52	5.37	1.05
Oral detection ability	5.13	1.31	5.25	1.08
Leadership competences				
Communicative Skills	5.88	1.09	5.53	1.00
Discourse and moderation skills	5.77	1.21	4.69	1.19
Diplomacy and ability to negotiate	5.43	1.43	4.96	1.11
Project management	5.38	1.34	4.78	1.22
Leadership ability	5.17	1.67	4.60	1.27
Self-confidence	5.05	1.29	4.57	1.40
Capacity for contact	4.93	1.40	4.88	1.44
Ability to persuade	4.78	1.40	4.70	1.10
Team competences				
Capacity for teamwork	5.83	1.28	5.65	1.06
Appreciation	5.77	1.29	5.55	1.04
Modesty	5.18	1.72	5.57	1.14
Friendliness	4.98	1.57	5.88	1.10
Willingness to compromise	5.10	1.48	5.36	1.08
Helpfulness	4.92	1.52	5.89	1.00
Perspective taking	4.98	1.36	5.48	1.11
Work competences				
Expert competence	5.93	1.26	5.36	1.05
Reliability	5.85	1.26	6.12	0.88
Capacity for criticism	5.85	1.15	5.11	1.05

(continued)

Table 2 (continued)

	Required		Actual	
	M	*SD*	*M*	*SD*
Conscientiousness	5.53	1.53	5.72	1.01
Pedagogical abilities	5.25	1.24	5.39	1.14
Achievement motivation/Proactivity	5.32	1.47	5.42	1.10

References

1. Aboelela S, Larson E, Bakken S, Carrasquillo O, Formicola A, Glied S, Haas J, Gebbie K (2007) Defining interdisciplinary research: conclusions from a critical review of the literature. Health Serv Res 42:329–346
2. Abdi H, Williams LJ (2010) Principal component analysis. Wiley Interdiscip Rev Comput Stat 2(4):433–459
3. Bayer Y (2011) Entwicklung und Evaluation eines innovativen interdisziplinären Forschungskonzepts in der Medizin am Beispiel des Graduiertenkollegs 1126 "Entwicklung neuer computerbasierter Methoden für den Arbeitsplatz der Zukunft in der Weichteilchirurgie". Dissertation, Ruprecht-Karls-Universität, Heidelberg
4. Besio Ch (2009) Forschungsprojekte - Zum Organisationswandel in der Wissenschaft. transcript Verlag, Bielefeld
5. Brandstädter S, Sonntag Kh (in prep.). Construction and validation of a model for interdisciplinary competency
6. Bronstein LR (2002) Index of interdisciplinary collaboration. Soc Work Res 26(2):113–123
7. Bronstein LR (2003) A model for interdisciplinary collaboration. Soc Work 48(3):297–306
8. Bruner C (1991) Thinking collaboratively: Ten questions and answers to help policy makers improve children's services. Education and Human Services Consortium, Washington, DC
9. Defila R, Di Giulio A, Drilling M (2000) Leitfaden – Allgemeine Wissenschaftspropädeutik für interdisziplinär-ökologische Studiengänge, Schriftenreihe "Allgemeine Ökologie zur Diskussion gestellt", IKAÖ
10. Epstein SL (2005) Making interdisciplinary collaboration work. In: Derry SJ, Schunn CD, Gernsbacher MA (eds) Interdisciplinary collaboration. Lawrence Erlbaum
11. Flanagan JC (1954) The critical incident technique. Psychol Bull 51(4):327–358
12. Haley H, Sidanius J (2005) Person-organization congruence and the maintenance of group-based social hierarchy: a social dominance perspective. Group Process Intergroup Relat 8 (2):187–203
13. Hall R, Stevens R, Torralba T (2002) Disrupting representational infrastructure in conversations across disciplines. Mind Cult Act 9(3):179–210
14. Harrison DA, Klein KJ (2007) What's the difference? Diversity constructs as separation, variety, or disparity in organizations. Acad Manage Rev 32:1199–1228
15. Horn JL (1965) A rationale and test for the number of factors in factor analysis. Psychometrika 30(2):179–185
16. Jehn KA, Northcraft GB, Neale MA (1999) Why differences make a difference: a field study of diversity, conflict, and performance in work groups. Adm Sci Q 44:741–763
17. Journet D (1993) Interdisciplinary discourse and "Boundary Rhetoric": The Case of S. E. Jelliffe. Writ Commun 10(4):510–541
18. Klein JT (1990) Interdisciplinarity. History, theory, & practice. Wayne State University Press, Detroit
19. Klein JT (2005) Interdisciplinary teamwork: the dynamics of collaboration and integration. In: Derry SJ, Schunn CD, Gernsbacher MA (eds) Interdisciplinary collaboration. Lawrence Erlbaum

20. Lai ER (2011) Critical thinking: a literature review. Pearson Assessment, Upper Saddle River, NJ
21. Laudel G (1999) Interdisziplinäre Forschungskooperation - Erfolgsbedingungen der Institution ‘Sonderforschungsbereich’. Edition Sigma
22. Mayring P (2010) Qualitative Inhaltsanalyse - Grundlagen und Techniken. Beltz Verlag
23. National Academy of Sciences (2004) Facilitating interdisciplinary research. National Academies S.26
24. Pecukonis E, Doyle O, Bliss DL (2008) Reducing barriers to interprofessional training: promoting interprofessional cultural competence. J Interprof Care 22(4):417–428
25. Reich SM, Reich JA (2006) Cultural competence in interdisciplinary collaborations: a method for respecting diversity in research partnerships. Am J Community Psychol 38:51–62
26. Repko A (2012) Interdisciplinary research – process and theory. Sage, London
27. Rogers Y, Scaife M, Rizzo A (2005) Interdiscipliarity: an emergent or engineered process? In: Derry SJ, Schunn CD, Gernsbacher MA (eds) Interdisciplinary collaboration. Lawrence Erlbaum, Mahwah, NJ
28. Schaufeli WB, Bakker AB, Salanova M (2006) The measurement of work engagement with a short questionnaire: a cross-national study. Educ Psychol Meas 66(4):701–716
29. Schoorman F, Mayer RC (2008) The value of common perspectives in self-reported appraisals: you get what you ask for. Organ Res Methods 11(1):148–159
30. Sonntag Kh, Schaper N (2006) Förderung beruflicher Handlungskompetenz. In: Sonntag Kh (ed) Personalentwicklung in Organisationen. Hogrefe, Göttingen
31. Sonntag Kh (2009) Kompetenztaxonomien und -modelle: Orientierungsrahmen und Referenzgröße beruflichen Lernens bei sich verändernden Umfeldbedingungen. Nova Acta Leopoldina NF 100, Nr. 364
32. Wanous JP, Reichers AE, Hudy MJ (1997) Overall job satisfaction: how good are single-item measures? J Appl Psychol 82(2):247–252
33. Windolf P (1995) Selection and self-selection at German Mass Universities. Oxford Rev Edu 21(2):207–231
34. Yalabik ZY, Popaitoon P, Chowne JA, Rayton BA (2013) Work engagement as a mediator between employee attitudes and outcomes. Int J Hum Resour Manag 24(14):2799–2823
35. Yukl G (2002) Leadership in organizations. Prentice-Hall, Upper Saddle River, NJ

Simulation-Based Scheduling and Risk Assessment of Complex Projects Under Uncertainty

Sebastian Terstegen, Andreas Petz, Christopher M. Schlick, and Sönke Duckwitz

Abstract

Complex and knowledge intensive innovation, engineering and service projects are important for companies to improve their market performance. Project planning and scheduling, resource assignment and cost planning set pointers for a successful project implementation. Due to complex interdependencies between project activities and resources, increasingly less project managers are able to deal with the complex and latent uncertainties of the planning process. Discrete event, Monte Carlo simulation seems to be a promising approach to comply with the requirements of realistic planning and scheduling of complex projects. In this paper an activity-oriented, personnel-integrated simulation model is presented facilitating project managers to model and simulate the schedule of weakly-structured complex projects like engineering and service projects and, thus, evaluate the strengths and weaknesses as well as opportunities and risks of these projects. Two explorative industry case studies in the chemical industry and metal processing industry are presented to demonstrate successful operational application of the simulation-based project planning and scheduling approach and the potential of respective project risk assessment.

Keywords

Project planning • Project scheduling • Risk assessment • Monte Carlo simulation • R&D project • Engineering project • Innovation project • Knowledge-intensive service • Design structure matrix

S. Terstegen (✉) • A. Petz • C.M. Schlick • S. Duckwitz
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany
e-mail: s.terstegen@iaw.rwth-aachen.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_27

1 Introduction

Activities in the area of research and development (R&D), e.g. new product development, engineering, innovation research or continual optimisation of production processes, are a key instrument for improving companies' market performance [2]. Aside, knowledge intensive services, e.g. engineering services, product-related and -accompanying services or business and management consulting, enable new and more competitive business models. They contribute to the creation of growth and employment and, thus, play a substantial role in national economies [11]. The great majority of these services is customised to the corresponding individual consumer and provided over a relatively long period of time with participation of a wide variety of stakeholders, employees and resources. An increasing number of R&D and service activities are now set up as projects using established methods of project planning and project controlling. I.e., R&D, engineering and service projects are structured according to time-restrictions and limited personnel, financial and technical resources as well as a project-specific organisation [5]. In this regard, the scope of successful project management is to define work requirements, quantity and quality of work and resources needed (referred to as project definition and planning) as well as to track project progress, compare actual to predicted outcome, analyse impact and make adjustments (referred to as project monitoring and controlling) [18]. Thereby, the project management process gives particular consideration to project planning. Project planning comprises the specification of objectives, goals and targets to be achieved at a certain time (referred to as performance planning) and plans showing when work packages and project activities will be started and completed (referred to as project scheduling) as well as resource assignment, cost and financial planning [18, 23]. In this way, project planning creates the necessary structural and organisational conditions for planned R&D and service projects of a company and sets pointers for a successful project implementation [30].

Due to complex interdependencies between project activities, human resources (employees), non-human resources, time resources and financial resources of a project, increasingly less project managers are able to deal with the complex and latent uncertainties of the planning process. Especially the evaluation of the completion date of a project, which is one of the most important objectives of project planning, is a highly complex planning problem and presents considerable challenges for project managers. Indeed, conventional project planning methods like Gantt charts and network analysis provide technical and operational support for project managers as for example Gantt charts can be used almost intuitively and give a quick overview of the project schedule. But these methods do not comply with the requirements placed upon a comprehensive project planning; at most they can be used for small, clear manageable projects.

Network analysis methods like Critical Path Method [8, 16], Metra Potential Method [17, 29] or Program Evaluation and Review Technique [21] are procedures for analysing, describing, planning, controlling and monitoring processes on the basis of graph theory [5]. With the help of network analysis methods it is possible to implement project schedules, define the critical path and identify critical activities in a project schedule that require enhanced monitoring during project implementation. Indeed, by including costs and resource capacities, network analysis is a valuable method for project scheduling, resource assignment and cost planning [27]. But using these methods is based on relatively strong assumptions, e.g., it is assumed that activity duration and logical relationship between project activities are known and deterministic [31] and the network diagram exhibits neither multiple and probabilistic branching nor repeating activities through feedback loops. Furthermore, Shtub et al. [31] criticise that project activities are identified as entities with certain beginning and ending points for each activity. Due to the fact that complex projects change in content over time, a network plan developed in the planning phase may be highly inaccurate in the implementation phase. Another criticism concerns the deterministic specification of logical relationships between project activities in advance as the ordering of activities is conditional on the outcome of the predecessor activities [31]. This is the reason why conventional project planning methods are limited for planning and scheduling complex and uncertain R&D and service projects.

Many reports can be found in literature that recommend to integrate Monte Carlo simulation approaches into project planning and scheduling as these approaches account explicitly for the effects of uncertainty [4, 15, 19, 31, 35]. The stochastic nature of activities in complex projects is represented by randomly generated discrete events and parameters of a project from perceived distributions [19]. So, discrete event, Monte Carlo simulation seems to be a promising approach to comply with the requirements of realistic planning and scheduling of complex projects. In this paper an activity-oriented, personnel-integrated simulation model is presented facilitating project managers to model and simulate the schedule of weakly-structured complex projects like R&D and knowledge-intensive service projects and, thus, evaluate the strengths and weaknesses as well as opportunities and risks of these projects. On the basis of industry case studies the potential practical application of these approaches is demonstrated.

2 Background and Related Work

In complex R&D, engineering and innovation projects concurrent engineering (CE) is a widely shared work methodology for reducing project duration and, thus, the time between when a new product is initially conceived and when it becomes available for sale (time-to-market) [24]. CE means to integrate and parallelise functions of product and process design, e.g. design engineering and manufacturing engineering functions. Requirements arising from the post-product lifecycle phase are considered at a very early stage. In this way, time and cost intensive modifications of product and process design in late phases can be prevented. Due to the emphasised parallelisation in CE, mutual dependent activities start based on informational assumptions. If the assumptions subsequently prove to be incomplete, erroneous or incorrect, affected activities have to be performed again in feedback loops.

CE also plays an important role in the context of service engineering, which is referred to here as the technical discipline of efficiently designing service provisions as well as planning and scheduling service projects. The aim of service engineering is to extend engineering approaches and experiences as well as corresponding models, methods and tools developed for conventional product and project management to the area of service projects [22]. As the service sector becomes increasingly important, service engineering methods can potentially support companies to develop innovative and complex service models as well as stable and efficient service processes that are adequate and proportionate to minimise risks and the costs of service provision [28]. Often models and methods originally developed for the management of R&D and engineering projects are extended to apply for the design and planning of service provisions and corresponding service projects. For example, for addressing the challenges of the market, it is often necessary to minimise time-to-market. Likewise, a service provider strives to design and schedule service projects in a way that the total period or duration of service provision will be minimised. Due to the assimilable planning approach, the engineering methods used for planning and scheduling R&D and engineering projects and processes are also suitable for designing service provisions and scheduling service projects and processes.

In connection with CE, matrix-based techniques are widely applied to analyse and manage complex projects. The Design Structure Matrix (DSM) supports project managers in modelling, visualising and analysing dependencies among the entities in their projects [20, 34]. In a DSM, a project can be decomposed into n distinct activities, which are entered as labels in the rows and columns of an $n \times n$ matrix. The values in a given cell indicate informational dependencies or coupling between the corresponding activities, resulting in feed-forward information flows or control flows respectively in the lower triangular matrix and feedback flows in the upper triangular matrix [7]. Hence, the strength of the DSM is its clear and compact representation, which allows the consideration of iteration loops and mutual activity dependencies. Interactions, interdependencies and interfaces between entities in a complex project can be captured easily in a DSM.

Despite the methodological support of the DSM, planning, scheduling and corresponding resource assignment of R&D and engineering projects and, likewise, finding the optimal design of service provision and scheduling of service activities are extremely complex tasks for project managers [10, 26]. The DSM method in itself and conventional project planning methods like Gantt charts and network analysis, as described before, do not ensure appropriate consideration of certain precedence and resource constraints and optimal use of resources, customer satisfaction and the attainment of project targets in an acceptable period of time. Instead, innovative approaches like discrete event, Monte Carlo simulation should be used to improve the design and scheduling of a complex project in order to forecast the total project duration as well as effects of changes to the project.

Several DSM-based simulation models can be found in literature. For example, Smith and Eppinger [32] developed a simulation model based on Marcov chains using a DSM to model interdependencies and iteration probabilities, that is, the probability of a feedback loop from a downstream activity to an upstream activity occurring during the execution of the downstream activity. Browning and Eppinger [3] developed a Monte Carlo simulation model enabling to predict the probability density function of the project duration. In addition to that, Yassine et al. [37] supplement this simulation model making use of a multi-project environment that considers resource constraints. The approach developed by Cho and Eppinger [4] allows the modelling of different workflow systems, like overlapping activities, and indicates the rework fraction during an iteration loop and the reduction of iteration probabilities resulting from increasing quality. Karniel and Reich [14] use different types of DSMs to model intra- and inter-activity iterations and developed a Petri net simulation model based on these matrices. The Monte Carlo simulation model developed by Gaertner et al. [9] considers a realistic decrease of iteration probabilities due to learning effects in feedback loops as well as the impact of product changes on the duration and costs of a project. Zhang et al. [38] model autonomous task scheduling behaviour based on information relation matrices and use it in an agent-based product development process simulation.

The common feature of all simulation models described above is that the simulation input data are modelled in DSMs. Though, these models as well as further simulation models found in the literature do not take into account characteristic properties of complex projects under uncertainty such as informational dependencies and feedback flows, mostly to reduce the model complexity. Only the simulation model described by Gaertner et al. [9] fully considers iterative workflow systems characterised by CE. Other models do not consider overlapping sequences of project activities, but only either sequential or parallel sequences. Mostly probabilistic multiple iterations and rework fractions due to iterations, which are specific to complex projects, are neglected in these models and cannot be simulated. Furthermore, a substantial deficit in these models is the activity-centred simulation approach. That means that the model dynamics are exclusively determined by the project activities, so that human and non-human resources are not explicitly taken into account. Due to the mentioned deficits the utility of the mentioned simulation models is significantly diminished.

To strive these shortcomings, the authors of this paper designed and developed a so called activity-oriented, personnel-integrated simulation model (according to the classification of organizational simulation approaches described by the VDI 3633 guideline [36]), which accounts for iteration loops, overlapping execution of project activities and assignment of human and non-human resources to project activities. Indeed, in an activity-oriented, personnel-integrated simulation model the model dynamics are determined by the activities, but human and non-human resources are explicitly considered. Activities can only be executed if the resources are available in the required quantity or capacity respectively. This kind of simulation approach allows a more realistic modelling and accurate simulation of the work flow of complex projects.

3 Personnel-Integrated Simulation Model

In the following a DSM-based, activity-oriented, personnel-integrated simulation model for the planning and scheduling of complex projects under uncertainty is presented.

In the context of service projects a certain service system definition described by Donabedian [6] and Hilke [12] was used for the modelling approach, which was assessed from three different perspectives: (1) potential dimension, (2) process dimension, (3) outcome dimension. Furthermore, according to Stachowiak [33] and Banks [1] a simulation model should be built upon specific characteristics considering only the partial representation of the real system and focusing on relevant characteristics for the specific purpose of the simulation study. For a comprehensive evaluation of a planned project only essential characteristics are attributed to the mentioned dimensions. The potential dimension comprises the organisational structure as well as human and non-human resources and their specific abilities and skills. The process dimension focuses on the process workflow and resource assignment. The outcome dimension focuses on the overall project outcomes quantified by key performance indicators. To model constructs and interdependencies between the dimensions both intra-domain matrices (DSM) and inter-domain matrices (DMM) are used (Fig. 1). The central dimension is the process dimension as it represents the dynamic part of the service. This dimension is influenced by the potential dimension and affects the outcome dimension.

The simulation model considers human and non-human resources. Human resources can be further differentiated in supplier's employees, provider's employees and customers. Relevant characteristics of the employees were identified from an exploratory survey [25]. Specific characteristics with significant impact on overall productivity are the ability (comprising qualification and experience) of the employees as well as their availability and cost rate. The relationship between human resources and organizational roles is modelled in the domain mapping matrix 'Employee-Role-DMM', a non-square matrix quantifying the ability of the given person on a scale from zero (none) to one (perfect). A role is a person independent organisational unit that combines in a goal oriented manner a series of abilities in order to run an activity. A role defines a profile of the working unit, thus enabling a flexible and scalable assignment of employees to activities.

The process dimension describes the sequence of activities for fulfilling a specified task defined by the performance planning. A process chain is thus composed of specific activities and their interdependencies. An activity is specified by an expected work effort quantified in full time equivalents. Because of the uncertainty induced by the customer a three-point estimation technique is used to construct an approximate probability density functions (PDF) of the effort based on the planners' experiences or best-guesses. It differentiates between a best case estimate D_b, a most likely estimate D_m and a worst case estimate D_w. Then, the effort is estimated by continuous-type random variables with a beta-distribution. The activity duration is automatically computed according to the parameterisation of the model by the simulation algorithm. Furthermore, in order to accurately model

		Resource Activity Mapping				Workflow Definition						Outcome Dimension			
		Role 1	Role 2	Role 3	Role 4	Activity 1	Activity 2	Activity 3	Activity 4	Activity 5	Activity 6	Duration [TU]	Costs [CU]	Effort [FTE]	…
		Activity-Role-DMM				Process-DSM						Activity-KPI-DMM			
Process Dimension	Activity 1	1										10	45	44	…
	Activity 2					0.2				1		5	78	66	…
	Activity 3			4			0.8					3	52	23	…
	Activity 4				2	1	0.5					4	98	11	…
	Activity 5		3					1	0.5		1	7	4	2	…
	Activity 6				1				1	1		8	56	78	…
		Employee-Role-DMM										Employee-KPI-DMM			
Potential Dimension	Employee 1	0.7										…	…	…	…
	Employee 2			0.8	0.5							…	…	…	…
	Employee 3		0.6									…	…	…	…

Fig. 1 Example of relationships between the considered elements of the simulation model

the influence of specific customers or stakeholders a minimum working time as well as a reasonable degree of task incompleteness can be assigned to specific activities. This is done by using four activity-based DSMs, referred to as process DSMs (see also chapter 4.2 as an example). In the first process DSM, the earliest start time (*EST*) of an activity has to be specified. Figure 2 shows an exemplary work process with activities 'A' to 'H' modelled in a DSM. The *EST* determines whether the activities are executed in parallel, sequentially, overlapped or coupled. If any dependencies are absent, two activities can be executed in parallel if all necessary information, employees or teams, and resources are available. In the event of a unidirectional dependency between two activities, the process starts with the information-generating activity (upstream activity). As soon as the necessary information is generated, the dependent activity (downstream activity) begins; both activities are executed sequentially. If the information is generated during the execution of the upstream activity at a certain point in time t_i, both upstream and downstream activities can be executed overlapped. The ratio to which the upstream activity has to be completed before the downstream activity can start is modelled in the DSM as quotient t_i/Sn_i, where Sn_i is the total duration of the upstream activity. In the event of a mutual dependency, two activities are processed sequentially or

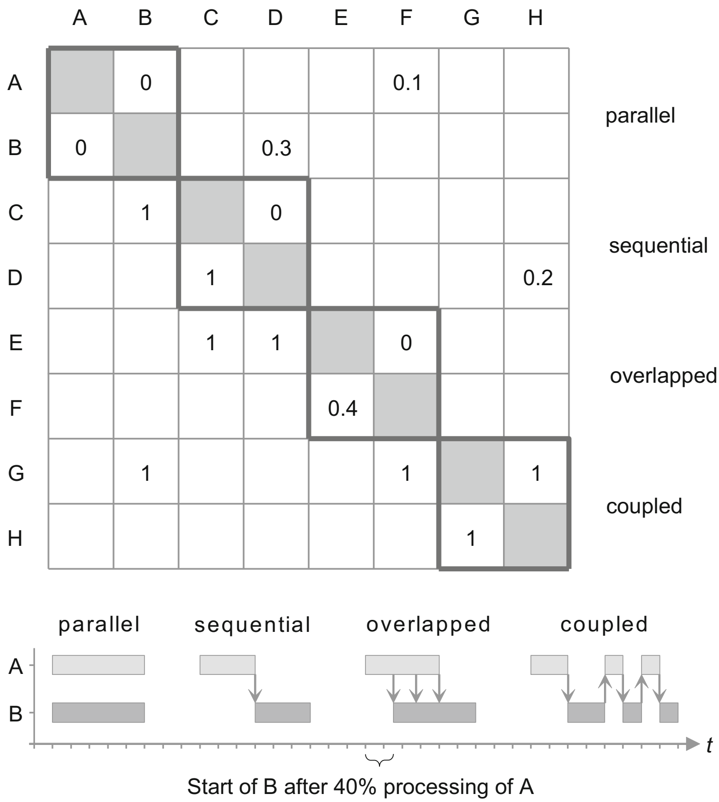


Fig. 2 DSM representing an exemplary work process (*top*). Framed cells in the DSM indicate different processing strategies. For overview purposes, zeros are not displayed except in framed cells. According to the IR/FAD convention (cf. [7]), reading across row ‘F’, for example, we see that activity ‘F’ is dependent on activity ‘E’. As the EST in row ‘F’ and column ‘E’ is 0.4, activity ‘F’ starts as soon as 40 % of activity ‘E’ is executed (that is, 40 % completion). Different processing strategies based on information dependencies modelled in the DSM, shown as bar graphs on a horizontal time scale as well as free graph notation (*bottom*)

overlapped in several feedback loops (iterations). This coupled process only terminates if a sufficient quality of information outputs is achieved for both activities. The different processing strategies described above enable the deduction and simulation of work processes based on the information dependencies modelled in the DSM. In the second process DSM iteration probabilities are modelled. They express the probability of iteration from a downstream activity to an upstream activity occurring during the execution of the downstream activity. The amount of rework is modelled in the third process DSM and defines how much of the primary work has to be performed during an iteration loop. The fourth process DSM models the factor of decrease of the iteration probability due to increasing quality.

In addition to planned iterations, which ensure the quality of the process, and unplanned iterations resulting from necessary corrections of incomplete, erroneous or incorrect assumptions as described above, the simulation can also consider changes and modifications to the project schedule. For example, these changes and modifications might stem from product changes, changing customer or client demands and requirements, changing legal regulations or the need for process optimisation. If a change to the project schedule is taken into account, a change vector has to be modelled. The change vector expresses the specific point in time within the project at which the change to the design occurs. Furthermore, the change vector expresses which project activities carried out so far (upstream activities) will possibly be affected by the change, and it indicates to what degree a certain activity is dependent on the change. The resulting rework due to a change in the process design is similar to the rework of upstream activities due to unplanned iteration loops.

For each activity the requirements regarding specific roles, requested suitability and quantitative demands are defined. Resources are assigned to the activities according to these requirements. The requirements are specified in the 'Activity-Role-DMM' as the quantity of the requested resource of a specific kind, thus closing the gap between activities and employees.

The outcome dimension represents the overall key performance indicators. Overall duration, cost and capacity utilisation are computed and visualised. These overall indicators consider not only resource characteristics like availability but also process characteristics like activity interrelations and additional effort due to overlapping or iteration induced rework. These indicators are to be evaluated in concordance with the project objectives. They serve as decision support for the optimisation of the project.

The simulation procedure usually starts with the definition of the project organisation structure and project breakdown structure. Project activities, resources and interdependencies between project activities and resources are modelled in according DSMs and DMMs. Furthermore, the user specifies the number of simulation runs and then initiates the simulation. Initially, the simulation identifies all activities that can be executed according to their precedence constraints, that means the information output of upstream activities enable execution of downstream activities. It follows from this approach that activities are executed either sequentially, simultaneously or partially overlapped. With a given probability, activities can be executed repeatedly in iteration loops due to coupling, uncertainty, discovery of errors or arrival of new information arising during project implementation. For each activity, the probabilities for iterations and the paths from each activity in the set to each other are modelled by the user and then described in the simulation using random variables. In case of iterations, the activity duration is increased according to the rework fraction and the activities are repeatedly executed for this duration. Iteration probabilities are reduced with each iteration loop and the coupled process terminates if a sufficient quality of information output is achieved by each of these activities. The variables modelled in the DSMs and DMMs are continuously adjusted accordingly during the simulation run. For each processing time event during the simulation run the simulation considers resource availability. If required resources in adequate capacity are not available for execution of a certain activity, the execution of the activity concerned will be suspended or interrupted. This simulation procedure is repeated until all activities have been executed completely. The total project duration corresponds with the interval between start of the simulation run and finish time of the last activity. During these simulation runs, the minimum, mean, modal and maximum value, standard deviation and variance as well as 5th and 95th percentile of the project duration and costs were calculated. Frequency distribution of the simulated project duration and costs (shown in the form of histograms as well as Gantt charts) allow detailed risk assessment in the run-up to important strategic and operative decisions. As a result, the simulation model can be used to iteratively improve and optimise the project schedule.

4 Industry Case Studies

In the following two explorative industry case studies on a plant development project in the chemical industry and a new product development project in the metal processing industry are presented. The case studies demonstrate successful operational application of the simulation-based project planning and scheduling approach and potential of respective project risk assessment.

4.1 Simulation of a Knowledge Intensive Engineering Service Project in the Chemical Industry

The case study concerns a subproject of an engineering service project in the chemical industry. It focuses on the development of a piping and instrumentation diagram. The case study presents early activities of the process design phase of a chemical plant development project. As starting point for the chemical process specification the project manager defines CAD design parameters and design rules. Based on a coarse chemical process plan the detailed process layout is drawn using CAD-software system. The finalised process layout has to be proofed and commented. Following the amendments the CAD-drawing is revised and finally released.

In this case study two different organisational scenarios were analysed. In the first scenario two process engineers are assigned to work on the activities, in the second scenario the two engineers are assisted by a technical drafter. The main emphasis of this case study is to determine the differences between the two organisational scenarios in terms of overall productivity. Figure 3 shows two different parameter settings of the engineering process with different resource allocations. In the square matrix on the left (process DSM), activity interrelations are modelled. Entries on the grey diagonal represent the estimates of the expected activity effort of a full time equivalent employee with 100 % ability. For example, considering activity 1 a skilled employee has to work at least four hours to finish the activity, most likely the employee has to work eight hours and 16 hours in a

									Setting 1			Setting 2		
Tasks		1	2	3	4	5	6	7	E1	E2	E3	E1	E2	E3
Defining CAD design rules	1	4/8/16							R	C	x	R	C	x
Developing chemical process	2	0.7	6/8/12	0.1	0.2				R	C	x	R	C	x
Drafting the process chart	3		0.1	14/20/25	0.5				x	R	x	x	R	x
Drawing process in CAD	4		0.7	1	8/10/12				C	R	x	C	C	R
Proof & amend CAD-drawing	5				1	4/8/10			C	R	x	C	R	C
Revision of CAD-drawing	6					1	3/5/6		C	R	x	C	C	R
Final release	7						1	2/4/5	C	R	x	C	R	C

Fig. 3 DSMs and DMMs representing two different case study scenarios (R stands for responsible, C stands for consulted, x stands for not assigned)

worst-case scenario. Entries below the diagonal represent the condition for activity activation. For example, to activate the activity 2, first activity 1 has to reach a degree of completion of 70 %. Entries above the diagonal represent iteration probabilities. The probability that activity 2 has to be iterated again after activity 3 is finished is 10 %.

The assignment of the employees is shown in the intra-domain matrix on the right-hand side of Fig. 3. Three different roles are considered: employee 1 (E1) is the project manager, employee 2 (E2) is a process engineer and employee 3 (E3) is a technical drafter. In this simplified case there is only one employee assigned for each role, thus the employee-role-DMM is obsolete. An employee can be accounted for a given activity; he is responsible for activity fulfilment. A consulted employee has to provide all necessary information to the responsible person.

Due to an overlapping processing of the activities a rework effort of 30 % from the current remaining effort is considered. Learning effects are taken into account by a decrease of the iteration probability of 50 %. Furthermore, the availability of the employee for the case study project is given and limited due to commitments in other projects. The project leader has a most likely availability for this project of only 20 % per working day (varying from 10 % in the worst-case scenario to 30 % in the best-case scenario), the process engineer of 30 % (varying from 20 % in the worst-case scenario to 40 % in the best-case scenario) and the technical drafter of 40 % (varying from 30 % in the worst-case scenario to 50 % in the best-case scenario). The communication effort is included by an additional effort of one hour per employee and week. Also considered but not depicted here is a reasonable incompleteness of activity 3 of 20 %.

The engineering service project was simulated considering two different organisational scenarios. For each scenario 10,000 independent simulation runs

Table 1 Simulation results of the engineering service project considering two different organisational scenarios (SD stands for standard deviation)

	Simulation scenario 1: without involvement of the technical drafter (E3)	Simulation scenario 2: with involvement of the technical drafter (E3)
Average project duration (effort-days)	29 (SD: 2.5)	26 (SD: 2.7)
Average project effort (effort-hours)	86 (SD: 6.0)	94 (SD: 6.5)
Average project costs (€)	7749 (SD: 533)	8058 (SD: 572)

were computed. The results are shown in Table 1. Following these results, first conclusions to the particular features of the specific project scenarios can be derived. A faster processing occurs in the second scenario, due to the involvement of all three employees. The average project duration is reduced by two effort-days. In turn, the total effort and costs increased. This is caused especially because of the higher communication effort due to additional interrelations. Furthermore, the process engineer is supported by the drafter, so that his utilisation decreases accordingly.

The relative capacity utilisation of the human ressources (E1 and E2) in the first scenario is high with a level of 65 % and 75 % respectively. The utilisation of the project manager even increases in the second scenario, the relative capacity utilisation of the technical drafter is only 25 %. Thus, the result under the given conditions shows that only 25 % of the originally scheduled time is used. This is possibly due to poor planning and allows other use of the technical drafter in further projects.

In particular, activities 4, 5 and 6 in the second scenario exhibit higher expenses as well as higher variations. This is due to activity sharing with the drafter. Thus, the second scenario has about 10 % higher deviation of time, effort and costs.

In summary, the simulation results show that the first scenario is robust regarding duration and costs (which is illustrated by the low variance), whereas it exhibits a high risk of capacity unavailability due to high capacity utilisation. The second scenario is more flexible with regard to resource allocation, thus, exhibiting high variability in service delivery.

4.2 Simulation of a Complex New Product Development Project in the Metal Processing Industry

This case study simulates a project for the customer-specific development of a cast iron part and corresponding design and production processes. The product designing procedure starts with a customer requesting a novel cast iron part. In the tendering stage, which takes place in close consultation and cooperation with the

Activity	D_b	D_m	D_w	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17	
1	1	1	1										1																								
												1	1																								
2	8	16	24	1	1																																
				.5	1																																
3	1	1	1			1	1																														
						1	1																														
4	8	12	16					1	1																												
								1	1																												
5	1	3	56								.05																										
										1	1																										
6	16	32	40							1											.2																
																				.3	1																
7	16	32	40							1											.2																
																				.3	1																
8	2	4	8											1	.5	1																					
														.2	1																						
9	3	56	112																1																		
																		1	.5																		
10	4	10	28															1													1						
																														.15	.25						
11	4	10	28															1													.75						
																														.15	.25						
12	4	10	28															1													.5						
																														.15	.25						
13	3	10	40																			1	1	1	1	1	1										
																						1		1		1											
14	1	4	8																									1	1								
																												.05	0								
15	1	2	16																											1							
16	4	16	56																													1					
17	27	40	100																															1			

Fig. 4 Project breakdown structure of the new product development project including the activity list with three-point estimates of activity duration (time units) and the process DSM. Each row and column of the process DSM contain the *EST* of an activity according to the upstream activity (*upper left quadrant*), iteration probability (*upper right quadrant*), amount of rework (*lower left quadrant*) and factor of decrease of the iteration probability (*lower right quadrant*)

customer, the technical feasibility of the requested part is estimated and conceivable alternatives are checked. Among other activities, in the development stage the manufacturing and casting technology is selected. Afterwards the manufacturing and casting preparation is implemented. The procedure terminates when the master sample is produced.

The main emphasis of this simulation study was to analyse various forms of the project organisation structure and project breakdown structure and how changes made to these structures affect the project duration. For reasons of confidentiality, in the following the project activities are presented in anonymous form and activity durations are specified as time units. The project breakdown structure and interdependencies between activities were modelled in corresponding DSMs. The project breakdown structure comprises a total of 17 activities which are executed either sequentially, simultaneously, partially overlapped or iteratively (Fig. 4). This explicitly demonstrates the complexity of the project and the difficulties in project planning and scheduling arising from this. The project breakdown structure particularly comprises large amount of coupled activities executed iteratively, especially in the tendering stage and development stage. Here it concerns planned iterations, which assist in the assurance of quality. The required amount of rework in iteration loops is essentially determined by the number of activities affected by these iterations. Thus, in the tendering stage even minor modifications by the customer cause large amount of rework. Project manager and project engineers used the

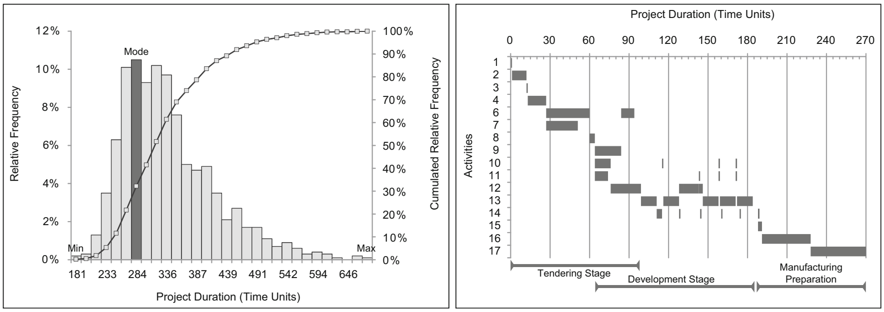


Fig. 5 Simulation results of the initial simulation experiment. Frequency distribution of the project duration (*left*), exemplary Gantt chart with average simulated project duration (*right*)

three-point estimation technique to determine the PDF of the duration of each activity based on their experiences or best-guesses. In the case study five departments of the company are considered. In each department one highly qualified employee is available to execute the activities assigned to the particular department. In addition to human resources, various non-human resources, such as e.g. computer, design software, FEM software and technical product documentation tools, are required in the analysed development project. But the non-human resources have not been decisively taken into account in the simulation study, since it could be assumed that these resources are available in the required capacity.

The initial simulation experiment resulted in an average project duration of 270 time units (TU) with a standard deviation (*SD*) of 79.9 TU. The minimum project duration was 181 TU, the maximum project duration was 680 TU. The frequency distribution of the simulated project duration (shown in Fig. 5 in form of a histogram) corresponds approximately to a logarithmic normal distribution, which means that an unanticipated long project duration is likely because the modal value of the frequency distribution is smaller than its mean value [13]. Accumulation of the project duration around the modal value of 270 TU is due to planned and unplanned iterations, which cause e.g. the amount of rework of approximately 80 TU in the development stage for each iteration loop.

Of particular note is a comparatively long duration of the two latest project activities for creating the production supplementary sheet and producing the master sample. The duration of these activities comprise nearly 30 % of the total project duration. Thus, these activities constitute a multitude of risks. Conversely, they offer a high potential for optimisation as project duration may be reduced by corresponding measures such as, for example, assigning additional resources or using efficient machinery and production technologies. Further measures to reduce project duration may be to optimise the project organisation structure and project breakdown structure from the perspective of CE as described in Sect. 1.2. For example, activities 10 and 11 could be executed simultaneously indeed, but in all simulation runs the activities are executed sequentially as only one employee (bottleneck resource) was available at a time and, thus, either activity 10 or activity 11 had to be suspended. So, the project duration could be reduced by 10 TU or 25 TU respectively through assigning a second employee to these activities.

The following simulation experiments analyse how changes made to the number of employees assigned to activities and project activities to be executed simultaneously in a multi-project environment would affect the project duration and the resource loading. Assigning additional employees to activities could potentially minimise the development period and project duration. The simulation provides accurate approximation of the percentage by which the project duration could be minimised. This can determine whether additional effort is to be considered worthwhile.

So, in the following simulation experiment the project organisation structure was modified, such that in each department three employees, instead of one employee, were available to execute activities assigned to the particular department. The simulation experiment resulted in a minimum project duration of 174 TU, an average project duration of 266 TU and a maximum project duration of 735 TU with a *SD* of 73.7 TU. Compared to the initial simulation experiment, on statistical average, an effect to the project duration could not be ascertained. The reason is, that, as mentioned before, only activities 10 and 11 offer potential to reduce the project duration as they could be executed simultaneously by two employees of the same department. Thus, with the help of simulation it was determined that effort involved in additional employees is not to be considered worthwhile.

Typically, the company receives several customer requests at a time, so that simultaneously several cast iron parts were developed or projects were running. For project managers the question arises how to meet additional requirements resulting from the multi-project environment, e.g. how the number of project activities to be executed simultaneously would affect the project duration and the resource loading. In the following simulation experiment three projects for the customer-specific development of cast iron parts according to the before mentioned project breakdown and organisation structure were simulated. Experience has shown that customer requests occur every 80 TU on average, hence, it was simulated that three projects were each started at time points 0, 80 and 160 and run simultaneously. As in the initial simulation experiment in each department one employee was available to execute activities assigned to the particular department. The simulation experiment resulted in a minimum total duration of all three projects of 654 TU, an average duration of 915 TU and a maximum duration of 1942 TU with a *SD* of 195.9 TU. Compared to the previous simulation experiments, which simulated single projects, the total project duration of each project in a multi-project environment was increased by 13 % on average as resources were not available in the required quantity or capacity and activities had to be suspended at certain points in time. Obviously, projects in the multi-project environment mutually influence each other by accessing the same resources. For example, the start of manufacturing and casting preparation of project 1 is delayed by 200 TU because required capacities of resources are tied up by development activities of projects 2 and 3 and blocked for the activities of project 1. In turn, the manufacturing and casting preparation

Table 2 Summary of simulation results of the project for the customer-specific development of a cast iron part

Simulation experiment	Minimum project duration (TU)	Average project duration (TU)	Maximum project duration (TU)	Standard deviation (TU)
Single project, one employee per department	181	270	680	79.9
Single project, three employees per department	174	266	735	73.7
Three projects in a multi-project environment, one employee per department	654	915	1,942	195.9
Three projects in a multi-project environment, three employees per department	360	542	938	77.2

activities of project 3 are delayed by 160 TU because required capacities of resources are tied up by activities of project 1.

Therefore, in a further simulation experiment the project organisation structure was once more modified, such that in each department three employees, instead of one, were available to execute activities assigned to the particular department in order to approximate the percentage by which project duration and resource loading could be minimised. The simulation experiment resulted in a minimum total duration of all three projects of 360 TU, an average duration of 542 TU and a maximum duration of 938 TU with a *SD* of 77.2 TU. Thus, it was determined that both project duration could be significantly minimised by higher resource capacities; effort involved in additional resources is to be considered worthwhile (Table 2).

5 Conclusions

Developing efficient and valid project schedules is often crucial for companies. The design and planning of the service processes must take due account of certain precedence relationships between activities and resource constraints. However, this is a complex cognitive task for project managers. Therefore, a discrete event, Monte Carlo simulation to plan and schedule weakly-structured complex projects was developed. The DSM-based simulation model allows the prediction of probability density functions for the total project duration and costs; it considers iterative workflow systems characterised by CE, overlapped execution of project activities and assignment of human and non-human resources to project activities. With respect to the industry case studies, the simulations performed so far have produced very good approximations of real project courses. The simulation results show a high correlation with historical data and expert knowledge. Besides, providing mean values for the total project duration and costs, the simulation model offers wide scope for reducing project duration and costs, minimizing unplanned iterations and rework and reducing risks associated with not meeting project deadlines and budgets. Although originally developed for the simulation of complex R&D and service projects, the simulation model is not limited to a particular field of application.

References

1. Banks J (1998) Handbook of simulation: principles, methodology, advances, applications and practice. Wiley, New York, NY
2. Bester H (2012) Theorie der Industrieökonomik, 6., korr. und erg. Aufl. Aufl. 2013. Springer-Lehrbuch. Springer, Berlin
3. Browning TR, Eppinger SD (2002) Modeling impacts of process architecture on cost and schedule risk in product development. Eng Manag IEEE Trans 49(4):428–442. doi:10.1109/TEM.2002.806709
4. Cho S, Eppinger SD (2005) A simulation-based process model for managing complex design projects. Eng Manag IEEE Trans 52(3):316–328. doi:10.1109/TEM.2005.850722
5. Deutsches Institut für Normung (2009) Begriffe, Januar 2009. Deutsche Norm, DIN 69901,5. Beuth, Berlin
6. Donabedian A (1980) The definition of quality and approaches to its assessment. His Explorations in quality assessment and monitoring, vol 1. Health Administration Press, Ann Arbor, MI
7. Eppinger SD, Browning TR (2012) Design structure matrix methods and applications. Engineering systems. MIT Press, Cambridge, MA
8. Fondahl JW (1987) The history of modern project management. Precedence diagramming methods: Origins and early development. Proj Manag J. 28(2)
9. Gaertner T, Rohleder N, Schlick C, Schulz L (2009) A simulation model for the planning and optimization of product development projects based on the design structure matrix. In: Zäh M (ed) 3rd International conference on changeable, agile, reconfigurable and virtual production (CARV 2009): Munich, Germany, October 5th–7th, 2009. Utz, München, pp 462–471
10. Gessler M (2011) Kompetenzbasiertes Projektmanagement (PM3): Handbuch für die Projektarbeit, Qualifizierung und Zertifizierung auf Basis der IPMA Competence Baseline Version 3.0, 4. Aufl. Qualifizierung. GPM, Deutsche Gesellschaft für Projektmanagement, Nürnberg
11. Götzfried A (2005) FuE im Unternehmenssektor der Europäischen Union. Statistik kurz gefasst, Wissenschaft und Technologie
12. Hilke W (1989) Dienstleistungs-Marketing: Banken u. Versicherungen, Freie Berufe, Handel u. Transport, nicht-erwerbswirtschaftlich orientierte Organisationen. Schriften zur Unternehmensführung, vol 35. Gabler
13. Huberman BA, Wilkinson DM (2005) Performance variability and project dynamics. Comput Math Organ Theory 11(4):307–332. doi:10.1007/s10588-005-5587-5
14. Karniel A, Reich Y (2007) Simulating Design Processes with self-iteration activities based on DSM planning. In: Systems engineering and modeling, 2007. ICSEM'07. International conference on, pp 33–41
15. Kausch B, Schneider N, Grandt M, Schlick C (2009) An integrative simulation model for project management in chemical process engineering. In: Bertelle C, Duchamp G, Kadri-Dahmani H (eds) Complex systems and self-organization modelling. Springer, Berlin, pp 211–232
16. Kelley JE, Walker MR (1960) Critical path planning and scheduling. Mauchly Associates, Ambler, PA

17. Kerbosch JAGM, Schell HJ (1975) Network planning by the extended metra potential method (EMPM). Technische Hogeschool Eindhoven
18. Kerzner H (2013) Project management: a systems approach to planning, scheduling, and controlling, 11th edn. Wiley, Hoboken, NJ
19. Licht T, Schmidt L, Schlick CM, Dohmen L, Luczak H (2007) Simulation-based multiple project management in engineering design. In: Krause F (ed) The future of product development. Springer, Berlin, pp 543–554
20. Lindemann U, Maurer M, Braun T (2009) Structural complexity management: an approach for the field of product design. Springer, Berlin
21. Malcolm DG, Roseboom JH, Clark CE, Fazar W (1959) Application of a technique for research and development program evaluation. Oper Res 7(5):646
22. Meyer K, Gerstner A, Thieme M (2012) Service Science and Service Engineering: Quo Vadis? In: Service and economic development: local and global challenges. The 22nd RESER international conference, Bukarest, 20–22. September 2012 (pp. 56). ISBN: 978–606-505–566-7
23. Motzel E (2010) Projektmanagement Lexikon: Referenzwerk zu den aktuellen nationalen und internationalen PM-Standards, 2., überarb. und aktualisierte Aufl. Wiley-VCH, Weinheim
24. Pennell JP, Winner RI (1989) Concurrent engineering: practices and prospects. In: Global Telecommunications Conference and Exhibition 'Communications Technology for the 1990s and Beyond' (GLOBECOM), 1989. IEEE, pp 647–655
25. Petz A, Duckwitz S, Schmalz C (2012) Productivity of services: an explorative study in the electrical and chemical engineering sector. Amfiteatru Econ 14(Special 6):635–652
26. Pietras CM, Coury BG (1994) The development of cognitive models of planning for use in the design of project management systems. Int J Hum Comput Stud 40(1):5–30. doi:10.1006/ijhc.1994.1002
27. Reichert T (2009) Projektmanagement: Die häufigsten Fehler, die wichtigsten Erfolgsfaktoren. Projektmanagement Klartext. Haufe, Freiburg i. Br.
28. Rosenkranz C (2009) Analyzing information flows in service networks. In: Thomas O, Nüttgens M (eds) Dienstleistungsmodellierung. Physica-Verlag HD, pp 35–52
29. Roy B (1965) Cheminement et connexité dans les graphes: Application aux problèmes d'ordonnancement = Routing and connectedness in networks: application to problems of scheduling = Verlaufsmöglichkeiten und Zusammenhang in Graphen: Anwendung auf Probleme der Terminplanung, [2e éd.]. Metra. Série spéciale, no 1. Revue Metra, Paris
30. Schelle H, Ottmann R, Pfeiffer A (2008) Projektmanager, 3rd edn. GPM Deutsche Gesellschaft für Projektmanagement, Nürnberg
31. Shtub A, Bard JF, Globerson S (2005) Project management: processes, methodologies, and economics, 2nd edn, Prentice-Hall international series in industrial and systems engineering. Pearson Prentice Hall, Upper Saddle River, NJ
32. Smith RP, Eppinger SD (1997) A predictive model of sequential iteration in engineering design. Manag Sci 43(8):1104–1120. doi:10.1287/mnsc.43.8.1104
33. Stachowiak H (2013) Allgemeine Modelltheorie, Softcover reprint of the hardcover 1. ed. 1973. Springer, Wien
34. Steward DV (1981) The design structure system: a method for managing the design of complex systems. Eng Manag IEEE Trans EM-28(3):71–74. doi:10.1109/TEM.1981.6448589
35. Tackenberg S, Duckwitz S, Schlick CM (2010) Activity- and actor-oriented simulation approach for the management of development projects. Int J Comput Aided Eng Technol 2 (4):414. doi:10.1504/IJCAET.2010.035395
36. VDI 3633. Simulation von Logistik-, Materialfluß- und Produktionssystemen: = Simulation of systems in materials handling, logistics and production 3633,3
37. Yassine AA, Whitney DE, Zambito T (2001) Assessment of rework probabilities for simulating product development processes using the design structure matrix. In: American Society of Mechanical Engineers (ed) Proceedings of the 2001 ASME Design Engineering Technical Conferences and Computers and Information in Engineering Conference: 13th

International Conference on Design Theory and Methodology. American Society of Mechanical Engineers, pp 1–9
38. Zhang X, Zhang S, Li Y, Schlick C (2012) Task scheduling behaviour in agent-based product development process simulation. Int J Comput Integr Manuf 25(10):914–923

Empowering Corporate Ageing Management by Interconnecting Existing Data: A Case Study from the German Automotive Industry

Michaela Kugler, Andrea Sinn-Behrendt, Ralph Bruder,
Grete Baumann, Lene Hodek, and Mathilde Niehaus

Abstract

In the context of demographic changes in Germany the question of how to maintain corporate competitiveness and employees' work ability becomes important. The case study in the automotive industry shows that in various corporate spheres of activity there are good initiatives but a strategic approach which interconnects existing processes in terms of a comprehensive ageing management is missing. Relevant corporate experts were therefore looking for support. The so-called FIT-model devised during the survey helps to make the issue in corporations more comprehensible. Five additional products were developed to support the organisation of performance indicators; to raise management's awareness of the issue; to communicate principles of ageing-appropriate work design; to reach the employees with the programmes of workplace health promotion and to improve the cooperation between companies and social insurance agencies. Initial feedback to the largely industry-independent results is positive.

Keywords

Work ability • Competitiveness • Demographic changes • Age management • Ageing management

M. Kugler (✉) • A. Sinn-Behrendt • R. Bruder
Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Germany
e-mail: sek@iad.tu-darmstadt.de

G. Baumann • L. Hodek • M. Niehaus
Unit of Labour and Vocational Rehabilitation, University of Cologne, Germany
e-mail: arbeit-reha@uni-koeln.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_28

1 Introduction

Demographic changes in industry necessitate a shift from short-term thinking and a youth-centred work and personnel policy to an ageing-oriented policy centred on a mature workforce as research published some 10 years ago pointed out [e.g. 2, 8]. The researchers described some pioneering work but were sceptical if and when this trend would be included in corporate strategic planning.

One industry that has been considering demographic change for a number of years is the German automotive industry. The workforce in the factories is roughly 45 years of age. Of the workers 5.5 % are regarded as disabled, the highest rate of all industries. For these reasons automotive companies have been considering measures to enable the workforce to age without health impairments, but despite some good initiatives to date no comprehensive process has been designed. Ageing management[1] that merges data across activity areas while also considering the employees' work history is needed.

Thus the goal of the project "Ageing healthily and well-qualified in the automotive industry. Participation and inclusion right from the beginning" (Gesund und qualifiziert älter werden in der Automobilindustrie. Partizipation und Inklusion von Anfang an – PINA) was to analyse the status quo and initiate the development of a process of systemic ageing management which might improve the interaction between divisions. The research was conducted as part of the Initiative New Quality of Work of the German Federal Ministry of Labour and Social Affairs by the Institute of Ergonomics and Human Factors of the Technische Universität Darmstadt and the Unit of Labour and Vocational Rehabilitation of the University of Cologne. The following corporations participated actively: Audi, Daimler, Evobus, Ford, Opel, Porsche, Volkswagen and Bosch, from the automotive parts industry. The project was supported by the human resources directors of these corporations as well as by the Association of Disability Councils of the German Automotive Industry.

Individual programmes were of interest but since management is important the interconnection between planning and evaluation was the main focus of the survey. A further focus was the area of internal communication as well as the interface between industry and agencies of prevention and rehabilitation (such as medical insurance companies, accident insurance companies and government and private pension schemes). To ensure that the data would be manageable, only those corporate spheres of activity relevant to the survey were consulted, viz. workplace health promotion; workplace (re)integration; job organisation; workplace design; working time; personnel and career development; further training and education. Transfer of knowledge was in the initial phase of interest; management, in the sense of team leadership, become part in the course of the investigation. Little time was spent on matters such as recruiting, remuneration or company pension schemes.

[1] In this article the term Ageing Management is used consistently.

The project surveyed a single industry only, thus the companies were comparable and important similarities became evident. But almost more significant is the fact that noticeable differences were found. Only this totality of findings made it possible to develop the measures presented here to strengthen ageing management in industry.

2 Approach

The project was loosely structured and filled with content as it progressed. On the basis of an enquiry into the status quo, themes emerged that were analysed in conjunction with the corporations. At the same time social security agencies were consulted.

For the survey into ageing management a qualitative approach was taken. As a first step (cf. Fig. 1) a partly standardised questionnaire was sent to all participating companies to gather data about employment structures, company agreements, projects to date and the people involved. From March to May 2012 interviews were conducted with experts in the relevant divisions in the companies about their activities. In addition to questions regarding work processes the focus in the approximately 60 interviews was on communication with other areas as well as on future plans and needs. The interviews were based on questionnaires which were specially designed for, and specific to, the spheres of activity. Depending on the company the discussions were also attended by representatives from the works council and/or the disabled employees' council. Furthermore, in most companies senior managers were available to discuss the corporation's overall strategy. The agenda for the discussions and the selection of interviewees were left to the project co-ordinators within the companies.

The companies' statements were augmented and compared with documentary material (internal company agreements, work manuals, flyers), then subjected to an initial accretion and compared with other statements to discover similarities,

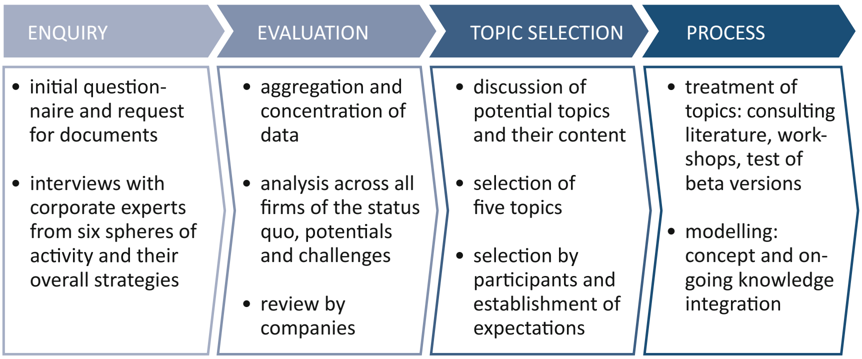


Fig. 1 Developing the project

differences, trends and challenges. The results were thematically integrated and presented to the companies to ensure accuracy. The final report, accompanied by suitable examples, was presented to the participants in November 2012.

The evaluation was three-dimensional: sphere of activity (all participating companies); the individual firms (across all spheres of activity) and an overall summary. The initial evaluation was done along the criteria of the questionnaires; in the course of the study further criteria, resulting from the analysis of the data, were added. Workplace health promotion, for example, was analysed on the basis of frequency, target groups and target group definition, as well as on ease of entry, budgets, people involved and performance indicators. In the area of workplace (re-) integration the framework (company agreements, areas of responsibility, triggers, people involved) was analysed. Also of interest was entry into the programme, the interconnections across divisions and again the performance indicators. Workplace design, to give another example, was evaluated on the basis of evaluation tools and the frequency and regularity of their application. In addition, aspects relating to ageing management (consideration of employees' abilities or adaptation of working conditions to accommodate ageing) were surveyed. The evaluation across all spheres of activity concentrated on the breadth of the measures, how stringently they were applied and how systematically they were interconnected. Comparing such measures across all participating companies led to innovative solutions and highlighted the potential for improvement. The differences due to size, product, production and corporate structure were taken into account.

The participants' reviews showed that this type of comparison, the case studies and real-life experiences, were infinitely valuable. As a result the format of the process was changed. Instead of strengthening ageing management by in-depth survey in each of four participating companies all eight companies were invited to participate in a study across four (later five) themes of ageing management. The topics, selected by all participants, were investigated by experts from three to six companies. The experts' participation was voluntary and depended on the topic's relevance to their expertise. The process comprised a number of steps. Initial workshops concentrated on needs and contents, followed by workshops to explore successful programmes and to review initial presentations. The comprehensibility and usefulness of beta versions of the products, generally manuals, were tested in participating companies. At the same time a comprehensive model was developed to reinforce the topic of ageing management and simplify its dissemination.

The findings of the above process and the resultant list of topics is described in Sect. 3, the outcome of the project in Sect. 4.

3 Findings

Despite the differences in corporate structure and size the results of the qualitative investigation were comparable and confirmed the initial assumptions. While all companies have good processes in place not one has an ageing management programme, as described in Sect. 1. Many of the processes are well structured but

Fig. 2 Challenges evident from the project development

they aren’t connected across divisions. The proposals and goals of the programmes are often similar, though they differ on details and the degree of implementation. Furthermore, different emphases and strengths were evident between the companies due partly to company culture, partly to the implementers. However, similar problems and challenges could be found across all spheres of activity. Figure 2 gives an overview of the initial results. They will be looked at in depth in the following pages which will also explain the choice of the topics that were analysed during the further survey.

3.1 The Findings in the Spheres of Activity

The major challenge companies face in the field of workplace health promotion is how to reach the workforce. This generally leads to a broadening of the target groups and a lower entry into the programmes [20]. In addition to programmes aimed specifically at production workers there is a trend towards offering medical check-ups to all employees. At the time of the investigation two companies already had such programmes in place; others are planning to introduce them even if the organisational aspects differ (in-house-examinations vs. external, intensive examinations in central rooms vs. a check-up in doctors’ rooms close to the factory). Further programmes need work in defining the target groups, others in implementing performance evaluation.

Introducing workplace (re)integration is often problematic because of friction between external and internal interfaces. The companies have well-planned programmes in place, the details of which obviously differ. However, external conditions, especially with regard to social security agencies, often make it difficult to (re)integrate personnel. Internally, lack of cooperation between divisions sometimes makes the work challenging, e.g. transferring a person from one division to another. Therefore, the managers of these programmes see one of their major jobs as raising their colleagues' awareness of the problems. Examples of functioning networking across divisions can be found in the field of integration management, which was examined in this context. Additionally, integration teams often develop forward-looking programmes for workplace design and health promotion. However, in order to implement them, management often has to be brought on board, something the teams try to achieve through performance and economic indicators.

Awareness and systematisation are also important issues for work design. In production, all participating companies use evaluation tools for (physical) work load but they differ in the frequency and quality of implementation. Some companies have programmes in place that consistently consider ergonomics in early phases of job planning, but information about workers' (dis)abilities and prognoses about their development is seldom utilised. This leads to a reduction in suitable work places as workers' abilities become impaired. For example, very few companies consider a change from a standing to a sitting job when designing a new production line. For office jobs central planning is largely confined to the office and desk, job organisation is left to the teams involved. The experts in the companies felt that the programmes could be improved if workplace assessment and data evaluation were applied more systematically, in addition to alerting (operating) managers to the value of the programmes. A valid question is what a workplace that is ageing-appropriate would look like. This is a field where the companies feel that research is needed as is the area of mental stress, how to recognise and reduce it.

In the area of working time there are many challenges such as communicating the working hours to target groups or their consistent implementation (that might touch on legal issues). Especially introducing ageing-appropriate shifts is difficult since not only management but also the workforce often has to be convinced that it is a good thing. Instead of a three-shift model most companies prefer a voluntary extended night shift. However, this model does not always have a time limit and in only one company does it have an age limit (discretionary). Short-term work-time accounts are common although different models are used. Long-term work-time accounts that would enable the introduction of age-appropriate working hours are rarely in place because of problems with accruals and tariff agreements; workers seldom make use of such programmes.

Personnel development is highly systematised in the firms, which all contain a variety of programmes. Nevertheless, expanding development perspectives is still a challenge [3]. Normally the annual employee appraisal is used to initiate on-the-job or advanced training; two companies are considering introducing this kind of appraisal and the chance for training throughout the year. All of the participating companies offer their managers, skilled workers and foremen the opportunity for further education, only two have on-the-job training that allows employees to progress from worker to manager. In the area of succession planning there are major differences. Flexibility in jobs is partly only offered intra-departmentally, a flexible career path is still not really possible. Health-oriented career planning is to date only reactively granted in terms of workplace (re)integration.

Advanced training throws up the question of how to ensure that necessary skills are acquired and how to strengthen the learning culture. To achieve this all companies continually expand their curricula, both technical and non-technical, though there are differences with regard to access and teaching format. In addition, there is a movement towards target group training (for certain jobs or areas). In one of the companies this leads to creating job clusters and defining requirement specifications for them. Yet, it is not clear, whether all employees can be reached in this way and how a more target group-specific didactics could be installed. Little can be said about the sphere of activity of knowledge transfer since very few companies have a manager for the process nor do they have explicit programmes in place. Answers regarding training indicate that experts in the companies would be used as trainers and also that programmes such as mentoring/shadowing and succession planning are in place. Other processes in the field of knowledge transfer could not be identified.

Despite the marked differences shown up by the survey, there are some striking similarities. Bringing (top and operational) management on board is a recurring concern across all areas of the survey, including that of a learning culture. Another topic of concern in all companies is how to reach employees in the fields of e.g. health promotion, working time or training. Very often the existing programmes and processes have not been thought through systematically. In some areas, e.g. job design, the problems can only be solved with the help of executives. No strategies on how to raise their awareness have yet been found.

3.2 Interconnection across Spheres of Activity

Analysing the degree of interdepartmental connectedness brought to light some fascinating results. It showed that to date networking was basically due to individuals (bilaterally or in work groups), not policy. Only a few instances of connecting programmes across departments were found, viz. stress-related measures in the workplace health programmes and/or taking into consideration disabled workers when planning a new production line. It is true that demographic change is being discussed in the industry but mainly by a few engaged individuals in departments of human resources, corporate health services or (re)integration and ergonomics managers. If resources are needed for operational reasons, e.g. to ensure that a production line is built in a particular plant, these "demographers" lose support [9]. Demographics-oriented projects which demand the interconnection of individuals can be found in different firms (either as stand-alone projects or as part of an overall strategy); but it has proven difficult to turn such networks into a permanent feature.

Consequently, although there are a number of good initiatives in the field of demographic change they are often stand-alone programmes. Except for some short-term projects there is no systemic attempt to build interconnections across departments nor are processes being developed to achieve this. Also for this, the support of management is regarded as crucial but seen as a challenge for which no strategic concept is to date discernible.

3.3 Thematic Focus

What was said above is reflected in the list of topics: identify key performance indicators and profitability, raise managers' awareness of the issues and secure their leadership in projects, describe the principles of ageing-appropriate work design, communicate the topic to the workforce and promote employee participation, improve cooperation with social service agencies. The fact that these topics are regarded as urgent may well be due to the project coordinators who come from departments of human resources or programmes for workplace health promotion and additionally often act as head of a demographic project. Even so the list shows that they look for support and programmes to communicate the topicality of the subject to their audience.

On all these topics data are available in the companies, meaning such programmes could be developed internally. However, due to the pressures of the day-to-day business little time can be devoted to such projects. The survey gave these people the opportunity to meet and profit from the experience of their peers in other firms. Existing programmes were conflated and from the resultant product manuals were developed which can help people in ageing management in their day-to-day work and thus strengthen their effectiveness. Concurrently it was thought necessary to develop a model into which individual steps can be fitted in order to enhance the systematisation and thus the awareness of the programme. The results of this work are described in Sect. 4.

4 Results

The section starts off briefly describing the background and contents of this model, followed by a review of further products. The centrepiece of these products is the so-called performance indicators matrix, which deals with key indicators in the areas of strategic and ageing management. To this end the spheres of activity covered by the investigation were reviewed from a different perspective. The matrix warrants its own sub-section; the other products are covered cursorily in the rest of the section.

4.1 The FIT Model

German, which divides the internationally used term "Age Management" into Altersmanagement (age management) and Alternsmanagement (ageing management) contains different definitions of the latter (cf. the list in [4]). Common to the

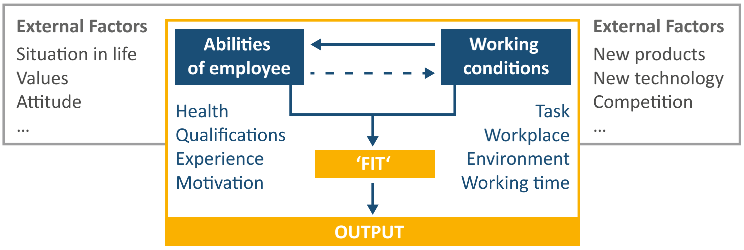


Fig. 3 The FIT model

definitions is the emphasis on maintaining the work ability of the employees over the course of their working life and on the measures taken to achieve this. Other authors use health management as a synonym for this complex [e.g. 18]. The major drawback of these approaches is that the connection to the company's goals tends to get lost. In manufacturing firms that define themselves through their products and production technology this easily reduces ageing management to a function of health promotion or human resources management that has dutifully to be performed. But ageing management does much more. It contributes to retaining the firm's or the factory's competitive edge globally (and also locally in the scramble for corporate funds) despite an ageing workforce. Maintaining the work ability of the employees is simply one aspect. To clarify this connection, the FIT model (cf. Fig. 3) was developed over the course of the project.

What may at first blush look like a profile comparison as used in (re)integration is in fact a general way of looking at a larger area of operations or a collective, ideally the whole of the workforce. The model raises the awareness of comprehensive connections and risks and thus makes managers more goal-oriented. It contrasts characteristics, strengths, skills and needs of the labour force, bundled under the term Abilities of employee, with the Working conditions and indicates that these two fields continuously have to be adjusted in order to achieve a good fit and thus an improved output for the firm.

Working conditions can either enhance or hamper labour's skills; on the other hand labour's needs can influence working conditions. In the model this is symbolised by the small arrows between the two sides. Additionally, both sides are subject to external factors. One's situation in life affects the working framework; values and attitudes influence what is expected of the job and the job situation. Customer demands, technology and the competition are decisive factors on the part of corporate leadership for job design. The plant's location as influenced by infrastructure or supply of skilled workers, etc. can play a further role.

The spheres of activity fit the model. Spheres of activity like health promotion or on-the-job-training originate with the work force, job design with management. (Re)integration functions directly in the FIT, even if mainly on the level of the individual. Leadership depends on context as regards either working conditions, the promotion of employees or the FIT. Comparing the two sides (abilities of employees and working conditions) dependencies become clear that can be used for forecasts which in turn may orient actions in the spheres of activity. Neither side has the upper hand; on the contrary continuous adjustment is needed, even without explicit reasons. Whereas other models show output as a direct result of skills and working conditions [e.g. 6] this continuous adjustment is what “makes it fit”. Output in this meaning is not a variable but part of doing a job. This will become clearer when looking at the key indicators (Sect. 4.2).

The process as described is built on data and decisions that cross divisional lines; interconnections are crucial. Ageing management is thus not a matter for one department only, it needs collaboration by different experts from various hierarchical levels [12, 9]. Only in this way can systems be developed that indicate where the two sides diverge, in order to work on them. The executive has to create an environment that fosters transparency, to introduce change when needed and to resolve potential conflicts. Junior and middle managers regularly have to review the situation in their areas of responsibility, review the need for action and, with input from experts, resolve the issues. The planning, executing and evaluation of the details fall in the purview of the managers in the spheres of activity. The works council and the disabled workers’ representative are important disseminators.

The contents of the FIT model are not new. Obviously, all companies have managers who work in this manner. What is new is pointing out the interdependencies and to demonstrate that ageing management is a common objective, part of the company’s goals. The FIT model is helpful in explaining these connections to all parties involved. The performance indicator matrix, described below, supports this approach.

4.2 The Performance Indicator Matrix

Key indicators are of utmost importance for management intervention in this complex field where vastly disparate business areas interact. However, because of the immense variety of factors that for example influence the health or motivation of the workforce, it is not possible to posit a linear cause and effect. Therefore, using a single indicator only, e.g. the absenteeism rate, is not meaningful; it is necessary to use a combination of indicators [cf. 22].

Appropriate sets of indicators are required on three levels: on the level of the operational divisions, the level of departmental management and on a comprehensive strategic level. The companies already collect data on all three of these levels. However, whenever certain programmes have to be strengthened because of demographics, companies start thinking about what additional data might be useful and what the competition does in this regard. This also became very clear in the course of the survey. To be of help in this situation and to be able to offer a catalogue raisonné against which firms can check their key indicators and expand their list if needed, a number of sources were evaluated. Data from the survey were discussed with experts from the participating companies, followed by a questionnaire which inter alia supplied information about how the figures are incorporated in the reporting system and used in day-to-day management. Since the result still did not suffice to present a systematic structure, the initial data were re-analysed and augmented by more information from the companies and from relevant literature. It became evident that the literature contains extensive sections on how to implement processes in the different spheres of activity, but very little on key indicators in these fields. The works consulted covered mainly the areas of human resources, health and ageing management; a bibliography is shown in Türk [21].

Most indicators were found in literature dealing with workplace health promotion [cf. 2] and on-the-job training; however, they are mainly confined to a description of the programmes. More extensive analyses, e.g. work area or age, as suggested by Langhoff [17] for ageing management, are little used. Books on corporate (re)integration programmes contain many ideas for introducing such projects, but few performance indicators (the exception being again Langhoff [17]). Performance indicators also are rare in the areas of personnel and career development; the same applies to job design. Literature dealing with leadership in a time of demographic change covers the topic health(y) management with very few criteria (for the executive) to ensure quality. Virtually no book discusses more than one sphere of activity at one time. Balanced score cards offer an interconnected approach [cf. 14] but their focus is different. Uhle and Treier [22] describe a holistic approach which deals at length with health monitoring and makes the case for performance indicators that connect hard data with the results from questionnaires, as well as so-called early and late indicators. But they, too, do not provide real-life indicators. On offer are aids for the monitoring of health management that correlate data from questionnaires on health with other company key indicators [10]. However, they look at only a few of the performance indicators important for ageing management and, because they are closed systems, cannot easily be combined with other ratios. Referring to the extended definition of health of the WHO, viz. health is not the absence of illness, but a state of physical, mental and social well-being [1], well-being is also looked at as a competitive factor; concrete numbers are not to be found.

For this reason a system of performance indicators for the various spheres of activity, useful for strategic management and business reporting, was developed. A first overview of the so-called performance indicator matrix and the underlying idea is contained in Fig. 4. The two sides of the FIT model here are characteristics that could largely be reduced to numbers. Their juxtaposition can point up a variety of qualitative and quantitative risks. One example would be if a high number of workers reach pensionable age, making it difficult to hire replacements. Demands posed by the job and workers' health may diverge sharply as may working hours or skills levels. This could result in underutilised potential, lack of skills or understaffing that may influence the quality of task fulfilment. Ageing management's goal is to recognise such risks and to take timeous counter-measures. The so-called FIT indicators provide a quick overview. When these indicators are given numerical goals they can furthermore be utilised as a framework for measures to be taken in the spheres of activity, increasing the fit and thus decreasing potential risks. The measures can be assessed by making use of the level or quality of the

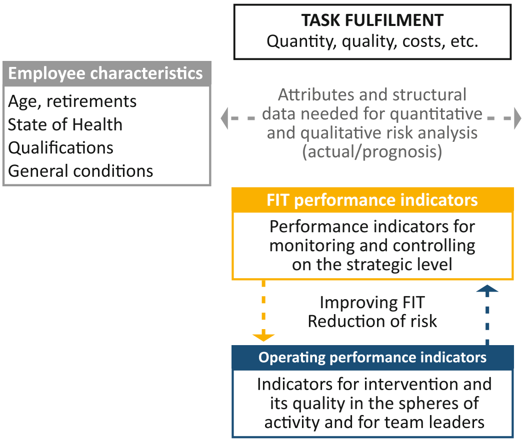


Fig. 4 Performance indicators for ageing management

action, shown in the overview as operating performance indicators. Examining more than one of the spheres of activity may show up a conflict of interests. In addition, it is useful to measure the level of task fulfilment as an overarching indication of quality as well as goals that may influence ageing management (e.g. productivity or costs) and to look for possible correlations with the other ratios. The definition of "task fulfilment" varies from firm to firm and product to product but will essentially be categorised under the headings quantity, quality, cost and other criteria.

The matrix of indicators largely covers the spheres of activity that were at the heart of the survey, with two exceptions. Working time, though part of the survey, was ignored since only a very limited number of performance indicators was discovered when analysing initial responses. Conversely (team) leadership, that was not part of the initial survey because of its decentralised organisation, has been included as it plays a crucial role in ageing management.

4.2.1 Operative Indicators

All eight companies already collect data in the various spheres of activity, if only for statistical purposes such as the workforce (e.g. its age structure); job demands and conditions (e.g. workplace assessments); rates of absenteeism and job satisfaction as initial indicators for the fit. In addition, in the spheres of activity data such as costs or utilisation of training courses are collected. However, collecting and analysing the data are not always sufficiently systematised. Number of courses offered, participation in them and similar data are collected but not interconnected. The numbers are furthermore rarely collated with other data. For example attendance of courses is seldom analysed from the point of view of age, operating department or length of service or skills. As an aid in getting such a process going for every sphere of activity, a list of possible indicators and their combinations was compiled and structured along the lines of initial input, measures, quality of measures, budget/costs and overarching goals [cf. 22]. However, for purposes of aggregated reporting (e.g. to human resource or factory management or for a report to a task force) the level of detail necessary for planning and evaluating such measures would be too comprehensive. For this reason an additional list was compiled which shows the data that can be used for condensed reporting. The topics were selected in conjunction with the companies participating in the survey.

4.2.2 FIT Indicators

In particular two FIT performance indicators are regularly used in industry, viz. absence rates as calculated from reports on sick leave and job satisfaction as measured by employee assessments. On their own these indicators are not really conclusive. Attendance at the work bench may be a first indicator, but it does not indicate anything about the actual state of health of the workforce, nor is it a meaningful measurement of the fit between workers' skills and work conditions. A survey of job satisfaction can provide additional information; however, the answers are snapshots of conditions at the time of the survey. Therefore in the course of the project a search for performance indicators that would augment these numbers was launched. A precondition was that the indicators should apply to both the workforce and management. To ensure that everything was taken into account the goals and the concomitant performance indicators of both sides were surveyed in depth. The process is shown in Fig. 5.

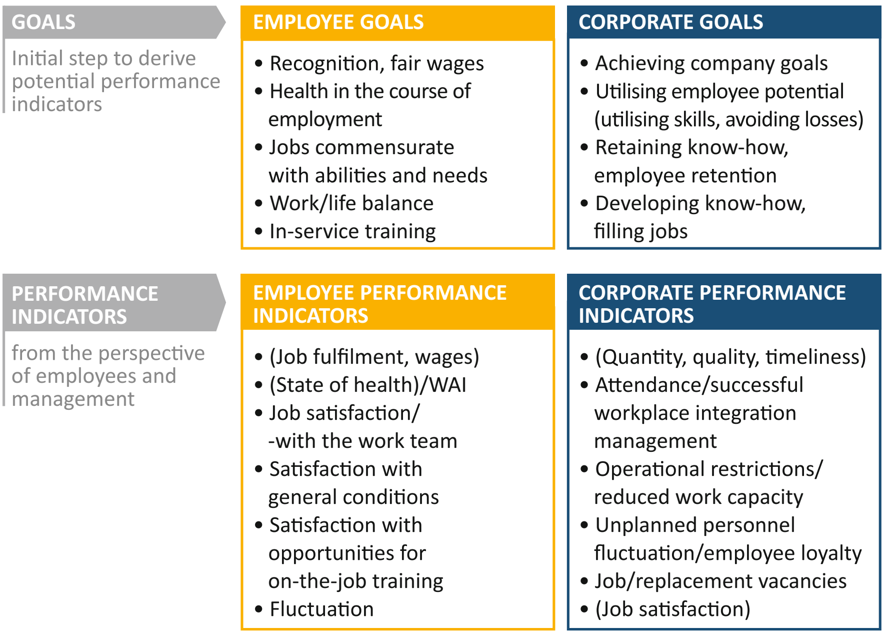


Fig. 5 Goals and potential performance indicators: employees and management

It was soon clear that though the needs were formulated from different perspectives and independently of the order in which they were listed (the importance and order of the workers' goals depend on phase of life, social environment and other factors) they can be seen as being intertwined. For example, the workers' wish (even if subconsciously) for work that measures up to their abilities, i.e. work that does not permanently demand too much or too little from them, can be seen as meshing with management's goal of utilising workers' potential to the utmost. This means avoiding outages or situations in which the workforce is unable to use its full potential. The Work Ability Index (WAI) [15] or an equivalent measure of self-assessment on the part of the workers, and the "reduced work capacity" supplies a set of indicators for a situation that occurs when job demands and the workers' abilities do not mesh.

A total of seven FIT indicators were identified: job attendance; long-term illness (in other words job disablement resulting in more than 6 weeks off work); a WAI (or similar index) score; reduced work capacity (expressed in work years or financially); the percentage of vacancies; the overall score of job satisfaction; unplanned fluctuation in the workforce. When combined these numbers give a good initial impression of a potential divergence. Further indicators, such as the average period of employment as an indication of loyalty or pensionable age were not considered since they are dependent on the age structure within a corporation and therefore difficult to use for goal-setting. It is, however, recommended to use the numbers to aid in the analysis of the FIT indicators. To get a value for "well-being" as mentioned earlier, a combination of FIT scores (e.g. job attendance, the WAI score and job satisfaction) can be utilised.

The usefulness of the performance indicator matrix as a basis for and assistance in structuring indicators for an individual company was proven in the course of the project.

4.3 Additional Components

Looking at the performance indicators it is evident that they implicitly are linked to other topics. Important criteria in the areas of health promotion and job training are target group(s) and successful communication with them; (team) leadership regarding ageing management requires data; successful (re)integration is the decisive criterion for (re)integration management, etc. However, a hierarchy can be seen in the topics (cf. Fig. 6).

The organisation of performance indicators underlies analysis and management, while managers who are passionate about the process are potent multipliers. Basics of ageing-appropriate job design are useful in raising awareness while at the same time addressing one sphere of activity. The same applies to the other two components. Employee participation was looked at during the survey of corporate health promotion, and cooperation with external agencies when looking at (re) integration management. These can be expanded with additional components. The basic approach of the project, used for the performance indicator matrix, namely to

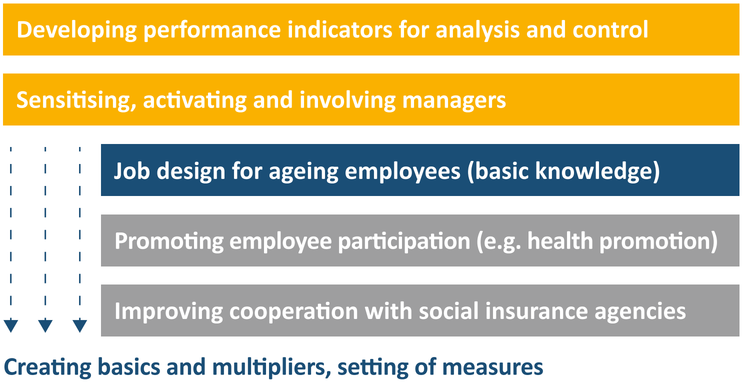


Fig. 6 Components for an improved ageing management

collect extant research, re-analyse and expand it, was also used to develop the other products that emerged from the project, e.g. a manual to raise the awareness and active participation of managers; guidelines for workplace design that is fair to ageing staff members; a check-list for improving programmes of workplace health promotion and guidelines to improve the cooperation between industry and social insurance agencies. The manual for managers is discussed in some detail while the other products are summarised. The products are addressed to the same groups as the indicator matrix, viz. internal and external managers in the spheres of activity as well as executives in charge of internal projects.

4.3.1 Manual for Managers

Team leaders play a crucial role in ageing management, not only as far as task control and team management are concerned, but also in influencing conditions in the workplace. They organise job assignments, are important movers in the area of staff and career development and may influence the participation of their staff in on-the-job training and workplace health promotion. Managers are role models, multipliers, sponsors, and organisers of the job structure and of inter-personal relationships, roles that they partially have to be taught. However, when thinking about leadership in the context of ageing management (or similar projects) the knee-jerk reaction is simply to provide information regarding demographic change [17].

That more is needed became clear when factors which were used successfully to raise the awareness of managers about ageing management were identified. One process in particular was striking, because it benefitted a number of components, viz. self-reflection, considering the team's (dis)abilities and discussions with work-force and experts. From this and other examples, factors that lead to the desired outcome were identified with the aid of experts from the fields of executive development, training, health management and human resources, as well as from books [e.g. 11]. The result is organised in an intervention model (Fig. 7) leading in three steps to increased appreciation of the problems and participation by managers.

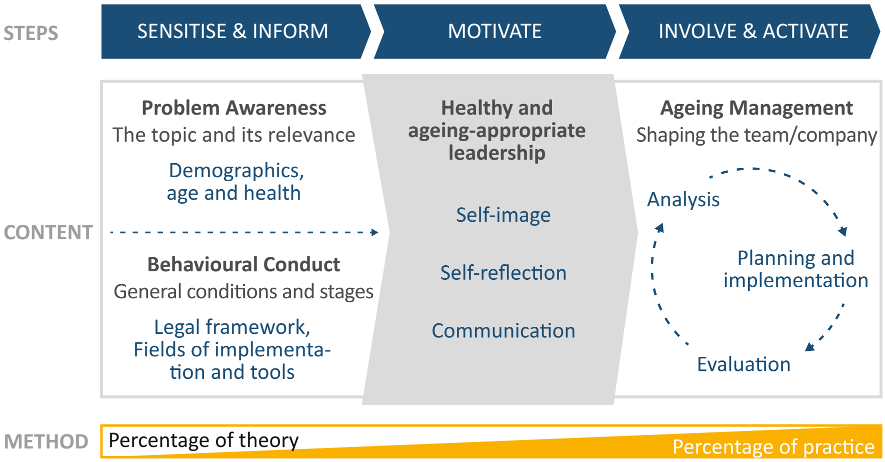


Fig. 7 Intervention model to sensitise/activate managers to ageing management

The first step (sensitise & inform) raises awareness and affects subsequent behaviour. The second step (motivate) aims at creating good team leadership through self-reflection and understanding one's role as leader. Both of these steps are often taught in management seminars and courses. However, of equal importance is a process that supports managers in playing an active role in ageing management. For this reason a third step (involve and activate) aims at integrating ageing management in the team process; this only works if it is part of company regulations and if experts are involved to support it. The steps, especially raising awareness and motivation, are indications only as the areas cannot be separated clearly. It is, however, crucial that all the criteria are considered, ideally as part of a total concept. This intervention model and the individual criteria are described comprehensively in a manual which is backed up by learning aids.

4.3.2 Further Components

People involved in ageing-appropriate workplace design face a number of questions. The participants in the survey acknowledged that the literature contains hints to this complex but deals only with single aspects. A reference that covers all aspects simply and comprehensively is missed. As part of the project therefore an introductory guide was written which starts off making clear that ageing does not automatically lead to a reduction of skills but to their variety increasing [cf. 7]. For this reason it, unlike other manuals, differentiates between ageing-related and ageing-unrelated changes. The model developed in the project covers ageing-appropriate workplace design in general; measures for possible ageing-relevant changes and measures for general disability.

A problem often raised in connection with workplace health promotion is how to communicate successfully with the workforce. Health-related programmes are used mainly by people who are health conscious anyway [cf. 5]. Obviously the individuals themselves decide which programmes they take up. However, the companies are responsible for the programmes and the conditions under which they are offered once it is clear which workers should be targeted and for what reason, something that is often neglected during planning. Therefore, with the aid of experts from three corporate health departments a check-list was compiled to address this topic. It starts by explaining how to define the problem and the concomitant target group(s) and moves on to describing which aspects of content and organisation have to borne in mind when designing such a measure. The check-list consists of two parts, a summary as a primary overview and secondly an expanded version with approximately forty items to consider.

If the physical limitations and disabilities of workers are so severe that corporate health departments alone cannot resolve them, corporations and the people concerned have to rely on outside agencies. Because of the framework of the German social insurance system this cooperation is not always harmonious. Different competencies, a variety of agencies and contact points for prevention and rehabilitation often complicate participation in a rehabilitation programme [13]. Many institutional participants try to smooth the process in their area of responsibility but this does not really improve the relationship between the corporate world and social agencies [19]. One approach is that the agencies and the corporation agree on a set of guidelines to be followed by both parties. This has been tried but was not often successful because the guidelines were not extensive enough. Therefore, a manual developed with the aid of experts from two corporations details what has to defined when considering such cooperation; an exemplary flowchart shows how cooperation could function in case of a rehabilitation measure from the moment a need is diagnosed until the aftercare has been organised.

5 Discussion and Outlook

The project was carried out in the automotive industry, innovator of processes but also known for complex and variegated products. Technology, vehicle design and variety of models are all subject to rapid changes, processes often product-driven. Ageing management in such an environment is not simple, let alone reconsidering existing strategies and programmes. However, in view of demographic change such measures are vital. To raise the awareness of this problem people who to date work in this field need support. The areas investigated in this project are the major components of such a process. The FIT model, performance indicators and raising awareness among managers help to elevate the profile of the issue in corporations and to gain more support for effective ageing management, leading to more interconnection of existing processes. The other products concentrate on aspects of primary health care as well as workplace (re)integration. Common to all approaches is the desire to structure and to systematise processes more even than has already been achieved in the corporations. The various components and programmes do not claim to offer turn-key solutions, as the corporations are too different; instead they highlight approaches that may help to solve problems. They offer suggestions which are useful in improving existing programmes and procedures and resolve questions regarding the strengthening of ageing management. What they make of the suggestions is up to the corporations themselves. Starting such an internal discussion a network will be automatically formed. This became clear during the work on the present project, which led to many useful discussions across divisions and company lines. The project was carried out in a particular industry, its results [cf. 16], however, are largely industry-independent. This has already been proven with non-automotive corporations. The project results have aroused great interest in industry generally and feedback has been very positive, even though to date no (authoritative) reports about its usefulness in real life have been received.

Acknowledgements From September 2011 to August 2015 PINA was sponsored by the Initiative New Quality of Work of the German Federal Ministry of Labour and Social Affairs and professionally supported by the Federal Institute for Occupational Safety and Health. The products that were developed as well as a brochure can be obtained from www.pina-projekt.de.

References

1. Antonovsky A (1979) Health, stress and coping: new perspectives on mental and physical well-being. Jossey-Bass, San Francisco
2. Badura B, Schellschmidt H, Vetter C (eds) (2003) Fehlzeiten-Report 2002. Demographischer Wandel: Herausforderung für die betriebliche Personal- und Gesundheitspolitik. Springer, Berlin
3. Baumann G, Saupe-Heide M, Hodek L, Niehaus M, Kugler M (2013) Weiterbildung und Qualifizierung von Mitarbeitern im "rehakritischen Alter" – Welcher Beitrag kann zum Erhalt der Beschäftigungsfähigkeit vor dem Hintergrund des demografischen Wandels geleistet werden? Eine Bestandsaufnahme in der Automobilindustrie. In: Stein R, Niehaus M, Stach M (eds) bwp@ Spezial 6 – Hochschultage Berufliche Bildung 2013, Fachtagung 05, 1–13. Online: http://www.bwpat.de/ht2013/ft05/baumann_etal_ft05-ht2013.pdf
4. Bögel J, Frerichs F (2011) Betriebliches Alters- und Alternsmanagement. Handlungsfelder, Maßnahmen und Gestaltungsanforderungen. BoD
5. Brandenburg U, Domschke JP (2007) Die Zukunft sieht alt aus. Herausforderungen des demografischen Wandels für das Personalmanagement. Gabler, Wiesbaden
6. Bubb H (1995) Qualitätsmanagement, Ergonomie und Zuverlässigkeit. Zeitschrift für Arbeitswissenschaft 49(3):157–164
7. Buck H (2002) Alternsgerechte und gesundheitsförderliche Arbeitsgestaltung – ausgewählte Handlungsempfehlungen. In: Morschhäuser M (ed) Gesund bis zur Rente – Konzepte gesundheits- und alternsgerechter Arbeits- und Personalpolitik. Broschürenreihe "Demographie und Erwerbsarbeit". Fraunhofer IAO – Institut für Arbeitswirtschaft und Organisation, Stuttgart
8. Busch R (ed) (2004) Alternsmanagement im Betrieb. Ältere Arbeitnehmer – zwischen Frühverrentung und Verlängerung der Lebensarbeitszeit. Hampp, München
9. Buss K-P, Kuhlmann M (2013) Akteure und Akteurskonstellationen alter(n)sgerechter Arbeitspolitik. WSI-Mitteilungen 5 (Schwerpunktheft "Altern in der Arbeitsgesellschaft"), pp 350–359

10. Fischer JE (2009) Gesundheitsmonitoring für Unternehmen. In: Kromm W, Frank G (eds) Unternehmensressource Gesundheit. Weshalb die Folgen schlechter Führung kein Arzt heilen kann. Symposium, Düsseldorf, pp 215–240
11. Franke F, Felve J (2011) Diagnose gesundheitsförderlicher Führung - Das Instrument "Health-oriented Leadership". In: Badura B, Ducki A, Schröder H, Klose J, Macco K (eds) Fehlzeiten-Report 2011. Führung und Gesundheit. Springer, Berlin, pp 3–13
12. Hasselborn HM, Ebener M (2014) Gesundheit, Arbeitsfähigkeit und Motivation bei älter werdenden Belegschaften. In: Badura B, Ducki A, Schröder H, Klose J, Meyer M (eds) Fehlzeiten-Report 2014. Erfolgreiche Unternehmen von morgen – gesunde Zukunft heute gestalten. Springer, Berlin, pp 75–84
13. Hodek L, Niehaus M (2013) Vernetzung von Arbeitswelt und Sozialversicherung: Zukunft der Arbeit für ältere gesundheitlich eingeschränkte und behinderte Arbeitnehmer. In: Stein R, Niehaus M, Stach M (eds) bwp@ Spezial 6 – Hochschultage Berufliche Bildung 2013, Fachtagung 05, 1–16. Online: http://www.bwpat.de/ht2013/ft05/hodek_niehaus_ft05-ht2013.pdf
14. Horváth P, Gamm N, Möller K, Kastner M, Schmidt B, Iserloh B, Kliesch G, Otte R, Braun M, Matter M, Pennig S, Vogt J, Köper B (2009) Betriebliches Gesundheitsmanagement mit Hilfe der Balanced Scorecard. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund, Berlin, Dresden
15. Ilmarinen J, Tuomi K (2004) Past, present and future of work ability. In: Ilmarinen J, Lehtinen S (eds) Past, present and future of work ability, vol 65, People and Work, Research Reports. Finnish Institute of Occupational Health, Helsinki, pp 1–25
16. Kugler M, Baumann G, Bruder R, Hodek L, Niehaus M, Sinn-Behrendt A (2015) Bausteine für ein verbessertes Alternsmanagement. Ergebnisse des Projekts "Gesund und qualifiziert älter werden in der Automobilindustrie - Partizipation und Inklusion von Anfang an (PINA)". Technische Universität Darmstadt & Universität zu Köln
17. Langhoff T (2009) Den demographischen Wandel im Unternehmen erfolgreich gestalten. Eine Zwischenbilanz aus arbeitswissenschaftlicher Sicht. Springer, Heidelberg
18. Rump J, Eilers S (2014) Personalrekrutierung und -entwicklung der Zukunft. In: Badura B, Ducki A, Schröder H, Klose J, Meyer M (eds) Fehlzeiten-Report 2014. Erfolgreiche Unternehmen von morgen – gesunde Zukunft heute gestalten. Springer, Berlin, pp 195–200
19. Schott T (2011) Rehabilitation – was ist und was nötig ist. Public Health Forum 19(4):2.e1–2.e4
20. Sinn-Behrendt A, Saupe-Heide M, Kugler M, Niehaus M, Bruder R (2013) Betriebliche Gesundheitsförderung (BGF) als Teil eines Alter(n)smanagements am Beispiel der deutschen Automobilindustrie. In: 53. Wissenschaftliche Jahrestagung von DGAUM, ÖGA, SGARM und ASU, 13.-16. März 2013, Bregenz, pp 486–489
21. Türk M (2014) Kenngrößen eines Alternsmanagements und ihre Nutzungsmöglichkeiten für die betriebliche Bewertung von Maßnahmen. Studienarbeit. Technische Universität Darmstadt, Institut für Arbeitswissenschaft
22. Uhle T, Treier M (2013) Betriebliches Gesundheitsmanagement. Gesundheitsförderung in der Arbeitswelt – Mitarbeiter einbinden, Prozesse gestalten, Erfolge messen (2 revised ed). Springer, Heidelberg

Consequences of Line Balancing of an Assembly Line on the Health, Satisfaction and Performance Indicators of Assembly Line Workers in a Case Study of an Automotive Manufacturer

Johanna Büttner, Barbara Deml, and Sebastian Neveling

Abstract

In assembly-line work the term cycle time refers to the period of time between the completion of one product and the completion of the next one. Against the context of work design this parameter is important as it seems to be related both to the physical and the mental well-being of workers as well as to their job satisfaction and their performance. In order to gain a better understanding of the underlying causal relationships a structural equation modeling approach was taken in this work. By this means it is possible to hypothesize on the interaction of variables and to test these hypotheses empirically. Within a field experimental study at an automobile manufacturing company it was shown that shorter cycle times lower the physical well-being of workers—but, however, this fact did not affect the overall performance outcome. Besides, it could have been shown that cycle-related working conditions in general—regardless of the length of cycle times—reduced the mental well-being of assembly workers. This result is important as the mental well-being did not only reinforce the physical conditions, but it turned out to be also relevant for the performance of the workers.

J. Büttner (✉)
Volkswagen AG, Wolfsburg, Germany
e-mail: johanna.buettner@volkswagen.de

B. Deml
ifab-Institute of Human and Industrial Engineering, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
e-mail: barbara.deml@kit.edu

S. Neveling
Volkswagen AG, Wolfsburg, Germany
e-mail: sebastian.neveling@volkswagen.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_29

Keywords
Structural equation model • Cycle time • Line balancing • Physical/psychological well-being • Performance

1 Introduction

The Toyota Production System (TPS) and its management principles of lean production have received a lot of attention both in the scientific community and in industry [39]. One of its core ideas is to avoid any kind of "waste" (muda). In the context of TPS "waste" is everything, which involves costs and which does not add any value to a product. The worst form of waste is over-production, which occurs when more products are produced than are required by the customer right now. To describe this problem in more detail, the following three terms takt time t_{tak}, cycle time t_{cyc}, and line balancing shall be regarded [22, 23]:

- The takt time t_{tak} considers the needs of the customer against the background of the production capacity. More formally speaking, it is the quotient of the production time, which is available per day, and the production units, which are needed per day. Let us assume that the customer requests 450 products and that a working shift accounts for 8 h (28.800 s). Thus, to fulfill the customer wish within a working shift, the takt time accounts for $t_{tak} = 64$ s for each product. In doing so, the customer is put at the heart of production in order to avoid any over-production.
- The cycle time t_{cyc} is the period between the completion of one product and the completion of the next one. This means for our example, if t_{tak} accounts for 64 s all required working steps—the assembly of all parts of the product—will have to be done in maximally 64 s. As t_{tak} has to be always less than or equal to t_{cyc}, there may be differences between these two times.
- Finally, the task of line balancing is to make sure, that the work content is distributed evenly on all workstations, so that the sum of all deviations is as low as possible. This in turn means low waste and high productivity [3].

On a first glance, it may be obvious that longer cycle times are also more desirable for the employees. This seems to be particularly true with the prospect of an aging workforce. From the perspective of production planners, however, it is often argued that the length of cycle time is also correlated to the amount of assembly errors. Although there is some empirical doubt [9], this argument in terms of performance is often used to implement short cycle times. Besides it is argued that shorter cycle times also allow for a faster training of new workforce.

Within this work the cycle time and particularly its impact on the physical and mental health, on job satisfaction, and the performance of employees shall be studied in more detail. To the best of our knowledge, there is no comprehensive study available, in which these effects of line balancing are examined in a systematic manner. Within this work a field experiment was carried out, in which three different types of line balancing are studied on two assembly-lines within the "Cockpit Assembly/Business Unit Plastics" at Volkswagen plant in Wolfsburg. As originally proposed by TPS, one of the assembly-lines is balanced so that the standardized cycle time accounted for 60 s and thus, one unit is made per minute; this condition shall be referred to as "one unit cycle per workstation (1-unit-cycle)". With all other major conditions being identical, the second assembly-line is balanced so that the standardized cycle time accounted for 300 s and 480 s, respectively, and thus one unit is made in either 5 or 8 min; these two conditions shall be called "five unit cycle per workstation (5-unit-cycle)" and "eight unit cycle per workstation (8-unit-cycle)", respectively. In this context should be mentioned that the work processes performed by a single employee are less fragmented due to the longer cycle time. As this study was carried out in a plant under real production conditions the workers could not be assigned randomly on the three experimental conditions and thus, strictly speaking a pseudo experiment was carried out. Besides, it could not be avoided that the 1/8-unit-cycle team consisted of eight workers, whereas the 5-unit-cycle team was made up by ten team members. With regard to the 5-unit-cycle existed a systematic difference, mainly the workers within this team worked in teams of two. Just the same, it has to be kept in mind that the 1-unit-cycle team rotated between the workstations, while this was not the case for the 5/8-unit-cycle team. However, this difference may not be over-estimated as the workers could only rotate between repetitive tasks that put stress on the hand-arm-system and on the upper back. Other interfering variables were not observed.

The reminder of the article is organized as follows: In order to gain a better understanding of how cycle time influences various human outcome parameters a literature study is carried out first (see Sect. 2). On this basis a causal model for the interaction of the various parameters is derived. More formally speaking this is done by structural equation modeling. Both the methodology and the model are explained next (see Sect. 3). In order to test our assumed structural equation models an empirical study is carried out. Finally the results are presented (see Sect. 4) and discussed (see Sect. 5).

2 Empirical Findings

The working conditions within holistic production systems have been studied by several researchers. Based on these empirical findings the following relationship between line balancing and human factors variables may be assumed (see Fig. 1):

a) There is a lot of empirical evidence for a direct relationship between *line balancing*, especially short cycle times, and *physical reactions* (e.g. musculoskeletal diseases, tendonitis) due to the resulting repetitive strain [e.g. 11, 12, 26, 34]. In particular, it could have been shown that the workload increases with the pace of work [20, 21]. A more recent study by Frieling et al. [10] concludes that cycle time reductions due to the introduction of a TPS approach do not offer any possibilities for short breaks in combination with a workload rate up to 97 %. In line with these findings, Kapellusch et al. [18]

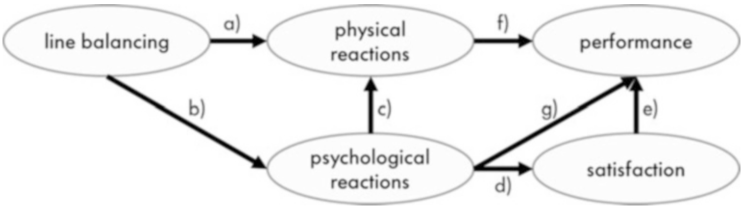


Fig. 1 Assumed relationship between cycle time and human factors variables

recommend to consider also risk factors, which facilitate musculoskeletal disorders, for deriving the optimal cycle time.

b) Melamed et al. [26] have also shown a direct relationship between *line balancing* and *psychological reactions* (e. g. stress, monotony, mental fatigue) of the workers.
c) Further on a study of Pütz [31] revealed that *psychological reactions* (e.g. stress) can in turn lead to *physical reactions* (e.g. back pain).
d) Kvarnström [19] found that *psychological reactions*, such as mental underload, while at the same time having physical overload due to short cycle times, leads to low *job satisfaction*.
e) The relationship between *job satisfaction* and *performance indicators* has been subject of many studies. The relationship between job satisfaction and performance indicators (especially productivity) is not clearly defined in the literature. On the basis of a literature review Judge et al. [17] conclude that the job satisfaction has an effect on the performance of workers. This relationship has been assumed in the present study.
f) Lin et al. [24] indicate that *physical reactions* due to physical load and time pressure are mainly responsible for a decrease of *performance indicators* (especially reduced product quality).
g) Udris and Frese [36] and Kvarnström [19] found that *psychological reactions* lead to a decrease of *performance indicators* (e.g. quality fluctuations, error rate, decrease of concentration)

Against the background of our work, it is to be assumed that within the 1-unit-cycle more negative aspects are to be observed than within the 5-unit-cycle or the 8-unit-cycle.

3 Methods

The methodology of this work is based on structural equation modeling (SEM). This is a statistical approach, with which a conceptual model, such as stated in Fig. 1, may be tested empirically [e.g. 6, 7]. As within this work three groups, 1-/5-/8-unit cycle, are regarded and a multi-group analysis is carried out. The above mentioned five constructs—line balancing, physical reactions, psychological reactions, satisfaction, and performance—cannot be measured directly. In order

Table 1 Structural equation model: constructs and corresponding indicators

Construct	Indicator	Derivation
	Exogenous indicators	
1) Line balancing	1) Time for regeneration	Adler et al. [1]
	2) Ergonomic deficits	Rimann and Udris [33]
	3) Perceived time pressure	Nübling et al. [29], Rimann and Udris [33], Weyer et al. [38]
	4) Working pace	Rimann and Udris [33]
	5) Perceived self-dependency	Nübling et al. [29], Rimann and Udris [33]
	6) Variety of work	Berggren [2], Rimann and Udris [33]
	7) Intensity of work	Rimann and Udris [33]
	8) Meaningfulness of work	May et al. [25], Rimann and Udris [33]
	9) Degree of repetitiveness	Rimann and Udris [33]
	Endogenous indicators	
2) Physical reactions	1) Musculoskeletal disorder	Tuomi et al. [35]
	2) Repetitive motion injury (RMI)	Own
	3) Absenteeism due to illness	Tuomi et al. [35]
	4) Productive efficiency	Tuomi et al. [35]
3) Psychological reactions	5) Mental saturation	Richter [32], Weyer et al. [38]
	6) Mental fatigue	Neuhaus [27], Richter [32]
	7) Stress	Neuhaus [27], Nübling et al. [29], Richter [32], Weyer et al. [38]
	8) Monotony	Neuhaus [27], Richter [32]
	9) Work ability	Tuomi et al. [35]
4) Satisfaction	10) Job satisfaction	Fischer and Lück [8], Nübling et al. [29]
	11) Commitment	Own
	12) Tendency for aversion	Burke [4], Herzberg et al. [13]
5) Performance	13) Selected performance indicators	Nübling et al. [29], Richter [32]
	14) Pulling andon cord	Own
	15) Error rate	Own

[Tabellenfußzeile - bitte überschreiben]

to make these latent variables measurable, observable variables, so called indicators, have to be derived first (see Table 1). In the language of SEM the indicators for the independent latent variable (line balancing) are also referred to as exogenous, whereas the dependent latent variables are called endogenous. Table 1 summarizes the constructs and corresponding indicators as well as their origin (for further details on the derivation of indicators and items see Büttner [5].

All of the 9 exogenous and 15 endogenous indicators are composed of various items, which at the end could be combined into one single value by either calculating mean or sum values [6, 7]. The items were obtained by presenting a questionnaire to the workers at the regarded assembly-lines (see Sect. 1). The survey consisted of 107 questions, which partially were taken from standardized questionnaires (e.g. SALSA – Rimann and Udris [33]; COPSOQ – Nübling et al. [29]; WAI – Tuomi et al. [35]). All questions were presented on a five-point Likert scale (for further details on the questionnaire see Büttner [5]).

Participation in the survey was voluntary. The questionnaire was answered by 68 assembly workers (M = 36.9, SD = 15.5 years) in the 1-unit-cycle team, 46 assembly workers (M = 33.5, SD = 17.2 years) in the 5-unit cycle team, and 31 workers (M = 33.7, SD = 16.9 years) in the 8-unit-cycle team. The response rate accounted for 61 %. The inquiry of gender and affiliation to the assembly line had to be removed from the questionnaire after consultation with the data protection commission, because the group size of less than five people could allow for identification of individuals.

4 Results

For data processing Cronbach's Alpha, Inter-Item-Correlations, and corrected Item-Scale-Correlations were derived. Nunally [30] recommends only the usage of items and sets of indicators with an Alpha value greater than 0.7. Values for Cronbach's Alpha smaller than 0.5 cannot be employed for further analysis. For the corrected Item-Scale-Correlation a cut off value of 0.5 and for the Inter-Item-Correlation coefficient a cut off value of 0.3 are mentioned [37]. In consequence, some indicators (degree of repetitiveness, meaningfulness of work, perceived self-dependency, absenteeism due to illness, mental saturation) as well as single items (questions with regard to productive efficiency and musculoskeletal disorder) had to be removed for the further analysis. Just the same, the model construct satisfaction had to be excluded from the subsequent evaluation.

All calculations, except for the two indicators error rate and pulling andon cord, are based on item parcels instead of individual items. Doing so enables an increased approximation to the normal distribution assumption and a reduction in model complexity [14].

Due to unacceptable values for Cronbach's Alpha the indicators repetitiveness (Cronbach's Alpha = .315) and meaningfulness of work (Cronbach's Alpha = .391) had to be removed for the further analysis. Both indicators showed also unacceptable values with regard to Inter-Item-Correlation coefficient (<.30) and corrected Item-Scale-Correlation coefficient (<.50).

Regarding the indicator perceived self-dependency the analysis showed that also after removing single items the value for Cronbach's Alpha is still below the cut-off value of 0.7 and the associated Inter-Item-Correlation coefficient has an unacceptable value (<.30). For this reason, the indicator perceived self-dependency was removed for the further analysis.

The analysis of the model construct satisfaction revealed Alpha values for the indicators commitment and aversion tendencies, which were below 0.7. Additionally, the items of these two indicators showed mostly inadequate corrected Item-Scale-Correlation coefficients (<.50). For this reason both indicators were excluded from further analysis. The claim by Weiber and Mühlhaus [37] that with regard to reflective models in a multi-construct-model constructs should be measured with at least two indicators was not applicable in this case. Moreover, the construct satisfaction was removed from the model. Similarly, the measure of the construct satisfaction is controversially discussed and the relationship between satisfaction and the model construct performance (especially productivity) is not clearly defined in the literature.

With regard to the construct physical reactions the indicator absenteeism due to illness had to be removed for further analysis (Cronbach's Alpha = .502; corrected Item-Scale-Correlation <.50). The value for Cronbach's Alpha improved (.729) for the indicator productive efficiency by removing one item. Thus, the indicator was changed and remained in the model for further analysis. Similarly, the value for Cronbach's Alpha for the indicator musculoskeletal disorder improved (.758) by removing one item.

For the construct psychological reactions the analysis showed an unacceptable Alpha value for the indicator mental saturation with 0.537. The corrected Item-Scale-Correlation coefficient has a below threshold value (<0.5) and the Inter-Item-Correlation coefficients of the indicator is below the claimed value of 0.3 in three of five items. Thus, the indicator mental saturation was removed from further analysis. For further details see Büttner [5].

Next for structural equation modeling IBM SPSS Amos was used and for each sample group the parameters of the model were estimated. In view of the indices of model fit further indicators were removed; the final model is shown in Fig. 2.

For small sample sizes ($n \leq 250$) Hu and Bentler [15] recommend to analyze the following fit indices: Chi-Square, Chi-Square/df, Standardized Root Mean Square Residual (SRMR), Incremental Fit Index (IFI), and Comparative Fit Index (CFI). When comparing the recommended values with the outcome of our study, the model fit can be judged as (very) good; only for IFI the here found results are slightly smaller than recommended. By this index the model of interest is compared with some alternative, such as the null or the independence model, whereby values close to 1 indicate a good model fit. As, however, the model may be accepted, if IFI is equal or greater to .90, we may proceed with our analysis. In conclusion, we can assume that the empirical variance-covariance-matrix is predicted correctly by the model, which can be interpreted by the absolute fit indices in Table 2 and the here suggested model reflects reality well.

In order to quantify the strength of the observed relationships the model path coefficients between the constructs are derived for the 1-/5-/8-unit cycle within a multi-group analysis (see Table 3). Thereby a higher value indicates a more negative impact of line balancing on the human outcome variables, for instance, the 1-unit-cycle (.539) causes more negative physical reactions than the 5-unit-cycle (.218) or the 8-unit-cycle (.197). Thus, as might be expected the 1-unit-cycle

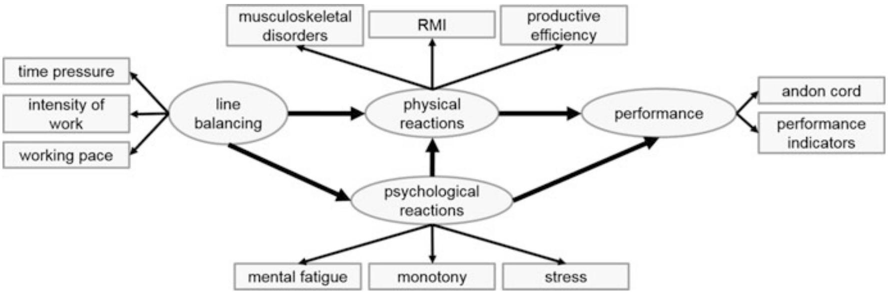


Fig. 2 Structural Equation Model for line balancing and human outcome variables. Latent variables are shown as ovals and observed variables are displayed as rectangles

Table 2 Fit indices for the structural equation model

Fit index	Chi-square	Chi-square/df	SRMR	IFI	CFI
Recommended value[a]	Small[b]	≤3	≤.10	≥.95	≥.90
Results	84.037	2.155	.066	.927	.925

[a]According to Hu and Bentler [15]
[b]As the value of Chi-square increases, the fit of an model becomes increasingly worse

Table 3 Summarized regression weights

	1-unit-cycle	5-unit-cycle	8-unit-cycle
a) Line balancing → physical reactions	.534	.218	.197
b) Line balancing → psychological reactions	.799	.715	.748
c) Psychological reactions → physical reactions	.255	.644	.167
d) Physical reactions → performance	.101	−.197	.161
e) Psychological reactions → performance	.898	.999	.828
Sum	2.587	2.379	2.101

The value range for a) – e) is between +1 and −1

is the most physically demanding configuration for the assembly workers here; this result is also confirmed by the overall sum of the path coefficients. In terms of physical health the assembly-line balancing seems to be crucial. For the psychological reactions, however, simply the fact of cycle-related work turned out to be essential, while it does not seem to be important how the line is balanced. This observation must not be underestimated as the relationship is much stronger than that for the physical reactions. Besides, it could have been shown that in contrast to physical reactions only psychological reactions affected the performance rate of assembly workers. Finally, as already suggested by Pütz [31], our results also support the finding that psychological symptoms may intensify physical ones.

5 Discussion

Within this work a structural equation model is presented that allows for a better understanding of how cycle time influences the performance rate of assembly workers. Due to complementary results within a multi-group analysis (1-/5-/8-unit-cycle teams), it can be assumed that the approach is valid. For sure, the generalization of results is always a sensitive issue and thus, it has to be pointed out that our observations refer to repetitive tasks that are stressful for the hand-arm-system and the upper back.

In our study it could have been shown that line balancing influences the physical well-being of workers and that this was best in our study, when the cycle time accounts for 8 min. Unfortunately, a rotation between workstations was only given in our 1-unit-cylce team. Thus, it is to be assumed that this fact had been even stronger if there would have been also a rotational organization between the 5/8-unit-cycle team. In addition to our work a further empirical study was carried out, which also examined the cycle length of the here regarded assembly-lines. By applying REBA 9.0—a computer based method for the evaluation and design of task contents—Iffländer et al. [16] recommended a minimum cycle length of 7 min. Although this is in line with our results, it has to be mentioned, that within our study the physical well-being did not influence the performance of workers.

The results for the 5-unit-cycle are not in line with the expected values. On the one hand it can be assumed that the work within a team of two workers had a positive effect on the results. On the other hand the work tasks within this team can be defined as more physically exhausting in than in the other two teams. For this reason, it can be concluded that the comparability between the 5-unit-cycle team and the other two teams is not given.

However, irrespective of the length of the cycle time, the psychological well-being of our participants was influenced by cycle-related working conditions in general. This outcome is rather important as the psychological well-being affected the overall performance and it had a reinforcing effect on the physical well-being of workers.

Furthermore the rotation concept within a team should also be assessed critically. The rotation within the various workstations is one of the elements of the holistic production system. Notbohm [28] defines job rotation as systematic and planned changes in the workstations, while the tasks within the individual work stations stay the same. If a task includes unilateral strain it should only be repeated for a certain time followed by a task in a different work station with physical strain of another part of the body. With regard to the cockpit assembly all workstations have similar work activities and can thus represent limited relief for the assembly worker. The workers have only the possibility to rotate between workstations which are all classified as cycle-related, repetitive and the hand-arm areas and the upper back are mainly burdened. It is advised to optimize the rotation as well as the team design. In the surveyed case it could be beneficial to rotate team members between assembly lines and material provision team. The greatest effect could be achieved by rotating with workstations without a fixed cycle time in order to minimize the amount of cycle-related work.

The results, opinions or conclusions of this dissertation are not necessarily those of the Volkswagen AG.

References

1. Adler P, Goldoftas B, Levine D (1997) Ergonomics, employee involvement, and the Toyota production system: a case study of NUMMI's 1993 Model Introduction. Ind Labor Relat Rev 50(3):416–437
2. Berggren C (1993) Alternatives to lean production: work organization in the Swedish auto industry, No. 22. Cornell University Press
3. Boysen N, Fliedner M, Scholl A (2008) Assembly line balancing: which model to use when? Int J Prod Econ 111:509–528
4. Burke RJ (1969–1970) Occupational and life strains, satisfaction and mental health. J Bus Admin 1:35–41
5. Büttner J (2015) Auswirkungen der Austaktung einer Automobil-Montagelinie auf menschliche Leistungsfaktoren, Otto-von-Guericke-Universität Magdeburg, Dissertationsschrift
6. Byrne BM (2001) Structural equation modeling with AMOS: basic concepts, applications, and programming. Erlbaum, Mahwah, NJ
7. Byrne B (2010) Structural equation modeling using AMOS, Basic concepts, applications, and programming, 2nd edn. Routledge, New York, NY
8. Fischer L, Lück HE (1972) Entwicklung einer Skala zur Messung von Arbeitszufriedenheit (SAZ). In: Psychologie und Praxis. H. 16, pp 64–76
9. Frieling E (Hrsg) (1997) Automobilmontage in Europa. Frankfurt/Main Campus
10. Frieling E, Kotzab D, Enriquez-Diaz A, Sytch A (2012) Mit der Taktzeit am Ende – die älteren Beschäftigten in der Automobilmontage, Ergonomia, vol 1. Auflage, Stuttgart
11. Fucini J, Fucini S (1990) Working for the Japanese: inside Mazda's American auto plant. Free Press, New York, NY
12. Garrahan P, Stewart P (1992) The Nissan Enigma-flexibility at work in a local economy. Mansell, London
13. Herzberg F, Mausner B, Snyderman B (1959) The motivation to work. Wiley, New York, NY
14. Hoyle RH (1999) Statistical strategies for small sample research. Sage, Thousand Oaks, CA
15. Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 6(1):1–55
16. Iffländer J, Schmicker S, Wassmann S (2013) Projektbericht, Erfassung psychischer Belastungen in der Cockpitmontage und Ableitung entsprechender Arbeitsgestaltungsempfehlungen – Projektnummer AZ-76988089, in Auftrag von der Volkswagen AG, durchgeführt von der Otto-von-Guericke-Universität Magdeburg – Fakultät für Maschinenbau – Institut für Arbeitswissenschaft, Fabrikautomatisierung und Fabrikbetrieb – Lehrstuhl für Arbeitswissenschaft und Arbeitsgestaltung in Kooperation mit der METOP GmbH – An-Institut der Otto-von-Guericke Universität Magdeburg
17. Judge T, Thoresen C, Bono J, Patton G (2001) The job satisfaction-job performance relationship: a qualitative and quantitative review. Psychol Bull 127(3):376–407
18. Kapellusch JM, Kapellusch JM, Garg A (2012) Conference proceeding: ergonomic cycle time: a proposed metric for the design of safe, productive manufacturing jobs. 10th International Conference on Manufacturing Research, Birmingham, England, 09/2012
19. Kvarnström S (1996) Stress prevention for blue-collar workers in assembly-line production, Conditions of work and welfare facilities branch. Working paper, Genf
20. Landsbergis PA, Cahill J, Schnall P (1999) The impact of lean production and related new systems of work organization on worker health. J Occup Health Psychol 4(2):108–130

21. Lewchuk W, Robertson D (1996) Working conditions under lean production: a worker based benchmarking study. Asia Pacific Bus Rev 2(4):60–81
22. Liker JK, Meier DP (2007) Praxisbuch der Toyota Weg: Für jedes Unternehmen. FinanzBuch Verlag, München
23. Liker JK (2007) Der Toyota Weg, 14 Managementprinzipien des weltweit erfolgreichsten Automobilkonzerns. FinanzBuch Verlag, München
24. Lin L, Drury C, Kim S-W (2001) Ergonomics and quality in paced assembly lines. Hum Factors Ergon Manuf 11(4):377–382
25. May DR, Gilson RL, Harter LM (2004) The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. J Occup Organ Psychol 77:11–37
26. Melamed S, Ben-Avi I, Luz J, Green M (1995) Objective and subjective work monotony: effects on job satisfaction, psychological distress, and absenteeism in blue-collar workers. J Appl Psychol 80(1):29–42
27. Neuhaus R (2009) KPB – Kurzverfahren Psychische Belastung. Ein Verfahren zur Beurteilung psychischer Belastungen, 2. überarbeitete Auflage, Wirtschaftsverlag Bachem, Köln
28. Notbohm G (1984) Mehrfachbelastungen und Gesundheitsrisiken am Arbeitsplatz. In: Gros E (ed) Anwendungsbezogene Arbeits-, Betriebs- und Organisationspsychologie. Verlag Angewandte Psychologie, Göttingen
29. Nübling M, Stößel U, Hasselhorn H-M, Michaelis M, Hofmann F (2005) Methoden zur Erfassung psychischer Belastungen – Erprobung eines Messinstrumentes (COPSOQ), Bremerhaven, Wirtschaftsverlag NW (Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Fb 1058)
30. Nunally JC (1978) Psychometric theory, 2nd edn. New York, NY
31. Pütz D (2010) Psychosoziale Einflussfaktoren bei Beschwerden am Bewegungsapparat. In: Gesellschaft für Arbeitswissenschaft e.V., Mensch- und Prozessorientierte Arbeitsgestaltung im Fahrzeugbau, Herbstkonferenz 2010 der Gesellschaft für Arbeitswissenschaft. Dortmund, GfA Press
32. Richter G (2000) Psychische Belastung und Beanspruchung – Stress, psychische Ermüdung, Monotonie, psychische Sättigung. In: BAuA (Hrsg), Arbeitswissenschaftliche Erkenntnisse: Forschungsergebnisse für die Praxis (Nr. 116), Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund, pp 1–13
33. Rimann M, Udris I (1997) Subjektive Arbeitsanalyse: Der Fragebogen SALSA. In: Strohm O, Ulich E (Hrsg), Unternehmen arbeitspsychologisch bewerten. Ein Mehr-Ebenen-Ansatz unter besonderer Berücksichtigung von Mensch, Technik und Organisation. Zürich, vdf Hochschulverlag, pp 281–298
34. Stewart P, Garrahan P (1995) Employee responses to new management techniques in the auto industry. Work Employ Soc 9(3):517–536
35. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A (2001) Arbeitsbewältigungsindex – Work Ability Index, 1. Auflage, Wirtschaftsverlag NW Verlag für neue Wissenschaft GmbH, Bremerhaven, Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin: Übersetzung, Ü 14
36. Udris I, Frese M (1999) Belastung und Beanspruchung. In: Hoyos Graf C, Frese D (Hrsg) Arbeits- und Organisationspsychologie – GL Lehrbuch. Psychologie Verlagsunion, Weinheim, pp 429–445
37. Weiber R, Mühlhaus D (2010) Strukturgleichungsmodellierung, Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS. SmartPLS und SPSS, Heidelberg
38. Weyer G, Hodapp V, Neuhäuser S (1980) Weiterentwicklung von Fragebogenskalen zur Erfassung der subjektiven Belastung und Unzufriedenheit im beruflichen Bereich (SBUS-B). Psychologische Beiträge 22:335–355
39. Womack JP, Jones DT, Roos D (1991) The machine that changed the world. Harper-Perennial, New York, NY

Sustainable Effects of Simulator-Based Training on Eco-Driving

Cornelia Lüderitz, Maria Wirzberger, and Katja Karrer-Gauss

Abstract

Simulation-based driver training offers a promising way to teach ecological driving behavior under controlled, comparable conditions. In a study with 23 professional drivers, we tested the effectiveness of such training. The driving behavior of a training group in a simulated drive with and without instructions were compared. Ten weeks later, a repetition drive tested the long-term effect training. Driving data revealed reduced fuel consumption by ecological driving in both the guided and repetition drives. Driving time decreased significantly in the training and did not differ from driving time after 10 weeks. Results did not achieve significance for transfer to test drives in real traffic situations. This may be due to the small sample size and biased data as a result of unusual driving behavior. Finally, recent and promising approaches to support drivers in maintaining eco-driving styles beyond training situations are outlined.

Keywords

Simulator-based training • Eco-driving • Sustainability • Professional drivers

C. Lüderitz (✉) • K. Karrer-Gauss
Institute of Psychology and Ergonomics, Chair of Human-Machine Systems, Technische Universität Berlin, Berlin, Germany
e-mail: cornelia.luederitz@mms.tu-berlin.de; karrer@mms.tu-berlin.de

M. Wirzberger
Institute for Media Research, E-Learning and New Media, Technische Universität Chemnitz, Chemnitz, Germany
e-mail: maria.wirzberger@phil.tu-chemnitz.de

B. Deml et al. (eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, DOI 10.1007/978-3-662-48661-0_30

1 Introduction

Within the ongoing debate on environmentally friendly and resource-saving driving the issue of eco-driving training is becoming increasingly important, especially for large road transport companies, including from an economical point of view. A cost analysis by Wittenbrink [13] showed an increase in the proportion of fuel costs during long-distance transport. For example, in 2007 fuel already accounted for more than 25 % of the total costs. Fuel consumption can be reduced in various ways, including improving vehicle technology, optimizing traffic infrastructure, but also through driver training [12]. In particular, teaching drivers a fuel-conserving driving style has been examined in several studies [cf. 10, 11]. Some studies use purely theory-based training while others apply real test drives to demonstrate an green driving style. Eco-driving can be described as a more efficient driving performance involving the application of eco-driving skills and techniques not only to reduce fuel consumption but also greenhouse-gas emissions [2]. These techniques include both theoretical and practical components. Especially practical components can be easily taught by using simulators as an additional environment for exercising technical skills [8]. Usually, within such training, skills like correct acceleration and gear changing behaviour to reduce engine speed, optimum braking characteristics, and the benefit of coasting phases are communicated [2]. Simulators for training eco-driving represent an alternative medium that offers many advantages. Specific training situations can be designed to effectively teach the required behaviour and can be repeated under controlled conditions as often as necessary. During simulator training, exact behavioural parameters of each driver in terms of timing and brake or accelerator pedal actuations can be detected online, providing instantaneous feedback.

In the present study, a simulator-based training concept is tested for its effectiveness in eco-driving training to professional truck drivers. Training evaluation focuses on three main issues:

1. Potential of the training, which refers to the degree to which a driver is potentially able to improve performance. It will be investigated to what extent driving behavior might be changed towards an eco-driving style by training.
2. Sustainability of the training, focusing on long-term effects. It will be examined whether the training leads to a measurable and successful application of eco-driving behavior in the longer term, in this case after a time span of 10 weeks.
3. Transfer of the training content to unfamiliar driving situations. On this account, two questions are considered: Will the trained behavior be used by drivers on unfamiliar routes within the simulator? Does the training have effects on real life driving behavior in the daily work of truck drivers?

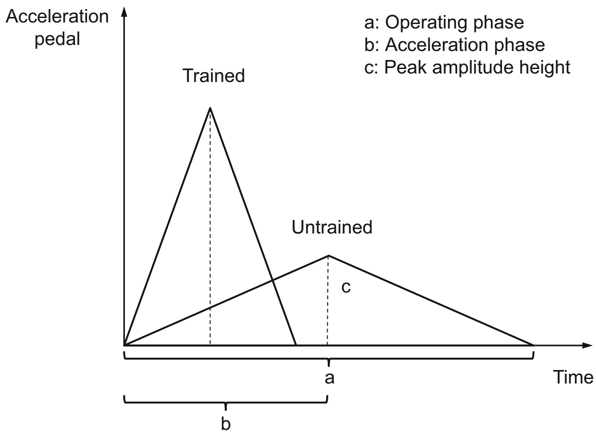
1.1 Hypothesis and Variables

To examine effectiveness of the training in terms of a green and cost-efficient driving style, various variables were assessed. According to previous studies, in which eco-driving trainings were evaluated on the basis of fuel consumption, the extent of fuel savings ranged from 7–27 % [e.g. 10, 11]. Hence, fuel consumption was taken as a key indicator for eco-driving performance, representing the overall potential and effectiveness of an eco-driving training concept. In addition, the time needed to complete the test drive was logged, since Thijssen et al. [12] report a significant increase in travel time after a training intervention for eco-driving, which would be a great disadvantage for truck companies, resulting in longer delivery times and thus additional costs. For this reason, the study also examined whether such a training intervention would increase travel time.

Acceleration performances of drivers were assessed by analyzing the profile of accelerator pedal usage, which is shown in Fig. 1 as triangles for trained and untrained drivers. There are three components of accelerator pedal activation indicating changes in terms of eco-driving style. The operating phase (a) reflects the overall time span of accelerator pedal usage, including acceleration phase (b), which describes the time participants need to accelerate to full depression of the accelerator pedal. The peak amplitude (c) was measured by coding the depression of the accelerator pedal from zero (no usage of acceleration pedal) to 255 (the maximum depression of pedal flooring the accelerator). It was assumed that with realization of an eco-driving style through training, operating phase as well as acceleration phase would be shorter, reflecting a more effective use of the accelerator pedal. In this context, a quicker acceleration performance together with a higher degree of pedal depression was expected. Hence, the peak amplitude should also increase. In this context, the absolute number of peaks for each drive was recorded. This should reveal information about the frequency of accelerator pedal use due to the smoothness of a participant's driving style. Trained participants would use the accelerator pedal less frequently in contrast to untrained drivers, resulting in a decreased number of peaks in accelerator activation.

Associated with acceleration performance, average speed was assessed as well as examined whether eco-driving training would lead to lower average speeds. Furthermore, coasting phases were measured. They were represented as time spans in which the vehicle was rolling and drivers neither used the accelerator pedal nor the breaks. It was hypothesized that trained drivers would use coasting phases to a greater extent than untrained drivers.

Fig. 1 Triangle profile of accelerator pedal activation to define acceleration performance variables



2 Methods

2.1 Participants

23 male truck drivers employed at a transport company participated in the study. 13 drivers ($M = 43.54$ years, $SD = 7.32$) completed the simulator-based training. The other 10 drivers ($M = 47.80$ years, $SD = 10.09$) served as a control group.

2.2 Simulation

Training evaluation was conducted using a "DriveSim mobile" simulator from SiFaT-Road Safety GmbH displayed in Figs. 2, 3 and 4. The air-conditioned, soundproof driver cabin conformed to a Mercedes-Benz Actros and included a shored motion system for simulating motor movements and braking. The driving scenario was back-projected onto a screen with 180° field of vision. Audible instructions from a male trainer were given promptly during the guided run regarding certain training events. The separate operator cabin displayed all driver views, a view to regulate the route, and dynamic displays showing the use of accelerator and braking pedal. The simulation emulated a four-wheeled goods vehicle with long wheelbase, drawbar trailer and a load of 32 barrels. The simulated route included urban traffic with crossroads and roundabouts as well as highways.

Fig. 2 Mobile truck simulator “DriveSim mobile” (SiFaT Road Safety GmbH)

Fig. 3 Trainer cabin of the mobile truck simulator (SiFaT Road Safety GmbH)

2.3 Procedure

As shown in Fig. 5, the inital training session was followed up by a validation run after an interval of 10 weeks.

Training sessions consisted of familiarization trial run and then two test drives, baseline (1) and guided drive (2). The main purpose was to demonstrate the potential of environmentally-motivated driving in terms of reduced fuel consumption after providing a short theoretical introduction, the previous driving behavior (baseline drive 1) was directly compared to a more environmentally aware driving style (guided drive 2) over the same route.

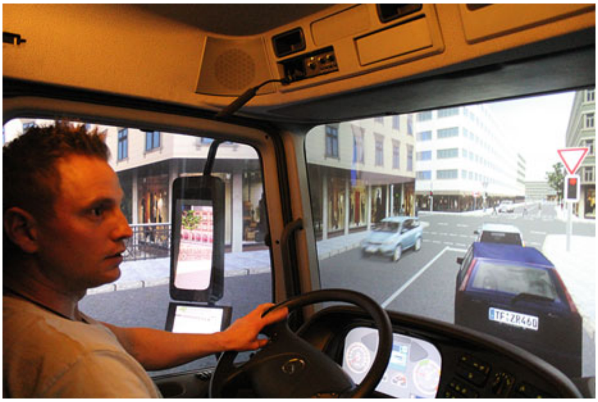


Fig. 4 Mercedes-Benz Actros driver cabin (SiFaT Road Safety GmbH)

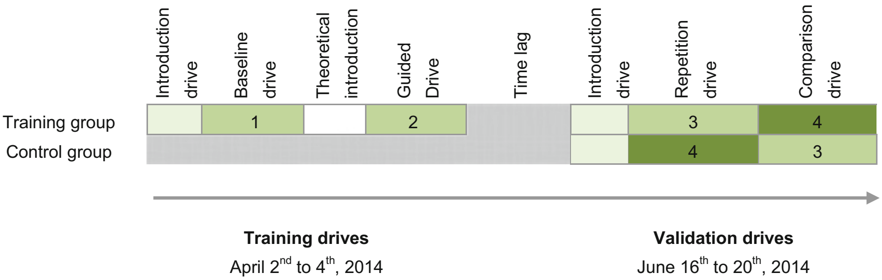


Fig. 5 Training and testing schedule

The follow-up validation drives aimed to examine long-term training effects. Trained drivers again completed the simulated route (repetition drive 3) to enable the comparison with their driving behavior from the baseline drive. At the same time, comparing repetition drive and guided drive, conducted within the training, provided information about how drivers exploit the fuel saving potential without instruction aid after a certain amount of time.

To avoid learning effects due to the repeated confrontation with the same route, another comparison between trained drivers and the control group of untrained drivers using a new route (comparison drive 4) was introduced. This made it possible to show the transfer of the acquired driving behavior to a new situation, containing elements similar to those from the previously completed route. In this way acquired knowledge was activated, but had to be applied to an unfamiliar route. Both routes were 8–10 km long.

To be able to assess the application of trained behavioral patterns in real-world traffic situations, fuel consumption from real driving data, was also inspected covering the time span from January to May 2014.

3 Results

3.1 Training Potential

Training potential was inspected in a within-subjects design, comparing fuel consumption within the baseline drive and guided drive by analyses of variance (ANOVAs). Travel time and fuel consumption were corrected with regard to a consistent starting point for each participant, i.e. the moment of first operation of the accelerator pedal. Means and standard deviations are displayed in Table 1. A significant effect for training potential was found, $F(1,12) = 23.404$, $p < .01$, with a lower fuel consumption in the guided drive. Travel time was examined in the same way, providing evidence for a significant effect in terms of shorter travel times in the guided drive compared to the baseline drive, $F(1,12) = 5.319$, $p < .05$. Considering the coasting phase revealed a significant effect for an increase in the amount of time within the guided drive, $F(1,12) = 67.122$, $p < .01$. Moreover, a significant effect for average speed arose in terms of higher average velocity in guided drive compared to baseline drive, $F(1,2) = 6.249$, $p < .05$.

Means and standard deviations of the acceleration performance variables are displayed in Table 2. ANOVAs revealed significant effects between baseline and guided drive for the number of peaks, $F(1,12) = 75.045$, $p < .01$, peak amplitude, $F(1,12) = 34.251$, $p < .01$, operating phase, $F(1,12) = 73.556$, $p < .01$ as well as acceleration phase, $F(1,12) = 48.400$, $p < .01$. The number of peaks in guided drive decreased in comparison to baseline drive, whereas the height of peak amplitude increased in the extent the accelerator pedal was depressed. Both

Table 1 Means and standard deviations of eco-driving variables within the training group on identical simulation tracks

	Fuel consumption (l/100 km)		Travel time (s)		Coasting phase (s)		Average speed (km/h)	
Track	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Baseline drive	30.18	2.67	570.92	25.85	59.15	39.01	45.76	2.07
Guided drive	25.64	1.00	555.62	18.16	191.62	39.01	46.80	1.78
Repetition drive	27.99	2.24	567.50	25.89	105.80	36.66	46.85	2.41

Table 2 Means and standard deviations for acceleration performance variables within the training group on identical simulation tracks

	Number of peaks		Peak amplitude (1-255)		Operating phase (s)		Acceleration phase (s)	
Track	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Baseline drive	66.78	9.58	164.81	12.20	430.46	43.49	272.46	27.52
Guided drive	38.62	9.64	190.00	16.30	299.78	46.67	208.08	17.48
Repetition drive	51.00	11.54	168.55	15.93	375.30	36.38	257.40	28.83

variables indicate a steady driving style, which was trained during the guided drive. The examination of the operation and acceleration phase revealed shorter durations for both phases in guided drive compared to baseline drive indicating a quicker accelerating performance of participants in guided drive in contrast to baseline drive.

3.2 Long-term Training Effects

To test long-term training effects, variances were analyzed in a within-subjects design to compare the results for the training group, both between baseline drive and repetition drive and between guided drive and repetition drive. Three participants of the training group did not complete the repetition drive. Therefore only ten datasets could be analyzed.

For fuel consumption there was a significant difference in the comparison between baseline and repetition drive, $F(1,9) = 7.766$, $p < .05$. Even after ten weeks fuel consumption remained at a lower level for trained participants, which can be attributed to the long-term training effect. However, fuel consumption was significantly lower in the guided drive compared to the repetition drive, $F(1,9) = 9.063$, $p < .05$. Likewise, there were significant effects on the coasting phase for baseline vs. repetition drive, $F(1,9) = 25.099$, $p < .01$, and the comparison between guided vs. repetition drive, $F(1,9) = 21.961$, $p < .01$. In the repetition drive, trained participants allowed the vehicle to coast longer than in the baseline drive, but not as much as in the guided drive, in which the coasting phase was longest. No significant long-term training effects could be shown for travel time (baseline vs. repetition drive: $F(1,9) = 0.029$, $p = .869$; guided vs. repetition drive: $F(1,9) = 1.855$, $p = .206$) or for average speed (baseline vs. repetition drive: $F(1,9) = 1.589$, $p = .239$; guided vs. repetition drive: $F(1,9) = 0.001$, $p = .976$).

Analyzing the results of acceleration performance variables, there were significant outcomes for the number of peaks in both comparisons (baseline vs. repetition drive: $F(1,9) = 22.257$, $p < .01$ and guided vs. repetition drive: $F(1,9) = 7.381$, $p < .05$). In the repetition drive, trained participants reached a lower number of accelerator pedal use compared to baseline drive; in comparison to guided drive, however, the number again increased significantly. For height of peak amplitude no significant effect could be found for baseline and repetition drive, $F(1,9) = 0.233$, $p = .641$, whereas the comparison between guided vs. repetition drive became significant, $F(1,9) = 13.109$, $p < .01$. Mean values for peak amplitude indicated a significant decrease in the extent of accelerator pedal depression in the repetition drive. In other words, in the repetition drive participants did not depress the accelerator pedal to the same intensity as in guided drive as shown in Fig. 6. The analysis of the operating phase revealed significant effects for both baseline vs. repetition drive, $F(1,9) = 55.981$, $p < .01$, and guided vs. repetition drive, F

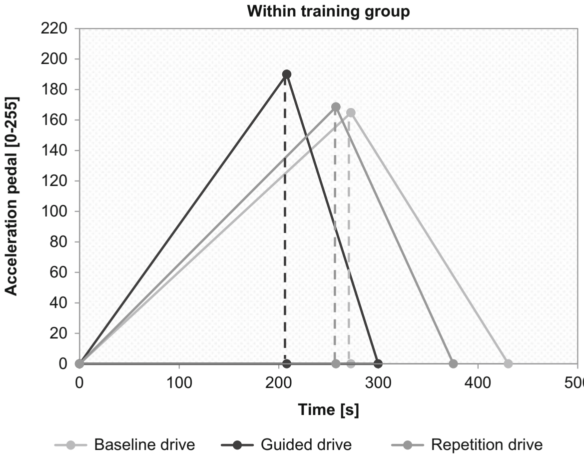


Fig. 6 Triangle diagram of acceleration performance variables displaying the potential and long-term effect of ecological training intervention

$(1,9) = 13.854$, $p < .01$. In the guided drive the duration of the operating phase in terms of acceleration activation was shorter in comparison to baseline drive. However, compared to the guided drive a significant increase was found in the repetition drive. The length of acceleration phase was only statistically significant for guided vs. repetition drive, $F(1,9) = 15.750$, $p < .01$, but not for baseline vs. repetition drive, $F(1,9) = 1.746$, $p = .219$. In the repetition drive, trained participants were not able to reduce the acceleration duration as in the guided drive. They roughly needed as much as time for acceleration as in baseline drive.

3.3 Transfer Effects

Transfer effects were assessed by comparing the results of the training group and the control group within comparison drive, displayed in Tables 3 and 4. Effects were tested by ANOVAs in a between-subjects design. The comparison of both groups yielded no significance for fuel consumption, $F(1,18) = 1.261$, $p = .276$, or travel time, $F(1,18) = 0.005$, $p = .943$. Although there was a tendency for longer coasting phase, shown by a higher mean value for training group, the difference was not significant, $F(1,18) = 1.727$, $p = .205$. Solely, average speed was significantly higher for training group compared to control group, $F(1,18) = 6.088$, $p < .05$.

For acceleration performance variables there were several significant differences in the number of peaks, $F(1,18) = 13.876$, $p < .01$, the peak amplitude, $F(1,18) = 16.967$, $p < .01$, and the operating phase as well, $F(1,18) = 15.229$, $p < .01$. The analysis of the acceleration phase did not reveal a significant effect for the comparison of control and training group, $F(1,18) = 0.019$, $p = .892$. Except for the acceleration phase, participants of the training group showed better acceleration performance compared to the control group in comparison drive, i.e. a shorter operation phase, less frequent use of accelerator pedal shown by fewer peaks as well as a higher degree of accelerator pedal depression, displayed in Fig. 7.

Table 3 Means and standard deviations for eco-driving variables in training and control groups

	Fuel consumption (l/100 km)		Travel time (s)		Coasting phase (s)		Average speed (km/h)	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Training group	32.95	1.77	524.00	28.83	70.10	26.10	40.62	2.37
Control group	31.71	3.01	523.00	64.13	57.60	14.96	37.44	3.31

Table 4 Means and standard deviations for acceleration performance variables in training and control groups

	Number of peaks		Peak amplitude (1-255)		Operating phase (s)		Acceleration phase (s)	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Training group	34.40	10.96	172.49	11.36	270.90	28.92	198.70	17.43
Control group	49.50	6.65	150.31	12.68	318.90	26.01	197.20	29.74

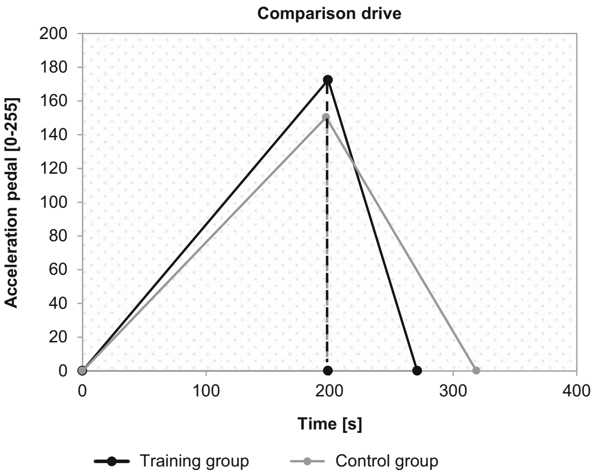


Fig. 7 Triangle diagram of acceleration performance variables displaying the transfer effect of ecological training intervention

Finally, driving behavior in real-world traffic situations was considered to provide evidence for transfer effects. For these purposes, fuel consumptions of the training group and control group were compared using available real driving data from January to May 2014, as shown in Table 5. Due to missing data, only nine datasets from each group could be analyzed. Data from both groups for March and April 2014 were tested by ANOVAs within a 2 × 2 mixed between-within-subjects design. There was a significant main effect for the factor time, $F(1,16) = 8.043$, $p < .05$, but not for group membership $F(1,16) = 0.000$, $p = 1.00$, i.e. fuel consumption declined after the training intervention, but not more than in the control group. A potential interaction effect did not achieve significance either, $F(1,16) = 0.371$, $p = .551$.

Table 5 Means and standard deviations of fuel consumption from real driving data

	Fuel consumption (l/100 km)									
	January		February		March		April		May	
Group	M	SD	M	SD	M	SD	M	SD	M	SD
Training group	32.7	1.19	31.62	2.44	31.99	1.43	29.79	3.89	31.44	1.92
Control group	32.5	2.15	30.83	2.57	31.60	2.51	30.18	2.12	31.24	2.11

4 Discussion

Even a single guided training session was able to demonstrate the potential and long-term effect of simulator-based training on driving behavior: Under instruction by a trainer, fuel consumption decreased by 16–21 %. This even exceeds other research findings which report up to 15 % fuel consumption decrease (as summarized in Camacho et al. [5]). In a repetition drive after a period of ten weeks fuel consumption was still reduced by about 8 %. Fuel savings were not obtained by longer driving times. On the contrary, driving time significantly decreased as well and remained below the level of the baseline drive in the repetition drive. Different to the results of Thijssen et al. [12], ecological driving training is not automatically prone to diminish the resulting benefit of reduced fuel consumption by means of increased travel times and in this vein associated with financial disadvantages. However, in their study, Thijssen and colleagues only refer to the training of skills concerning braking and coasting behavior and do not take into account the variety of other behavioral elements of an ecological driving style, such as behavioral patterns during an acceleration situation.

Considering acceleration performance in terms of the assumed long-term effects in this study, trainees could improve their acceleration behavior for several eco-driving techniques due to the simulator-based training intervention. Moreover, in comparison to an untrained control group, they deployed the learned skills on an unknown route, as indicated by significant transfer effects. Nevertheless, they did not accomplish the same acceleration performance as they were trained in the guided drive. In this case, one single guided training session might be not enough to practice and internalize the knowledge, skills and attitudes of ecological driving.

For fuel savings, the anticipated transfer effect could not be confirmed. Neither in the comparison drive nor in real traffic situations was significantly less fuel consumed by the trained group compared to the baseline drive or to the control group, respectively. However, factors such as weather and order situation could not be controlled for during the collection of real traffic driving data. Another potential influence on the results consists in motivational aspects. Saving costs due to reduced fuel consumption is primarily of interest for the company [12]. Promoting an environmental motivation to drivers thus plays an important role with respect to the further application of the acquired knowledge.

During the test drives some exceptional driving errors may have led to biases in the data, e.g. drivers failing to stop at a red traffic light. These aspects might confound comparisons between control and training group. General validity of results is limited as well by means of small sample size and drop outs due to simulator sickness.

5 Conclusions and Outlook

Receiving broad attention by politics and society, eco-driving comprises an often discussed topic. Besides technical and strategic variables like regular maintenance, appropriate loading, or route choice [1], the most effective way to decrease fuel consumption and erosion continues from the individual driver. By means of training, an environmentally friendly way of operating the vehicle should be communicated and internalized. However, the question arises, what exactly constitutes a desirable driving style. Traditionally, behavioral outcomes like reductions in fuel consumption or a more reasonable use of brakes or accelerator are regarded as indicators for a successful switch over to eco-driving. Nevertheless, to achieve long-term changes it is necessary to influence drivers' attitude as well via highlighting the individual environmental responsibility. In doing so, the critical goal consists in creating a solid environmental awareness that is—in the long term—able to contribute to climate protection and pollution reduction.

Moreover, although performance within the training situation is adequate in most studies, mainly based on the potential of simulator training to demonstrate drivers the direct effect of behavioral changes [8], training success seems to diminish over time and reveals a high variance between drivers [3]. To preserve constructive behavior even within real-world settings, a promising approach might involve assistance systems that remind drivers to constantly apply eco-driving [5, 6, 9]. While such strategies have definitely proven of value, the potential disadvantages have to be considered as well. Continuous feedback messages may put additional perceptual and cognitive demands on the driver. Recently, Jamson et al. [7] inspected the effect of either visual or haptic feedback on eco-driving in a simulator-based study. They reported visual feedback to be superior, but at the same time to significantly distract drivers' attention and increase their experienced workload. Haptic feedback, on the other hand, had less effects on performance, but in line with Birrell et al. [4] in this case no increase in subjective workload was reported.

With these results in mind, a less-is-more approach might be a wise solution to keep drivers focused on the key objective of safely reaching the destination. With increased environmental awareness, achieved by goal setting assistant systems, each driver would develop responsibility for protecting our environment, and therefore apply an environmentally friendly driving style to prevent harmful pollution.

References

1. Alam MS, McNabola A (2014) A critical review and assessment of Eco-Driving policy & technology: benefits & limitations. Transp Policy 35:42–49
2. Baric D, Zovak G, Perisa M (2013) Effects of eco-drive education on the reduction of fuel consumption and CO_2 emissions. Promet – Traffic Transportation 25:265–272
3. Beusen B, Broekx S, Denys T, Beckx C, Degraeuwe B, Gijsbers M, Scheepers K, Govaerts L, Torfs R, Panis LI (2009) Using on-board logging devices to study the longer-term impact of an eco-driving course. Transp Res D 14:514–520
4. Birrell SA, Young MS, Weldon AM (2013) Vibrotactile pedals: provision of haptic feedback to support economical driving. Ergonomics 56:282–292
5. Camacho T, Quintal F, Scott M, Kostakos V, Oakley I, da Penteada C (2011) Towards egocentric fuel efficiency feedback. In: Proceedings of the 2nd International Workshop on Persuasion, Nudge, Influence, and Coercion through mobile devices, March 8, 2011, Vancouver, Canada
6. Daun TJ, Braun DG, Frank C, Haug S, Lienkamp M (2013) Evaluation of driving behavior and the efficacy of a predictive eco-driving assistance system for heavy commercial vehicles in a driving simulator experiment. In: 16th International annual conference on intelligent transportation systems, October 6–9, 2013, The Hauge, The Netherlands
7. Jamson SL, Hibberd DL, Jamson AH (2015) Drivers' ability to learn eco-driving skills; effects on fuel efficient and safe driving behaviour. Transp Res C 58:657–668. doi:10.1016/j.trc.2015.02.004
8. Mitsopoulos-Rubens E, Lenné MG, Salmon PM (2013) Effectiveness of simulator-based training for heavy vehicle operators: what do we know and what do we still need to know? In: Proceedings of the 2013 Australasian Road Safety Research, Policing & Education Conference, August. Brisbane, Queensland, pp 28–30
9. Staubach M, Scheibitz N, Köster F, Kuck D (2013) Evaluation of an eco-driving support system. Transp Res F 27:11–21
10. Strömberg HK, Karlsson ICM (2013) Comparative effects of eco-driving initiatives aimed at urban bus drivers – Results from a field trial. Transp Res D 22:28–33
11. Symmons MA, Rose G (2009) Ecodrive training delivers substantial fuel savings for heavy vehicle drivers. In: 5th International driving symposium on human factors in driver assessment, training, and vehicle design, June 22–26, 2009, Big Sky, Montana
12. Thijssen R, Hofman T, Ham J (2014) Ecodriving acceptance: an experimental study on anticipation behavior of truck drivers. Transp Res F 22:249–260
13. Wittenbrink P (2011) Transportkostenmanagement im Straßengüterverkehr: Grundlagen -Optimierungspotentiale - Green Logistics. Gabler Verlag, Wiesbaden