30 Methods for Measuring Mental Stress and Strain

Martin Schütte

1 Introduction

It is generally agreed that the measurement of work stress and strain provides indication of the appropriateness of work design, i.e. information which is necessary for verifying whether working conditions are in accordance with the criteria of human-centred design (LUCZAK 1998). Accordingly, such an analysis represents an important pre-condition for the evaluation of work-systems. Furthermore an assessment of mental workload is required by legislation, such as the "Machinery Directive" or the "EU Directive on the minimum health and safety requirements for work with display screen equipment". Both directives contain the obligation to measure mental stress of operators or employees.

However, a universally accepted concept of mental workload is missing and rather different conceptions concerning human information processing exist. The single resource theory assumes that the processing of information presupposes the supply of an adequate quantity of energy. Various internal processes which may take place in sequence or in parallel receive energy from a single undifferentiated pool, whereas an allocation policy mechanism governs the supply of resources (KAHNEMANN 1973). Other approaches, as e.g. the discrete stage model, assume that the information processing system is composed of various elements possessing a discrete function in the information transfer of the system. A processing stage can only be passed through if the preceding stage is completed (STERNBERG 1969). Accordingly, human information processing is therefore assumed to be strictly serial. In contrast, the multiple resource theory is based on the assumption that the capacity or energy necessary for various internal activities during information processing is divided into specific reservoirs (WICKENS 1984). The cognitive energetical stage model (SANDERS 1983) integrates the structural and energetical approaches of human information processing. The model considers three energetical supply mechanisms with the two more basic mechanisms arousal and activation are linked to input and output stages of information processing. The

C.M. Schlick (ed.), *Industrial Engineering and Ergonomics*,
DOI 10.1007/978-3-642-01293-8_30,

basal mechanisms are coordinated und supervised by a third mechanism, the effort system. An evaluation mechanism mediates the information about the basal mechanisms. Although the concept of situational awareness (ENDSLEY 1995) does not aim at the identification of different information processing mechanisms, it accentuates the control of attention during control of complex technical systems. This concept does not aim at the identification of several information processing mechanisms. Situational awareness encompasses the perception, comprehension and anticipation of events. Working memory and attention pertain to those factors most limiting situational awareness. Therefore, situational awareness represents an agent, facilitating the coordination and processing of information coming from different sources.

2 Classification of Measurement Procedures

A multidimensional concept of mental workload is consistently postulated because the complexity of working conditions goes along with demands on different human information processing mechanisms. For an analysis of mental workload different investigation levels such as task demands, individual modality of task completion and the resulting human related consequences must be taken into consideration. Several measurement procedures are available measuring different aspects of mental workload. Only some procedures explicitly refer to specific theoretical models of human information processing.

In order to facilitate the selection of an suitable measurement procedure a classification system is suggested considering the terminology as proposed in DIN EN ISO 10075-1. According to this standard, all assessable external influences affecting an employee mentally are called mental stress. The immediate consequences of mental stress within a person depending on the individual habitual and actual preconditions as well as the particular coping styles are called strain. Correspondingly, the measurement procedures were classified into approaches for measurement of stress and strain.

3 Measurement of Mental Stress

Regarding the measurement instruments aiming at the determination of stress, two main groups can be distinguished. Approaches developed in engineering often refer to the technical performance of work or the work functions. Psychological approaches rather concentrate on the occurring interactions between the employee and his/her work.

3.1 Approaches Developed in the Field of Engineering

Methods attributable to engineering disciplines (Table 30.1) for documentation of mental stress often use quantitative parameter compatible to those for describing technical systems (SHERIDAN & FERELL 1974).

Table 30.1: Methods for measuring mental stress (SCT: Supervisory and control tasks, ST: Supervisory tasks, ATC: Air traffic control tasks, HCI: Human computer interaction, PT: Piloting tasks

Name of the approach	Literature	Tasks	Stress related parameter
Information theory	Senders 1964	SCT	Relative proportion of monitoring tasks
Control Theory	Rouse 1980	SCT	Linearity and time related stationarity of the human transfer behavior
Queuing Theory	Rouse 1980	ST	Relation between required and available time
Fuzzy Theory	Zadeh 1965	SCT	Quality of the used rules
Algorithm Theory	Matern 1983	SCT	Number of operations and decisions
Markov-Chains	Haccou & Meelis 1992	ACT, HCI	Information content of operation sequences
Production Systems	Anderson 1983	PT	Number of productions Number of arguments of the if component
TTA	Stone et al. 1987	PT	Relation between required and available time
TLAP	Parks & Boucek 1989	PT	Complexity of displayed information
SWAM	Linton et al. 1977	PT	Time related load of the human information channels
Workload Index	North & Riley 1989	PT	Requirements concerning attention Number of simultaneously feasible tasks
TASCO	Ellison & Roberts 1985	PT	Task difficulty Number of failures during task completion
SWAS	Holley 1989	PT	Number of simultaneously feasible tasks
HOS	Harris et al. 1987	PT	Frequency of task occurrence Task execution time
SIMWAM	Kirkpatrick et al. 1984	PT	Relation between required and available time Number of accomplished tasks

The advantage is that items for measurement directly refer to the same aspects of technical components that should be considered for design improvements. The vast majority of methods can be applied in settings, where the human operator performs monitoring and control operations.

Due to an increase in automation, for example in aviation and transportation those methods designed for the analysis of monitoring and control tasks will become more important.

Some of the instruments such as TTA, TLAP, SWAM, W/Index and TASCO can be used for predicting the level of stress for a planned workplace. It is assumed that adequate/optimal positioning of displays and control elements can ensure optimal task performance in given time. Accordingly, these approaches provide information about the feasibility of tasks and time constraints during task accomplishment. The above mentioned instruments assume that task execution happens successively and the human information processing system is a central processor with limited channel capacity. The only exception is the W/Index which is based on the multiple resource theory. An application of the simulation tools like SWAS, HOS and SIMWAM require a detailed task analysis. They additionally intend to reproduce essential characteristics of the human operator by considering variations in task execution times or parameters related to human perception, cognition and motor function. The simulation approaches mainly provide information about time related aspects of task feasibility. SWAS is based on the multiple resource theory whereas the other simulation procedures assume that tasks are to be completed sequentially.

3.2 Psychological Methods

The psychological methods for measuring demands of mental tasks are analyzed with respect to the commonly used criteria reliability and validity (Table 30.2). However, the instruments should also be evaluated with regard to their sensitivity and diagnosticity. A measurement procedure differentiating between for example low, moderate and high stress levels meets the criterion of sensitivity. An instrument responsive to different kinds of stress satisfies the criterion of diagniosticity. Both aspects point to the application area of an instrument and facilitate the selection of an instrument adequate for an intended measurement for instance in a factory. The psychological measurement procedures depicted in Table 30.2 are not only suitable for monitoring and control or supervisory tasks but also for office work and manufacturing.

The SWAT, NASA-TLX as well as the Cooper-Harper-Scale, Bedford-Scale and ZEIS-Scale are suited for determining the temporal course of stress levels since these instruments can be applied without much effort. Therefore these scales can be used for repeated measurements. SWORD and Workload Profile only allow a retrospective measurement of stress due to the high costs for the application of the procedures. The remaining approaches such as the DTV-AET, TAI, TBS-GA, VERA-G, ATAA, FEMA and VAB are especially useful for the analysis of the current state of a workplace and for checking the effects of accomplished (re-)design measures.

Table 30.2: Psychological methods for measuring mental stress (*black circle*: criterion was analyzed, *white circle*: criterion was not analyzed)

Name of the approach	Literature	Application Area	Reliability	Validity
DTV-AET	Haider & Rohmert 1981	Office work	●	●
TAI	Frieling et al. 1993	Universal	●	○
TBS-GA	Rudolph et al. 1987	Mental tasks with and without computer support	●	●
VERA-G	Resch 1994	Mental tasks	○	○
ATAA	Wächter et al. 1989	Tasks in production	●	●
FEMA	Tielsch et al. 1993	Industrial workplaces equipped with new technologies	●	●
VAB	Debusmann et al. 1991	Office work	○	●
Stress Questionnaire	Pfendler & Schubert 1985	Supervisory and control tasks	○	●
SWAT	Reid & Nygren 1988	Determination of Mental Load	●	●
NASA-TLX	Hart & Staveland 1988	Determination of mental Load	●	●
Workload Profile	Tsang & Velazquez 1996	Determination of mental load	●	●
Cooper-Harper Scale	Cooper & Harper 1969	Piloting tasks	●	●
Bedford Scale	Roscoe 1987	Piloting tasks	○	●
ZEIS	Käppler 1993	Vehicle guidance	●	●
SWORD	Vidulich et al. 1991	Piloting tasks	●	●

The psychological approaches are mainly based on the stress-strain concept (DIN EN ISO 10075-1) or the so-called action-regulation theory. Exemptions are the Bedford Scale which relies on the single channel concept, and the Workload

Profile which refers to the multiple resource theory. Consequently, the psychological methods mainly provide information about the feasibility of actions or potential for personality development.

Most of the procedures are analyzed with respect to their reliability and validity. However, only the SWAT, NASA-TLX and the Workload Profile were also tested regarding their sensitivity and diagnosticity (RUBIO et al. 2004).

The number of relevant dimensions of mental stress may justify the assumption that only multidimensional methods are adequate for the measurement of mental stress. Nevertheless, the findings of several studies aiming at the identification of the overall stress level of piloting tasks showed that one- or multi-dimensional methods provide widely comparable results of measurement (HENDY et al. 1993). This finding might be astonishing but corresponds to other experiences in using one-dimensional methods for recording the level of mental stress (SCHÜTTE et al. 1994). Instruments can generate meaningful results as long as the rater is capable to integrate the different stress related aspects in an overall assessment of stress. One-dimensional approaches are especially useful for initial screenings of mental stress (HART & WICKENS 1990). If the measurements are carried out to obtain detailed information about the working conditions multi-dimensional instruments should be applied only. Since all instruments are based on subjective perceptions and evaluations of the respondents, it should be kept in mind that answers are possibly subjected to biases resulting e.g. from the tendency to choose only average rating categories. One way to reduce biases could be to inform the rater about problems in using ratings scales (PFENDLER & SCHWEINGRUBER 1996). Despite the above mentioned disadvantages the application of questionnaires is considered as absolutely essential since it must be assumed that those aspects of working conditions which an employee perceived as stressful in fact represent stressful facets (GOPHER & BRAUNE 1984).

4 Measurement of Mental Strain

Approaches for measuring mental strain can be divided into two different groups. On the one hand physiological parameters are used for recording the intensity and time course of strain. On the other hand questionnaires and rating scales are available documenting the individual feeling with regard to mood, fatigue, effort etc.

4.1 Physiological Methods

The registration of physiological parameters is based on the assumption that they not only provide an indication of the information processing mechanisms involved in task execution but also point to costs and effort during task accomplishment. Initially, one-dimensional models such as the concept of central activation were most popular in this field (LUCZAK 1987). Currently, theories concerning attention are favoured; especially the cognitive-energetical stage model procuring an integrative heuristic for the analysis of mental strain (MANZEY 1998).

Starting from the assumption that mental strain is influenced by processes located in the central nervous system it seems plausible to use strain indicators that cover changes in the activity of cortical or sub-cortical neuronal formations assumed to reflect particular perceptual, cognitive and motor related task demands (MANZEY 1998). While imaging techniques allow for identification of inactive and active brain areas in high resolution regarding space, the registration of the EEG permits the registration of brain activity in high resolution regarding time. Furthermore this approach is well examined concerning methodological problems such as the avoidance as well as detection or clearing up of artefacts. The spontaneous activity of the EEG and differences between the amplitudes of various event related potentials (ERP) are discussed as possible strain indicators.

As regards spontaneous brain activity experimental findings verify that a rise of the frontal-central theta activity correlates substantially with the level of task demands. Accordingly this parameter fulfils the criterion of sensitivity. However, theta activity does not seem to be influenced by a particular kind of task demands. Theta activity presumably represents a more general activation mechanism comparable to the effort dimension of the cognitive energetical stage model (MANZEY 1998).

Regarding event-related potentials the P-300 is presumed to be an adequate strain indicator. The P-300 reflects voltage changes occurring not earlier than 300 milliseconds and above the parietal cortex with an amplitude maximum after the presentation of a stimulus (MANZEY 1998). Indications militating in favour of the sensitivity of the P-300 result from studies based on dual task paradigm. An increase of the demands of the primary task should lead to a decrease of the amount of resources available for execution of the secondary task. Experimental findings are in accordance with these assumptions. If two tasks had to be accomplished simultaneously, the amplitude of the P-300 for relevant secondary task events declined when the demands of the primary task were raised. The P-300 also satisfies the criterion of diagnosticity since the described changes of the P-300 amplitude appeared only if the secondary task is characterized by perceptual and cognitive demands. Even if this parameter permits apparently the recording of mental strain it must be taken into consideration that the analysis is very laborious (filtering and averaging of signals etc.). Additionally this approach is more suited for experiments than for field studies.

Furthermore heart-rate-variability (HRV) was discussed as another promising parameter indicating the level of mental strain, since it could be shown, that the time between the R-waves of the electrocardiogram decreases when task demands increase (KALSBEEK & ETTEMA 1963). Various algorithms were proposed for the calculation of HRV (e.g. LUCZAK 1979; LUCZAK & LAURIG 1973) but techniques of spectral analysis are currently nearly used by default. Three frequency bands could be identified having relation to thermoregulation (0.02–0.06 Hz), blood pressure (0.07–0.14 HZ) and respiration (0.15–0.50 Hz). The performance of mentally demanding tasks lead to a power reduction in all three frequency bands. However, the biggest changes appear in the central band (0.07–0.14 Hz), which accordingly is considered as the best indicator for mental strain. Experimental findings demonstrate that the 0.1 Hz component of HRV differentiates between resting and non-resting conditions only. The component does not discriminate between different levels of task difficulty (JORNA 1992) and is therefore not conform to the criterion of sensitivity. Further results indicate that the 0.1 Hz component presumably reflects variations in unspecific, general activation. This component more likely indicates emotional than mental strain (NICKEL & NACHREINER 2002).

Regarding the human visual system the frequency of spontaneous eye blinks and the pupil diameter are considered as mainly influenced by mental strain. Several experimental findings show that the frequency of eye blinks depends on the difficulty of the particular task. An increase in task difficulty is accompanied by a decrease of the blink rate. The parameter is apparently responsive to different stress levels and can be considered sensitive. Even if the blink rate is unaffected when an auditory task presentation takes place (e.g. CASALI & WIERWILLE 1983), the diagnosticity of this parameter is not completely clarified. It is required to carefully analyze whether a particular task provokes eye movements which are usually accompanied by blinks (HARGUTT 2001).

Similar to blink rate, a clear relationship seems to exist between stress level and pupil diameter. An increase in task difficulty comes along with an increase in pupil diameter (e.g. RÖßGER 1999). This relation occurred irrespective of the particular kind of task. Therefore pupil diameter meets the criterion of sensitivity, but is not in accordance with the criterion of diagnosticity. Based on these results it is assumed that this parameter bear probably upon the effort dimension of the cognitive energetical stage model. The measurement of the pupil diameter requires relatively constant environmental conditions in order to avoid disturbing factors such as difference in illuminance or accommodation processes on pupil size. Accordingly, the method is more suited for experiments than for field studies.

4.2 Methods for Recording the Subjectively Perceived Level of Strain

Measurement instruments recording the subjectively perceived level of strain are part of the commonly used repertoire of methods utilized for the analysis of mental strain. This is because ratings are regarded as the most direct procedures for evaluating the human costs during task accomplishment (HART & WICKENS 1990). Rating procedures are also considered as efficient and as non intrusive in application (PFENDLER & SCHWEINGRUBER 1996, WIERWILLE & EGGEMEIER 1993). The approaches available can be classified in one- and multi-dimensional. One-dimensional methods are based on the assumption that the number of different components relevant in subjective perceived strain can be reduced to a single dimension involved in information processing (BARTENWERFER 1978). Multi-dimensional instruments capture different aspects of mental strain and provide additional information about the structure of subjectively experienced strain.

The effort scale (ZIJLSTRA & VAN DOORN 1985) is a one-dimensional rating procedure which refers to the effort system of the cognitive energetical stage model. The instrument is composed of 7 judgement anchors describing different effort levels (see EILERS et al. 1986 for a German language version). The anchors are scaled according to the magnitude estimation method and possess a satisfying reliability. The validity of the effort scale was evaluated in experimental and field studies using mainly signal detection tasks of different levels of difficulty. Amongst others also time of day and the frequency of critical signals was taken into account (EILERS et al. 1989, EILERS et al. 1990). None of these experiments provided evidence for the validity of the scale. In another experiment on logical reasoning the participants were asked to estimate their experienced effort subsequent to the accomplishment of easy and difficult tasks. The length and the sequence of sections with easy and difficult demands were experimentally varied. As a result, subjectively perceived effort was substantially influenced by the sequence and also the duration of task sections. This finding demonstrates that the effort scale is responsive to different stress profiles (SCHÜTTE 1999). Using tasks of the AGARD-STRES Battery for generating different kinds and levels of mental stress it could be shown, that the effort scale differentiates between low and high stress levels. This result militates in favour of the sensitivity of the effort scale (SCHÜTTE 2001, SCHÜTTE 2002). Nevertheless the diagnosticity of the effort scale requires further investigation.

Regarding the multi-dimensional approaches, the EZ-scale (NITSCH 1974) is a highly frequently used method. The instrument measures the actual general condition of a person with regard to mood. The questionnaire consists of 40 items mapping two dimensions namely the readiness for action (motivation) and the capacity to act (strain), of which are both divided into subscales. Since the actual mood-related state of a person shows fluctuation throughout the day, the reliability of the EZ scale is hardly determinable. The validity of the questionnaire was investigated in several studies. The findings verify that the scale differentiates between relaxed and tense situations (e.g. examinations). The EZ-scale was also used in pharmacological tests; the measurements discriminated between the reference situation (placebo) and conditions with the administration of different doses of a tranquillizer (NITSCH 1976). Furthermore, the EZ-scale maps the strain-related differences between variably organized office work (FRIELING et al. 1979). These results demonstrate the validity of the approach, but more detailed studies concerning the sensitivity and diagnosticity are not available. Since some items of the EZ scale were difficult to comprehend for some raters a modified version of the scale was developed (APENBURG 1986). The reliability (inner consistency) of the modified version reaches satisfying values between 0.80 and 0.92 which are in accordance with measurement precision class 2 ($r \geq 0.8$) respectively 1 ($r \geq 0.9$) according to DIN EN ISO 10075-3. The findings of a study comparing strain of six occupational groups provide evidence for diagnosticity as the groups differ in their average strain level. Nevertheless, studies giving more detailed information about both the diagnosticity and sensitivity of the modified version of the questionnaire are not available.

The questionnaire for measuring fatigue, monotony, satiation and stress (BMS II) was explicitly developed for assessing strain levels during the accomplishment of monitoring and control task performance. The instrument is based on a model assuming that strain during task performance depends on the intensity and level of task demands as well as the individual qualification and habitual particularities of activation (PLATH & RICHTER 1984). In addition the model supposes that strain has positive consequences. Task performance may facilitate learning processes and can lead to (re-)forming of work related attitudes which both improve the psychological regulatory processes concerning task completion and motivation. Negative consequences of strain may also be expected leading to impairments such as fatigue, monotony, satiation and stress. Accordingly, the development of the questionnaire was based on several postulations. The instrument should not only capture negative but also positive conditions of mood. The items should map task demands and the instrument should measure actual mood conditions. The questionnaire should allow for repeated measures resulting in the development of two widely comparable alternative versions of the questionnaire. Each version comprises 10 items for each of the four scales fatigue, monotony, satiation and stress. The method of successive intervals was used for calculating a scale-value for each item reflecting the degree of the particular characteristic the item captures (e.g. fatigue). When applying the method the respondents only have to indicate whether a particular item adequately describes the actual condition of mood (yes / no answer). The individual level of e.g. fatigue will be determined by summing up the scale values of all items answered affirmatively. Subsequently the average value will be calculated. Reliability was assessed by correlations between the two alternative forms of the questionnaire. Based on the post test measurements obtained in various experiments the correlation coefficients take values between 0.77 (scale monotony) and 0.88 (scale fatigue). Taking into consideration the recommendations of DIN EN ISO 10075-3 the reliability of the scales corresponds to precision class 3 ($r \geq 0.7$) respectively 2 ($r \geq 0.8$). The validity of the instrument was proved in several experiments and field studies. The findings verify that the questionnaire discriminates between conditions of low and high stress and also between different kinds of stress. These outcomes militate in favour of the sensitivity and diagnosticity of the instrument. Furthermore it could be shown that an increasing number of failures during task completion goes along with the perception of satiation and stress (in control task settings) and that prolonged reaction times are attended by higher stress and fatigue (in choice reaction time task settings). Field studies verify that a small number of machine operations lead to fatigue. In monitoring tasks or multi operation situations the feeling of satiation and stress dominates. All in all these results confirm the validity of the questionnaire. Furthermore the current level of strain can be compared with threshold values indicating the necessity of work redesign. The BMS II is also applicable during the planning phase of work places since the questionnaire is integrated in the REBA tool (JORDAN et al. 1997).

The BLV measures changes in strain (KÜNSTLER 1980). The approach has the advantage to provide information about an individual strain level and about compensatory behaviour. The development of the instrument was based on the assumption that strain destabilizes the biological equilibrium which a person attempts to restore. Four possible behavioural patterns can be distinguished (1) break off of the particular task, (2) lowering the level of achievement, (3) increase of effort and (4) variation of mode of task accomplishment. The questionnaire comprises in total 46 items which constitute the four scales perceived momentary performance capability, fatigue, tension and achievement motivation. The reliability (test-half coefficient) varied between 0.91 (tension) and 0.98 (momentary performance capability); i.e. above the lower limit of 0.90 as recommended by DIN EN ISO 10075-3 for high precision measurements (precision class 1: $r \geq 0.9$). The validity was proved in a study comparing the measurement results of the BLV with the measurement results of the BMS II. The calculated correlation coefficients take values between 0.31 and 0.68 indicating that the BLV conforms to the criterion of internal validity. Further support for the validity of the instrument is provided by a field study analyzing the strain of air-traffic-controllers in shift time (KÜNSTLER & NELTE 1983). Substantial changes of strain were detectable during night shifts and two different forms of compensatory behaviour could be observed. Some air-traffic-controllers attempted to maintain their reactivity by increasing their tension, despite the existence of fatigue. For the other air-traffic-controllers the experience of fatigue was dominant. This was accompanied by reduced motivation, perceived momentary performance capability and tension symptoms which are typical for the condition of monotony. No studies are available on sensitivity and diagnosticity of the BLV.

One reason for diversity of questionnaires available and their differences concerning strain dimensions could be assumed in procedures chosen for item collection and selection as well as psychometric methods used for scale development. However, this would ignore the fact that strain related dimensions such as fatigue or activation can be identified even if different techniques of data acquisition and analysis methods are applied (HAIDER 1961, HAMPEL 1977, HECHELTJEN & MERTESDORF 1973). It is therefore more likely that the diversity of instruments is due to different theoretical conceptions underlying the questionnaires. The EZ-scale is based on a model that assumes a strong relation between mood and demands as well as motivating elements of a task. The BLV is based on a concept that assumes strain to be the result of complex decision and evaluation processes and provoked by changes in the individual work capabilities. The BMS II is based on a model that differentiates between the subjective representation of task demands and the supporting or impairing consequences. The development of most of the questionnaires was not influenced by models of human information processing. This might have been because in these models performance parameter such as task execution time or the number of failures during task completion are more suited for the determination of the human related consequences of information processing. However, the cognitive energetical stage model (SANDERS 1983) considers strain related dimensions and provides a theoretical basis of only few instruments such as the effort scale. One possible explanation could be that this model seems to be more adequate for experimentally generated simple tasks allowing a precise manipulation of stimulus intensity, signal quality, stimulus-response comparability and time uncertainty.

5 Conclusion

Most of the described instruments are analyzed with respect to their reliability and validity. Regarding strain measures one essential aspect of validity is the confirmation that an instrument is sensitive to both the duration and intensity of stress. Furthermore, there should also be a check of the validity of the statistical decision, the internal and external validity of the experimental design as well as construct validity (NACHREINER et al. 1987). Although there are most often some studies providing evidence for sensitivity and diagnosticity, systematic analyses is mostly missing. However, such an investigation is very important since it provide information about the scope of application of a measurement procedure and is a helpful specification for practitioners looking for an instrument adequate for their purpose of measurement. Furthermore, it should be noted that reliability studies often erroneously assume that the person represents the object of measurement. Therefore calculated coefficients reflect the reliability for discriminating between persons based on the measurements. Only a few studies exist (e.g. SCHÜTTE 2001, SCHÜTTE 2002, SCHÜTTE & NICKEL 2002) reporting the reliability for the discrimination between different stress conditions based on the measurements. One elegant way for determining the person or condition related reliability of measures of stress or strain is the application of generalizability theory (GT; BRENNAN 2001). GT has the advantage to allow for an estimation of the magnitude of multiple sources of measurement error and thus facilitating the separation of major error sources. In G-theory every measurement is regarded as a sample taken from a universe of possible measurements which could have been taken for an object of measurement (e.g. stress level). This universe is named the universe of admissible observations and is characterized by various variables (facets) which represent possible sources of variability in the measured scores (e.g. persons). Each specific instance of a facet is called condition. The application of G-Theory requires two different steps, namely the realization of a G- and a D-study. At first the G-study has to be accomplished. The G-study aims at the estimation of the magnitude of the sources of variability as defined in the universe of admissible observations. Subsequently a D-study (Decision study) has to be accomplished and is based on the results of the G-study. The D-study gives information about the generalizability of the measurements. For that purpose two different reliability analogous parameters can be computed. The so-called relative G-coefficient (ρ^2) indicates on a 0-to-1 scale, how well an observed score is likely to locate the objects of measurement relative to other members of the corresponding population. The absolute G-coefficient (ϕ) indicates also on a 0-to-1 scale how well an observed score is likely to locate the objects of measurement independent from others in the respective population. Furthermore, the D-study provides not only indications concerning the costs required for an application of the measurement procedure (e.g. optimal number of conditions of a facet) but also information about alternative measurement designs bringing about reliable scores. It was criticized that the application of this procedure requires controlled experiments and neglects field studies. G-theory do not exclusively presuppose experiments as also several field studies have been conducted e.g. in the domain of education (GILLMORE et al. 1978). However, if an instrument does not provide reliably measured values in an experiment, testing the method in a field setting can be omitted (NACHREINER & SCHÜTTE 2002).

Some instruments such as SWAT or NASA-TLX (REID & COLLE 1988, COLLE & REID 2005) possess guideline or threshold values indicating whether the strain level has reached a critical level and by that indicating that performance decrements could occur. However, the proposed values give no information about the admissible stress level and duration. Future research should therefore establish critical or threshold values to provide an orientation for human related criterions of work design.

6 References

Anderson JT (1983) The architecture of cognition. Havard University Press, Cambridge.

Apenburg E (1986) Befindlichkeitsbeschreibung als Methode der Beanspruchungsmessung. Zeitschrift für Arbeits- und Organisationspsychologie, 30: 3–14.

Bartenwerfer H (1978) Ermüdung durch psychische Belastung. Zeitschrift für Arbeitswissenschaft, 32: 12–14.

Brennan RL (2001) Generalizability Theory. Springer, Berlin.

Casali JG, Wierwille WW (1983) A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. Human Factors, 25: 623–641.

Colle HA, Reid GB (2005) Estimating a mental workload redline in a simulated air-to-ground combat mission. The International Journal of Aviation Psychology, 15: 303–319.

Cooper GE, Harper RP (1969) The use of pilot rating in the evaluation of aircraft handling qualities, NASA-TN-D-5153. National Aeronautics and Space Administration, Washington.

Debusmann E, Licher A, Mehrtens M (1991) Aufgaben im Büro analysieren und bewerten. Verlag TÜV Rheinland, Köln.

Eilers K, Nachreiner F, Hänecke K (1986) Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung. Zeitschrift für Arbeitswissenschaft, 40: 215–224.

Eilers K, Nachreiner F, Böning E (1989) Zur subjektiven Skalierung psychischer Beanspruchung – Teil 1: Experimentelle Überprüfung der Validität verankerter Relativurteile. Zeitschrift für Arbeitswissenschaft, 43: 217–223.

Eilers K, Nachreiner F, Böning E (1990) Zur subjektiven Skalierung psychischer Beanspruchung – Teil 2: Überprüfung der Validität verankerter Relativurteile in einer Felduntersuchung. Zeitschrift für Arbeitswissenschaft, 44: 24–29.

Ellison MG, Roberts BB (1985) Time based analysis of significant coordinated operations (TASCO): A cockpit workload analysis technique. In: Human Factors Society (Edt.) Proceedings of the Human Factors Society – 29th Annual Meeting – 1985. Human Factors Society, Santa Monica, 774–778.

Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. Human Factors, 37: 32–64.

Frieling E, Maier W, Michallik U, Schwan R (1979) Belastung, Beanspruchung und Zufriedenheit bei unterschiedlich organisierten Schreibtätigkeiten. Zeitschrift für Arbeitswissenschaft, 33: 216–223.

Frieling E, Facaoaru C, Benedix J, Pfaus H, Sonntag K (1993) Tätigkeitsanalyseinventar – Theorie, Auswertung, Praxis – Handbuch und Verfahren. Ecomed Verlagsgesellschaft, Landsberg.

Gillmore GM, Kane MT, Naccarato RW (1978) The generalizability of student ratings of instruction: Estimation of the teacher and course components. Journal of Educational Psychology, 15: 1–13.

Gopher D, Braune R (1984) On the psychophysics of workload: Why bother with subjective measures? Human Factors, 26: 519–532.

Haccou P, Meelis E (1992) Statistical analysis of behavioural data – An approach based on time-structured models. Oxford University Press, Oxford.

Haider M (1961) Direkte und indirekte Selbsteinstufungen der Auswirkungen von freier Arbeit und Fließbandarbeit. Psychologie und Praxis, 5: 1–11.

Haider E, Rohmert W (1981) Anforderungsermittlung im Bereich der Daten- und Textverarbeitung mit einem arbeitswissenschaftlichen Erhebungsverfahren zur Tätigkeitsanalyse (DTV-AET). International Archives of Occupational and Environmental Health, 46: 203–217.

Hampel R (1977) Adjektiv-Skalen zur Einschätzung der Stimmung (SES). Diagnostica, 23: 43–60.

Hargutt V (2001) Das Lidschlussverhalten als Indikator für Aufmerksamkeits- und Müdigkeitsprozesse bei Arbeitshandlungen, unpublished doctoral thesis. Julius-Maximilians-Universität, Würzburg.

Harris RM, Iavecchia HP, Ross LV, Shaffer SC (1987) Microcomputer human operator simulator (HOS-IV). In: Human Factors Society (Edt.) Proceedings of the Human Factors Society – 31st Annual Meeting – 1987. Human Factors Society, Santa Monica, 1179–1183.

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. North-Holland, Amsterdam, 139–183.

Hart SG, Wickens CD (1990) Workload assessment and prediction. In: Booher HR (Ed.) Manprint – An approach to systems integration. Van Nostrand Reinhold, New York, 257–296.

Hecheltjen KG, Mertesdorf F (1973) Entwicklung eines mehrdimensionalen Stimmungsfragebogens. Gruppendynamik, 4: 110–122.

Hendy KC, Hamilton KM, Landry LN (1993) Measuring subjective workload: When is one scale better than many? Human Factors, 35: 579–601.

Holley CD (1989) A model for peforming system performance analysis in pre-design. In: McMillan GR, Beevis D, Salas E, Strub MH, Sutton R, van Breda L (Eds.) Applications of human performance models to system design. Plenum Press, New York, 91–99.

Jordan P, Pohlandt A, Hacker W, Richter P (1997) REBA – Rechnergestütztes Verfahren zur psychologischen Tätigkeitsbewertung. In: Landau K, Luczak H, Laurig W (Eds.) Software-Werkzeuge zur ergonomischen Arbeitsgestaltung. Verlag Institut für Arbeitsorganisation e.V., Bad Urach, 34–52.

Jorna PGAM (1992) Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. Biological Psychology, 34: 237–257.

Käppler WD (1993) Beitrag zur Vorhersage von Einschätzungen des Fahrverhaltens. VDI-Verlag, Düsseldorf.

Kahnemann D (1973) Attention and effort. Prentice-Hall, Englewood Cliffs.

Kalsbeek JWH, Ettema JH (1963) Continuous recording of heart rate and the measurement of perceptual load. Ergonomics, 6: 306–307.

Kirkpatric M, Malone TB, Andrews PJ (1984) Development of an interactive microprocessor based workload evaluation model (SIMWAM). In: Human Factors Society (Ed.) Proceedings of the Human Factors Society – 28th Annual Meeting – 1984. Human Factors Society, Santa Monica, 78–80.

Künstler B (1980) Psychische Belasdtung durch die Arbeitstätigkeit – Theoretisches Rahmenkonzept der Entwicklung eines Fragebogens zum Belastungserleben. Probleme und Ergebnisse der Psychologie, 74: 45–66.

Künstler B, Nelte B (1983) Zur Analyse der psychischen Beanspruchung von Flugsicherungspersonal. Verkehrsmedizin und ihre Grenzgebiete, 30: 89–97.

Linton PM, Jahns DW, Chatelier PR (1977) Operator workload assessment model: An evaluation of a VF/VA-V/STOL system. In: North Atlantic Treaty Organization (Ed.) Methods to assess workload, AGARD Conference Proceedings No. 216. Advisory Group for Aerospace Research & Development, Neuilly sur Seine.

Luczak H (1979) Fractioned heart rate variability. Part II: Experiments on superimposition of components of stress. Ergonomics, 22: 1315–1323.

Luczak H (1987) Psychophysiologische Methoden zur Erfassung psychophysischer Beanspruchungszustände. In: Kleinbeck U, Rutenfranz J (Eds.) Arbeitspsychologie – Wirtschafts-, Organisations- und Arbeitspsychologie, Band 1. Hogrefe, Göttingen, 185–259.

Luczak H (1998) Arbeitswissenschaft. Springer, Berlin.

Luczak H, Laurig W (1973) An analysis of heart rate variability. Ergonomics, 16: 85–97.

Manzey D (1998) Psychophysiologie mentaler Beanspruchung. In: Rösler F (Ed.) Ergebnisse und Anwendungen der Psychophysiologie – Enzyklopädie der Psychologie, Band 7. Hogrefe, Göttingen, 799–864.

Matern B (1983) Psychologische Arbeitsanalyse. VEB Deutscher Verlag der Wissenschaften, Berlin.

Nachreiner F, Schütte M (2002) Zur Messung psychischer Belastung und Beanspruchug. Zeitschrift für Arbeitswissenschaft, 56: 1–3.

Nachreiner F, Müller GF, Ernst G (1987) Methoden zur Planung und Bewertung arbeitspsychologischer Interventionsmaßnahmen. In: Kleinbeck U, Rutenfranz J (Eds.) Arbeitspsychologie – Wirtschafts-, Organisations- und Arbeitspsychologie, Band 1. Hogrefe, Göttingen, 360–439.

Nickel P, Nachreiner F (2002) The suitability of the 0.1 Hz component of heart rate variability for the assessment of mental workload in real and simulated work situation. In: de Waard D, Brookhuis KA, Moraal J, Toffetti A (Eds.) Human factors in transportation, communication, health, and the worklplace. Shaker Publishing, Maastricht, 317–334.

Nitsch JR (1974) Die hierarchische Struktur des Eigenszustandes – Ein Approximationsversuch mit Hilfe der Binärstrukturanalyse. Diagnostica, 20: 142–164.

Nitsch JR (1976) Die Eigenzustandsskala (EZ-Skala) – Ein Verfahren zur hierarchisch-mehrdimensionalen Befindlichkeitsskalierung. In: Nitsch JR, Udris I (Eds.) Beanspruchung im Sport. Limpert, Bad Homurg, 81–102.
North RA, Riley VA (1989) W/INDEX: A predictive model of operator workload. In: McMillan GR, Beevis D, Salas E, Strub MH, Sutton R, van Breda L (Eds.) Applications of human performance models to system design. Plenum Press, New York, 81–89.
Parks DL, Boucek GP (1989) Workload prediction, diagnosis, and continuing challenges. In: McMillan GR, Beevis D, Salas E, Strub MH, Sutton R, van Breda L (Eds.) Applications of human performance models to system design. Plenum Press, New York, 47–63.
Pfendler C, Schubert KP (1985) Assessment of operator input-load with highly automated engine control tasks. In: Willumeit HP (Ed.) Human decision making and manual control. North-Holland, Amsterdam, 173–183.
Pfendler C, Schweingruber J (1996) Übersicht über rechnergestützte Methoden zur Beanspruchungsermittlung. Forschungsgesellschaft für angewandte Naturwissenschaften e.V., Wachtberg.
Plath HE, Richter P (1984) Ermüdung-Monotonie-Sättigung-Stress. Psychodiagnostisches Zentrum der Humboldt-Universität zu Berlin, Berlin.
Reid GB, Colle HA (1988) Critical SWAT values for predicting operator workload. In: Human Factors Society (Ed.) Proceedings of the Human Factors Society – 32nd Annual Meeting – 1988. Human Factors Society, Santa Monica, 1414–1418.
Reid GB, Nygren TE (1988) The subjective workload assessment technique: A scaling procdure for measuring mental workload. In: Hancock PA, Meshkati N (Eds.) Human mental workload. North-Holland, Amsterdam, 185–218.
Resch M (1994) VERA-G. In: Volpert W (Ed.) Humanwissenschaft der Arbeit – ein Rückblick. Institut für Humanwissenschaft in Arbeit und Ausbildung der Technischen Universität Berlin, Berlin, 61–64.
Rößger P (1999) Das Verhalten des Pupillendurchmessers bei einfachen Entscheidungen. In: Rötting M, Seifert K (Eds.) Blickbewegungen in der Mensch-Maschine-Systemtechnik. Pro Universitate Verlag, Sinzheim, 163–172.
Roscoe AH (1987) In-flight assessment of workload using pilot ratings and heart rate. In: Roscoe AH (Ed.) The practical assessment of pilot workload, AGARDograph No. 282. Advisory Group for Aerospace Research & Development, Neuilly sur Seine, 78–82.
Rouse WB (1980) System engineering models of human-machine interaction. North-Holland, New York.
Rubio S, Diaz E, Martin J, Puente JM (2004) Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and Workload Profile methods. Applied Psychology: An International Review, 53: 61–86.
Rudolph E, Schönfelder E, Hacker W (1987) Tätigkeitsbewertungssystem – Geistige Arbeit. Psychodiagnostisches Zentrum – Sektion Psychologie der Humoldt-Universität zu Berlin, Berlin.
Sanders AF (1983) Towards a model of stress and human performance. Acta Psychologica, 53: 61–97.
Senders JW (1964) The human operator as a monitor and controller of multidegree of freedom systems. IEEE Transactions on Human Factors in Electronics, 5: 2–5.
Schütte M (1999) Mental strain and the problem of repeated measurement. Ergonomics, 42: 1665–1678.
Schütte M (2001) Diagnostizität und Sensitivität von Beanspruchungsmessverfahren. Zeitschrift für Arbeitswissenschaft, 55: 39–48.
Schütte M (2002) Bestimmung der bedingungsbezogenen Messgenauigkeit der Anstrengungsskala. Zeitschrift für Arbeitswissenschaft, 56: 37–45.

Schütte M, Ge S, Hertting-Thomasius R (1994) An analysis of the subjective strain perceived by parcel-deliverers. International Journal of Industrial Ergonomics, 13: 15–23.

Schütte M, Nickel P (2002) Multivariate Analyse der bedingungsbezogenen Messgenauigkeit von Beanspruchungsindikatoren. Zeitschrift für Arbeitswissenschaft, 56: 55–66.

Sheridan TB, Ferell WR (1974) Man-Machine-Systems: Information, control, and decision models of human performance. MIT Press, Cambridge.

Sternberg S (1969) The discovery of processing stages: Extension of Donders' method. Acta Psychologica, 30: 276–315.

Stone G, Gulick RK, Gabriel RF (1983) Use of task timeline analysis to assess crew workload. In: Roscoe AH (Ed.) The practical assessment of pilot workload, AGARDograph No. 282. Advisory Group for Aerospace Research & Development, Neuilly sur Seine.

Tielsch R, Hofmann A, Häcker H (1993) FEMA – Fragebogen zur Erfassung mentaler Arbeitsbelastungen – Erste Ergebnisse einer Validierungsstudie im industriellen Bereich. Zeitschrift für Arbeits- und Organisationspsychologie, 37: 86–94.

Tsang PS, Velazquez VL (1996) Diagnosticity and multidimensional subjective workload ratings. Ergonomics, 39: 358–381.

Vidulich MA, Ward GF, Schueren J (1991) Using the subjective workload dominance (SWORD) technique for projective workload assessment. Human Factors, 33: 677–691.

Wächter H, Modrow-Thiel B, Roßmann G (1989) Persönlichkeitsförderliche Arbeitsgestaltung – Die Entwicklung des arbeitsanalytischen Verfahrens ATAA. Hampp, München.

Wickens CD (1984) Processing resources in attention. In: Parasuraman R, Davies DR (Eds.) Varieties of attention. Academic Press, Orlando, 63–102.

Wierwille WW, Eggemeier FT (1993) Recommendations for mental workload measurement in a test and evaluation environment. Human Factors, 35: 263–281.

Zadeh LA (1965) Fuzzy sets. Information and Control, 8: 338–353.

Zijlstra F, van Doorn L (1985), The construction of a scale to measure subjective effort, unpublished report. Universität Groningen, Groningen.

Laws

Council Directive of 29 May 1990 on the minimum safety and health requirements for work with display screen equipment (fifth individual directive within the meaning of article 16 (1) of directive 87/391/EEC), L156, Official Journal of the European Union, 21. 6. 1990.

Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (recast), L157/24, Official Journal of the European Union, 9. 6. 2006.

Standards

DIN EN ISO 10075-1 (2000) Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung – Teil 1: Allgemeines und Begriffe. Beuth, Berlin.

DIN EN ISO 10075-3 (2004) Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung - Teil 3: Grundsätze und Anforderungen an Verfahren zur Messung und Erfassung psychischer Arbeitsbelastung. Beuth, Berlin.