

RESEARCH

Open Access



Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain

Andreas Rausch^{1*}, Jürgen Seifried², Eveline Wuttke³, Kristina Kögler³ and Steffen Brandt⁴

*Correspondence:

andreas.rausch@
uni-bamberg.de

¹ Business Education,
University of Bamberg,
Kapuzinerstrasse 16,
96047 Bamberg, Germany
Full list of author information
is available at the end of the
article

Abstract

Background: To measure higher-order outcomes of vocational education and training (VET) we developed a computer-based assessment of domain-specific problem-solving competence. In modeling problem-solving competence, we distinguish four components of competence: (1) knowledge application, (2) metacognition, (3) self-concept, and (4) interest as well as thirteen facets of competence, each of which is assigned to one of the four components.

Methods: With regard to ecological and content validity, rather than apply highly structured items (e.g. multiple choice items), we developed three authentic problem scenarios and provided an open-ended problem space in terms of an authentic office simulation. The assessment was aimed at apprentice industrial clerks at the end of a 3-year apprenticeship program and focused on the domain of controlling (i.e., support of managerial decisions, cost planning, cost control, cost accounting, etc.). The computer-based office simulation provided typical tools (e.g., email client, spreadsheet software, file system, notebook, calculator, etc.). In order to assess the non-cognitive components in our competence model, we implemented an integrated measurement of self-concept and interest that we refer to as 'Embedded Experience Sampling' (EES). Test-takers are requested to spontaneously answer short prompts (EES items) during the test that are embedded in typical social interactions in the workplace. The empirical section is based on a study with 780 VET students from three commercial training occupations in Germany (industrial clerks and apprentices from two similar VET programs). The focus of the contribution is on testing a theoretically derived competence model based on item response theory, the implemented scoring methods and reliability of the instrument. Fine-grained response patterns from automated codings and human ratings were condensed into one partial credit item for each scenario and each of the facets in the cognitive component 'knowledge application'.

Results: The multidimensional Rasch analysis revealed satisfactory EAP/PV reliabilities, which are between .78 and .84 for the 'knowledge application' facets and between .77 and .85 for the non-cognitive facets. Furthermore, the achievement differences between the industrial clerks and their comparison groups are as assumed.

Conclusions: In our study, we introduced an innovative method to measure non-cognitive facets of problem-solving competence in the course of complex problem

scenarios. Furthermore, by using authentic problem scenarios and providing an open-ended and authentic problem space, our assessment of domain-specific problem-solving competence focuses on ecological validity but also ensures reliability.

Keywords: Computer-based assessment, Problem-solving competence, Vocational education and training, Office work, Office simulation, Non-cognitive facets, Embedded experience sampling, Item response theory, Partial credits, Automated scoring

Introduction

Assessing outcomes of educational efforts in terms of competence has a long tradition in fields of general education and gained particular recognition through international large-scale assessments such as Programme for International Student Assessment (PISA). However, respective efforts to measure domain-specific vocational and professional competences are still rare. In 2011, the German Federal Ministry of Education and Research (BMBF) launched the research initiative ‘Technology-Based Assessment of Skills and Competencies in Vocational Education and Training (ASCOT)’. The initiative aimed at the development of computer-based instruments for the assessment of domain-specific competences in selected vocations in Germany on the basis of authentic work and business processes. The work presented in this paper is part of a research project entitled ‘Modelling and measuring domain-specific problem-solving competence of industrial clerks (DomPL-IK)’.¹ In the following, we want to highlight two innovative features of our competence measurement in the business domain: (1) With problem-solving competence we address higher-order competences instead of just knowledge reproduction. Therefore, we did not only develop complex problem scenarios within an authentic office simulation but also provided an open-ended and authentic problem space for working on these problems rather than apply highly structured items (e.g., multiple choice items). The analysis of the participants’ complex behavior patterns was based on a theoretically driven competence model and on item response theory (IRT). (2) As an alternative to relying on detached self-report questionnaires, we implemented an integrated measurement of non-cognitive facets of competence (i.e., facets of self-concept and interest) that we refer to as ‘Embedded Experience Sampling’ (EES): Test-takers in a complex problem-solving task are requested to stop at certain times during the test and spontaneously answer short prompts regarding their actual experience of the problem situation (e.g., ‘Your colleague Julian visits your office: *Hi, how are you? I heard you have to deal with a rather large task. Well, I just wanted to ask how you are doing.*’; answers were to be given on a four-point Likert scale, e.g. from ‘At the moment, I feel not at all confident’ = 1 to ‘... very confident’ = 4). The project is located in commercial vocational education and training (VET). Nevertheless, the approach is applicable in other domains as well.

A study with nearly 800 VET students was conducted in 2014. This paper provides an overview of the theoretical modeling of domain-specific problem-solving competence, the development of problem scenarios in the field of controlling, the computer-based test environment, and the implementation of EES. Particular attention is given to the analysis of reliability and validity of the developed competence assessment based on the

¹ The project is funded by the German Federal Ministry of Education and Research under Grant No. 01DB081119-01DB1123.

empirical study. Finally, we discuss limitations, possible applications and advancements of the assessment.

Background

Domain-specific problem-solving competence

According to common definitions, a person is confronted with a problem when he or she has a goal but—in contrast to facing a simple task/routine task—does not immediately know what is needed to reach the desired goal (Duncker 1945; Newell and Simon 1972; OECD 2013). Thus, whether a situation is perceived as a task or a problem depends on an individual's prior experience, knowledge and skills (Dörner 1987; Mayer 1994; Funke et al. in print). However, even for routine tasks one may not always immediately recognize all necessary operations. It may take time or additional information to consider what to do without ever considering the situation as a problem. This challenges the clarification of the term 'problem.' Hence, in addition to an initial 'state of not knowing,' we suggest that problems are also characterized by the affective response to this initial 'state of not knowing.' A tendency towards negative emotional responses then indicates a problem situation (i.e., a significant discrepancy between an actual and a desired state), whereas the absence of such an initial negative emotional state would indicate that goal achievement is either not significant enough (e.g., the goal can easily be abandoned) or not considered too challenging (e.g., the goal can easily be achieved). This perspective is also found in the problem definition by Jonassen and Hung (2012) who suggest two critical attributes of a problem, namely the existence of an unknown and the need to determine the unknown. Thus, experiencing negative emotions indicates that an individual really cares about solving the problem (Op't Eynde et al. 2006) or finding the unknown, respectively. Furthermore, the problem solver might (and should) try to actively down-regulate such tendencies towards negative emotional responses (Dörner and Wearing 1995; Funke 2012; Funke et al. in print). The effects of emotions on achievement behavior are ambiguous (Carver and Scheier 2014). Positive moods at a medium level of activation were found to facilitate adequate, planned, and reflective problem-solving behaviours in a study by Reither and Stäudel (1985), whereas negative emotions increased the tendency to avoid a problem by shifting attention to easier tasks (Schwarz and Bless 1991). However, Sperring et al. (2005) and Barth and Funke (2010) showed that negative feedback from the problem environment triggered negative affect which in turn might enhance problem solving. Still, from a perspective of emotion regulation (Gross 1998), it is important to regulate these negative emotions even if they represent valuable feedback on the progress of problem solving (Hannula 2015).

In line with Weinert (2001) the attribution of competence should be based on dealing with complex situations. The complexity of a problem situation is defined by the number and interconnectedness of variables, number of conflicting goals, lack of transparency, self-reinforcing tendencies and time pressure (Dörner 1996; Funke 2003). With regard to dynamics, Leutner et al. (2005) distinguish dynamic problem solving from analytic problem solving. Dynamic problems require exploration by means of manipulating variables, observing effects, and drawing conclusions. The MicroDYN approach is the most common psychometric instrument for dynamic problem solving and was also applied in PISA. The participants explore linear systems, usually consisting of three independent

variables and three dependent variables, by manipulating the independent variables and enter their insight in a causal diagram (i.e. knowledge acquisition). Afterwards they have to manipulate the independent variables to achieve a given array of target values (i.e. knowledge application). The participants are usually confronted with seven to nine tasks, each lasting about a maximum of 5 min (Greiff et al. 2013a; Schoppek and Fischer 2015). In contrast, our own approach builds on analytic problem solving, in which relevant information is presented or can be derived by deductive reasoning (Leutner et al. 2005), which also resembles information problem solving as, for instance, referred to by Brand-Gruwel et al. (2009). We do not follow Leutner's et al. (2005) opinion that only dynamic problems are complex problems. Likewise, Schoppek and Fischer (2015) argue that problems within the MicroDYN approach lack many of the above characteristics of complex problems. Furthermore, analytic problems can apparently possess all further features of complex problems, too.² In addition, we argue that the degree of complexity of a problem is to some extent subjectively perceived and may also vary frequently while working on the problem. Any attempts to objectively predefine the complexity of problems have to be based on the anticipation of a target group's problem-solving competence.

Following Fischer and Neubert (2015) we consider problem-solving competence as a combination of knowledge, skills, abilities, and other components ('KSAO approach') rather than a single ability as within the MicroDYN approach (Greiff et al. 2013a).³ In the context of problem solving, domain-specific knowledge refers to declarative, procedural, conditional and other types of knowledge, which is relevant in problem situations within a particular domain and thus, domain-specific (Ackerman 2000; Nokes et al. 2011; Woolfolk 2005). Thus, by including knowledge in the definition of problem-solving the construct becomes domain-specific. Although domain-specific knowledge plays an important role, problem solving is also enhanced by non-cognitive factors such as self-confidence, perseverance, motivation, interest, frustration tolerance and the like (Frensch and Funke 1995; Schoppek and Fischer 2015; Sugrue 1995). Similarly, Kanfer and Ackerman (2005) consider knowledge, skills and abilities, motivation, personality, and self-concept as components of work competence. In summary, we follow Herl et al. (1999, p. 2) who state that in order '... to be a successful problem solver, one must know something (content knowledge), possess intellectual tricks (problem-solving strategies), be able to plan and monitor one's progress towards solving the problem (metacognition), and be motivated to perform (effort and self-efficacy).'

Based on extensive literature research, we developed a model of domain-specific problem-solving competence (for more information concerning the development of the competence model see Rausch and Wuttke 2016) that comprises 13 facets of competence, which are assigned to four components—(1) knowledge application,⁴ (2) metacognition, (3) self-concept, and (4) interest—and aligned along an ideal problem-solving process, whilst recognizing that complex problem solving is rarely a linear process (Fig. 1).

² Hence, we resist equating complex problem-solving competence with performance derived from working on MicroDYN items.

³ Although the MicroDYN approach distinguishes between knowledge acquisition and knowledge application, the two components are highly correlated ($r = .74$; Greiff et al. 2013b).

⁴ We use the term 'knowledge application' in a broad sense which implies both, knowledge acquisition and knowledge application in the sense of Greiff et al. (2013a).

Four components of competence	Thirteen facets of competence			
(A) Knowledge application (Cognition)	Identifying needs for action and information gaps	Processing information	Coming to well-founded decisions	Communicating decisions appropriately
(B) Action regulation (Metacognition)	Planned (well-structured) action	Persistence (focused action)	Retrospective action control	
(C) Self-concept (Expectancies)	Situational confidence in one's competence	Ambiguity / Uncertainty tolerance	Situational confidence in one's solution	
(D) Interests (Valences)	Personal interest in the problem context / content	Maintaining positive and active emotional states	Interest in the progress of / in learning from the problem	

Fig. 1 Model of domain-specific problem-solving competence (Rausch and Wuttke 2016, p. 177)

Furthermore, we refer to the facets of the first two components (A and B in Fig. 1) as cognitive facets and to the facets of the last two components (C and D) as non-cognitive facets. While cognition usually refers to ‘cold’ information processing (Collins and Smith 1994), quite often the term non-cognitive is a ‘residual category’ (Funke et al. in print, p. 8) and ‘comes by default to describe everything else’ (Duckworth and Yeager 2015, p. 238). We follow this distinction between cognitive and non-cognitive facets whilst recognising that many constructs such as self-concept imply both, cognitive and non-cognitive processes.

In contrast to generic dispositions (e.g. intelligence), competence is considered to be domain-specific. People are usually more competent in one domain while being less competent in others (e.g., accounting, baseball, chess). Following Weinert (2001), the underlying constructs of competence in different domains are comparable, although the performance differs substantially between the different domains. Although the performances of preparing a tender letter or setting up a CNC machine are very different, a high self-concept in the respective domain usually enhances one’s performance. Hence, the proposed competence model is not restricted to one domain, but can be easily be adapted to other domains; still it is domain-specific as opposed to domain-general approaches of problem-solving competence. However some of the facets in Fig. 1 might be more domain-specific while others might be more general since different components of problem-solving competence vary in their degree of generalizability (Fischer and Neubert 2015; Funke et al. in print). The 13 competence facets facilitate the development of problem scenarios to measure domain-specific problem-solving competence.

Development of authentic problem scenarios in the domain of controlling

A valid measurement of domain-specific competence builds on the requirements of a particular domain, that is the bundle of tasks that one is expected to solve. Our research focuses on the problem-solving competence of industrial clerks, which is the fifth most common of 328 state-recognized vocational training programs in the German dual

system of VET.⁵ Certified industrial clerks usually work in back-office departments of industrial or service companies. Thus, the qualification is roughly comparable to a Bachelor's degree in business administration. Further education and professional development can lead to lower or middle management positions. Although routine tasks are still an important part of office work, many of those repetitive processes have become automated or outsourced in recent decades (Autor et al. 2003; Frey and Osborne 2013). Thus, employees in back offices of industrial and service companies are increasingly confronted with the remaining non-recurrent problem cases.

VET programs claim to prepare individuals for a broad range of workplace requirements. Consequently, vocational curricula comprise several domains. With regard to the vocational competences of industrial clerks we focused on 'operative controlling',⁶ which is an important part of the curriculum followed by apprentice industrial clerks, as well as being a relevant domain of business administration in general. Further insight was derived from the content analyses of vocational training regulations, textbooks, a survey of workplace demands on employees in controlling departments (in cooperation with the European Competence Center for Applied Research on Medium-Sized Enterprises at the University of Bamberg/Germany; Becker et al. 2012), a diary study on problem solving in office work (Rausch et al. 2015) and an interview study on typical tasks and requirements in the domain of controlling with teachers, workplace trainers, VET students, and employees in the domain of controlling (Eigenmann et al. 2015). The findings from these domain analyses form the basis for the development of authentic problem scenarios. Appendix (Table 6) gives an overview of the studies and main findings during the phase of domain analyses.

To ensure authenticity, all problem scenarios are embedded in a model company, which is based on a real-life medium-sized bicycle manufacturer. We developed three complex and authentic problem scenarios, each of which demands for various steps of researching, evaluating and processing information, decision making, and communicating a proposed solution within 30 min. The built-in complexity of the scenarios was designed with regard to typical characteristics of complex problems (see above) and in anticipation of the target group's professional knowledge and problem-solving competence (based on our domain analysis). Scenario 1 requires a deviation analysis of budget and actual costs. The participants have to calculate budget costs, absolute and relative deviations in a spreadsheet application, identify relevant deviations, investigate the diverse reasons of these deviations in a large number of business documents, and propose adjustments for future budgeting in an email to their supervisor. In scenario 2 the participants must carry out a supplier selection by calculating acquisition prices and applying a value analysis, and scenario 3 concerns a make-or-buy decision. Besides a variety of scenario-specific business documents of various types (invoices, letters, bids, notes, etc.), a comprehensive archive containing short explanations of relevant and

⁵ Vocational education and training (VET) is a highly significant education sector in Germany; there are approximately as many new training contracts in VET as there are first-year students in higher education each year. Apprenticeship programs within the German dual VET system usually take three years and are characterized by a combination of workplace learning in the training company and classroom-based learning in state-run vocational schools.

⁶ On the level of operative controlling, typical activities concern the supply of information to support managerial decisions, cost planning, cost control, cost accounting and periodic reporting.

irrelevant technical terms, which constitutes an ‘open-book testing’, is available. As with real-life problem solving, the participants can look up information that they do not know by heart but—of course—none of the documents within the test environment provides a complete solution to the problem scenario. Furthermore—just as in real life—many documents provide irrelevant, conflicting and misleading information. In addition, in two of three scenarios the participants receive an email with distracting information (e.g., a listing of income of industrial clerks in different regions of Germany), which is also irrelevant for the problem but may be tempting to read. The participants cannot consult information outside of the software environment.

The problem scenarios allow for an ecologically valid assessment of domain-specific problem-solving competence with respect to curricular requirements, workplace requirements and authentic problem presentation. Scenarios are specified as a set of XML-files, which can be implemented into the computer-based office simulation with a minimum of programming expertise.

Computer-based office simulation

The participants register with the software using a predefined password, choose a last name from a given list and enter a first name, by which they are addressed during the following scenarios. The model company is then introduced via a slideshow with short subtitles. The slideshow is followed by a tutorial introducing the participants to the features of our custom-built office simulation Technology-Based Domain-Specific Learning Assessment (TeBaDoSLA). The tutorial is highly structured and ensures that all participants master the relevant features of the software. The software provides the typical features of an office environment such as a file system with hierarchical folder structure, a file-viewer, an email client, a calculator, a notepad and a clock that shows the remaining time for 3 s when clicked on. The core of the office simulation is a spreadsheet application. It provides most of the common functions of standard software such as Microsoft® Excel®. Altogether, an authentic task environment for the holistic processing of the problem scenarios without any artificial fragmentation was designed. Thus, not only the problem scenarios but also the open problem space (i.e., the entirety of possible system states and available operators; Newell and Simon (1972); see also ‘outcome space’, Wilson et al. 2012) were developed with regard to ecological validity. Figure 2 shows a screenshot of the office simulation software.

The test environment records each valid mouse click and keystroke with time stamps. The resulting log-file data enable detailed process analyses which are designated to reveal the metacognitive strategies of the participants. However, log-file analyses are not part of the current paper; instead we focus on the components A, C and D of our competence model (see Fig. 1).

Implementation of EES

Although non-cognitive facets of problem-solving competence are prevalent in contemporary theoretical modeling, they are often neglected in measurement approaches. Focusing on only cognitive variables is often legitimated with reference to Weinert, who suggested analyzing cognitive and non-cognitive facets separately (Klieme and Leutner 2006, p. 880; Klieme et al. 2008, p. 9). From our perspective, disregarding non-cognitive

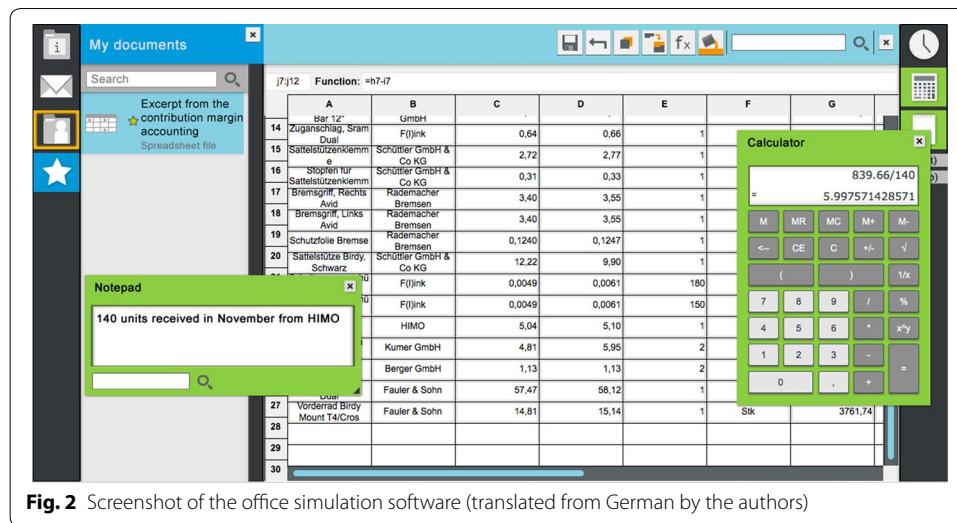


Fig. 2 Screenshot of the office simulation software (translated from German by the authors)

facets does not do justice to Weinert's approach, since he claimed that '... it would not be useful to restrict attention to cognitive and metacognitive competencies if one is concerned with success in broad fields of action across a variety of tasks (e.g., in school, in social institutions, or in a profession)' (Weinert 2001, p. 61). However, if non-cognitive facets are measured at all, the method of choice is usually self-report questionnaires. This poses methodological problems: while the tasks to be solved are highly concrete and embedded in a certain context, questionnaires on non-cognitive facets such as domain-specific self-concept or interests are usually phrased very universally. The use of different methods—task-specific performance vs. universal self-reports—leads to weak empirical relationships between cognitive and non-cognitive facets, which are often misinterpreted as a low impact of non-cognitive variables (Dermitzaki et al. 2009; Sembill et al. 2013). In a pilot study of our project with 100 VET students, no significant correlations were found between the cognitive component of domain-specific problem solving and neither work-related self-efficacy ($p = .188$; n.s.) nor vocational interest ($p = -.026$; n.s.). While the cognitive component of domain-specific problem solving was measured on the basis of three complex scenarios (similar to our approach presented in this paper), the non-cognitive components were measured by universal self-report questionnaires (Rausch under revision). Wittmann and Süß (1999) refer to the 'Brunswik symmetry' (named after Brunswik 1952) as an explanation for such phenomena. The Brunswik symmetry suggests the every level of generality at the predictor side has its symmetrical level of generality at the criterion side. Maximum predictability can only be obtained when predictor and criteria are symmetrical (see also Ackerman and Beier 2006). This is apparently not the case when predicting specific task performance by very broad self-evaluated personality traits.

We developed an approach to measure non-cognitive facets of competence during problem solving—referred to as EES. 'Embedded Experience Sampling' (EES) builds on the 'Experience Sampling Method' (ESM) introduced by Csikszentmihalyi and colleagues (Hektner et al. 2007) and similar methods of data-collecting 'in situ' such as the 'Continuous State Sampling Method (CSSM)' introduced by Sembill and colleagues

(2002). Test-takers are requested to stop at certain times during the test and spontaneously answer short prompts (EES items) regarding their actual experience of the problem situation. These EES events are embedded into the problem situation in a way that resembles common social interaction in the workplace. In doing so, we aim to reduce the artificiality of otherwise isolated questions that are usually administered in supplemental questionnaires. Closed-ended questions were used in order to spare the test-takers the time and effort they would need to write down their answers. Furthermore, closed-ended prompts improve the comparability of the answers and facilitate EES in large-scale assessments. Thus, a participant's answer is largely pre-specified (e.g., 'Hi Julian, that's very nice of you. At the moment, I feel ...'). The EES items are rated on a Likert-scale (e.g., from 1 = *not nervous at all* to 4 = *very nervous*). EES focuses on non-cognitive constructs such as interest, attitudes, commitment, self-concept and so on that are not possible to observe or infer otherwise. In our research, EES serves to measure the non-cognitive facets of competence presented in Fig. 1 (components C and D). The EES events are integrated into the office simulation. EES events pop up at predefined times during the problem scenarios. Figure 3 shows the screenshot of an EES event within the office simulation. Participants need to answer four closed EES items before they can get back to the problem scenario.

From measuring non-cognitive facets within the problem-solving process, a better ecological validity than from administering more unspecific retrospective self-report questionnaires that are separated from context, is assumed. In addition, bias due to social desirability (Harley 2016) might decrease in EES compared to retrospective self-reports, due to the concurrent cognitive load and time pressure during the problem-solving process (Stodel 2015). In group discussions and one-to-one interviews, the participants of a pilot study reported that they liked the idea of the EES. They experienced the specified situations as quite realistic as those were occurrences that they encountered in their everyday working environment. Interestingly enough, they reported that they did not elaborate on what would be 'good answers' but instead answered spontaneously, as was requested.

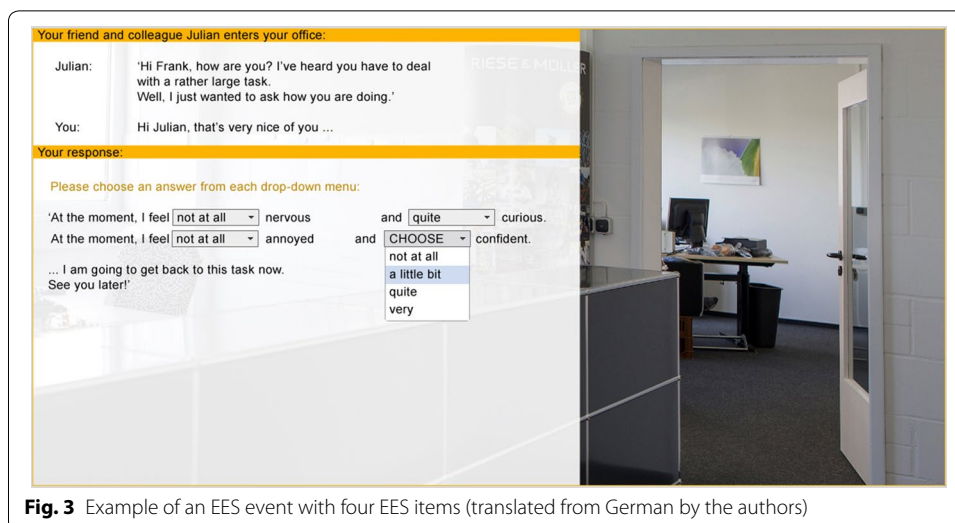


Fig. 3 Example of an EES event with four EES items (translated from German by the authors)

In PISA 2006, for instance, for the measurement of interest ‘in situ’ short ratings of interest in scientific domains were requested directly after particular test items in the field of science (Drechsel et al. 2011). However, these items were not embedded into the ‘storyline’. An approach similar to ours is the ‘affect self-report device’ applied to the game-based learning environment ‘Crystal Island’. During their interaction with the learning environment, participants received an in-game prompt asking them to report on their cognitive and emotional states. These status updates were described as part of an in-game social network (Sabourin and Lester 2014). Another example is the ‘Belief Meter’ within the computer-based learning environment ‘BioWorld’. Medical students report their confidence in their final diagnosis as a percentage (0–100 %) on the ‘Belief Meter’ during problem solving (Jarrell et al. 2016). However these in-game self-reports were not designed to assess facets of competence. Aside from these recent and inspiring works, we did not find more similar approaches.

Research questions

In the empirical section the focus lies on the reliability of the assessment. First, we analyze whether the above approach allows for a reliable measurement of the cognitive facets in the competence component *knowledge application*. Furthermore, we analyze whether the above approach allows for a reliable measurement of the non-cognitive facets in the competence components *self-concept* and *interest*.

While the scenarios were developed with respect to industrial clerks (IC), they were also administered to IT-systems management assistants (ITMA) and merchants in wholesale and foreign trade (MWFT). Their apprenticeship programs are similar to that of industrial clerks. However, the domain addressed in the problem scenarios (‘controlling’, see above) is of less significance in the curricula of ITMA and MWFT apprentices. Given a valid assessment of domain-specific competence, IC apprentices are supposed to outperform the comparative samples. This was also confirmed in a previous pilot study (Wuttke et al. 2015).

Methods

Sample

The main study took place between April and September 2014. The sample was approached via vocational schools but participation was voluntary both on the school level and on the individual level of each student. A total of 786 VET students from various German federal states participated in the study, of which six were excluded from the analyses due to missing data (due to either lack of willingness or technical malfunction of the test software). All of the remaining 780 participants (50.1 % female) were in the second or third year of a 3-year commercial apprenticeship program and showed a typical right skewed age distribution ($M = 21.3$ years; $SD = 2.69$; $min = 17$; $max = 44$). Of the total sample, 537 were enrolled in an apprenticeship program to become industrial clerks (IC), 106 were apprentice IT-systems management assistants (ITMA), and another 137 were apprentice merchants in wholesale and foreign trade (MWFT).

Procedure

All data were collected in computer-equipped classrooms in vocational schools. At the beginning of the data collection sessions the researchers introduced the project, and the agenda. They also provided information about anonymity, data protection, and ethical factors and emphasized that participation was voluntary. All participants provided written, informed consent before completing any of the assessments. Before and after the problem scenarios, the participants completed several self-report questionnaires including scales on vocational interest, work-related self-concept, and several antecedents of apprenticeship success (Baethge-Kinsky et al. 2016) as well as further tests of general cognitive ability (German version of Cattell's Culture Fair Test developed by Weiss 2006), domain-specific content knowledge (based on test items from final exams), literacy and numeracy (Ziegler et al. 2016). Nevertheless, these instruments are not in the centre of attention in this paper.

When participants registered in the computer-based office simulation, they were introduced to the underlying model company and the features of the software, before working on the three problem scenarios. Each problem scenario was followed by a short questionnaire intended to assess test motivation, self-assessed quality of the problem-solving process, self-assessed quality of the proposed solution and so forth. Altogether, the procedure lasted 5 h. In the following, we focus on the internal consistency and internal validity of the assessment of domain-specific problem-solving competence.

Results

Reliability of the cognitive facets measured by content analyses

By providing a very open problem space we aimed at ecological validity, as the given problems were designed like real-life scenarios without clear instructions. In the end, the estimation of competence scores for each facet is based on only three stimuli. However, scoring such complex and open-ended responses is laborious—especially in large-scale assessment—and may also impair the reliability of the assessment (Wilson 2008). The scoring process was carried out in three steps as, for instance, suggested by Bennett et al. (2003) in the context of assessing problem solving in technology-rich environments (TRE) within the National Assessment of Educational Progress (NAEP) in the United States. The three steps comprise two levels of coding followed by an IRT analysis.

1. On the first level of coding, the participants' solutions were analyzed on the basis of fine-grained category systems according to the qualitative content analysis approach by Mayring (2014). Graduate students were trained to rate the categories. They used an additional software ('Rating Suite') to display the participants' solutions and rated them according to the coding guide. The coding guide provided definitions, coding rules and examples for the coding of each category in each scenario. The categories were designed against the background of domain-specific quality standards which were identified during the domain analysis (see Appendix Table 6). Some categories were identical for all three scenarios (e.g., all categories in facet 4 'communicating the decision appropriately') while most of them were scenario-specific (e.g., coding which of the relevant documents were used). Altogether, the category systems for the three scenarios comprised 97 categories (22 for the first, 34 for the second and 41 for

Table 1 Hierarchical decomposition of the facets of knowledge application

Competence facet	Decomposition of the competence facet
A1: Identifying needs for action and information gaps	1.1 Identifying needs for action 1.2 Identifying sources of information
A2: Processing information	2.1 Applying algorithms (e.g., calculating procedures) 2.2 Using domain-specific tools (e.g., functions in a spreadsheet software)
A3: Coming to a well-founded decision	3.1 Quality of arguments 3.2.1 Reflection of one’s own solution 3.2.2 Reflection of applied methods 3.2.3 Reflection of side and long-term effects
A4: Communicating the decision appropriately	4.1 Using domain-specific language (e.g., technical terms) 4.2 Meeting communication standards (e.g., complimentary close in an email) 4.3 Structuring one’s statement (e.g., paragraphs) 4.4 Meeting formal standards (orthography and punctuation)

the third scenario), each of which corresponded to an item (in our case we denote these as level-one-items) and was assigned to one of the four facets of *knowledge application* (see Fig. 1). Table 1 shows the hierarchical decomposition (top down).

Human raters assessed, for instance, the quality of arguments (category 3.1; Table 1) but many level-one-items were scored automatically on the basis of log-files, for instance whether relevant documents were found (category 1.2) and many of the calculations in the spreadsheet (category 2.1). Altogether, automated and human rating resulted in 97 mostly dichotomous level-one-items across the three scenarios. For each item, a higher value indicates a higher quality of the solution. Dual coding enabled an enhancement of the coding guide and the training of the raters based on the inter-rater-reliability for each item.

- On the second level of our two-level coding we aggregated the 97 level-one-items from the fine-grained coding process into one partial credit item for each competence facet and each scenario ($4 \times 3 = 12$ partial credit items—which we denote as level-two-items). For this purpose, the response patterns in the level-one-items of one competence facet and one scenario were extracted and ordered by the sum score of the items. Thus, a low sum score is a first indicator of a low quality of the solution. Subsequently, experts rated each response pattern with regard to the quality of the solution as compared to other response patterns. Experts not only decided on cutoff values between lower and higher partial credits but they also defined weightings or necessary preconditions with regard to the content of the problem scenario. Assigning credit points to each response pattern resulted in one (level 2) partial credit item per facet and per scenario, each of which had four to seven categories. Thus, the estimation of competence scores for each competence facet is based on only three items. Nevertheless, these partial credit items provide rich information (e.g., 3 partial credit items with 5 categories each equal 12 dichotomous items). Besides the strong qualitative verification of the dimensionality that comes along with the assignment of the

partial credits, the main reason to include just one item per scenario in the IRT analysis is to avoid local item dependence (LID). A major problem in evaluating complex scenarios is the strong local dependence of the items that refer to the same scenario, and the corresponding LID is known to bias the reliability, item difficulty estimates, as well as variance and covariance estimates, as has been shown by many authors (see, e.g., Brandt 2012; Sireci et al. 1991; Wainer et al. 2007; Yen 1993). A further option to consider the scenario-based LID might have been to model the observed dependencies, for example, via a hierarchical model such as the Rasch testlet model (Wang and Wilson 2005). The given covariance structure of the testlet specific factors, however, typically is not as proposed by the model (which supposes that they are uncorrelated), and furthermore, the covariances can change depending on the considered (sub-)sample. Such changes then lead to changes for the calculation of the general factor, making the calculation of the latter sample dependent. This also the reason why these models are not used in the known large scale assessments, such as the Programme for International Student Assessment (PISA) or the National Assessment of Educational Progress (NAEP). We therefore also preferred an approach that avoids LID by the design of the underlying items instead of an approach based on modeling.

3. In a third step, the four cognitive competence facets of the competence component *knowledge application* were, at first, analyzed separately in order to investigate the fit of the constructed partial-credit items. For all item as well as step parameters the calculated in-fit values ranged between .95 and 1.05, that is the items show good fit.

Thereafter, the test was analyzed using a four-dimensional partial credit model (Masters 1982) including background information such as gender, age, vocation, intelligence, the answer data from the non-cognitive facets, and other relevant variables. All calculations were conducted using the R package TAM (Kiefer et al. 2015). Table 2 shows the EAP/PV reliabilities (on the diagonal) and latent correlations between the competence facets.

The EAP/PV reliabilities of the four cognitive facets are satisfactory; compared to the pilot study (Wuttke et al. 2015) they increased considerably. The latent correlations between the facets are medium on average and reflect the multidimensionality of the competence component ‘knowledge application’. The multidimensionality of the construct is further supported by the comparison of the likelihoods of the unidimensional model and the multidimensional model, respectively. While the unidimensional shows a deviance (equals $-2 \text{ Log-Likelihood}$) of 16,178.8, the multidimensional model shows a

Table 2 EAP/PV reliabilities and latent correlations of the facets of knowledge application

	(A1)	(A2)	(A3)	(A4)
(A1) Identifying needs for action and information gaps	.84			
(A2) Processing information	.64	.79		
(A3) Coming to a well-founded decision	.61	.31	.82	
(A4) Communicating the decision appropriately	.43	.15	.51	.78

deviance of 16,058.6, which results in a significant Chi Square test ($df = 9$) as well as AIC and BIC model fit values (16,283 vs. 16,181 and 16,504 vs. 16,440, respectively) in favor of the multidimensional model.

Reliability of the non-cognitive facets measured by EES

For an integrated measurement of non-cognitive facets, embedded experience sampling (EES; see above) was used. Appendix (Table 7) provides an overview of the EES events, the respective competence facets (see Fig. 1), and the EES items, which were the same for all three problem scenarios.

Initially a six-dimensional partial credit model (Masters 1982) including all non-cognitive facets was calibrated. Facet D3 (*Interest in the progress of/in learning from the problem*), however, showed insufficient reliability (EAP/PV reliability = .30) and was excluded. The final estimation therefore only included five dimensions and was estimated using various variables as background information (compare estimation of the cognitive facets above). Table 3 shows the EAP/PV reliabilities (on the diagonal) and latent correlations between the five remaining non-cognitive competence facets.

The EAP/PV reliabilities of the five non-cognitive facets are satisfactory. The latent correlations between the non-cognitive facets are slightly higher than for the cognitive facets, they can still be considered as moderate though with only one correlation being larger than .70 (between facet C1 and C3). For a conference paper focused on the EES approach, we also calculated the correlations between the non-cognitive facets as measured by EES and similar constructs measured by universal questionnaires (Rausch et al. 2016). We only found only small correlations between both the facets in component C and work-related self-efficacy ($.18 < r < .27$) and the facets of component D and vocational interest ($.10 < r < .25$).

Table 3 EAP/PV reliabilities and latent correlations of the non-cognitive facets

	(C1)	(C2)	(C3)	(D1)	(D2)
(C1) Situational confidence in one's competence	.85				
(C2) Ambiguity/uncertainty tolerance	.57	.77			
(C3) Situational confidence in one's solution	.72	.46	.84		
(D1) Personal interest in the problem context/content	.57	.62	.38	.80	
(D2) Maintaining positive and active emotional states	.51	.39	.45	.45	.78

Table 4 Correlations between cognitive and non-cognitive facets of competence

	(A1) Identifying	(A2) Processing	(A3) Deciding	(A4) Communicating
(C1) Situational confidence in one's competence	.40	.34	.35	.25
(C2) Ambiguity/uncertainty tolerance	.38	.29	.25	.20
(C3) Situational confidence in one's solution	.43	.37	.45	.27
(D1) Personal interest in the problem context/content	.30	.21	.22	.16
(D2) Maintaining positive and active emotional states	.23	.21	.20	.15

Correlations between cognitive and non-cognitive facets

Table 4 shows the latent correlations between the cognitive facets and the non-cognitive facets based on the plausible values (the answer data from the cognitive data was included as background information in the estimation of the model for the non-cognitive data and vice versa; these via plausible values calculated correlations are therefore also latent correlations).

In general, the correlations are all positive and of small to medium size. However, the correlations also show certain tendencies considering the relationship of the dimensions. For facet A4 ('communicating the decision appropriately') and facet C1 the correlation is significantly smaller (according to the Fisher *r*-to-*z* transformation) than for A1, A2, and A3 with C1; the same holds for facet C3. For the remaining three non-cognitive facets the differences in the correlations are not (statistically) significant, however, all values show the same tendency. Averaging across the non-cognitive facets the correlation of these with A4 is also significantly smaller than with A1, A2, and A3. For the correlations between the facets of domain-specific self-concept (C1, C2, and C3) and the facets of knowledge application (A1 through A4) a similar tendency can be observed. While not all correlations between self-concept and knowledge application are significantly larger than the correlations of domain-specific interest (D1 and D2) and knowledge application, averaging across the corresponding correlations again results in significantly smaller relationships between the facets of interest and knowledge application than between the facets of self-concept and knowledge application. When checking the correlations in the subgroups of different vocations, we found that in the subgroup of merchants in wholesale and foreign trade the correlations between A4 (communicating the decision appropriately) and the non-cognitive facets were smaller (some of them zero). Possible explanations will be discussed below.

Differences between VET students of different vocations

In a first step, a differential item functioning (DIF) analysis was conducted in order to investigate whether the test included items that were particularly unfair for one of the vocations. Using the R package TAM again facet models were calibrated, which yielded the differences in the item difficulties for each of the three groups. The size of DIF effects is typically categorized into three different categories (Zieky 1993, 2003):

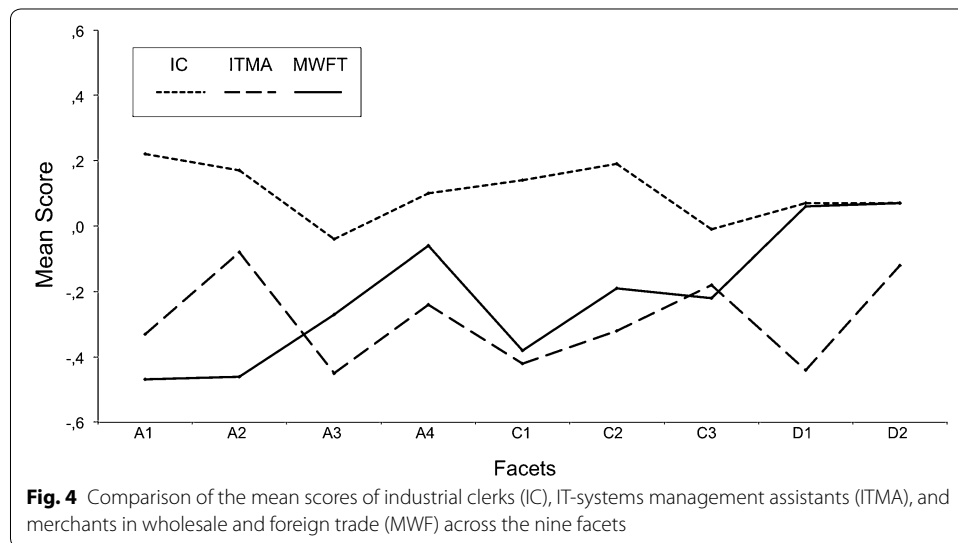
- Negligible effect: $<.43$ Logits
- Light to moderate effect: $\geq .43$ and $<.63$ Logits
- Moderate to large effect: $\geq .63$ Logits

All of the DIF effects for the items of facet A1, A2, A4, C2, C3, D1, D2 were negligible, only facet A3 had two items with light effects, and facet C1 had one item with a light effect. Due to the small effect sizes, we decided to nevertheless include them in the comparison of the groups. In the second step, the competences of the three training vocations were compared. Figure 4 graphically displays these results, and Table 5 gives more details on results particularly considering the significance of the differences. All calculations were based on plausible values (Fig. 4).

Table 5 Group differences for competence facets between IC and ITMA and between IC and MWFT

	IC (n = 523)		ITMA (n = 106)		t	p	MWFT (n = 138)		t	p
	M	SD	M	SD			M	SD		
(A1)	.22	.04	-.33	.09	-5.68	.000	-.47	.07	-8.33	.000
(A2)	.17	.04	-.08	.09	-2.59	.010	-.46	.06	-8.59	.000
(A3)	-.04	.03	-.45	.08	-4.96	.000	-.27	.07	-3.33	.001
(A4)	.10	.04	-.24	.08	-3.61	.000	-.06	.08	-1.8	.072
(C1)	.14	.05	-.42	.12	-4.46	.000	-.38	.10	-4.78	.000
(C2)	.19	.04	-.32	.10	-4.71	.069	-.19	.08	-4.18	.000
(C3)	-.01	.04	-.18	.09	-1.82	.069	-.22	.08	-2.42	.016
(D1)	.07	.05	-.44	.11	-4.02	.000	.06	.10	-.04	.970
(D2)	.07	.02	-.12	.06	-3.08	.002	.07	.05	.05	.960

IC industrial clerks, ITMA IT-systems management assistants, MWFT merchants in wholesale and foreign trade, (A1) identifying needs for action and information gaps, (A2) processing information, (A3) coming to a well-founded decision, (A4) communicating the decision appropriately, (C1) situational confidence in one's competence, (C2) ambiguity/uncertainty tolerance, (C3) situational confidence in one's solution, (D1) personal interest in the problem context/content, (D2) maintaining positive and active emotional states



As hypothesized, the VET students in an apprenticeship program to become industrial clerks outperform the comparison groups. However, only small to medium effect sizes were found. The largest effects were found for the cognitive facets A1 ‘Identifying needs for action and information gaps’ and A2 ‘Processing information’. We regard these differences as an indicator of curricular validity of our assessment, which was developed to primarily meet the curricular requirements of industrial clerks (IC). The domain of controlling and thus, the contents of our problem scenarios, are part of the curricula of IT-system management assistants (ITMA) and merchants in wholesale and foreign trade (MWFT), too, but play a minor role.

Conclusions

In this paper, a computer-based assessment of domain-specific problem-solving competence in the field of commercial vocational education and training was presented. Based on a multi-faceted model of problem-solving competence (Rausch and Wuttke 2016), the development of the assessment focuses on ecological validity, which refers to the congruence between behaviors observed in test environments and real life, and content validity with regard to the competence which is actually required in practice. Therefore, authentic problem scenarios on the basis of extensive domain analyses (curricula analysis, textbook analysis, interview and diary studies, etc.; Eigenmann et al. 2015) were developed. Assumed differences in the performance of apprentice industrial clerks and comparative groups support the assumption of curricular validity of the three problem scenarios in the field of controlling.

We did not only develop authentic problem scenarios but also provided an open-ended problem space for working on these problems within an authentic office environment instead of applying highly structured items (e.g., multiple choice items). Expanding the problem space for the test takers (i.e. reducing experimental control) resulted in very heterogeneous behavior patterns and solutions. Nevertheless, statistical tests and indices based on item response theory demonstrate the reliability of the measurement of cognitive competence facets. We applied a three-step method (similar to Bennett et al. 2003): (1) Fine-grained results from a highly structured content analysis were condensed into (2) partial credit items on the basis of consensual expert judgments. (3) Finally, these partial credits were subject to psychometric scaling using a multidimensional Rasch model (a publication with a more detailed description of the procedure is in preparation).

Besides cognitive facets of problem-solving competence, we also consider non-cognitive facets of competence (e.g., self-concept, interest) to play a role in problem solving in the workplace. Therefore, content validity also calls for the measurement of these non-cognitive facets of problem-solving competence. However, we argue against the use of prevalent self-report questionnaires. Instead, we developed a method—EES—to measure non-cognitive facets of problem solving ‘in situ’. Test-takers are requested to stop at certain times and spontaneously answer short prompts (EES items) regarding their actual experience of the problem situation. Again, aiming at ecological validity, these EES events are embedded into the problem situation in a way that resembles common social interaction in the workplace. Statistical tests and indices based on item response theory demonstrate the reliability of the measurement of non-cognitive competence facets across the three problem scenarios. However, only five of the six non-cognitive facets could be measured reliably. Facet D3 (*Interest in the progress of/in learning from the problem*) showed a very low EAP/PV reliability and had to be excluded from the analysis. To our mind, this is due to our approach to ask for several competing activated motives (see Appendix Table 7), which did not work out as we anticipated.

The correlations between the four cognitive and the five remaining non-cognitive facets were all positive and showed moderate effect sizes. In a pilot study, we assessed the cognitive facets of problem-solving competence in a similar way as in the present study and found smaller (even zero) correlations with non-cognitive facets, which we then measured by universal self-report questionnaires (Rausch under revision). In the

present study, the correlations between the cognitive facet A4 (*communicating the decision appropriately*) and several of the non-cognitive facets were remarkably smaller in the subgroup of Merchants in wholesale and foreign trade (MWFT) than in the other subgroups. Although the MWFT performed poorer in the cognitive facets A1, A2 and A3, they still managed to produce an appropriate email reply with regard to domain-specific language, communication standards, structure and formal standards. Apparently, the non-cognitive facets such as self-concept are more linked to the 'core processes' of problem solving. Table 5 shows further interesting differences between the three training programmes that, due to lack of space, cannot be discussed in detail.

We want to emphasize that we did not model non-cognitive facets as mere explanatory or even confounding factors of the 'true cognitive competence' but as competence facets in their own right. Decomposing domain-specific problem-solving competence into various facets and, at the same time, providing an integrated measurement offers opportunities for a differentiated assessment of competence profiles and individualized interventions (Herl et al. 1999; Sugrue 1995). We also postulated metacognitive facets of problem solving, which as yet have not been addressed. We plan to identify metacognitive patterns on the basis of the log-files that are already available from the present study. Inspiring research on pattern recognition in log-files is available, for instance, for the game-based learning environments 'Crystal Island' (Sabourin et al. 2013) and 'Betty's Brain' (Biswas et al. 2014). A further limitation of our current approach is the absence of a social component of problem solving since cooperation and collaboration is a major way of solving work-related problems in real life (Rausch et al. 2015). It would be an exciting challenge to integrate cooperative and collaborative features into authentic problem scenarios and hence, into an authentic office simulation. Furthermore, the current degree of automated codings could be advanced in order to reduce the efforts of human coding. Finally, this would also increase the opportunities for dissemination into practice.

Authors' contributions

All authors made substantial contributions to conception and design of the study and the acquisition of data. SB and AR took responsibility of data analysis. All authors were involved in the interpretation and discussion of the results. While AR has undertaken the task of drafting the manuscript, all authors have been revising it critically and approved the final version to be published. All authors agree to be accountable for all aspects of the work. All authors read and approved the final manuscript.

Author details

¹ Business Education, University of Bamberg, Kapuzinerstrasse 16, 96047 Bamberg, Germany. ² University of Mannheim, Mannheim, Germany. ³ University of Frankfurt, Frankfurt, Germany. ⁴ Art of Reduction, Kiel, Germany.

Acknowledgements

This research was supported by grants from the German Federal Ministry of Education and Research (BMBF) under Grant Nos. 01DB081119-01DB1123. The authors declare that the funding had no influence on the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Competing interests

The authors declare that they have no competing interests.

Appendix

See Tables 6 and 7.

Table 6 Overview of studies and main findings during domain analyses

Study	Sample	Main findings
Content analysis of training regulations and curricula	Training regulations for industrial clerks (IC), merchants in wholesale and foreign trade (MWFT) and IT-systems management assistants (ITMA) Curricula from the federal states of Germany for the above training occupations	High significance of controlling throughout the curriculum of industrial clerks The curricula of MWFT and ITMA also include controlling contents but to a lesser extent
Content analysis of textbooks	Three standard textbooks for commercial training occupations in VET	Curricular significance of controlling contents is also represented in textbooks. However, strong emphasis on content knowledge rather than competence-orientation Majority of exercises offers only low complexity, instead clear action sequences and guidance is provided
Diary study on everyday problem solving in office work	n = 8 VET students n = 5 skilled employees in operative controlling	The majority of participants is confronted with problems during office work several times a week The occurrence of complex problems is mostly going along with time pressure and certain risks for the enterprise The emotional response to problems is particularly negative when individual is pressed for time The problems to be solved by apprentices are less complex High significance of problem solving for workplace learning and individual growth
Interview study on typical tasks and requirements in the field of operative controlling	n = 16 skilled employees in operative controlling n = 14 VET students (industrial clerks) n = 7 teachers in vocational schools	VET students are entrusted with operative rather than strategic controlling tasks Problem-solving competence is considered an important skill in the field of controlling High quality of solutions to problems includes correctness and plausibility of calculations as well as successfully dealing with time pressure and a well-targeted and clear communication of results High significance of spreadsheet processing and Enterprise Resource Planning (ERP) software

Table 7 Overview of EES events, competence facets, and EES items

EES event (point of time)	Competence facet (see Fig. 1)	EES items (translated from German and condensed)
Short email response after receipt of the task (after 3 min)	Situational confidence in one's competence (C1)	C1_1: Sender of the task requests a first quick estimation Answer from 1 = 'I do not know yet what to do here' to 4 = 'I know exactly what to do here'
	Personal interest in the problem context/content (D1)	D1: Sender of the task asks whether tasks like this are interesting to the apprentice Answer from 1 = 'Tasks like this are not interesting to me' to 4 = 'Tasks like this are very interesting to me'
Phone call of sender of the task (after 10 min)	Situational confidence in one's competence (C1)	C1_2: Sender of the tasks requests a further estimation Answer from 1 = 'I am afraid I will not be able to cope with it, but I will do my best' to 4 = 'I can definitely cope with it and give my best'
	Ambiguity/uncertainty tolerance (C2)	C2: Sender of the task asks whether the apprentice likes to work on comprehensive tasks like this Answer from 1 = 'I do not like to work on such comprehensive tasks' to 4 = 'I very much like to work on such comprehensive tasks'
Short visit of a colleague (after 20 min)	Maintaining positive and active emotional states (D2)	Friend enters the office asks how one is doing
		D2_1: from 1 = 'not at all' to 4 = 'very nervous' (–)
		D2_2: from 1 = 'not at all' to 4 = 'very curious'
		D2_3: from 1 = 'not at all' to 4 = 'very irritated' (–)
Short request from the sender of the task after receipt of the solution (after submission or after 30 min)	Situational confidence in one's solution (C3)	C3: Sender of the task asks how confident the apprentice is about her/his solution and whether the solution has to be checked before its implementation Answer from 1 = 'Unfortunately, I did not arrive at a solution at all' over 2 = 'I am afraid you should check everything in detail because I assume I made some mistakes' to 5 = 'I think, I found a proper solution that you do not have to check in detail, again'
	Interest in the progress of/in learning from the problem (D3)	Participants are to check two of the following statements for his email answer. 'Working on tasks like this, ... D3_1: ... I am always a bit anxious that I might not solve it.' (distractor) D3_2: ... I feel that I am accepted as a full team member'. (distractor) D3_3: ... I can always learn something interesting' D3_4: ... I have the opportunity to demonstrate my skills'. (distractor) D3_5: ... I am afraid to make a fool of myself if I fail'. (distractor) D3_6: ... I wish someone would explain to me how to make it better afterwards'

All prompts and items were the same for all of the three problem scenarios; (–) indicate inverse items

Received: 25 February 2016 Accepted: 11 July 2016

Published online: 19 July 2016

References

- Ackerman PL (2000) Domain-specific knowledge as the “dark matter” of adult intelligence: Gf/Gc, personality and interest correlates. *J Gerontol Psychol Sci* 55(2):69–84
- Ackerman PL, Beier ME (2006) Methods for studying the structure of expertise: psychometric approaches. In: Ericsson KA, Charness N, Feltovich PJ, Hoffman RR (eds) *The Cambridge handbook of expertise and expert performance*. Cambridge University Press, Cambridge, pp 147–165
- Autor DH, Levy F, Murnane RJ (2003) The skill content of recent technological change: an empirical exploration. *Quart J Econ* 118(4):1279–1333
- Baethge-Kinsky V, Baethge M, Lischewski J (2016) Bedingungen beruflicher Kompetenzentwicklung: institutionelle und individuelle Kontextfaktoren (SiKoFak) [Conditions for the development of vocational competencies: institutional and individual context factors (SiKoFak)]. In: Beck K, Landenberger M, Oser F (eds) *Technologiebasierte Kompetenzmessung in der beruflichen Bildung—Ergebnisse aus der BMBF-Förderinitiative ASCOT [Technology-based measurement of competencies in VET—findings from the BMBF research initiative ASCOT]*. Bertelsmann, Bielefeld, pp 265–299
- Barth CM, Funke J (2010) Negative affective environments improve complex solving performance. *Cogn Emot* 24(7):1259–1268
- Becker W, Ebner R, Brandt B, Holzmann R (2012) Anforderungen an den Controller [Demands on the controller]. *Bamberger Betriebswirtschaftliche Beiträge*, 185. Band. Otto-Friedrich-Universität Bamberg, Bamberg
- Bennett RE, Jenkins F, Persky H, Weiss A (2003) Assessing complex problem solving performances. *Assess Educ* 10(3):347–359
- Biswas G, Kinnebrew JS, Segedy JR (2014) Using a cognitive/metacognitive task model to analyze students learning behaviors. In: Schmorrow DD, Fidopiastis CM (eds) *Foundations of augmented cognition—advancing human performance and decision-making through adaptive systems*. Proceedings of the 8th international conference on AC 2014, Crete, Greece, 2014. Springer, Heidelberg, pp. 190–201
- Brand-Gruwel S, Wopereis I, Walraven A (2009) A descriptive model of information problem solving while using internet. *Comput Educ* 53:1207–1217
- Brandt S (2012) Robustness of multidimensional analyses against local item dependence. *Psychol Test Assess Model* 54:36–53
- Brunswik E (1952) *The conceptual framework of psychology*. University of Chicago Press, Chicago
- Carver CS, Scheier MF (2014) The experience of emotions during goal pursuit. In: Pekrun R, Linnenbrink-Garcia L (eds) *International handbook of emotions in education*. Routledge, New York, pp 56–72
- Collins A, Smith EE (1994) Cognitive science. In: Eysenck MW (ed) *The Blackwell dictionary of cognitive psychology*. Blackwell, Cambridge, pp 66–71
- Dermitzaki I, Leondari A, Goudas M (2009) Relations between young students’ strategic behaviours, domain-specific self-concept, and performance in a problem-solving situation. *Learn Instr* 19(2):144–157
- Dörner D (1987) Denken und Wollen. Ein systemtheoretischer Ansatz [Cognition and volition. A system-theoretical approach]. In: Heckhausen H, Gollwitzer PM, Weinert FE (eds) *Jenseits des Rubikon [Beyond the rubicon]*. Springer, Berlin, pp 238–250
- Dörner D (1996) *The logic of failure: recognizing and avoiding error in complex situations*. Perseus, New York
- Dörner D, Wearing A (1995) Complex problem solving: toward a (computersimulated) theory. In: Frensch PA, Funke J (eds) *Complex problem solving. The European perspective*. Lawrence Erlbaum, Hillsdale, pp 65–99
- Drechsel B, Carstensen CH, Prenzel M (2011) The role of content and context in PISA interest scales—a study of the embedded interest items in the PISA 2006 science assessment. *Int J Sci Educ* 33(1):73–95
- Duckworth AL, Yeager DS (2015) Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. *Educ Res* 44:237–251
- Duncker K (1945) On problem solving. *The American Psychological Association*, Washington
- Eigenmann R, Siegfried C, Kögler K, Egloffstein M (2015) Aufgaben angehender Industriekaufleute im Controlling: Ansätze zur Modellierung des Gegenstandsbereichs. [Prospective industrial clerks’ tasks in controlling: approaches to domain modeling]. *Zeitschrift für Berufs- und Wirtschaftspädagogik* 111:417–436
- Fischer A, Neubert JC (2015) The multiple faces of complex problems: a model of problem solving competency and its implications for training and assessment. *J Dyn Decis Mak* 1:1–14
- Frensch PA, Funke J (1995) Definitions, traditions, and a general framework for understanding complex problem solving. In: Frensch PA, Funke J (eds) *Complex problem solving. The European perspective*. Lawrence Erlbaum, Hillsdale, pp 3–25
- Frey CF, Osborne MA (2013) The future of employment: How susceptible are jobs to computerisation? http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf
- Funke J (2003) *Problemlösendes Denken [Problem-solving thinking]*. Kohlhammer, Stuttgart
- Funke J (2012) Complex problem solving. In: Seel NM (ed) *Encyclopedia of the sciences of learning*. Springer, New York, pp 682–685
- Funke J, Fischer A, Holt D (in print) Competencies for complexity: problem solving in the 21st century. In: Care E, Griffin P, Wilson M (eds) *Assessment and teaching of 21st century skills (volume 3)*. Springer, New York
- Greiff S, Wüstenberg S, Holt DV, Goldhammer F, Funke J (2013a) Computer-based assessment of complex problem solving: concept, implementation, and application. *Educ Tech Res Dev* 61:407–421
- Greiff S, Wüstenberg S, Molnár G, Fischer A, Funke J, Csapó B (2013b) Complex problem solving in educational contexts—Something beyond g: concept, assessment, measurement invariance, and construct validity. *J Educ Psychol* 105(2):364–379

- Gross J (1998) The emerging field of emotion regulation: an integrative review. *Rev Gen Psychol* 2:271–299
- Hannula MS (2015) Emotions in problem solving. In: Cho SJ (ed) Selected regular lectures from the 12th international congress on mathematical education. Springer, New York, pp 269–288
- Harley JM (2016) Measuring emotions: a survey of cutting edge methodologies used in computer-based learning environment research. In: Tettegah SY, Gartmeier M (eds) Emotions, technologies, design, and learning. Elsevier, Amsterdam, pp 89–114
- Hektner JM, Schmidt JA, Csikszentmihalyi M (2007) Experience sampling method—measuring the quality of everyday life. Sage, Thousand Oaks
- Herl HE, O'Neil HF Jr, Chung GK, Bianchi C, Wang S, Mayer R, Lee CY, Choi A, Suen T, Tu A (1999) Final report for validation of problem-solving measures. Technical report No. 501 at the Center for the Study of Evaluation (CSE), National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education & Information Studies, University of California, Los Angeles
- Jarrell A, Harley JM, Lajoie SP (2016) The link between achievement emotions, appraisals, and task performance: pedagogical considerations for emotions in CBEs. *J Comput Educ*. doi:10.1007/s40692-016-0064-3
- Jonassen DH, Hung W (2012) Problem solving. In: Seel NM (ed) Encyclopedia of the sciences of learning. Springer, New York, pp 2680–2683
- Kanfer R, Ackermann PL (2005) Work competence. A person-oriented perspective. In: Elliot AJ, Dweck CS (eds) Handbook of competence and motivation. Guilford Press, New York, pp 336–353
- Kiefer T, Robitzsch A, Wu M (2015) TAM: test analysis modules (version 1.3) [R]. <http://www.cran.r-project.org/package=TAM>
- Klieme E, Leutner D (2006) Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen—Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG [Competence models for assessing individual learning outcomes and evaluating educational processes—Description of a recently approved DFG priority program]. *Zeitschrift für Pädagogik* 52(6):876–903
- Klieme E, Hartig J, Rauch D (2008) The concept of competence in educational contexts. In: Hartig J, Klieme E, Leutner D (eds) Assessment of competencies in educational contexts. Hogrefe, Göttingen, pp 3–22
- Leutner D, Funke J, Klieme E, Wirth J (2005) Problemlösekompetenz als fächerübergreifende Kompetenz [Problem-solving competence as cross-curricular competence]. In: Klieme E, Leutner D, Wirth J (eds) Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie [Students' problem-solving competence. Diagnostic approaches, theoretical foundations and empirical results of the German PISA study 2000]. VS Verlag, Wiesbaden, pp. 11–19
- Masters GN (1982) A Rasch model for partial credit scoring. *Psychometrika* 47:149–174
- Mayer RE (1994) Problem solving. In: Eysenck MW (ed) The Blackwell dictionary of cognitive psychology. Blackwell, Oxford, pp 284–288
- Mayring P (2014) Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt; 2014. <http://www.nbn-resolving.de/urn:nbn:de:0168-ssaoar-395173>. Accessed 12 Jan 2016
- Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs
- Nokes TJ, Schunn CD, Chi MTH (2011) Problem solving and human expertise. In: Grøver Aukrust V (ed) Learning and cognition in education. Elsevier, Oxford, pp 104–111
- OECD (2013) PISA 2015. Draft collaborative problem-solving framework. Organisation for Economic Cooperation and Development (OECD), Paris. <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>. Accessed 25 Nov 2015
- Op't Eynde P, De Corte E, Verschaffel L (2006) Accepting emotional complexity: a socio-constructivist perspective on the role of emotions in the mathematics classroom. *Educ Stud Math* 63:193–207
- Rausch (under revision) Dispositional predictors of problem solving in the field of office work
- Rausch A, Wuttke E (2016) Development of a multi-faceted model of domain-specific problem-solving competence and its acceptance by different stakeholders in the business domain. *Unterrichtswissenschaft* 44(2):164–189
- Rausch A, Schley T, Warwas J (2015) Problem solving in everyday office work—A diary study on differences between apprentices and skilled employees. *Int J Lifelong Educ* 34(4):448–467. doi:10.1080/02601370.2015.1060023
- Rausch A, Seifried J, Kögler K, Brandt S, Eigenmann R, Siegfried C (2016) Measuring non-cognitive facets in computer-based problem-solving assessments by using Embedded Experience Sampling (EES). Full paper presented at the AERA meeting in Washington
- Reither F, Stäudel T (1985) Thinking and action. In: Frese M, Sabini J (eds) Goal directed behavior: the concept of action in psychology. Lawrence Erlbaum, Hillsdale, NJ, pp 110–122
- Sabourin JL, Lester JC (2014) Affect and engagement in game-based learning environments. *IEEE Trans Affect Comput* 5(1):45–56
- Sabourin JL, Mott B, Lester JC (2013) Discovering behavior patterns of self-regulated learners in an inquiry-based learning environment. In: Lane HC, Yacef K, Mostow J, Pavlik P (eds) Artificial intelligence in education—proceedings of the 16th international conference of AIED 2013, Memphis, TN, USA, July 9–13, 2013. Springer, Heidelberg, pp 209–218
- Schoppek W, Fischer A (2015) Complex problem solving—single ability or complex phenomenon? *Front Psychol* 6:1–4
- Schwarz N, Bless B (1991) Happy and mindless, but sad and smart? The impact of affective states on analytic reasoning. In: Forgas J (ed) Emotion and social judgment. Pergamon, London, pp 55–72
- Sembill D, Wolf KD, Wuttke E, Schumacher L (2002) Self-organized learning in vocational education—foundation, implementation, and evaluation. In: Beck K (ed) Teaching-learning processes in vocational education. Peter Lang, Frankfurt, pp 267–295
- Sembill D, Rausch A, Kögler K (2013) Non-cognitive facets of competence—theoretical foundations and implications for measurement. In: Beck K, Zlatkin-Troitschanskaia O (eds) From diagnostics to learning success: proceedings in vocational education and training. Sense Press, Rotterdam, pp. 199–212. doi:10.1007/978-94-6209-191-7_15
- Sireci SG, Thissen D, Wainer H (1991) On the reliability of testlet-based tests. *J Educ Meas* 28:237–247
- Spering M, Wagener D, Funke J (2005) The role of emotions in complex problem-solving. *Cogn Emot* 19(8):1252–1261

- Stodel M (2015) But what will people think? Getting beyond social desirability bias by increasing cognitive load. *Int J Mark Res* 57(2):313–321
- Sugrue B (1995) A theory-based framework for assessing domain-specific problem-solving ability. *Educ Meas Issues Pract* 14(3):29–35
- Wainer H, Bradlow ET, Wang X (2007) *Testlet response theory and its applications*. Cambridge University Press, Cambridge
- Wang WC, Wilson M (2005) The Rasch testlet model. *Appl Psychol Meas* 29:126–149
- Weinert FE (2001) Concept of competence: a conceptual clarification. In: Rychen DS, Salganik LH (eds) *Defining and selecting key competencies*. Hogrefe and Huber, Seattle, pp 45–65
- Weiss RH (2006) *CFT 20-R*, 4th edn. Hogrefe, Göttingen
- Wilson M (2008) Cognitive diagnosis using item response models. *J Psychol* 216(2):74–88
- Wilson M, Bejar I, Scalise K, Templin J, William D, Irribarra DT (2012) Perspectives on methodological issues. In: Griffin P, McGaw B, Care E (eds) *Assessment and teaching of 21st century skills*. Springer, Dordrecht, pp 67–141
- Wittmann WW, Süß HM (1999) Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In: Ackerman PL, Kyllonen PC, Roberts RD (eds) *Learning and individual differences: process, trait, and content determinants*. Am Psychol Assoc, Washington DC, pp 77–104
- Woolfolk A (2005) *Educational psychology*, 9th edn. Pearson, Boston
- Wuttke E, Seifried J, Brandt S, Rausch A, Sembill D, Martens T, Wolf K (2015) Modellierung und Messung domänenspezifischer Problemlösekompetenz bei angehenden Industriekaufleuten—Entwicklung eines Testinstruments und erste Befunde zu kognitiven Kompetenzfacetten [Modeling and measuring domain-specific problem-solving competence of prospective industrial clerks—development of an instrument and first results regarding cognitive facets of competence]. *Zeitschrift für Berufs- und Wirtschaftspädagogik* 111(2):189–207
- Yen WM (1993) Scaling performance assessments: strategies for managing local item dependence. *J Educ Meas* 30:187–213
- Ziegler B, Frey A, Seeber S, Balkenhol A, Bernhardt R (2016) Adaptive Messung allgemeiner Kompetenzen (MaK-adapt) [Adaptive measurement of general competencies (MaK-adapt)]. In: Beck K, Landenberger M, Oser F (eds) *Technologiebasierte Kompetenzmessung in der beruflichen Bildung—Ergebnisse aus der BMBF-Förderinitiative ASCOT* [Technology-based measurement of competencies in VET—Findings from the BMBF research initiative ASCOT]. Bertelsmann, Bielefeld, pp 33–54
- Zieky M (1993) Practical questions in the use of DIF statistics in test development. In: Holland PW, Wainer H (eds) *Differential item functioning*. Erlbaum, Hillsdale, pp 337–347
- Zieky M (2003) *A DIF primer*. Educational Testing Service, Princeton

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
