International Journal for Research in Vocational Education and Training (IJRVET)
2023, Vol. 10, Issue 1, 46–67
https://doi.org/10.13152/IJRVET.10.1.3

Effectiveness of Collaboration in VET: Measuring Skills for Solving Complex Vocational Problems With a Multidimensional Authentic Technology-Based Assessment

Jessica Paeßens, Beifang Ma, Esther Winther

University of Duisburg-Essen, Institute for Vocational Education and Training, Universitätsstraße 2, 45141 Essen, Germany

Received: 06 September 2022, Accepted: 16 February 2023

Abstract

Context: Dealing with professional complexity has been of scientific interest in the research field of vocational education and training for decades. So far, there is a lack of empirical evidence regarding how professional complexity finds its way into learning processes in VET. A common option is to model complexity through authentic simulations and/or problem-solving tasks. This study does both: Complex problem-solving tasks are integrated into an authentic office simulation and are expanded to include collaborative elements, using computer-based agents. Collaboration is used to improve learning, but it is also an authentic representation of current work processes, so we ask whether apprentices perform better in the individual or in the collaborative test setting.

Methods: To compare skills used for complex problem-solving tasks within an authentic business simulation, the test settings are systematically varied between individual and collaborative test settings. The test environment is a technology-based assessment (TBA). Test tasks reflect a complex professional problem; they are presented in an office simulation, in which apprentices can collaborate with a computer-based agent. Multi-group confirmatory factor analysis is conducted to test for measurement invariance across test settings, and the two-dimensional Rasch-model incorporating between-item multidimensionality with

*Corresponding authors: jessica.paessens@uni-due.de, beifang.ma@uni-due.de; The two authors contributed equally to this work.

ISSN: 2197-8646
https://www.ijrvet.net

correlated dimensions is used for ability estimation. We also conduct ANOVA tests to determine if there is a statistically significant difference regarding the problem-solving ability between the individual and collaborative test settings.

Findings: The study provides various findings: First, collaboration helps apprentices to deal with complexity. Second, to solve complex problem-solving tasks, two bundles of skills are activated: Cognitive and social skills. The two-dimensional construct of skills with correlated dimensions showed better fit than the unidimensional construct. The scalar measurement invariance was established after excluding three items. ANOVA tests confirmed that the collaborative setting enhances the problem-solving ability of learners significantly regarding both, cognitive skills and social skills, with cognitive skills being fostered more.

Conclusion: The findings suggest the validation of the two-dimensional construct with cognitive and social skills in economic domains. The results show the effectiveness of collaboration with a computer-based agent. In the practice of vocational training and education, teachers can use digitalized collaborative approaches to enhance learning.

Keywords: Multidimensional Construct, Collaboration, Performance Measurement, Economic Domain, Web-Based Simulation, VET, Vocational Education and Training

1 Designing Learning Processes in VET: The Increasing Importance of Collaboration

Increasingly rapid changes in technology and the use of new information and communication technologies are factors that explain the emergence of teamwork and collaboration in work processes. Consequently, academic inquiries into the transversal competencies which are necessary to cope with complex requirements at the workplace suggest a re-modeling of the content of vocational training in line with the transformations in the practical field. In this section, we outline our theoretical conceptualization of skills needed to solve complex problems individually and collaboratively.

1.1 Modeling the Complexity of Practice for Learning Processes in VET

Work processes are complex, and their complexity is increasing. This development can be explained by increasing rationalization through technical progress and the use of new information and communication technologies (Achtenhagen et al., 1992). Consequently, work processes are changing. Current research programs at national and international level are looking at future technologies and show that "digitization of work", or "Industry 4.0", is leading to a change in work and production processes across all industries (e.g., Hasenbeck, 2019; Hirsch-Kreinsen, 2017). The current, especially digitally driven, transformations have far-reaching effects on existing organizational structures and processes, on customer relationships, and on business models, which are being realigned according to the digitalization of products, services, and processes (Sczogiel et al., 2019). Transformation in business and work processes has an impact on commercial workplaces: Jobs are changing and with that, the competency requirements as basis for professionally adequate action are changing, too (Brötz et al., 2014). In a complex environment, teamwork and collaboration become more important (Schlicht, 2019). A closer look at the developments of Industry 4.0 provides opportunities for making vocational education and training (VET) more attractive (Esser, 2015). In this context, learning processes in vocational training should be made to focus more on competency development (Anselmann et al., 2022; Spöttl & Windelband, 2021) and require apprentices to make enterprise-specific decisions in more complex and authentic scenarios. Regarding the resulting consequences for content-related training priorities, teamwork skills, negotiation skills, and the willingness to learn and collaborate (Ahrens & Spöttl, 2015) are particularly relevant – these transversal competencies are at the core of the framework curricula for vocational training.

The few empirical studies on the re-modeling of the content of vocational training in line with the transformations in the practical field that are currently available focus especially on participation and communication as conditions for success in the construction of learning processes. The study by Schlicht (2019), for example, points to the following learning situations that are gaining in importance:

- Learning situations that emphasize communication and cooperation in business processes. Communication and cooperation are highly relevant for shaping social and business relationships, for entrepreneurial success, and for the personal development of experts and leaders.

- Learning situations that address typical (industry-specific) problem situations linked to the current and future (projected) developments of a sustainable economy.

Learning processes should consider both the digital and employee-related changes that are currently taking place in real work processes.

1.2 Skills Needed to Solve Complex Problem-Solving Tasks

Complex work and training situations are, e.g., work processes in which, in addition to routine activities, economic decisions must be taken, and work results presented and reflected on. Work is increasingly done in project teams and across departmental boundaries. Such work requirements show characteristics of complex problems. Thus, coping with complex requirements is not only about accuracy and speed, but above all about acting strategically and coping with a "strategic moment". This includes

– the ability to control cognitive operations,

– the availability of heurisms, and

– the "wisdom" of the problem solver (Dörner, 1986).

Furthermore, a complex problem demands the operational intelligence of the problem solver. This includes, e.g., a balanced elaboration and negotiation of goals and self-management. Solving a complex problem thus implies the efficient interaction between a solver and the situational demands and requires cognitive and social skills and knowledge, as well as emotional and personal regulation (Frensch & Funke, 1995). Against this background, we model and assess two sets of skills for solving a complex problem: *Cognitive and social skills* (Funke, 2003; Hesse et al., 2015, Table 1). The cognitive skills focus on the complex problem-solving process itself, and in particular on task regulation and knowledge building in the solution process; the social skills address the interpersonal or interactional skills for successful work processes and are primarily reflected in individual opportunities for participation, perspective taking, and social regulation (Andrews-Todd & Kerr, 2019; Davier et al., 2018; Hesse et al., 2015).

Table 1: Cognitive and Social Skills Needed to Solve Complex Problem-Solving Tasks (Hesse et al., 2015)

Cognitive skills	**Social skills**
– Task regulation	– Participation
– Learning and knowledge construction	– Perspective taking
	– Social regulation

To solve complex problems, both skills must be applied, regardless of whether the problem is dealt with individually or collaboratively. Both dimensions are correlated. The relevance of social skills for cognitive processes is quite obvious, which has also been referenced in team research (Salas et al., 2017). With reference to these findings, it can be assumed that collaboration has a positive effect on the solution quality of a complex problem. On the one hand, the competencies of several people are needed for complex work processes, and on the other hand, the focus is increasingly on professional solutions developed collaboratively through changed forms of learning and work. Accordingly, collaboration in complex work processes can be defined as *the potential to act in a cognitively and socially appropriate manner in specific, technically complex problem situations.*

So far, complex problem scenarios in VET have usually been modeled as situations that had to be mastered alone. To add more authenticity to the learning tasks here, problem scenarios are expanded to include collaborative components. This can be done in different ways; in this study, computer-based agents are used as simulated colleagues, who interactively offer (standardized) support. Regarding the changes in commercial workplaces, it is helpful that technical progress in the field of computer technology, network technology, and telecommunication (can) create new possibilities for cooperation and more efficiency in the collaborative processing of a task (Barkley et al., 2014; Borghoff & Schlichter, 2000; Haake & Pfister, 2010). For example, computer-supported cooperative learning seeks to make working on a collaborative task more efficient and easier (Barron, 2003; Dillenbourg & Traum, 2006). A special kind of collaboration is that between learners and computer-based agents. The first technical implementation of such collaboration was in PISA 2015 (He et al., 2017). The computer-based agents communicated with the learner in text-based chats and simultaneously evaluated the learner's problem-solving skills based on the given, pre-formulated answers. Thus, the agents were able to observe performance, knowledge, skills, and psychometric abilities and enabled a standardized large-scale assessment (LSA; Graesser et al., 2017).

2 Goal and Research Question of the Study

The studies on problem-solving in complex problems (Funke, 2003) and collaborative problem-solving (Hesse et al., 2015) clearly indicate that problem-solving is a multidimensional construct. In the present study, we assume a two-dimensional construct that differentiates between cognitive and social skills. The cognitive skill bundle is operationalized via task regulation and knowledge construction; the social skill bundle includes participation, perspective taking, and social regulation as indicators.

The studies on the didactics of complex problems in VET (Rausch et al., 2017; Seifried et al., 2016; Winther & Klotz, 2016; Winther, 2011) show that learning in complex learning arrangements can improve vocational competencies in economic domains. So far, however, there are no findings on how well learners can handle complex, vocational tasks. The studies (summarized for ASCOT: Beck et al., 2016) show that apprentices have difficulties in grasping complexity and translating it into task solutions. Transferred to the assessments, this means that test items are often not used to their full potential because the apprentices cannot fully resolve the complexity. One may assume that learners are better at dealing with complexity when they collaborate. Thus, in this study, collaboration takes place with simulated colleagues in the form of computer-based agents.

Against the background of the theoretical considerations, two research questions are addressed:

1. Can the two-dimensional construct of cognitive and social skills for complex problem-solving be empirically represented?

2. Can the complexity of vocational situations be dealt with more successfully with collaborative support from a computer-based agent?

3 Measurement Environment: Using a Business Simulation as TBA

To answer the research questions, a complex learning arrangement is used. Commercial tasks for learning and measurement are deployed through the office simulation LUCA. LUCA computer-based agents could offer learners different kinds of hints/scaffolds (learning environment) and at the same time record and measures learners' performance (measurement environment). In this section, we describe how test items are integrated in LUCA and how LUCA is used both for individual and collaborative settings.

3.1 Understanding the Task "Supplier Selection"

LUCA is an adaptive learning environment for enabling complex learning processes (Rausch et al., 2021). LUCA simulates real everyday working life and thus supports commercial learning (learning environment). The office simulation can also be used as a measurement environment. For this purpose, complex problem-solving tasks are set as performance tests. The present study refers to the complex problem-solving task "supplier selection" as an example.

In the authentic problem scenario "supplier selection", the apprentices are required to select a suitable supplier from several offers. The task contains a large amount of information and additional attachments that must be viewed and processed. The task "supplier selection" is a domain-typical problem scenario in the business context (design principles of complex scenarios, see Ma et al., 2022; Paeßens & Winther, 2021; Paeßens et al., 2022; Paeßens & Winther, 2023b). The scenario can be placed in the value creation processes of companies and here in particular in the area of purchasing. The vocational requirement is to decide for a supplier, which is worked out by the apprentices in the office simulation in various substeps. The apprentices first view various offers and then prepare a utility value analysis to evaluate the offers. Supplier selection has a high curricular and practical relevance in the field of commercial-vocational education, involves various economic procedures, and focuses on reasoned decisions; thus, it is particularly suitable as an illustrative example for a complex vocational situation.

In this authentic vocational task, the complexity of vocational activity was constructed in such a way that it can be experienced and processed by the apprentices. Working with complexity is systematically but equally varied for both test settings. As theoretically described, both test settings have in common that cognitive and social skills are necessary for solving complex problems (cf. Hesse et al., 2015). The cognitive and social skills involved in both, individual and collaborative problem-solving are measured with validated test items (for individual problem-solving: Paeßens et al., 2023a; for collaborative problem-solving: Paeßens & Winther, 2023b). Table 2 outlines how cognitive and social skills are modeled as theoretically described sets of sub-performances and measured through the different indicators. In the present study, only task regulation is tested as an indicator for cognitive skills.

Table 2: Indicators Measuring Sub-Performance for Cognitive and Social Skills

Primarily cognitive sub-performance	Primarily social sub-performance
Task regulation: Screening different attached offers	Social regulation: Reasons for choosing a supplier
Task regulation: Screening relevant information for the benefit analysis	Social Regulation: Written presentation (considering paragraphs, greetings, salutations, and general politeness)
Task regulation: Attaching the modified benefit analysis to the email	Perspective taking: Transfer of the weighting into the benefit analysis
	Participation: Considering the test result sent by email later on

The different indicators testing the corresponding sub-performances are collected in standardized task items and then embedded in various complex problem scenarios. Therefore, the cognitive and social skill sets are evaluated based on the responses of learners to task items. For example, in the problem scenario "supplier selection" for social regulation (a sub-performance of social skills), learners' skill-level is evaluated according to the criteria of quantity and quality of the reasons for choosing a supplier. To obtain a better insight into the tasks in the applied technology-based assessment (TBA), an indicator within the complex task for perspective taking (a sub-performance of social skills) is presented in Figure 1. Learners work through complex problem scenarios in an authentic office environment where various applications (email client, spreadsheet, ERP, etc.) are available. The department manager "Timur Demir" explains in his email with which weighting the factors price, delivery time, and quality should be considered in the benefit analysis. The learners are expected to transfer the information into their own mental model and ultimately into the benefit analysis for other colleagues. Learners in both, the collaborative and the individual test settings need to be able to understand information from other colleagues and their intentions and then integrate this into their work process. This item shows that, in addition to cognitive skills, social skills are also necessary for complex problem-solving in both test settings.

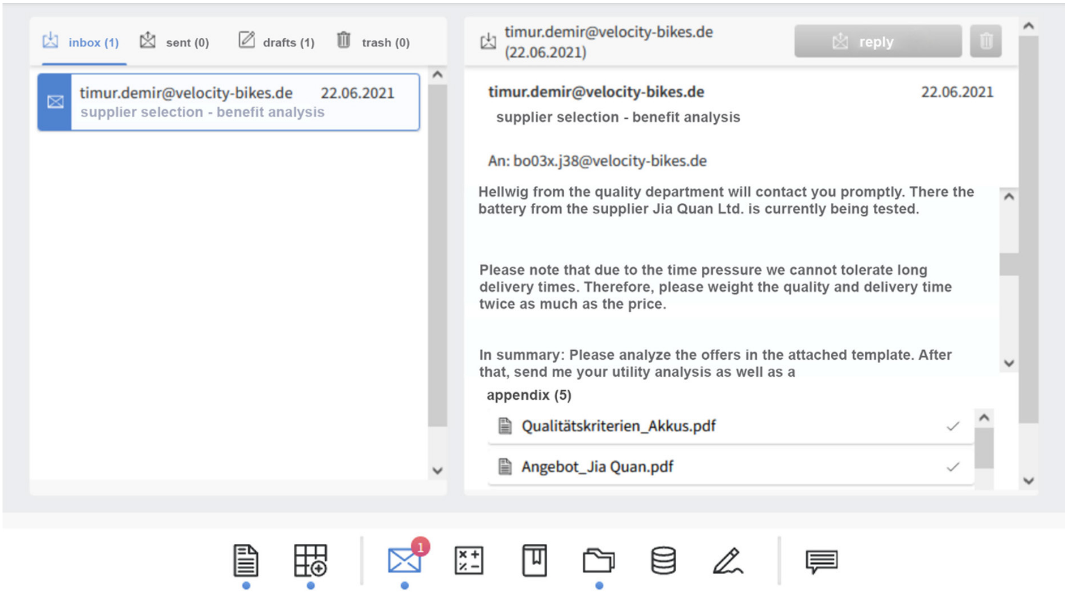

Figure 1: Representation of the Primarily Social Sub-Performance "Perspective Taking" in the Office Simulation LUCA

3.2 Expanding the Task by Collaborative Interaction

In this study, the existing items from the technology-based office simulation are used in two different test settings. The sub-performances are equivalent in content in both test settings. The collaborative test setting is expanded with various interaction formats that provide general as well as specific solution hints and concretely address both cognitive and social skills. Learners are given guidance via computer-based agents. For example, overlays in the office simulation (see Figure 2) simulate instructor interactions or contacts with colleagues. When learners respond to the computer-based agent, they receive procedural information by email that addresses, e.g., specific perspective taking, prompts for individual activities in the overall collaborative process, or suggests general problem-solving heuristics.

Figure 2: Simulated Interaction with a Colleague "Rene Eßer" and Response of the Computer-Based Agent in the Form of an Email From the Colleague

To sum up, the measurement environment is based on a business simulation, in which day by day work tasks have to be mastered. The performance of the apprentices in solving the task "supplier selection" is recorded. This problem-solving task requires both cognitive and social skills. While the apprentices have to master the complex problem-solving task on their own in the individual test setting, they are supported by a computer-based agent in the collaborative test setting.

Thus, the computer-based agent provides on the one hand interactions with learners to support problem-solving processes and on the other hand a diagnostic function (for the use of this particular strength of LUCA, see Paeßens & Winther, 2023c). The measurement environment fulfills the requirements of individual and collaborative learning processes in the way that

- Vocational learning takes place largely in social situations – namely, in a business process in which third parties are involved.

- Sub-performance can be assigned to both cognitive and social skills, which are important for problem-solving of complex problems.

- Computer-based agents are used in collaborative test settings to help deal with complexity.

4 Research Design

The theoretical conceptualization emphasizes that skills needed to solve complex problems is a multidimensional construct composed of two key dimensions, one of which is cognitive skills and the other one is social skills. In this section, we adapt the theory-driven framework of cognitive and social skills involved in complex problem-solving into statistical model to process the data statistically.

4.1 Model and Hypothesis

The hypothesized two-dimensionality of the social and cognitive abilities involved in problem-solving in individual and collaborative settings is illustrated in the model below (Figure 3). In this model, the cognitive dimension and the social dimension are two distinct latent dimensions. That is, there are learners who might have high ability in one dimension but not in the other. However, according to the theoretical framework, the two dimensions are related (for the relation between the two dimensions, see Hesse et al., 2015; chapter 1).

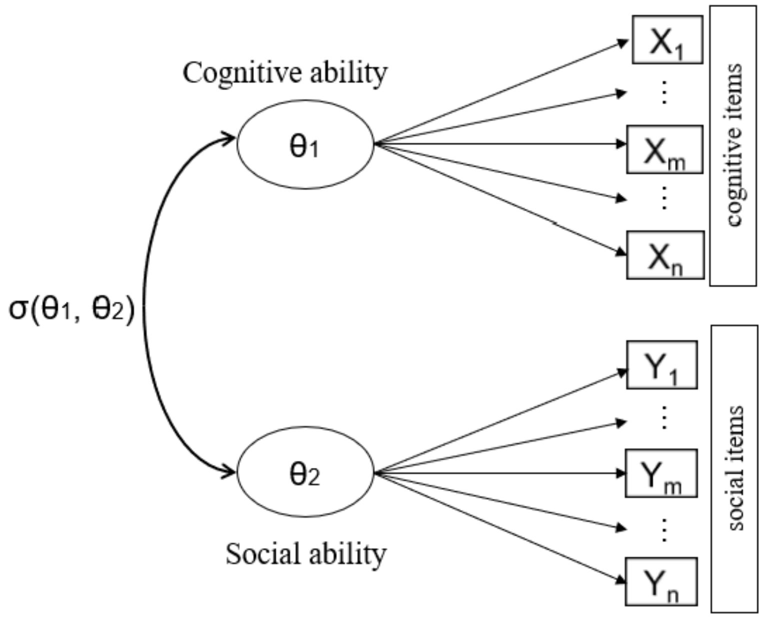

Figure 3: Schematic Illustration of the Two-Dimensional Rasch-Model Incorporating Between-Item Multidimensionality (adapted from Hartig & Höhler, 2009)

To test this latent two-dimensional ability construct, the digital authentic assessment described above was used. The assessment is designed to measure the abilities in the cognitive and social dimensions involved in solving the different items. Multidimensional IRT models (MIRT) can be employed to investigate the construct validity of tests with multiple dimensions (Baghaei, 2013; Embretson, 1980; Field, 2013; Janssen & De Boeck, 1999; Santelices & Caspary, 2009; Wilson & Moore, 2011) and assess learners' abilities separately for the dimensions involved.

Therefore, we state the following two hypotheses:

1. The theoretical two-dimensional construct of the test form is valid. To confirm this hypothesis, the two-dimensional construct should be compared with the unidimensional construct regarding the fit indices.

2. Learners show significantly higher ability in both, the cognitive and the social dimension in the collaborative test setting than in the individual test setting. To test this hypothesis, the comparability between the collaborative and the individual test settings should first be confirmed. After the comparison of learners' abilities in both test settings, it should also be analyzed which dimension will be influenced more strongly by the test setting.

4.2 Sample

505 commercial apprentices from Germany participated in the study. 240 were male, 240 female, 3 divers, and 22 values for sex are missing due to technical problems. The age of the learners ranged from 17 to 42 years (M = 20.92; SD = 2.41). These data were collected at 21 vocational schools and in 28 classes in North Rhine-Westphalia. All participants were in their first year of training. 250 of them worked in the individual test setting and 255 were assigned to the collaborative test setting. The assignment was random at class level.

4.3 Measures

The 11 items within the task are developed according to theory-based design principles and are empirically validated (Paeßens et al., 2023a; Paeßens & Winther, 2023b). The construction of the items is presented in Table 2. The item "participation" in the social dimension, for example, refers to the sub-performance how a problem-solver considers a test result from a supplier that will be sent to them by email later on. The 11 items are embedded in the web-based office simulation and a questionnaire. Seven items belong to the social dimension and four items to the cognitive dimension. Equal test items were used in two test settings; in the individual test setting, the learners processed the supplier selection alone, while in the collaborative test setting, agents were available as a group partner to interact with the learners.

5 Statistical Analyses and Results

The analysis presented in this section begins with testing the first hypothesis and involves two primary steps: (1) Comparing model fit between the Unidimensional Rasch model and the two-dimensional Rasch model —and then (2) testing for measurement invariance between individual and collaborative test settings. The primary purpose of these two steps is to examine how the theoretical two-dimensional construct of the test form performs. The analysis then proceeds to the examination of the second hypothesis that the complexity of vocational situations can be dealt with more successfully with collaborative support from a computer-based agent.

5.1 Comparing Model Fit: The Unidimensional Rasch Model and the Two-Dimensional Rasch Model

This step aims to test the hypothesis that this is a two-dimensional construct. For this purpose, a unidimensional and a multidimensional Rasch model are applied to the test results and we compare the fits of these models. The results of the ANOVA test are is summarized in Table 3. The two-dimensional model with a significantly smaller model deviance, larger log likelihood, smaller AIC, and smaller BIC fits better to the data compared to the unidimensional model. The theoretical construct of the cognitive and social dimensions is thus, statistically confirmed.

Table 3: Model Fit Statistics for the Unidimensional and Multidimensional Models

Model	loglike	Deviance	Npars	AIC	BIC	Chisq	df	p
uni	2697.273	5394.545	12	5418.545	5469.240	187.222	2	<.001
multi	-2603.662	5207.323	14	5235.323	5294.467			

5.2 Testing for Measurement Invariance

To test for measurement invariance, we ran five rounds of Multi-Group Confirmatory Factor Analysis (MG-CFA) using the R statistics package. The main outputs of model fits are summarized in Table 4. Firstly, we ran an MG-CFA without cross group equality constraints; this configural model shows a good fit (CFI = 0.975, RMSEA = 0.053, SRMR = 0.061) according to Kline (2010). Following configural invariance, we tested for metric invariance with equality of the factor loadings across groups. Although model fit indices (CFI = 0.963, RMSEA = 0.061, SRMR=0.077) show that the metric model is acceptable, a change of 0.012 in CFI (ΔCFI > 0.01) when comparing the metric model and the configural model implies non-invariance according to Chen (2007). Chen (2007) recommends the following alternative cutoff criteria for model comparisons due to the sensitivity of the most commonly used χ2 test for goodness of fit (Cochran, 1952) to sample size: For testing loading invariance with a sample size larger than 300, a change of ≥ -.010 in CFI, supplemented by a change of ≥.010 in RMSEA or a change of ≥.030 in SRMR would indicate non-invariance; for testing intercept or residual invariance, a change of ≥ -.010 in CFI, supplemented by a change of ≥ .010 in RMSEA or a change of ≥ 0.10 in SRMR would indicate non-invariance.

Table 4: Model Fit Incidents for 3 Levels of Measurement Invariance

	Configural model	Metric model	Partial metric model	Scalar model	Partial scalar model
CFI	0.975	0.963	0.966	0.928	0.958
RMSEA	0.053	0.061	0.059	0.084	0.071
SRMR	0.061	0.077	0.075	0.078	0.066

This study shows that partial metric invariance is achieved across the test settings when releasing only one item of social dimension among the total 11 items. Then we tested for scalar invariance by constraining the factor loadings and the intercepts across groups. Partial scalar invariance is established when releasing two additional intercept parameters for two items. A total of three items were excluded from further analyses. The reasons why they have significant influence on measurement invariance is discussed further down. The established measurement invariance implies that the instrument assesses the psychometric equivalence of a construct across test settings.

5.3 Comparing Abilities: The Individual Group and the Collaborative Group

Before person abilities are estimated using the multidimensional Rasch model, we examine correlations between the two dimensions and reliabilities of the two dimensions. As Table 5 shows, a correlation coefficient of .342 is considered moderate correlation between the cognitive and social dimensions in the individual test setting and is smaller than a correlation coefficient of .542 in the collaborative test setting, which corresponds to a large effect size. This finding supports the theoretical consideration that the correlation between cognitive and social dimensions in the collaborative test setting is stronger than in the individual test setting.

Table 5: Correlations and Reliabilities of Dimensions and Average Person Abilities on Dimensions

		EAP Reliability		**Person abilities (logits)**	
	Correlations between dimensions	Cognitive dimension	Social dimension	Cognitive dimension	Social dimension
Individual test setting	0.342	0.690	0.591	-0.075	-0.208
Collaborative test setting	0.542	0.634	0.681	1.044	0.188

The EAP reliability of dimensions ranges from 0.591 to 0.690 (see Table 5). Note that the EAP reliability is sensitive to the length of the test, and two items from the social dimension as well as one item from the cognitive dimension were excluded after testing the measurement invariance. We used the Spearman-Brown formula to predict the reliability of the original test. After Spearman-Brown correction, the lowest EAP reliability is .68, slightly less than 0.7. Since the test is not used as a psychometric scale or an individual diagnosis, this value is considered appropriate for the empirical structuring and description of the ability model.

We can see in Table 5 that the leaners in the collaborative group show higher abilities in both, the cognitive and social dimensions. Two rounds of one-way analysis of variance were conducted to test whether the means of the individual group and the collaborative group differ statistically significantly. The summary of ANOVA shows that the choice of test setting has a significant impact both on the ability of the cognitive dimension (F (1,503) = 38.41, $p < .001$, partial $\eta^2 = .071$, n = 505) and on the ability of the social dimension (F (1,503) = 7.64, $p < .001$, partial $\eta^2 = .015$, n = 505). The partial eta square is converted here into the effect size f according to Cohen (1992). For the cognitive dimension, the effect size is f = .276 and corresponds to a medium effect according to Cohen (1988). Regarding the social dimension, the effect size is f = .123 and corresponds to a medium effect according to Cohen (1988).

6 Conclusion and Discussion

Work processes are complex and are increasing in complexity through transformations. This complexity must be taught in the VET processes. For this purpose, we offer complex learning arrangements and complex vocational tasks. The web-based office simulation LUCA is a complex learning arrangement and, as it were, a TBA; commercial competency development can be fostered and the performance in typical commercial situations can be assessed with it. Two test settings were implemented in LUCA for this study:

- Test setting 1: Solving an authentic problem with high complexity alone;

- Test setting 2: Solving an authentic problem with high complexity supported by a computer-based agent.

In test setting 1, apprentices must deal with the complexity of the problem on their own; in test setting 2, they receive specific support. It is important that this support is not classic solution support, but rather simulates authentic collaboration situations. Test setting 2 thus extends authenticity. To be able to meaningfully compare both test settings, an indicator model is chosen; in this case, components of complex problem-solving (cf. Hesse et al., 2015). Complex problem-solving - whether handled with individually or collaboratively - requires the knowledge necessary for problem-solving (regarding both, task regulation and knowledge construction) and social skills (perspective taking, social regulation, and identified participation). While in test setting 2, the cognitive and social aspects are fostered by the agent, this support in dealing with complexity does not exist in test setting 1. The hypotheses are confirmed: (1) The test settings both measure complex problem-solving. The construct is two-dimensional in both test settings. Cognitive and social skills are necessary to solve a complex vocational task. (2) Higher performance is shown in the collaborative test setting.

The results confirm that the demands of a complex problem are both cognitive and social and learners are better able to deal with complexity through modeled collaboration than when they solve the task without a tutor.

Furthermore, the cognitive dimension is more sensitive to the change of the test setting. This result can be interpreted such that by listening to other viewpoints, considering other positions, and reconsidering own ideas, more cognitive processes are activated to gain a more complete understanding as a group. For the practice of VET, the findings imply that (1), collaboration is effective if the task involved is designed based on didactic principles, especially in terms of its complexity and (2), a collaborative approach actively engages learners to process and synthesize information both in the cognitive and the social dimensions. Regarding the findings in the collaborative test setting, we need to discuss to which extent the learners collaborate. Overall, it can be stated that the interaction with an agent represents a basic form of collaboration. The agent responds adaptively to learner input and, thus, simulates collaboration. To evaluate social embeddedness, Braunstein et al. (2022) develop a theory-driven framework and adapt it for the office simulation LUCA (Rausch et al., 2021). Within the office simulation, the agent can be classified as a social interaction. The use of computer-based agents makes an empirical and future-relevant contribution to VET. Besides this empirical goal, web-based office simulations also have an impact on the practice of VET. VET should respond to digitization, which is changing the world of work and will have employees undergo further training to be able to act competently within the new business processes in companies and on the labor market (Hirsch-Kreinsen, 2017; Dengler & Matthes, 2015). Whether at school or at the workplace: Collaboration should be facilitated and integrated into learning and working processes.

The complex LUCA environment has also important implications for the way researchers and practitioners organize collaboration. Firstly, as a learning environment, LUCA can provide instructional interventions, learners receive different kinds of hints and scaffolds from the computer-based agent to construct their collaboration process in the form of human-to-agent interaction. Secondly, a strength of LUCA is that it combines a specific pedagogy with integrated measures, so it is also an innovative assessment tool. Collaboration at the highest level – between colleagues or learners as a form of human-to-human interaction (see framework Braunstein et al., 2022) – requires different tools for measuring interaction. The tools must be able to evaluate the collaboration process in a standardized and, ideally, automated way. Motivational and emotional factors influence the collaboration process. Various approaches impressively show that attributing individual performance in a group process is challenging (vignettes in King et al., 2004, forced decisions in Salgado & Táuriz, 2014, third-party evaluations in Connelly & Ones, 2010; Oh et al., 2011, Situational Judgment Test in Motowidlo et al., 1990; Weekley et al., 2004; Whetzel & McDaniel, 2009, cooperative games and simulations in Griffin, 2017; Hao et al., 2017, multiplayer games in Zhu & Bergner, 2017).

In PISA 2015, collaboration with automated agents was standardized for learners (Graesser et al., 2018; He et al., 2017). The innovation here was that individual performance, particularly social and cognitive skills, could be observed at individual level in a group setting (He et al., 2017; Organisation for Economic Co-operation and Development, 2017). Standardization seems necessary in an LSA, although standardization could also be achieved using an external test instrument to measure collaboration. The use of automated assessments, such as those generated by RIFF, has innovative potential. Automated evaluation promises to observe learners' collaboration performance in real time and make it accessible to empirical assessment. While it has, so far, only been possible to realize the attribution of individual performances with the help of agents in collaborative assessments (cf. Graesser et al., 2017; He et al., 2017), in RIFF, learners can collaborate with each other "for real". Instructional designs have been developed in RIFF that can be used to build and develop technical as well as collaboration skills for commercial apprentices (see Paeßens & Winther, 2021; Paeßens et al., 2022; findings of this study). Collaboration is not the ideal modus operandi in every learning and working situation. The problem has to be complex and different learners have to bring in their specific knowledge to work out a solution. In the future, it should be worked out which complex (sub) tasks are suitable for collaborative settings to foster learners' abilities in the collaboration process as well as in the collaborative performance (first approaches in Ma et al., 2022). Another limitation of the study is that we limited the number and range of items for this first-time validation. We recommend that future research extends the items by taking domain-specific and domain-related competencies into account.

Acknowledgement

The project Problem Solving Analytics in Office Simulations (PSA-Sim), on which this publication is based, is funded by the German Federal Ministry of Education and Research under the grant number: 21AP008B. The authors are responsible for the content of this publication.

Ethics statement

The authors state that they have heeded the ethical principles in this submission.

References

Achtenhagen, F., John, E., Preiß, P., Tramm, T., Schunck, A., & Seemann-Weymar, H. (1992). Grundgedanken des Projektansatzes — Lernen, Denken, Handeln in komplexen ökonomischen Situationen — Forschungsidee und Projektkonzeption. In F. Achtenhagen, E. John, P. Preiß, T. Tramm, A. Schunck & H. Seemann-Weymar (Eds.), *Lernhandeln in komplexen Situationen* (pp. 1–27). Gabler. https://doi.org/10.1007/978-3-322-90323-5_1

Ahrens, D., & Spöttl, G. (2015). Industrie 4.0 und Herausforderungen für die Qualifizierung von Fachkräften. In H. Hirsch-Kreinsen, P. Ittermann & J. Niehaus (Eds.), *Digitalisierung industrieller Arbeit* (pp. 184–205). Nomos. https://doi.org/10.5771/9783845283340-172

Andrews-Todd, J., & Kerr, D. (2019). Application of ontologies for assessing collaborative problem solving skills. *International Journal of Testing*, *19*(2), 172–187. https://doi.org/10.1080/15305058.2019.1573823

Anselmann, S., Harm, S., & Faßhauer, U. (2022). Input from the grassroots level — reflecting challenges and problems for VET professionals in Germany. *International Journal for Research in Vocational Education and Training*, *9*(2), 239–268. https://doi.org/10.13152/IJRVET.9.2.5

Baghaei, P. (2013). Development and psychometric evaluation of a multidimensional scale of willingness to communicate in a foreign language. *European Journal of Psychology of Education*, *28*, 1087–1103. https://doi.org/10.1007/s10212-012-0157-y

Barkley, E., Cross, K., & Major, C. (2014). *Collaborative learning techniques: A handbook for college faculty*. Wiley.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/S15327809JLS1203_1

Beck, K., Landenberger, M., & Oser, F. (2016). T*echnologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT*. wbv. https://doi.org/10.3278/6004436w

Borghoff, U., & Schlichter, J. (2000). Computer-supported cooperative work. In *Computer-supported cooperative work* (pp. 87–141). Springer. https://doi.org/10.1007/978-3-662-04232-8

Braunstein, A., Deutscher, V., Seifried, J., Winther, E., & Rausch, A. (2022). A taxonomy of social embedding. A systematic review of virtual learning simulations in vocational and professional learning. *Studies in Educational Evaluation*, *72*. https://doi.org/10.1016/j.stueduc.2021.101098

Brötz, R., Kaiser, F., Kock, A., Krieger, A., Noack, I., Peppinghaus, B., & Schaal, T. (2014). *Gemeinsamkeiten und Unterschiede kaufmännisch-betriebswirtschaftlicher Aus- und Fortbildungsordnungen: Abschlussbericht.*

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cochran, W. (1952). The $\chi 2$ test of goodness of fit. The Annals of mathematical statistics, 23(3), 315–345. https://doi.org/10.1214/aoms/1177729380

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Lawrence Erlbaum. https://doi.org/10.4324/9780203771587

Cohen, J. (1992). Statistical power analysis. *Current directions in psychological science*, *1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Connelly, B., & Ones, D. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*, 1092–1122. https://doi.org/10.1037/a0021212

Davier, A. von, Zhu, M., & Kyllonen, P. (2017). *Methodology of educational measurement and assessment: Innovative Assessment of Collaboration*. Springer. https://doi.org/10.1007/978-3-319-33261-1

Dengler, K., & Matthes, B. (2015). *Folgen der Digitalisierung für die Arbeitswelt: Substituierbarkeitspotenziale von Berufen in Deutschland*. IAB-Forschungsbericht 11/2015. https://doku.iab.de/forschungsbericht/2015/fb1115.pdf

Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences*, *15*(1), 121–151. https://doi.org/10.1207/s15327809jls1501_9

Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*. *32*(4), 290–308.

Embretson, S. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, *45*, 479–494. https://doi.org/10.1007/BF02293610

Esser, F. (2015). "Wer vorsieht, ist Herr des Tages" – Digitalisierung erfordert vorausschauendes Handeln. In Bundesinstitut für Berufsbildung (Eds.), Lernen für die digitale Wirtschaft. *BWP*, *6*(3), p. 3.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 77–151). Cambridge University Press.

Frensch, P. A., & Funke, J. (1995). Complex problem solving: The European perspective. Lawrence Erlbaum. https://doi.org/10.4324/9781315806723

Funke, J. (2003). *Problemlösendes Denken*. Kohlhammer. https://doi.org/10.11588/diglit.18679#0015

Graesser, A., Fiore, S., Greiff, S., Andrews-Todd, J., Foltz, P., & Hesse, F. (2018). Advancing the science of collaborative problem solving. *Psychological science in the public interest: A Journal of the American Psychological Society*, *19*(2), 59–92. https://doi.org/10.1177/1529100618808244

Graesser, A., Dowell, N., & Clewley, D. (2017). Assessing collaborative problem solving through conversational agents. In A. von Davier, M. Zhu & P. Kyllonen (Eds.), *Innovative assessment of collaboration. Methodology of educational measurement and assessment* (pp. 65–80). Springer. https://doi.org/10.1007/978-3-319-33261-1_5

Griffin, P. (2017). Assessing and teaching 21st century skills: Collaborative problem as a case study. In A. von Davier, M. Zhu & P. Kyllonen (Eds.), *Innovative assessment of collaboration. Methodology of educational measurement and assessment* (pp. 113–134). Springer. https://doi.org/10.1007/978-3-319-33261-1_8

Haake, J., & Pfister, H. (2010). Scripting a distance-learning university course: Do students benefit from net-based scripted collaboration? *International Journal of Computer-Supported Collaborative Learning*, *5*(2), 191–210. https://doi.org/10.1007/s11412-010-9083-7

Hao, J., Liu, L., Davier, A. von, & Kyllonen, P. (2017). Initial steps towards a standardized assessment for collaborative problem solving (CPS): Practical challenges and strategies. In A. von Davier, M. Zhu & P. Kyllonen (Eds.), *Innovative assessment of collaboration. Methodology of educational measurement and assessment* (pp. 135–156). Springer. https://doi.org/10.1007/978-3-319-33261-1_9

Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, *35*(2-3), 57–63. https://doi.org/10.1016/j.stueduc.2009.10.002

Hasenbeck, F. (2019). *Macht die Digitalisierung alles komplexer?* Roundtable.

He, W., von Davier, Matthias, Greiff, S., Steinhauer, E., & Borysewicz, P. (2017). Collaborative problem solving measures in the Programme for International Student Assessment (PISA). In A. von Davier, M. Zhu, & P. Kyllonen (Eds.), *Innovative assessment of collaboration. Methodology of educational measurement and assessment* (pp. 95–112). Springer. https://doi.org/10.1007/978-3-319-33261-1_7

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In E. Care, & P. E. Griffin (Eds.), *Assessment and teaching of 21st century skills: Educational assessment in an information age* (pp. 37–56). Springer. https://doi.org/10.1007/978-94-017-9395-7_2

Hirsch-Kreinsen, H. (2017). Digitalisierung industrieller Einfacharbeit. *Arbeit*, *26*(1), 7–32. https://doi.org/10.1515/arbeit-2017-0002

Janssen, R., & De Boeck, P. (1999). Confirmatory analyses of componential test structure using multidimensional item response theory. *Multivariate Behavioral Research*, *34*, 245–268. https://doi.org/10.1207/S15327906Mb340205

King, G., Murray, C., Salomon, J., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–207. https://doi.org/10.1017/S000305540400108X

Kline, R. (2010). *Principles and practice of structural equation modeling*. The Guilford Press. https://doi.org/10.1080/10705511.2012.687667

Ma, B., Paeßens, J., & Winther, E. (2022). Wir machen das zusammen oder: Worin unterscheiden sich individuelle und kollaborative Aufgabenbearbeitungen in der kaufmännischen (Berufs-)bildung? *Berufsbildung*, *3*(22), 49–53. http://u.wbv.de/bb2203w013

Motowidlo, S., Dunnette, M., & Carter, G. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*(75), 640–647. https://doi.org/10.1037/0021-9010.75.6.640

Organisation for Economic Co-operation and Development. [OECD] (2017). *PISA 2015 Assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving*. OECD Publishing. https://doi.org/10.1787/9789264281820-en.

Oh, I., Wang, G., & Mount, M. (2011). Validity of observer ratings of the five-factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, *96*(4), 762–773. https://doi.org/10.1037/a0021832

Paeßens, J., Ma, B., & Winther, E. (2022). *Von wo aus arbeitest du? Analog, digital und hybrid lernende Gruppen in der beruflichen Bildung*. Weiterbildung - Zeitschrift für Grundlagen, Praxis und Trends (5), 35–37.

Paeßens, J., & Winther, E. (2021). Kollaboratives Problemlösen in kaufmännischen Geschäftsprozessen. In E. Wittmann, D. Frommberger & U. Weyland (Eds.), *Jahrbuch der berufs- und wirtschaftspädagogische Forschung 2021* (pp. 67–82). Budrich. https://doi.org/10.25656/01:23408

Paeßens, J., Winther, E., Grünkemeyer, T., Schäufele, F., Braunstein, A., Deutscher, V., Gentner, S., Mayer, C., Seifried, J., Brandt, S., Ludwig, S., & Rausch, A. (2023a). *Ganz schön tricky: Wie kaufmännische Auszubildende in der Bürosimulation LUCA berufliche Komplexität bearbeiten* (können). [Manuscript submitted for publication]. Vocational and Educational Training, University Duisburg-Essen.

Paeßens, J., & Winther, E. (2023b). *Kollaboration, aber sinnvoll: Wie sich mit komplexen Problemszenarien berufliche und kollaborative Kompetenzen erfassen lassen*. [Manuscript submitted for publication]. Vocational and Educational Training, University Duisburg-Essen.

Paeßens, J., & Winther, E. (2023c). *Kollaboration als transversale Kompetenz in der kaufmännischen Ausbildung*. [Manuscript submitted for publication]. Vocational and Educational Training, University Duisburg-Essen.

Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die web-basierte Bürosimulation LUCA–Funktionen, Einsatzmöglichkeiten und Forschungsausblick. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *117*(3), 372–394. https://doi.org/10.25162/zbw-2021-0017

Rausch, A., Kögler, K., Frötschl, C., Bergrab, M., & Brandt, S. (2017). Problemlöseprozesse sichtbar machen: Analyse von Logdaten aus einer computer-basierten Bürosimulation. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *113*(4), 569–594. https://doi.org/10.25162/zbw-2017-0024

Salas, E., Reyes, D., & Woods, A. (2017). The Assessment of teamp. In A. von Davier, M. Zhu & P. Kyllonen (Eds.), *Innovative assessment of collaboration: Methodology of educational measurement and assessment* (pp. 21–36). Springer. https://doi.org/10.1007/978-3-319-33261-1_2

Salgado, J., & Táuriz, G. (2014). The five-factor model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, *23*(1), 3–30. https://doi.org/10.1080/1359432X.2012.716198

Santelices, M., & Caspary, K. (2009). Development of a multidimensional measure of academic engagement. *Journal of Applied Measurement*, *10*(4), 371–393.

Schlicht, J. (2019). *Kommunikation und Kooperation in Geschäftsprozessen: Modellierung aus pädagogischer, ökonomischer und informationstechnischer Perspektive*. wbv.

Sczogiel, S., Schmitt-Rueth, S., Malapally, A., & Williger, B. (2019). *Future Digital Job Skills: Die Zukunft kaufmännischer Berufe*. IHK Nürnberg für Mittelfranken.

Seifried, J., Rausch, A., Kögler, K., Brandt, S., Eigenmann, R., Schley, T., Siegfried, C., Egloffstein, M., Küster, J., Wuttke, E., Sembill, D., Martens, T., & Wolf, K. (2016). Problemlösekompetenz angehender Industriekaufleute – Konzeption des Messinstruments und ausgewählte empirische Befunde (DomPL-IK). In K. Beck, M. Landenberger & F. Oser (Eds.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung – Ergebnisse aus der BMBF-Förderinitiative ASCOT* (pp. 119–138). Bertelsmann. https://doi.org/10.3278/6004436w

Spöttl, G., & Windelband, L. (2021). The 4th industrial revolution–its impact on vocational skills. *Journal of Education and Work*, *34*(1), 29–52. https://doi.org/10.1080/13639080.2020.1858230

Weekley, J., Ployhart, R., & Harold, C. (2004). Personality and situational judgment tests across applicant and incumbent settings: An examination of validity, measurement, and subgroup differences. *Human Performance*, *17*(4), 433–461. https://doi.org/10.1207/s15327043hup1704_5

Whetzel, D., & McDaniel, M. (2009). Situational judgment test: An overview of current research. *Human Resource Management Review*, *19*(3), 188–202. https://doi.org/10.1016/j.hrmr.2009.03.007

Wilson, M., & Moore, S. (2011). Building out a measurement model to incorporate complexities of testing in the language domain. *Language Testing*, *28*(4), 441–462. https://doi.org/10.1177/0265532210394142

Winther, E., & Klotz, V. K. (2016). Berufliche Kompetenzniveaumodellierung – Eine Blaupause für die kaufmännische Domäne. *Unterrichtswissenschaft*, *44*(2), 131–146.

Winther, E. (2011). Kompetenzorientierte Assessments in der beruflichen Bildung – Am Beispiel der Ausbildung von Industriekaufleuten. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *107*(1), 33–54. https://doi.org/10.25162/zbw-2011-0002

Zhu, M., & Bergner, Y. (2017). Modeling collaboration with social network models. In A. von Davier, M. Zhu & P. Kyllonen (Eds.), *Innovative assessment of collaboration: Methodology of educational measurement and assessment* (pp. 303–314). Springer. https://doi.org/10.1007/978-3-319-33261-1_19

Biographical Notes

Jessica Paeßens is a PhD student and her research focuses on collaborative problem-solving among commercial apprentices. The research, which is both empirical and relevant for (school) practice, focuses on the extent to which collaboration can be empirically measured at individual and group level. Measuring collaboration at the individual level is challenging; she uses innovative digital tools from practice for it. With the help of these tools, she explores how collaboration processes and the resulting performance are related and which tasks are suitable for collaborative problem-solving.

Beifang Ma is a PhD student and her current work focusing on psychometrics and educational measurement includes the structural equation modeling and validating of instruments in the research area of competence assessments in vocational education and training. She works on a project investigating the growth of professional competence and decisive context factors influencing it in vocational education and training. She aims at characterizing the trajectories and causal relationships of professional competence development to optimize the quality of training and education.

Esther Winther, Dr rer. pol. habil., is Full Professor of Vocational Education and Training, her main research interests are in the areas of empirical teaching-learning research with focus on psychometric competence modeling and measurement, the development of training and continuing education programs, and the conception of innovative and digital teaching and learning scenarios for professional and operational fields of action. As a project participant in large national research networks (DFG priority program, SAW research funding, BMBF research initiatives ASCOT and ASCOT+), she develops innovative and workplace-oriented assessments that record the status quo and the foreseeable development of vocational competences.