Vocations and Learning (2024) 17:459–485
https://doi.org/10.1007/s12186-024-09350-5

ORIGINAL PAPER

Domain-Linked and Domain-Specific Competence: a Validation Study of a Two-Dimensional Model of Economic Vocational Competence in Germany

Beifang Ma[1] · **Maximilian Krötz**[2] · **Esther Winther**[1]

Received: 20 July 2023 / Accepted: 6 July 2024 / Published online: 24 July 2024
© The Author(s) 2024

Abstract
Modeling vocational competence is increasingly crucial for monitoring and enhancing the quality of Vocational Educational Training (VET), particularly in the context of ongoing international comparative studies known as "large-scale assessments" of vocational education and training. This study endeavors to provide well-structured and guideline-compliant empirical evidence for the validation of the two-dimensional construct of economic vocational competence, advancing beyond the current state of research. A sample of 1438 first-year apprentices from two federal states in Germany participated as test-takers. The authentic assessment framework comprised 24 items, assessing two dimensions of vocational competence: domain-linked competence and domain-specific competence in the business/commercial domain. Measurement invariance was assessed across (1) federal states and (2) versions of test booklets, and the Multidimensional Random Coefficient Multinomial Logit model was employed to examine the quality of the two-dimensional vocational competence construct. The results supported the validity of the structure, highlighting the differentiation between domain-linked competence and domain-specific competence. This provides a more substantively accurate representation of trainees' vocational competence compared to a unidimensional model.

Keywords Vocational Competence · Construct Modeling · Multidimensional Model · Domain-Linked Competence · Domain-Specific Competence

✉ Beifang Ma
beifang.m@gmail.com

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

[2] Area Economic and Business Education, University of Mannheim, L 4, 1, Mannheim 68161, Germany

Introduction

The Concept of Competence

Modeling and diagnosing vocational competence have gained scientific interest in Germany since Weinert (2001) introduced comparative performance measurement in schools (Rüschoff, 2019). In 2007, the Federal Institute for Vocational Training (BIBB) outlined requirements for an international comparative study in VET, and the OECD is currently developing the foundation for the International Vocational Education and Training Assessment (PISA-VET) (BMBF, 2023). Existing vocational competence models face criticism for their broad definitions that include social, volitional, and motivational aspects, making it difficult to measure and compare competencies. Furthermore, in many instances, the definitions of vocational competence extend well beyond the vocational context. This ambiguity necessitates rigorous construct validation, which is currently insufficient (Rüschoff, 2019). One of our research aims is to address this conceptual gap.

The first step in constructing the competence model is to establish an operational definition to translate competence constructs into measurable observations. According to the definition by Klieme and Leutner (2006) that describes "competence as a cognitive disposition that is learnable and functionally related to specific situations," we understand competence as follows: (1) It encompasses only cognitive aspects, excluding motivational and volitional elements; (2) It differs from performance, which refers to what is actually done under existing circumstances (Messick, 1984). Competence is a potential cognitive function to act appropriately in various situations, which is not directly observable but can be inferred from observed behaviors (Winther, 2010; Winther & Achtenhagen, 2009). This definition supports using item-response theory to calculate response probabilities based on test-taker traits and item characteristics; (3) Competence is realized through interaction with specific performance requirements in situations (Connell et al., 2003; White, 1959), which span a range from general to specialized scopes (Klieme et al., 2008). This characteristic is essential for defining vocational competence in VET.

Vocational Competence in VET

PISA's findings on general education, though not covering VET-centric systems like Germany's, demonstrate how competence-oriented empirical data can drive educational reforms (Ertl, 2006). VET should not be assessed by the same criteria as academic education, given that company-based learning constitutes a significant part of vocational training. Recent research has focused on modeling vocational competence structures and levels to better understand professional skills development. Instruments have been developed for various professions, including commercial and technical training (Abele et al., 2014; Nickolaus et al., 2008, 2012, 2015), bank clerks (Lehmann & Seeber, 2007; Rosendahl & Straka, 2011), and industrial clerks (Deutscher & Winther, 2018; Winther & Achtenhagen, 2009). Most constructs deal

with vocational specialized competencies, with some including cross-domain competencies diagnosed in the area of vocational training (Rüschoff, 2019). For example, Winther (2011) distinguished between mathematical and literacy facets of general competence in commercial training, while Ziegler et al. (2016) measured general reading, math, and science competence in vocational contexts. Nine percent involve transversal competencies like social-communicative skills (Dietzen et al., 2016; Döring et al., 2016). This study focuses on occupational specialized competence, distinguishing it from general or transversal competence.

The Conception of Domain-Linked Competence and Domain-Specific Competence

VET research shows that cross-domain competence, including general basic knowledge and self-regulatory skills, influences occupation-specific competence across various training occupations (e.g., Lehmann & Seeber, 2007; Winther & Achtenhagen, 2008). However, this influence is limited and cannot be generalized. For example, basic mathematical skills correlate with performance in mathematically focused professional situations like controlling but not in general business administration (Winther, 2010; Winther & Achtenhagen, 2008), metacognitive knowledge and self-regulatory skills have low predictive power for commercial profession skills (Seeber, 2008; Winther, 2006). Therefore, in VET, domain-linked competence, also known as job-related literacy, is assumed to be more predictive than cross-domain general competence. Domain-linked competence pertains to general aspects relevant to a specific professional domain, linking cross-domain competence to specific situations (Winther et al., 2013). Additionally, domain-specific competence, which involves specific rules, principles, skills, and action plans related to a particular subject matter, addressing typical and concrete requirements in professional situations within specific occupational groups (Deutscher & Winther, 2016).

To explain the relationship between general competence, domain-linked competence, and domain-specific competence, Gelman & Greeno's theory (1989) provides valuable insights. Domain specificity, rooted in content central to a particular occupation, contrasts with domain-relatedness, which encompasses content supporting the occupational field but also applies to general or basic education in a broader context. For example, financial literacy, as defined by OECD (2020), includes knowledge and skills essential for making significant financial decisions—from selecting bank accounts and mortgages to investments and retirement planning. While financial literacy and digital literacy are crucial 21st-century skills applicable to daily life for everyone and serves as a foundational educational concept in commercial professions, it lacks the subject-specific focus necessary for economic occupations. One assumption, which has been empirically validated (Aprea et al., 2016), is that literacy significantly promotes ongoing learning in commerce-related professions.

Greeno et al. (1984) introduced the distinction between domain-linked and domain-specific competence. They proposed a framework within the domain of counting sets of objects, identifying three components of competence: (1) Conceptual competence involves an implicit understanding of general principles within the task domain, such as cardinality, one-to-one correspondence, and order. (2) Procedural competence encompasses understanding general principles related to goals and

actions within the task domain. For example, counting is linked to number because it involves determining the number of objects in a set. It also includes understanding relationships between necessary conditions and actions, such as equality in forming equal sets as a prerequisite for counting. (3) Utilizational competence focuses on understanding the relationships between specific task features and performance requirements. For instance, it considers how objects to be counted are arranged in a straight line within the task setting.

Applying Greeno et al.'s (1984) conceptual framwork to VET research provides a valuable framework for analyzing problem-solving competence within specific professional domains. Similar to the concrete task setting for counting objects, actions in commercial domains involve three key components: activating declarative knowledge of general principles (conceptual competence), selecting and executing action schemata based on task-specific logic (procedural competence), and integrating specific requirements of a defined situation by aligning general concepts and action principles with situational features (utilizational competence).

Greeno et al. (1984) described the implicit cognitive processes involved in domain-specific tasks, distinguishing between understanding general principles in a domain and applying them to specific task settings. Applying this framework to VET research offers several advantages: (1) It breaks down problem-solving effectiveness in job-related situations into distinct components rather than treating it as a singular construct. Utilizational competence, which addresses specific situational demands, develops gradually through vocational training and professional experience, while occupational literacy (conceptual and procedural competence) forms the general competence used across occupational fields and is acquired through general education. (2) This framework's dimensions can theoretically apply to all stages of vocational training, including initial stages where trainees may lack specific work experience but possess a foundational level of general competence. These stable competence dimensions provide a structural basis for developmental research tracking changes over time in vocational training contexts.

However, this conceptualization of competence primarily serves as a framework for curriculum design rather than as the foundation for a psychometric model to measure competence. The distinction between domain-linked and domain-specific competence is pivotal in constructing a psychometric unique to the VET field. Domain-linked competence includes conceptual competence and procedural competence. In the commercial trainings, it can be interpreted as so-called commercial core competencies, "which are required for practicing a profession in all commercial occupational fields and can therefore form a basis for commercial training and further education standards, albeit with different intensity depending on specified commercial occupational fields" (Brötz et al., 2009, p 0.19). For commercial professionals, domain-linked competence includes economic literacy and numeracy, essential skills within the commercial domain (Winther, 2010). For instance, an item in our assessment tool aimed at evaluating domain-linked competence presents a straightforward exchange-rate calculation. In this simulated scenario (see Sect. 3.2 for details), participants assumed the role of a businessperson and were required to calculate the expected USD payment based on the Euro amount specified in a contract, using the exchange rate also specified in the contract. This calculation involves applying

general mathematical operations to an economic context (currency exchange), leveraging provided information on both the euro total and exchange rate. In contrast, domain-specific competence is essential for addressing challenges in specific, narrow domains. In commercial settings, employees draw on domain-specific competence to navigate complex economic relationships and execute business transactions according to established protocols (Winther, 2010). An example of a domain-specific item in our assessment involves selecting a logistics company. Beyond basic mathematical calculations, test takers must analyze various aspects of the task scenario, considering factors such as company conditions, customer needs, and logistics company offerings like quotes, discounts, efficiency, and payment terms.

Domain-Linked Competence and Domain-Specific Competence as Two Separable but Related Dimensions of the Vocational Competence Construct

The distinction between domain-linked and domain-specific competences is crucial for VET, as trainees face both overlapping and activity-specific demands outlined in the Vocational Training Act (BBiG) and the Trade and Crafts Code of Germany (HwO) (Liedtke & Seeber, 2015; Reinisch & Götzl, 2013). The psychometric superiority of this two-dimensional structure is empirically supported (Klotz & Winther, 2015). Differentiating these competences is key to analyzing transitions between general cognitive abilities and specialized skills in VET (Winther, 2010). Contrasting with general education where cross-domain competence forms the cognitive foundation for specialized knowledge and skills (Leutner et al., 2004; Weinert, 2001), in VET, domain-linked competence serves as an "intermediate level" for successful action in a specific occupational area, (Deutscher & Winther, 2016) (see Fig. 1). Empirical evidence indicates that the concept of domain-linked competence has high predictive power for developing domain-specific competence in economic contexts (Achtenhagen & Winther, 2008; Winther et al., 2013). In commercial vocational training, this aligns with expectations, as general economic skills are essential for acquiring domain-specific skills in commercial fields (Deutscher & Winther, 2016).

In previous research, a competence development model for VET was conceptually developed for industrial clerks, based on domain-linked and domain-specific competence (Winther, 2011; Klotz & Winther, 2015). The transition from general schooling to VET is marked by domain-linked competence, reflecting trainees' prior numeracy and literacy skills acquired form general education. During VET, trainees acquire content from both competence dimensions, with domain-specific competence developing more rapidly and eventually becoming predominant by the end of training. This process continues throughout the training period, culminating in a vocational competence set where domain-specific competence is dominant. According to Klieme and Leutner (2006), a competence model with a structurally stable internal framework throughout training forms the basis for longitudinally establishing the hypothesized distinct growth trajectories of domain-linked and domain-specific competence. Hence, in addition to the first character emphasizing cognitive aspects and the second character focusing on vocational specialized competence, the third character of our competence model centers on distinguishing internal dimensions.

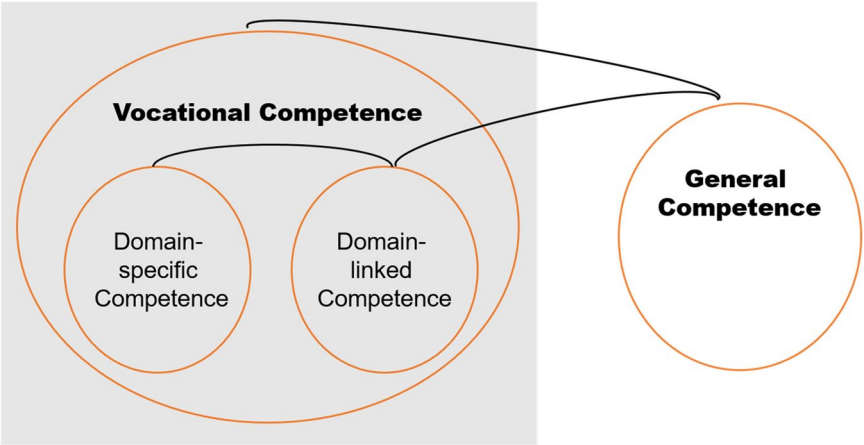


Fig. 1 Integrated Competence Model with General Competence and Vocational Competence. *Note.* Adapted from "Zusammenhänge zwischen allgemeinen und beruflichen Kompetenzen in der kaufmännischen Erstausbildung [Connections between general and vocational competencies in initial commercial training]," by E. Winther, J. Sangmeister, and A. K. Schade, in R. Nickolaus, J. Retelsdorf, E. Winther & O. Köller (Ebds.), *Mathematisch-naturwissenschaftliche Kompetenzen in der beruflichen Erstausbildung: Stand der Forschung und Desiderata* (Zeitschrift für Berufs- und Wirtschaftspädagogik—Beihefte; Band Beiheft 26) (p. 139 – 157), 2013, Franz Steiner Verl

Research Aim and Model

Guidelines for Assessing the Validity of the Vocational-Economic Competence Construct

Besides clearly and precisely defined models, a transparent, guideline-compliant validation procedure is crucial for advancing competence models in VET. There's a notable need for validity evidence based on internal structure and external relations, often missing in 78% and 88% of competence validation studies in Germany's initial VET from 2001–2017 (Rüschoff, 2019; Rohr-Mentele & Forster-Heinzer, 2021). Moreover, studies that do provide validity evidence often neglect measurement invariance across test occasions (e.g., Deutscher & Winther, 2016). According to the Standards for Educational and Psychological Testing (AERA, APA & NCME, 2014), internal structure evidence includes not only dimensionality but also measurement invariance across subgroups. This study aims to address the research gap by providing both empirical validity evidence based on internal structure and external relations and offer a well-structured validation process beyond existing research.

(1) Validity evidence based on test content: According to the Standards, important validity evidence is derived from logical or empirical analysis of the relationship between test content and the intended construct. The current assessment, adapted from validated domain-specific and domain-linked tasks, was developed by identifying relevant work activities and processes in the VET curriculum.

We assume the test content adequately represents the content domain and will not re-examine it in this study (see Sect. 3.2 and Deutscher & Winther, 2016 for more details).

(2) Validity evidence based on response processes: The Standards emphasize the importance of ensuring that judgments of test-takers' performance are based on appropriate standards and not influenced by irrelevant factors (e.g., handwriting quality in a written essay). Our competence construct includes only vocational competence, excluding general competence. We used a classic paper-and-pencil test to eliminate the influence of irrelevant general skills, such as digital literacy. Additionally, two versions of test booklets with different item orders were assigned to test-takers to minimize the influence of neighboring students and control for order effects. The Standards also addressed the evidence based on response processes relies on observers or judges recording and/or evaluating test takers' performances. In our study, raters underwent training sessions to familiarize themselves with scoring criteria and procedures. A subset of assessments was independently scored by three raters, achieving a Cohen's kappa of 0.84, which indicates high interrater reliability and ensures the consistency and accuracy of scoring procedures throughout the study.

(3) Validity evidence based on internal structure: We formulated hypotheses to test the internal structure of our construct. First, we hypothesized that the theoretical two-dimensional structure (domain-linked and domain-specific competence) would statistically outperform a unidimensional structure (H1a) and that the two dimensions would correlate positively (H1b). We also hypothesized that trainees would possess more domain-linked competence than domain-specific competence due to their prior general schooling (H1c). Additionally, we assumed measurement invariance across test versions (H2a) and federal states (H2b), in line with the Standards' expectation of invariance over occasions.

(4) Validity evidence based on relations to external variables: We investigates correlations between vocational competence scores and external performance-related variables, hypothesizing that:

- Vocational competence (both domain-linked and domain-specific) positively correlates with the final grades trainees aim to achieve in their training (H3a).

- Vocational competence positively correlates with trainees' self-evaluation of their overall performance in their training (H3b).

- The average grade from the last attended school before entering the training program correlates more strongly with domain-linked competence than with domain-specific competence at T0, as domain-linked competence mediates the relationship between general and domain-specific competence (H3c).

- Both domain-linked (H3d) and domain-specific competence (H3e) positively correlate with the concurrent average grade in vocational school, with a stronger correlation for domain-linked competence (H3f).

- Both domain-linked (H3g) and domain-specific competence (H3h) positively correlate with the concurrent average grade in the training company, with a stronger correlation for domain-specific competence (H3i).

Generalizability

Besides the validity, it is important to consider the generalizability of our research. While vocational training is firmly established and unique in Germany (BMBF, n.d.), VET is also a common focus internationally, aimed at preparing students for the workforce. Therefore, the insights from this research on vocational competence can be generalized internationally in several ways:

- Internationally Standardized Assessment: Developing an internationally standardized assessment for VET outcomes, such as PISA-VET (OECD, 2024), requires a valid vocational competence model adaptable to the diverse VET systems and training occupations across countries. The model studied here fits this requirement, as it is based on Greeno and colleagues' studies characterizing competence for cognitive tasks without specific learning objectives or curricula.
- Validation Procedure: Presenting a thorough and well-structured validation procedure is crucial for developing and validating VET assessments worldwide (Rüschoff, 2019; Rohr-Mentele & Forster-Heinzer, 2021).
- Globalization and Bilateral Cooperation: Globalization has led to increased bilateral cooperation with Germany in VET (GOVET, n.d.). Understanding Germany's dual education system can help other countries enhance their vocational training systems and foster international collaboration in designing more effective vocational education programs. This study, based on German vocational training objectives and curricula, aims to contribute to these efforts.

Psychometric Modeling of Vocational-Economic Competence

Domain-specific and domain-linked competences are two key psychometric properties of vocational-economic competence, defined as latent variables in the Multidimensional IRT (MIRT) model. There are two types of MIRT models: between-item and within-item multidimensionality (Adams et al., 1997). In between-item models, each item belongs to only one dimension, affecting the probability of a correct response on that dimension alone. In contrast, within-item models allow items to load onto multiple dimensions, meaning responses rely on abilities from several dimensions simultaneously (Hartig & Höhler, 2009). Additionally, in between-item models, the dimensions can correlate (Hartig & Höhler, 2009). In this study, each assessment item tests either domain-specific or domain-linked competence, loading onto one dimension. Thus, we chose the between-item model with correlated factors to model the empirically supported correlation between domain-specific and domain-linked competences (Fig. 2).

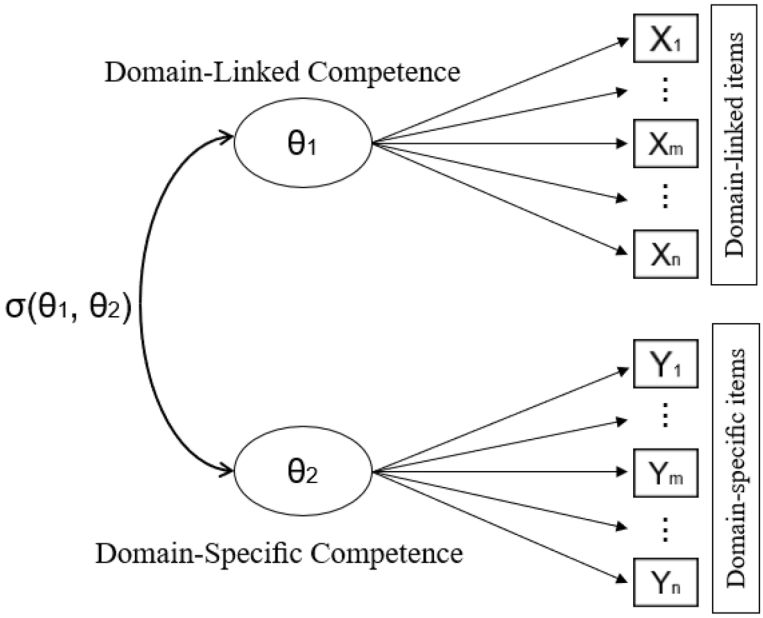


Fig. 2 Two-Dimensional Rasch Model with Between-Item Multidimensionality. *Note*. Adapted from "Multidimensional IRT models for the assessment of competencies," by J. Hartig, and J. Höhler, 2009, *Studies in Educational Evaluation, 35*(2–3), 57–63

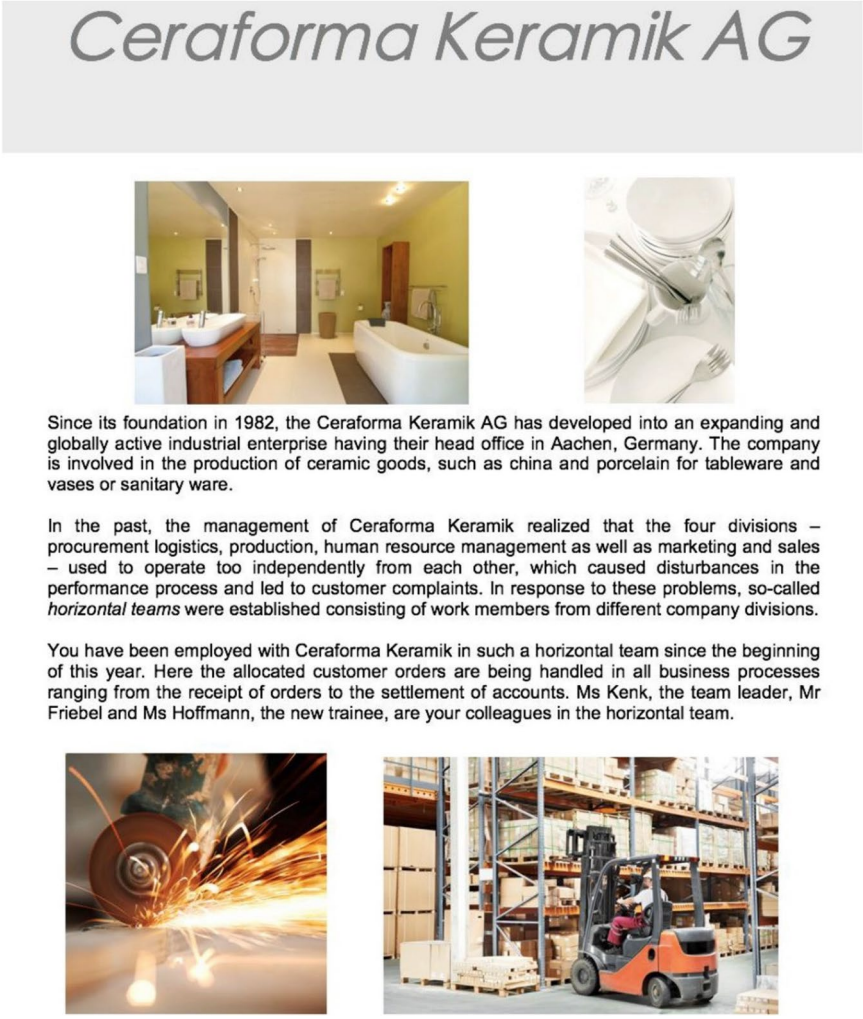
Methods

Samples

A total of 1438 commercial trainees from North Rhine-Westphalia (NRW; $n = 621$) and Baden-Württemberg (BW; $n = 817$), Germany, participated in this study. The sample comprised 837 females and 593 males, aged between 16 and 51 years (M = 20.96; SD = 2.83). Data collection spanned from October 2019 to December 2021, with competence assessments administered annually at the beginning of each apprentice year. For validating the competence structure, only data from the initial time point of data collection was utilized in this study.

Materials

Test items used in this study were adapted from a previously validated prototype of a competence-oriented authentic assessment designed for economic domain (Deutscher & Winther, 2018; Klotz, 2015), all based on scenarios from the simulated company CERAFORMA. These assessments replicate authentic business processes (see Fig. 3) with three key features: (1) structured complexity across three cognitive levels (Greeno et al., 1984); (2) vocational authenticity in realistic work situations and tasks; and (3) a process-oriented approach reflecting company operations and economic interrelations across departments. Items in the competence test



Ceraforma Keramik AG

Since its foundation in 1982, the Ceraforma Keramik AG has developed into an expanding and globally active industrial enterprise having their head office in Aachen, Germany. The company is involved in the production of ceramic goods, such as china and porcelain for tableware and vases or sanitary ware.

In the past, the management of Ceraforma Keramik realized that the four divisions – procurement logistics, production, human resource management as well as marketing and sales – used to operate too independently from each other, which caused disturbances in the performance process and led to customer complaints. In response to these problems, so-called *horizontal teams* were established consisting of work members from different company divisions.

You have been employed with Ceraforma Keramik in such a horizontal team since the beginning of this year. Here the allocated customer orders are being handled in all business processes ranging from the receipt of orders to the settlement of accounts. Ms Kenk, the team leader, Mr Friebel and Ms Hoffmann, the new trainee, are your colleagues in the horizontal team.

Fig. 3 A Business Simulation According to a Real Company as Test Environment

have various formats, such as open questions, multiple/dual choice, calculating tasks, and reasoning tasks.

The assessment framework consists of 24 items, with 11 items aimed at measuring domain-linked competence and 13 items targeting domain-specific competence. This paper-and-pencil assessment utilized two versions of test booklets (A/B), each with a different order of test items. The assignment of booklet versions (A or B) was randomized among participants to control for order effects. More information about test items see Appendix.

Results

Checking of Unidimensionality Versus Multidimensionality

Before calculating trainees' competence based on the theoretical model, two essential steps validate the construct. First, we determine the superiority of the two-dimensional model through model-fit analysis (Sect. 4.1). Second, we establish measurement invariance (Sect. 4.2). This section begins with dimensionality analyses of economic-vocational competence, crucial for subsequent measurement invariance testing.

The model fit of the two-dimensional construct was evaluated using the between-item multidimensional IRT model outlined in Sect. 2.3. This analysis compared the fit of the multidimensional Rasch model with a unidimensional model assumed to represent vocational-economic competence using the NRW dataset. Significance of the change in the -2LL statistic was assessed using a chi-square distribution, with detailed results presented in Table 1.

The deviance difference (178.10) is statistically significant at the 0.001 level, supporting the theoretical two-dimensional model as appropriate.

Additionally, when comparing the fit of the multidimensional Rasch model with the unidimensional model using the BW dataset (Table 2), significantly lower values of Deviance, AIC, and BIC indices indicate that the two-dimensional model aligns better with the test results. Therefore, H1a is confirmed.

CFA to Verify the Model of Vocational-Economic Competence

Two sets of CFA were conducted for two federal states respectively. The results of model fit statistics are summarized in Table 3.

The CFI cut-off scores are above 0.95. The values of RMSEA and SRMR are lower than cut-off scores of 0.05 and 0.08. The measured variables represent the factor structure excellently both in NRW data and BW data.

Table 1 The Comparison of Model Fit (NRW Dataset)

Model	Deviance	N of Parameter	AIC	BIC
unidimensional model	12443.54	23	14378.47	12591.46
multidimensional model	12265.44	25	12315.44	12426.22

Likelihood Ratio Chi-Squared $\chi^2 = 178.1053$, df = 2, $p < .001$

Table 2 The Comparison of Model Fit (BW Dataset)

Model	Deviance	N of Parameter	AIC	BIC
unidimensional model	14326.47	26	14378.47	14500.82
multidimensional model	14137.22	28	14193.22	14324.98

Likelihood Ratio Chi-Squared $\chi^2 = 189.2487$, df = 2, $p < .001$

Table 3 Goodness of Fit Measures for Datasets NRW and BW

	χ^2	df	*p*	RMSEA	CFI	SRMR
NRW	309.405	219	< .001	0.026	0.976	0.052
BW	277.300	291	< .001	0.018	0.989	0.033

Testing for Measurement Invariance

Measurement invariance assesses whether a psychometric construct, such as domain-specific and domain-linked competence in this study, has the same meaning across different groups (H2a and H2b). It ensures that scores of measured variables (e.g., scores for domain-specific and domain-linked competence) can be interpreted accurately across subsamples categorized by contextual conditions (e.g., booklet versions A and B, trainees from NRW and BW), or concretely, the differences in assessed scores accurately reflect differences in domain-specific and domain-linked competence, rather than differences due to group membership or version assignment. Measurement invariance testing proceeds through three sequential steps: (1) Fit the configural model to confirm consistent basic factor structure across groups. (2) Test for metric invariance by constraining factor loadings to equality across groups and comparing model fit. (3) Test for scalar invariance by further constraining item intercepts to equality and comparing model fit. These analyses will be conducted using Multi-Group Confirmatory Factor Analysis (MGCFA) in R.

Measurement Invariance Across Assessment Versions

- chi-square model fit test

Firstly, we ran MGCFA to estimate the same model for respondents with version A ($N_A = 699$) and version B ($N_B = 739$) separately. Table 4 displays the results of testing for metric, scalar, and strict measurement invariance.

The model fit indices show that the chi-square is rejected, which is common with large samples. However, the RMSEA, CFI, and SRMR indicate that metric, scalar, and strict measurement invariance are supported (see Hu & Bentler, 1999).

- chi-squared model fit difference tests

The metric model was compared to the configural model. The results indicate that the chi-square difference test was not statistically significant ($\Delta\chi^2 = 26.06$,

Table 4 Goodness of Fit Measures for factorial invariance – version A / B

	constrained parameters	χ^2	Df	*p*	RMSEA	CFI	SRMR
Configural Invariance	factorial structure	563.249	438	<.001	0.020	0.985	0.037
Metric Invariance	factor loadings	589.307	460	<.001	0.020	0.985	0.039
Scalar Invariance	factor loadings, intercept	642.572	482	<.001	0.022	0.981	0.040

df = 22, p = 0.25), suggesting equivalent fit of the metric model to the data. However, the chi-square difference test between the metric and scalar models was significant ($\Delta\chi^2 = 121.32$, df = 24, $p < 0.001$). Despite favorable model fit indices, this significant result indicates a lack of scalar invariance for the hypothesized construct of vocational-economic competence (see Chen, 2008). After releasing intercept constraints for two items across groups, partial scalar measurement invariance is achieved ($\Delta\chi^2 = 30.906$, df = 20, $p = 0.06$). This partially confirms H2a. Detailed reasons for the significant impact of these items on model fit are discussed in the Discussion section.

Measurement Invariance Across Federal States

- chi-square model fit test

The chi-square model fit test was conducted to determine whether measurement invariance between NRW ($N_{NRW} = 621$) and BW ($N_{BW} = 817$) was achieved. The model fit indices are reported in Tables 5 and 6.

The results show that the fit indices from the configural, metric, and scalar models are aligned with Hu and Bentler's guidelines (Hu & Bentler, 1999) for good model fit.

- chi-squared model fit difference tests

To test for measurement invariance across federal states, chi-square tests were performed to compare the fit of the metric model to the fit of the configural model, and the fit of the scalar model to the fit of the metric model. The following table shows the main outputs of the comparisons.

The Chi-square test to find the differences between the configural, metric, and scalar models for federated states provided p-values less than 0.05. It implies a non-invariance across multiple groups in different federated states. H2b is not confirmed.

Table 5 Goodness of Fit Measures for Factorial Invariance – NRW / BW

	constrained parameters	χ^2	df	*p*	RMSEA	CFI	SRMR
Configural Invariance	factorial structure	586.705	438	<.001	0.022	0.983	0.040
Metric Invariance	factor loadings	657.732	460	<.001	0.024	0.978	0.042
Scalar Invariance	factor loadings, intercept	847.589	482	<.001	0.032	0.959	0.057

Table 6 Chi-Square Difference Test for Factorial Invariance – NRW / BW

	χ^2	df	*p*
Metric Invariance – Configural Invariance	71.028	22	< .001
Scalar Invariance – Metric Invariance	189.857	22	< .001

Considering this non-invariant result, the theoretical hypothesized model of vocational-economic competence will be analyzed for NRW and BW separately.

Calculating Task Difficulty and Person Ability

Following checks for dimensionality and measurement invariance, we computed task difficulty and person ability using MRCMLM as described in Sect. 2.2. Table 7 presents descriptive statistics from separate analyses conducted for each federal state. Due to non-invariance, comparisons of trainees' competence between the two federal states will not be conducted.

Key insights from the (co)variance statistics in the table indicate that domain-specific and domain-linked competence correlate positively both in NRW data ($r(619) = 0.230$) and in BW data ($r(815) = 0.406$), confirming H1b. Contrary to expectations, participants showed higher competence in the domain-specific dimension than in the domain-linked dimension. Although EAP reliabilities for both competences are generally low, they are considered acceptable for a competence test rather than a psychological measurement, as discussed in detail in the Discussion section.

- Distribution of item thresholds and person abilities on domain-specific and domain-linked dimension

To gain deeper insights into competence allocation across two dimensions, the relationship between item difficulty and the distribution of person ability was visualized using the Wright Map (Figs. 4 and 5), also known as a person-item map. The Wright Map displays item parameters (right panel) alongside the distribution of person parameters along the latent dimension (left panel).

In the item panel, difficult items are positioned higher on the scale, indicating they require higher ability from test takers to endorse correctly, whereas easier items are lower on the scale. As a result, test takers with higher competence levels are depicted at the top of the map.

Several similarities can be noted between the two figures. Firstly, excluding outliers, the item difficulty distribution effectively spans the range of person ability distribution, affirming the assessment's overall accuracy. Secondly, person abilities

Table 7 Descriptive Statistics for the Two-dimensional Model Using NRW and BW Datasets Respectively

	Mean of person abilities (logits)		Covariance between dimensions	Variance		EAP Reliability	
	DL	DS		DL	DS	DL	DS
NRW	-0.061	0.155	0.222	0.881	1.055	0.606	0.532
BW	-0.038	0.260	0.552	1.260	1.466	0.658	0.571

Domain-specific competence as DS and domain-linked competence as DL for short

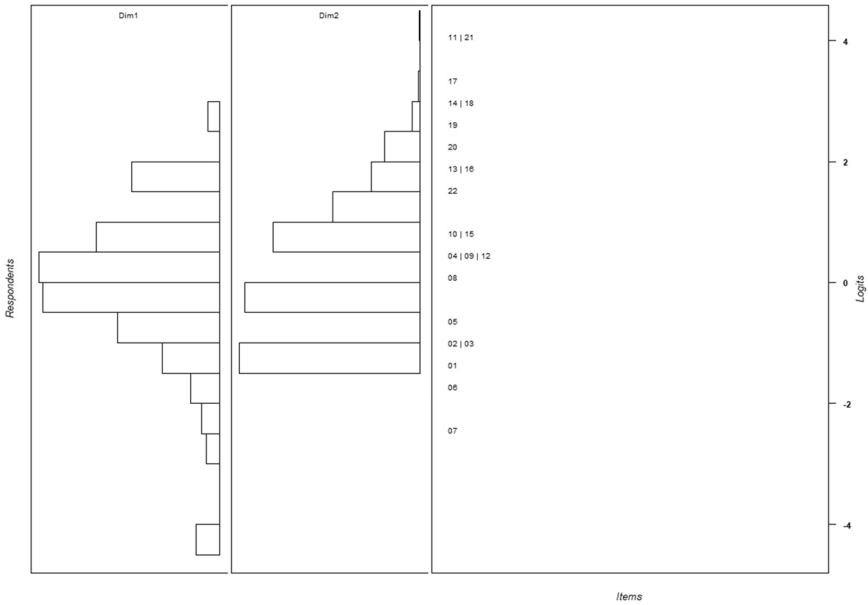


Fig. 4 Map of Latent Two-Dimensional Distribution of Item Thresholds and Person Abilities (NRW Dataset). *Note.* Dim 1 and Dim 2 stand for domain-linked and domain-specific dimensions respectively. Item 1 to item 10 test domain-linked competence and item 11 to item 22 belong to the domain-specific dimension

in the domain-linked dimension (Dim1) exhibit broader dispersion compared to the domain-specific dimension (Dim2), supporting the distinction between these dimesnions. Thirdly, histograms in the left panel show domain-linked competence values between 0 and 1 logits are most common, while domain-specific competence values between -1 and -2 logits prevail, indicating a greater proficiency in domain-linked tasks among test takers upon entering VET. While the mean values of domain-linked competence in lower, the insight from the graphics partially confirms the hypothesis (H1c) that participants initially possess more domain-linked competence. Lastly, both federated states show a similar distribution of item difficulty across the assessment, with items 01, 02, 03, 05, 06, and 07 being easier, items 04, 08, 09, 10, 12, and 15 moderate, and items 11, 13, 14, 16, 17, 18, 19, 20, and 21 more challenging relative to participants' abilities.

However, some considerations need to be addressed. While items cover the entire scale, indicating a comprehensive measure of competence, the difficulty parameters for domain-linked and domain-specific items only partially overlap with the latent trait parameters of test takers.

For the domain-linked dimension:

- In NRW, 27.1% of test takers are above the item difficulty range, and 4.5% are below it.
- In BW, 6.2% are above the range, and 1.7% are below it.
- It indicates a ceiling effect.

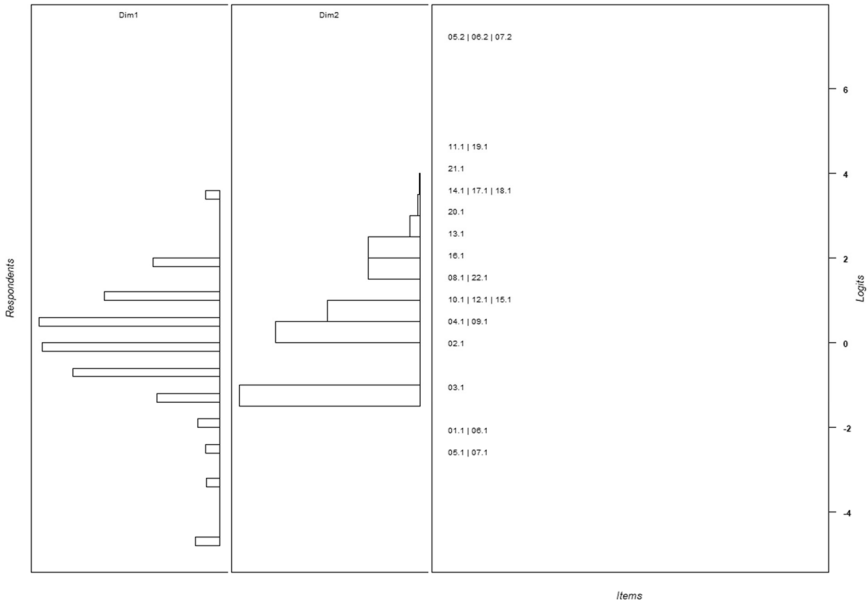


Fig. 5 Map of Latent Two-Dimensional Distribution of Item Thresholds and Person Abilities (BW Dataset)

For the domain-specific dimension:

- In NRW, 52% of test takers are below the item difficulty range, and the range covers 47.8% of their ability.
- In BW, 39.3% are within the item difficulty range, and 60.7% are below it.
- It indicates that there are few items that appears to be very difficult for test takers. The items are those located at the upper end portion of the Wright map in Figs. 4 and 5. Item 11 ($xsi_{_NRW} = 4.39$; $xsi_{_BW} = 3.93$) and 21 ($xsi_{_NRW} = 3.97$; $xsi_{_BW} = 3.99$) are two of the most difficult items for test takers from both federal states.; item 11 requires calculation of the total gross margin via order result or break-even point, and item 21 requires the calculation of the optimal order quantity.

Relations to External Variables

After testing the internal construct, validation evidence based on relations to external variables will be examined through correlations between the construct of interest and external variables measuring similar constructs. The results are summarized in Table 8.

Both domain-linked and domain-specific competences significantly correlate with trainees' targeted final grades and self-evaluations across all time points, confirming H3a and H3b. Only domain-linked competence correlates with the average grade

Table 8 Correlation Between Vocational Competence and External Variables

			Vocational competence					
			T0		T1		T2	
			DL	DS	DL	DS	DL	DS
External Variables	**T0**	Self-Evaluation of Overall Performance	0.161**	0.215**				
		Desired Final Grade	0.160**	0.203**				
		Average Grade of Last Attended School	0.109*	0.048				
	T1	Self-Evaluation of Overall Performance			0.176**	0.307**		
		Desired Final Grade			0.139**	0.219**		
	T2	Self-Evaluation of Overall Performance					0.257**	0.299**
		Desired Final Grade					0.271**	0.288**
		Current Average Grade in Company Training					0.147**	0.367***
		Current Average Grade in Vocational School					0.313***	0.297***

Domain-specific competence as DS and domain-linked competence as DL for short

$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

from trainees' previous general school attendance, confirming H3c. Additionally, both competences positively correlate with concurrent average grades in vocational school and training companies, confirming H3d, H3e, H3g, and H3h. Domain-linked competence shows a stronger correlation with vocational school grades compared to domain-specific competence, confirming H3f, while domain-specific competence correlates more strongly with training company grades, confirming H3i. While self-evaluation, desired final grade, and previous general school grade correlate with vocational competence, they have weaker effects. Performance in vocational school and company training, closely related to vocational competence, shows moderate effects, aligning with theoretical expectations.

Discussion

In this study, we addressed gaps in vocational competence research (see Rüschoff, 2019) by constructing a clear competence model focused on cognitive aspects and vocational competence, and conducting a comprehensive validation process. As recommended by the Standards (AERA, APA & NCME, 2014), we provided validity evidence based on internal structure, confirming the superiority of a two-dimensional structure (H1a), the correlation between dimensions (H1b), and the higher level of domain-linked competence compared to domain-specific competence (H1c). We also examined measurement invariance across test versions (H2a) and federal states (H2b). H1a and H1b are confirmed, H1c and H2a are partially confirmed, while H2b is not confirmed. Additionally, we provided validity evidence based on the relation with external variables (H3a-H3i), with all hypotheses concerning relations to external variables confirmed.

In general, The MIRT analysis results supported the hypothesized structure of vocational-economic competence in both NRW and BW test data. Established configural invariance between the two federal states suggests functional and structural consistency across groups, implying that the construct exists in all groups studied and that indicators relate to the basic model structure (Fontaine, 2011; van de Vijver & Leung, 1997). This indicates that commercial trainees from both federal states conceptualize vocational-economic competence similarly.

An ambiguous result of the study is its non-invariance between the datasets from two the federal states at the level of metric invariance. Despite good model fit coefficients, the significant difference between the configural and metric models indicates that the strength of relations between specific items and their respective latent dimensions varies across groups. A possible reason is that participants, at an early learning stage, accessed economic knowledge differently before VET, influencing their responses to VET curriculum-based items. They acquired economic competence may be more closely related to their previous informal learning experiences (e.g., through previous attendance at a commercial school) than to didactics in VET. Consequently, participants responded differently to the items designed according to the framework of the VET curriculum. This non-invariance may disappear in later stages of training and will be followed and monitored in follow-up research. Nevertheless, the two-dimensional model better structures vocational-economic

competence across both datasets. Another possible reason for this non-invariance is that the amplitudes for people ability (ranging from -4.67 to 3.51 logits for the BW dataset and from -4.20 to 2.85 logits for the NRW dataset) are much larger than those for item difficulty (ranging from -2.27 to 2.62 logits for the BW dataset and from -2.31 to 0.73 logits for the NRW dataset) in the domain-linked dimension. A ceiling effect may lead to the distortion of measurement.

For the domain-specific dimension, the scale seems adequate to measure the middle and upper ranges of the latent variable. This means that, at least in this sample, the domain-linked items may not be reliable to assess high levels of domain-linked competence, and the domain-specific items may not be reliable to assess low levels of corresponding competence. Another unexpected result is that the average value of domain-specific competence is higher than domain-linked competence (see H1c), and this is also attributed to the limited item difficulty ranges in both dimensions. Judging from the figures, H1c has indeed been confirmed, as the mode of domain-specific competence is between -2 and -1, whereas the mode of domain-linked competence occurs between 0 and 1. In order to improve measurement quality, it would be necessary to extend the range of item difficulty to cover competence at low, middle and high levels.

The two test items excluded during testing of partial scalar invariance are the final two items of version B. One of these items involves a complex calculation process, requiring participants to write down the detailed calculation steps for the task calculation of the defect rate in sink production, while the other requires making a judgment based on the cost analysis for quartz suppliers and providing a detailed rationale for that judgment. This type of question, demanding a thoughtful process and extensive written work, often creates pressure for participants, especially when these items are the last in the test, and participants respond under time constraints, potentially leading to non-invariance of the item intercepts.

The reliabilities of items in both dimensions do not meet our expectations. We assume that, similar to other instruments (Brüggemann & Nordmeier, 2018; Rutsch, 2016; Terzer, 2012; Wellnitz, 2012; Woitkowski, 2015), reliability for the competence test is low but acceptable. Reliability depends on the homogeneity of items and the similarity of the object being measured. Psychological assessments typically achieve high reliability by measuring specific traits, while competence tests cover a broad range of domain-specific knowledge and varied formats (e.g., open questions, multiple choice, calculations, reasoning), which can reduce reliability. Future studies should explore the impact of changing response formats on reliability. Additionally, in validation studies, reliability is just one aspect of validation evidence. Considering the test results for dimensionality and measurement invariance, we assume that this reliability is acceptable for this validation study. Given the low reliabilities, we interpret the test scores cautiously, focusing not only on assessment, dimensional levels, but also on item level. Item-level interpretation, based on content-related criteria, helps describe more accurately what learners know. For example, over 50% of trainees can solve domain-linked items 01, 04, 05, 06, 07, and 09, mostly related to 'formulating business emails according to DIN-Norm.' This example shows what previous knowledge the trainees come with and what further learning in the job may depend on.

The generalizability of our study's results requires further consideration. Given the German dual education system's emphasis on combining classroom and on-the-job training through apprenticeships, vocational education systems in other countries, like the U.S., typically focus more on classroom-based learning with fewer formal apprenticeships. We anticipate that while the dimensionality of vocational-economic competence holds across countries, the level of domain-specific competence may vary.

In recent years, vocational competence research has focused on modeling competence structures. Nickolaus et al., (2012, 2015), Abele et al. (2014), Lehmann & Seeber (2007), Rosendahl and Straka (2011), Winther and Achtenhagen (2009), Klotz & Winther (2015), and CoSMed project researchers (Seeber et al., 2016; Dietzen et al., 2016) have contributed insights across various occupational domains. These models are complementary rather than competitive, reflecting the specificity of each occupational domain in vocational education and training (VET). For example, computer simulations of an engine to measure the fault diagnosis skills of prospective automotive mechatronics engineers (see e.g., Gschwendtner et al., 2010) cannot be adapted to measure the vocational competence of industrial apprentices. Additionally, vocational competence is a unitary concept; these constructs of vocational competence may represent different perspectives of observation and illuminate distinct access to measure vocational competence. This distinction allows the intended interpretation of vocational competence for the proposed use. In the commercial domain, for example, the two-dimensional competence structure of Rohr-Mentele and Forster-Heinzer (2021) comprised of basic commercial knowledge and skills is developed for apprentices of all commercial branches and independently of a specific commercial occupation. Achtenhagen and Winther (2008) developed a competence structure model with action-based and skill-based dimensions, which fits advanced learners. As far as the model in this experiment is concerned, as the first step of longitudinal validation research, all the information gained from the present study will serve as an important piece of evidence for constructing vocational competence development. We hope in further longitudinal experiments to test whether the dimensionality is universal and stable over the duration of the training and add a developmental perspective to the dimensionality. Furthermore, we are looking forward to conducting more validation studies in other vocational domains and in other countries to test the generalizability of the construct.

The modeling and assessment of vocational competence have gained prominence recently, particularly within Germany's dual VET system, which facilitates a smooth school-to-work transition and bolsters economic competitiveness globally (Fürstenau et al., 2014; Rüschoff, 2019). To sustain its effectiveness, the system continuously updates to meet evolving societal and economic needs (Deissinger & Hellwig, 2005). International assessments like PISA VET provide crucial benchmarks, evaluating Germany's VET against global standards. These assessments inform policymakers about system strengths, weaknesses, and areas for improvement, guiding strategic modernization efforts. Furthermore, the outcomes of competence assessments in the dual VET system can be "translated" into concrete measures, actions, or plans where necessary to facilitate the competency development of trainees and employees (BMBF, 2023). The ongoing modernization efforts, alignment with

global standards, and international benchmarking through assessments like PISA VET collectively contribute to the system's resilience and effectiveness. The continuous translation of competence assessments into actionable plans underscores the commitment to the dynamic development of vocational competencies within the workforce. Implementing standardized measurements to gauge achievements across different proficiency levels within VET on a global scale creates prospects for collaboration among international VET sectors. This collaborative approach allows institutions to share best practices, exchange resources, and develop unified standards that enhance the quality and effectiveness of vocational training programs worldwide.

Appendix

Item Examples

Item 11 (domain-specific item): Your supervisor, Ms. Kenk, received a call from Mr. Schwienert, who is willing to accept the offer only if the price is set at 579.00 EUR per unit plus VAT. Ms. Kenk asks you to assess whether accepting this order would still be profitable for Ceraforma. For the upcoming team meeting, calculate the total contribution margin considering Mr. Schwienert's requested price, with variable costs of 395.00 EUR per unit and fixed order costs of 144,000.00 EUR

- Descriptive Statistics:
  - xsi_BW: 4.39
  - xsi_NRW: 3.93
- Factor Loading:
  - λ_BW: 0.79***
  - λ_NRW: 0.82***

Item 12 (domain-specific item):

After consulting with the Marketing/Sales department, it was determined that 25 tons of quartz need to be ordered for production. The following binding offers were received:

	Silica Glass GmbH	**Mineral Geifert AG**
List Price	525.00 EUR/ton (plus VAT)	470.00 EUR/ton (plus VAT)
Discount	4% for orders over 20 tons	3% for orders over 30 tons
Cash Discount	2% if paid within 10 days	
Payment Terms	60 days net	30 days net
Delivery Costs	100.00 EUR flat fee	0.60 EUR/100 kg ex-works
Minimum Order Quantity		10 tons
Offer Validity	Until 10th April	Until 25th March

Determine the most cost-effective supplier for 25 tons of quartz and provide reasons for your decision.

- Descriptive Statistics:
  - xsi_BW: 0.90.
  - xsi_NRW: 0.48.
- Factor Loading:
  - λ_BW: 0.84***
  - λ_NRW: 0.77***

Item 8 (domain-linked item).

After delivering 2,600 washbasins, 80 were reported by the customer as defective. The defects mainly occurred at the transitions to the purchased, chrome-plated metal parts and in the purchased components themselves. To reduce the defect rate and maintain quality leadership, Mrs. Kenk wants to expand the internal quality initiative. Three Measures to Reduce Defect Rate:

- Descriptive Statistics:
  - xsi_BW: 1.47.
  - xsi_NRW: 0.16.
- Factor Loading:
  - λ_BW: 0.78***
  - λ_NRW: 0.74***

Item 10 (domain-linked item).

After communicating your calculated price to Mrs. Kenk, she negotiates a contract with Miller Ltd. for 100 bathtubs at a total price of €180,000. You are asked to inform the accounting department of the expected payment in USD. Calculate the payment in USD using the following current exchange rate (note: round your result to two decimal places).

- Descriptive Statistics:
  - xsi_BW: 0.71.
  - xsi_NRW: 0.49.
- Factor Loading:
  - λ_BW: 0.85***
  - λ_NRW: 0.88***

Acknowledgements The project „Competence development in vocational enculturation processes "(INTERCONNECT), on which this publication is based, is funded by the German Research Foundation (DFG) under the grant number: WI 3597 / KL 3076. The authors are responsible for the content of this publication.

Authors' contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Beifang Ma, Maximilian Krötz and Esther Winther. The first draft of the manuscript was written by Beifang Ma. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials The datasets generated and analyzed during the current study are not publicly available due the fact that they constitute an excerpt of research in progress but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate In this non-interventional study (Survey Research), the data is completely anonymous with no personal information being collected. All subjects participated voluntarily. The participants provide their verbal informed consent to participate in this study. The Declaration of Helsinki was adequately addressed.

Consent for publication Not applicable.

Competing interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abele, S., Walker, F., & Nickolaus, R. (2014). Zeitökonomische und reliable Diagnostik beruflicher Problemlösekompetenzen bei Auszubildenden zum Kfz-Mechatroniker [Time-efficient and reliable diagnostics of vocational problem-solving skills in automotive mechatronics trainees.]. *Zeitschrift Für Pädagogische Psychologie, 28*, 167–179. https://doi.org/10.1024/1010-0652/a000138

Achtenhagen, F., & Winther, E. (2008). Wirtschaftspädagogische Forschung zur beruflichen Kompetenz-entwicklung [ecnomic education research for vocational competence development]. In BMBF (Eds.), *Kompetenzerfassung in pädagogischen Handlungsfeldern: Theorien, Konzepte und Methoden* (pp. 117–140). Bonn: BMBF.

Adams, R. J., Wilson, M. R., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1–23. https://doi.org/10.1177/0146621697211001

AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing: National Council on Measurement in Education*. Washington DC: American Educational Research Association.

Aprea, C., Wuttke, E., Breuer, K., Keng Koh, N., Davies, P., Greimel-Fuhrmann, B. & Lopus, J. (2016): Financial Literacy in the 21st Century: An Introduction to the International Handbook on Financial Litercy. In: C. Aprea, E. Wuttke, K. Breuer, N. Keng Koh, P. Davies, B. Greimel-Fuhrmann & J. Lopus (Eds.): *International Handbook on Financial Literacy. Springer.* https://doi.org/10.1007/978-981-10-0360-8_1

Brötz, R., Peppinghaus, B., Schapfel-Kaiser, F., & Brings, C. (2009). Gemeinsamkeiten und Unterschiede kaufmännisch-betriebswirtschaftlicher Berufe (GUK) – Ausgangspunkte und Ziele des

Forschungsprojekts [Similarities and differences between commercial and business administration professions (GUK) - starting points and goals of the research project]. In R. Brötz & F. Schapfel-Kaiser (Eds.), *Anforderungen an kaufmännisch-betriebswirtschaftliche Berufe aus berufspädagogischer und soziologischer Sicht. Berichte zur beruflichen Bildung* (pp. 19–43). Bielefeld: Bertelsmann.

Brüggemann, V., & Nordmeier, V. (2018). Naturwissenschaftliches Denken im Lehramtsstudium – Computeradaptive Leistungsmessung [Scientific thinking in teacher training courses – computer-adaptive performance measurement]. *PhyDid B - Didaktik Der Physik - Beiträge Zur DPG-Frühjahrstagung*, *1*. Abgerufen von https://ojs.dpg-physik.de/index.php/phydid-b/article/view/892

Bundesministerium für Bildung und Forschung [BMBF] (2023). Internationale Kompetenzmessung: Berufsbildungs-PISA [International competence measurement: Vocational training PISA]. Online available: https://www.bmbf.de/bmbf/de/europa-und-die-welt/pisa-in-der-berufsbildung/pisa-in-der-berufsbildung_node.html

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Connell, M. W., Sheridan, K., & Gardner, H. (2003). On abilities and domains. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 126–155). Cambridge University Press. https://doi.org/10.1017/CBO9780511615801.007

Deissinger, T., & Hellwig, S. (2005). Apprenticeships in Germany: Modernising the Dual System. *Education and Training, 47*(4–5), 312–324.

Deutscher, V., & Winther, E. (2016). Zur Entwicklung domänenverbundener und domänenspezifischer Kompetenz im Ausbildungsverlauf: Eine Analyse für die kaufmännische Domäne [On the development of domain-linked and domain-specific competence during VET: An analysis for the commercial domain]. *Zeitschrift für Erziehungswissenschaft, 19*(4). https://doi.org/10.1007/s11618-016-0687-1.

Deutscher, V., & Winther, E. (2018). A Conceptual Framework for Authentic Competence Assessment in VET: A Logic Design Model. In S. McGrath (Ed.), *Handbook of Vocational Education and Training: Developments in the Changing World of Work* (pp. 317–338). Cham.

Dietzen, A., Tschöpe, T., Monnier, M., & Srbeny, C. (2016). Berufsspezifische Messung sozialer Kompetenzen auf der Basis eines Situational Judgment Tests bei Medizinischen Fachangestellten im Projekt CoSMed [Job-specific measurement of social competencies based on a situational judgment test for medical assistants in the CoSMed project]. In: K. Beck, M. Landenberger& F. Oser (Eds.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT. Wirtschaft - Beruf - Ethik vol. 32* (pp. 225–241). Bielefeld: Bertelsmann.

Döring, O., Weyland, U., W, E., Nauerth, A., Hartig, J., Kaspar, R., Möllers, M., Rechenbach, S., Simon, J., Worofka, I., & Kraus, K. (2016). Technologiebasierte Messung beruflicher Handlungskompentenz in der Pflege älterer Menschen: Kompetenzmodellierung und Testverfahrensentwicklung [Technology-based measurement of vocational competence in the care of older people: competence modeling and test procedure development]. In A. Dietzen, R. Nickolaus, B. Rammstedt & R. Weiß (Eds.), Kompetenzorientierung - Berufliche *Kompetenzen entwickeln, messen und anerkennen* (pp. 117–132). Bielefeld: Bertelsmann Verlag. Education, Vol. 10 No. 2, pp. 160 74

Ertl, H. (2006). Educational Standards and the Changing Discourse on Education: The Reception and Consequences of the PISA Study in Germany. *Oxford Review of Education*, *32*(5), 619–634. http://www.jstor.org/stable/4618685

Fontaine, J. R. J. (2011). A fourfold conceptual framework for cultural and cross-cultural psychology: Relativism, construct universalism, repertoire universalism and absolutism. In F. J. R. van de Vijver, A. Chasiotis, & S. M. Breugelmans (Eds.), *Fundamental questions in cross-cultural psychology* (pp. 165–189). Cambridge University Press. https://doi.org/10.1017/CBO9780511974090.008

Fürstenau, B., Pilz, M., & Gonon, P. (2014). The dual system of vocational education and training in Germany – What can be learnt about education for (other) professions. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International Handbook of Research in Professional and Practice-based Learning* (pp. 427–460). Springer.

Gelman, R., & Greeno, J. G. (1989). On the nature of competence: Principles for understanding in a domain. In L. B. Resnick (Eds.), *Knowing and learning: Essays in honor of Robert Glaser* (pp. 125–186). Hillsdale: Erlbaum Associates.

German Office for International Cooperation in Vocational Education and Training (GOVET). (n.d.). Welcome to GOVET. Retrieved from https://www.govet.international/en/54894.php

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology, 16*, 94–143.

Gschwendtner, T., Geißel, B., & Nickolaus, R. (2010). Modellierung beruflicher Fachkompetenz in der gewerblich-technischen Grundbildung [Modeling professional expertise in basic industrial and technical education]. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes (Zeitschrift für Pädagogik, 56. Beiheft, S.258–269).* Weinheim: Beltz.

Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation, 35*(2–3), 57–63.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Hogrefe & Huber Publishers.

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 for balancing educational processes. Description of a newly established priority program of the DFG]. *Zeitschrift für Pädagogik 52*(6), 876–903.

Klotz, V. K. (2015). *Diagnostik beruflicher Kompetenzentwicklung: Eine wirtschaftsdidaktische Modellierung für die kaufmännische Domäne* [Diagnosis of professional development. A didactic modeling for the commercial domain]. Berlin:Springer.

Klotz, V. K., & Winther, E. (2015). Kaufmännische Kompetenz im Ausbildungsverlauf – Befunde einer pseudolängsschnittlichen Studie [Commercial competence during VET - findings from a pseudo-longitudinal study]. *Empirische Pädagogik, 29*(1), 61–83.

Lehmann, R., & Seeber, S. (Eds.). (2007). *ULME III. Untersuchung von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen [ULME III. Research of the performance, motivation and attitudes of students in the final classes of vocational schools].* Behörde für Bildung und Sport der Freien und Hansestadt Hamburg. Hamburger Institut für berufliche Bildung HIBB), Hamburg.

Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen [Problem-solving]. In PISA-Konsortium Deutschland (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 147– 175). Münster: Waxmann.

Liedtke, M., & Seeber, S. (2015). Modellgeltungstests und Einflussfaktoren auf differentielle Itemfunktionen in einem computergestützten Assessment für kaufmännische Berufe [Model validity tests and influencing factors on differential item functions in a computer-aided assessment for commercial occupations]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik, 111*(2), 242–267.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement, 21*(3), 215–238.

Nickolaus, R., Abele, S., Gschwendtner, T., Nitzschke, A., & Greiff, S. (2012). Fachspezifische Problemlösefähigkeit in gewerblich-technischen Ausbildungsberufen – Modellierung, erreichte Niveaus und relevante Einflussfaktoren [Subject-specific problem-solving ability in commercial-technical training occupations - modeling, levels achieved and relevant influencing factors]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik, 108*(2), 243–272.

Nickolaus, R., Gschwendtner, T., & Geißel, B. (2008). Entwicklung und Modellierung beruflicher Fachkompetenz in der gewerblich-technischen Grundbildung [Development and modeling of professional expertise in basic industrial and technical education]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik, 104*(1), 48–73.

Nickolaus, R., Nitzschke, A., Maier, M., Schnitzler, A., Velten, S., & Dietzen, A. (2015). Einflüsse schulischer und betrieblicher Ausbildungsqualitäten auf die Entwicklung des Fachwissens und die spezifische Problemlösekompetenz [Influences of school and company training qualities on the development of specialist knowledge and specific problem-solving competence]. *Zeitschrift Für Betriebs- Und Wirtschaftspädagogik, 111*(3), 333–358.

OECD. (2020). *OECD/INFE 2020 international survey of adult financial literacy.* OECD Publishing. https://doi.org/10.1787/bb2cd70c-en

OECD. (2024). PISA Vocational Education and Training (VET): Assessment and Analytical Framework, PISA. Paris: OECD Publishing. https://doi.org/10.1787/b0d5aaf9-en

Reinisch, H., & Götzl, M. (2013). Berufsgruppenbildung im Bereich kaufmännisch-betriebswirtschaftlicher Berufe aus historischer Sicht [Professional group formation in the area of commercial and business administration occupations from a historical perspective]. *Berufsbildung in Wissenschaft Und Praxis, 42*(3), 20–23.

Rohr-Mentele, Silja & Forster-Heinzer, Sarah. (2021). Practical validation framework for competence measurement in VET: a validation study of an instrument for measuring basic commercial knowledge and skills in Switzerland. *Empirical Research in Vocational Education and Training. 13*. https://doi.org/10.1186/s40461-021-00122-2.

Rosendahl, J., & Straka, G. (2011). Kompetenzmodellierung zur wirtschaftlichen Fachkompetenz angehender Bankkaufleute [Competence modeling for the economic expertise of aspiring bank clerks]. *Zeitschrift Für Betriebs- Und Wirtschaftspädagogik, 107*(2), 190–217.

Rüschoff, B. (2019). Methoden der Kompetenzerfassung in der beruflichen Erstausbildung in Deutschland. Eine systematische Überblicksstudie [Methods competence measurement in initial vocational training in Germany. A systematic overview study]. Wissenschaftliche Diskussionspapiere, 206. Bonn: Bundesinstitut für Berufsbildung.

Rutsch, J. (2016). *Entwicklung und Validierung eines Vignettentests zur Erfassung des fachdidaktischen Wissens im Leseunterricht bei angehenden Lehrkräften.* Dissertation *[Development and validation of a vignette test to assess subject-didactic knowledge in reading lessons for future teachers]*. Pädagogische Hochschule Heidelberg.

Seeber, S., Schumann, M., Ketschau, T., Rüter, T., & Kleinhans, J. (2016). Modellierung und Messung von Fachkompetenzen Medizinischer Fachangestellter (CoSMed) [Modeling and measurement of professional competencies of medical assistants (CoSMed)]. In F. Oser, M. Landenberger, & K. Beck (Eds.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung – Resultate aus dem Forschungsprogramm ASCOT* (pp. 205–223). Bielefeld: Bertelsmann.

Seeber, S. (2008). Ansätze zur Modellierung beruflicher Fachkompetenz in kaufmännischen Ausbildungsberufen [Approaches to modeling professional expertise in commercial training occupations]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik, 104*, 74–97.

Terzer, E. (2012). *Modellkompetenz im Kontext Biologieunterricht - Empirische Beschreibung von Modellkompetenz mithilfe von Multiple-Choice Items [Model competence in the context of biology lessons - Empirical description of model competence using multiple-choice items]*. Dissertation. Humboldt-Universität zu Berlin, Berlin. Mathematisch-Naturwissenschaftliche Fakultät I.

van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Sage Publications, Inc.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe & Huber Publishers.

Wellnitz, N. (2012). Kompetenzstruktur und -niveaus von Methoden naturwissenschaftlicher Erkenntnisgewinnung [Competence structure and levels of scientific knowledge acquisition methods]. In A. Sandmann & P. Schmiemann (Eds.), *Biologie lernen und lehren* (Bd. 2). Berlin: Logos.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*(5), 297–333. https://doi.org/10.1037/h0040934

Winther, E. (2011). Das ist doch nicht fair! – Mehrdimensionalität und Testfairness in kaufmännischen Assessments [That's not fair! – Multidimensionality and test fairness in commercial assessments]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 107*(2), 218–238. https://doi.org/10.25162/zbw-2011-0012.

Winther, E. (2006). *Motivation in Lernprozessen. Konzepte in der Unterrichtspraxis von Wirtschaftsgymnasien [Motivation in learning processes. Concepts in the teaching practice of business high schools]*. Wiesbaden: DUV.

Winther, E. & Achtenhagen, F. (2008). ‚Konzeptuale Kompetenz' und ‚Selbstregulation' als Grundlagen einer berufsbezogenen Kompetenzforschung ['Conceptual competence' and 'self-regulation' as the foundations of job-related competence research]. In D. Münk, P. Gonon, K. Breuer & T. Deißinger (Eds.), *Modernisierung der Berufsbildung. Neue Forschungserträge und Perspektiven der Berufs- und Wirtschaftspädagogik* (pp. 100–110). Opladen: Barbara Budrich.

Winther, E., Sangmeister, J., & Schade, A. K. (2013). Zusammenhänge zwischen allgemeinen und beruflichen Kompetenzen in der kaufmännischen Erstausbildung [Correlations between general and vocational competencies in initial commercial training]. In R. Nickolaus, J. Retelsdorf, E. Winther, & O. Köller (Eds.), *Mathematisch-naturwissenschaftliche Kompetenzen in der beruflichen Erstausbildung: Stand der Forschung und Desiderata* (Zeitschrift für Berufs- und Wirtschaftspädagogik

- Beihefte; Band Beiheft 26) (pp. 139 – 157). Franz Steiner Verl. http://www.steiner-verlag.de/programm/fachbuch/berufspaedagogik/reihen/view/titel/59600.html.

Winther, E. (2010). *Kompetenzmessung in der beruflichen Bildung [Competency measurement in vocational training]*. W. Bertelsmann Verlag.

Winther, E., & Achtenhagen, F. (2009). Skalen und Stufen kaufmännischer Kompetenz [Scales and levels of commercial competence]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik, 105*(4), 521–556.

Woitkowski, D. (2015). Fachliches Wissen in der Hochschulausbildung – Konzeptualisierung, Messung, Niveaubildung [Specialized knowledge in higher education – conceptualization, measurement, level formation]. In H. Niedderer, H. Fischler, & E. Sumfleth (Eds.), *Studien zum Physik- und Chemielernen.* (Vol. 185). Berlin: Logos.

Ziegler, B., Frey. A., Seeber, S., Balkenhol, A., & Bernhardt, R. (2016). Adaptive Messung allgemeiner Kompetenzen (MaK-adapt) [Adaptive measurement of general competences (MaK-adapt)]. In K. Beck, M. Landenberger & F. Oser (Eds.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (pp. 33–54). Bielefeld: wbv.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.