

RESEARCH

Open Access



Evaluating the pedagogical content knowledge of pre- and in-service teachers of business and economics to ensure quality of classroom practice in vocational education and training

Christiane Kuhn^{1*}, Alicia C. Alonzo² and Olga Zlatkin-Troitschanskaia¹

*Correspondence:

christiane.kuhn@uni-mainz.de

¹ Department of Business and Economics Education, Johannes Gutenberg University, Jakob Welder-Weg 9, 55099 Mainz, Germany
Full list of author information is available at the end of the article

Abstract

Background: Research on teachers' professional knowledge and skills is becoming increasingly important. However, valid assessments of teachers' knowledge and skills are scarce. In this study, we report on our development and validation (following the international Standards for Educational and Psychological Testing) of a situation-based tool to assess pre-service and in-service teachers' pedagogical content knowledge (PCK) of business and economics in vocational education and training.

Methods: We present a conceptual model and empirical evidence indicating the validity of test content and internal structure of the new developed assessment. We conducted conceptual analyses, interviews and standardized ratings with experts, and piloted the tool with 338 pre-service and in-service teachers.

Results: Our results indicate that our assessment tool is valid and reliable to measure case-related PCK (a central component of situation-oriented knowledge) of pre-service and in-service teachers.

Conclusions: A video- and performance-based assessment tool for assessing the strategic component of PCK is developed and currently being validated. Together, these tools can serve to assess the professional knowledge and skills of business and economics teachers.

Keywords: Pedagogical content knowledge, Teacher education and training, Video- and performance-based assessment, Business and economics teaching

Relevance and background

Empirical research on teachers' professional knowledge and skills is becoming increasingly important for practice in teacher education and especially in vocational education and training (De Wever et al. 2016). Several debates on educational policies have ensued, and discourse on this topic has featured political catchphrases such as 'outcome orientation' (Cochran-Smith et al. 2012; European Commission 2007). The central theme of the discussion on outcome orientation is that both the educational system and teaching can

become more successful by focusing less on curricula and teaching objectives, and more on empirically assessed outcomes, such as student learning and resulting knowledge and skills. This also applies to the domain of teacher education.

Empirical findings on the acquisition and promotion of professional knowledge and skills of pre-service and in-service teachers provide indications about the quality and effectiveness of teacher education programs (Blömeke et al. 2011; Fritsch et al. 2015; Kleickmann et al. 2012), which will affect instructional quality and, ultimately, student achievement (Baumert et al. 2010; Hill et al. 2005). Preliminary empirical research in this area, especially on teaching science and mathematics (e.g., Rohaan et al. 2009) has shown that assessments of professional knowledge and skills during teacher education and professional practice require appropriate field-specific assessment instruments, as they determine the quality of analyses and results and, therefore, their consideration in educational policies and practices (Mislevy and Haertel 2006; Zlatkin-Troitschanskaia et al. 2015). However, objective, reliable, and valid assessments of teachers' knowledge and skills across institutions are still scarce in most countries today (Darling-Hammond et al. 2013).

It is undisputed that teachers' professional knowledge is key to their professional performance (Desimone 2009; De Wever et al. 2016). Assessment of teachers' professional knowledge and skills often is based on Shulman's (1986) description of teacher knowledge, which comprises general pedagogical knowledge, subject matter knowledge, curricular knowledge, and pedagogical content knowledge (PCK). Various studies have investigated teachers' general pedagogical knowledge (Gold and Holodynski 2015; König et al. 2011); however, domain-specific facets of teachers' professional knowledge have been assessed in only a few subjects, mainly in mathematics (see review in Depaepe et al. 2013) and science (de Jong and van Driel 2004; Schmelzing et al. 2013; Veal et al. 1999). The lack of assessment of teacher knowledge, particularly PCK, in subjects with more complex curricular layouts may be due to challenges of agreeing on common teaching methods and content structures (Carlson 1990).

This is particularly an issue in the field of business and economics. On the one hand, business and economics is one of the most popular fields of study in vocational education and training (Organisation for Economic Co-operation and Development 2014; German Federal Statistical Office 2014). Economic and financial literacy is becoming increasingly important at all stages of education as well as of professional and social life, not least because of the need to better understand and handle the consequences of momentous economic events such as the recent financial crisis (Blinder 2015; Council for Economic Education [CEE] 2011). Accordingly, business and economics has also been introduced as a school subject in several countries, including the United States and Germany. On the other hand, in secondary education, business and economics is still rarely taught systematically or may be organized very flexibly, for example, according to fields of professional practice in commercial vocational education in Germany. Hence, from the point of view of teacher education, teachers of business and economics need to be prepared for a rather diverse set of content areas, teaching methods, and contexts, and be able to specialize depending on their group of learners and type of school. This diversity poses great challenges to standardized assessment of teachers' domain-specific knowledge. In this paper, we use the example of PCK in business and economics to show how these challenges can be addressed, step by step, in the development of tools for assessing teachers' PCK.

First, we present the current state of research on the valid assessment of teachers' PCK of business and economics. Next, we present our newly developed assessment instrument and discuss its strengths and limitations, including evidence of its reliability and validity for assessing case-related PCK. Finally, we describe a follow-up study currently under way, which validates a complementary performance-based tool to assess the strategic component of PCK of business and economics using video vignettes, and discuss uses of both tools and implications for valid assessment of the professional knowledge of business and economics teachers.

Current state of research on valid assessment of pre- and in-service teachers' PCK of business and economics

Recent research indicates that success in studying business and economics depends significantly on students' prior knowledge of business and economics acquired at secondary or vocational schools (Brückner et al. 2015; Gill and Gratton-Lavoie 2011; Walstad 2013). While this prior knowledge can be appropriately assessed, teachers' PCK of business and economics, which shapes the quality of instruction and likely influences student learning, has not yet been assessed directly, largely due to a lack of appropriate instruments for valid assessment (Zlatkin-Troitschanskaia et al. 2015).

In addition to curricular and subject matter knowledge, teachers' PCK is a crucial part of their domain-specific professional knowledge (Baxter and Lederman 1999; Berry et al. 2008). PCK is recognized widely as an important indicator of students' learning outcomes (Grossman et al. 1989; Hill et al. 2005). According to Shulman (1986), PCK includes 'the ways of representing and formulating the subject that make it comprehensible to others,' and 'an understanding of what makes the learning of specific topics easy or difficult' (p. 9). Therefore, appropriate tools to analyze teachers' PCK should have a strong orientation towards practical situations of everyday classroom teaching of the particular subjects.

Although there are some internationally validated instruments to assess pre-service and in-service teachers' knowledge of business and economics (Walstad et al. 2007; Förster et al. 2015), instruments to assess teachers' PCK of business and economics are not yet available internationally, or focused on accounting (Fritsch et al. 2015). The Mexican Higher Education Exit Assessments Test (EGEL) on accounting education by the National Center for the Evaluation of Higher Education is used nationally to assess learning outcomes of pre-service accounting teachers (Vidal Uribe 2013). In some states in the United States, pre-service teachers must pass the Praxis II test on business education to obtain the qualification to teach business administration (Educational Testing Service [ETS] 2011). However, this test assesses predominantly subject matter knowledge and some general pedagogical aspects rather than subject-specific requirements for classroom teaching of business and economics. In this paper, we present a newly developed and validated, situation-based instrument to assess pre-service and in-service teachers' PCK of business and economics.

Aim of the study

The aim of this study is to enhance teacher training in business and economics and to advance early-stage research on teachers' PCK. We address conceptual, methodological, and practical challenges of reliable and valid assessment of PCK, including how to

systematically design a test, how to confirm criteria of psychometric quality, and how to undertake comprehensive validation (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education [AERA, APA, and NCME] 2014). Regarding validation criteria, we present evidence of the two criteria of test content and internal structure that are fundamental for all tests. Our validation methods include conceptual analysis, interviews with and ratings by experts, as well as a quantitative assessment of pre-service and in-service business and economics teachers.

Validating the assessment of PCK of business and economics

Validation criteria

Test validity is ‘the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests’ (AERA et al. 2014, 11). Very early in the test development process, developers must not only define the construct to be assessed, but also create a validity framework that aligns theoretical and empirical evidence of the test scores and the conclusions to be drawn from them (Pellegrino et al. 2001; Kane 2006).

In our validation analyses, we examined the extent to which the theoretical construct explained the test scores and the underlying item response processes. Following the international Standards for Educational and Psychological Testing (AERA et al. 2014), we built a comprehensive validity argument based on theoretical and empirical evidence of (1) test content, (2) response processes, (3) internal structure, and (4) relationships to other variables. While our validation process addressed all four criteria and related hypotheses, in this paper, we focus on the criteria of test content and internal structure (e.g., Carlton 2012; Knupp and Harris 2012; Wilson 2005), which provide a vital indication of the test’s usability in business and economics teacher education and professional teaching practice. Analyses of response processes and relationships to other variables have been reported in Kuhn (2014) and Kuhn et al. (2013, 2014), and are summarized at the end of the paper.

Conceptual framework for PCK of business and economics

Theoretical model of the structure of PCK

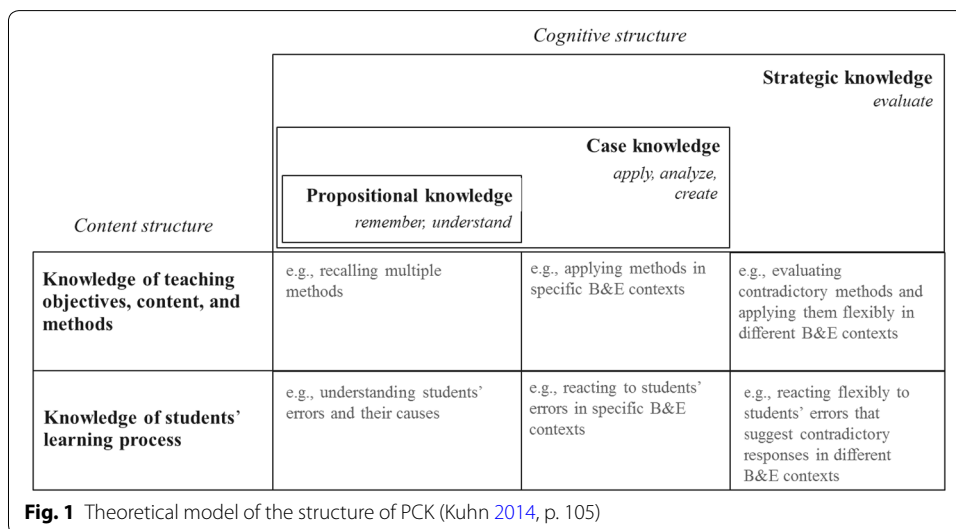
All validity analyses should begin with a clear construct definition. Our construct was based on Shulman’s definition of PCK (Shulman 1986) and a general consideration of modeling and measurement of knowledge and practical skills (Shavelson 2013). Our multi-perspective analyses involved a thorough examination of similar studies in other fields to maintain compatibility with other PCK assessment instruments (e.g., Blömeke et al. 2011; Rohaan et al. 2009; Schmelzing et al. 2013). General standards for teaching and teacher education, guidelines and curricula for business and economics teacher education and professional teaching practice, and international business and economics content standards were examined (e.g., National Council for Accreditation of Teacher Education 2008, Standard 1b; Council for Economic Education [CEE] 2010).

The theoretical construct of knowledge generally consists of two facets: content structure and cognitive structure (Bromme 1995; Shulman 1986). Hence, our comprehensive theoretical model of the construct of PCK of business and economics specified (1) the content structure with regard to subject-specific areas of teaching knowledge and (2)

the cognitive structure in terms of types of knowledge and corresponding cognitive processes (see Fig. 1).

First, consistent with similar studies in other domains (e.g., van Driel et al. 1998), our model defines PCK as consisting of two content components. The first content component is knowledge of teaching objectives, content, and methods, although others have also referred to a narrower category of knowledge of instructional representations. The interdependence of teaching objectives, content, and methods represents the traditional European discourse on didactics (e.g., Bromme 1995; Hopmann and Riquarts 1995). A teaching objective for business and economics, for example, might be to teach students to weigh decisions by considering economic and ecological factors; content might refer to the area of sales; and an appropriate method might be a role-play. These may refer to a mid-term perspective (i.e., with respect to curriculum) and also to a short-term perspective (i.e., with respect to lesson planning). The second content component is knowledge of students' learning process, also referred to as knowledge of learning difficulties, which includes awareness of common errors students make and their causes. Both content components are embedded in a broader context, which, for the vocational sector specifically, is characterized by a high heterogeneity (i.e., with respect to students' prior knowledge), among other things (Kuhn 2014).

Second, our comprehensive theoretical model characterizes the cognitive structure of PCK of business and economics in accordance with Shulman's (1986) differentiation of three types of teacher knowledge: propositional, case, and strategic knowledge. Following Anderson and Krathwohl (2001; Shulman 1987), we specified types of knowledge with regard to underlying cognitive processes, organized hierarchically according to complexity. Propositions present complex ideas in an abstract, simplified way. Thus, propositional PCK presents general principles of effective subject-specific teaching. However, propositional PCK is detached from specific contexts, providing little indication as to how it can be applied in practical teaching situations (Shulman 1986). In our model, we associated teachers' propositional knowledge with the cognitive processes of remembering and understanding. Thus, to demonstrate this most basic level



of knowledge, teachers need to remember and understand various general propositions about teaching and learning, such as knowledge of subject-related teaching principles and their significance for teaching, knowledge of certain key areas of business and economics teaching, or knowledge of student difficulties and their potential causes (Shulman 1987).

To be able to apply general propositions in specific contexts, a second type of knowledge is needed. For Shulman (1986), case knowledge is embedded in context-specific instances of practice. In our model, we linked teachers' case knowledge to the cognitive processes of analyzing, applying, and creating. We assume that case knowledge enables teachers to analyze and apply general propositions about teaching, learning, and subject matter in this field to specific business and economics teaching contexts and to create their own solutions for their lessons (for example, the didactic principle of conflict-orientation can be represented in a vocational class in retail business by a typical argument acted out by students playing the salesperson and the customer).

Teachers must be able to effectively handle many complex classroom situations that might not be comparable to previous cases and may require choosing among different, even contradictory, teaching principles (Fenstermacher 1994). This requires not only general propositional knowledge and context-specific case knowledge, but also strategic knowledge. Strategic knowledge enables teachers to evaluate and weigh different propositions and cases and to act flexibly in varying contexts, also described as wisdom of practice (Shulman 1986, 13). In our model, we associated teachers' strategic knowledge with the cognitive process of evaluating different ideas in complex classroom situations (for example, the teacher recognizes that the principle of conflict-orientation has a different relevance for students of a commercial high school discussing "economics vs. ecology" than it has in the aforementioned case, and thus requires a different application).

Test-specific definition of PCK of business and economics

Due to practical limitations of testing time, we could not operationalize the entire model of PCK of business and economics in one test; therefore, we narrowed the construct for this test; while retaining the content distinction between teaching objectives, content, and methods and students' learning process, we focused on the two aspects of lesson planning and reacting to students' statements. With regard to cognitive structure, we focused on case knowledge because case knowledge and the associated processes of analyzing, applying, and creating, which are considered important for subject-specific classroom teaching. Furthermore, item contexts were written to represent practical situations in classroom teaching of business and economics.

Test development was guided by the following assumptions about teaching business and economics in Germany; however, we also argue for why the test might be appropriate for use in other countries. In general secondary education in Germany, business and economics concepts are taught mostly in other subjects (e.g., geography, political and social studies). Most business and economics teachers in Germany work in commercial vocational and commercial upper secondary schools, where business and economics content is taught systematically. The curricular structure of business and economics is very diverse, organized in subjects at commercial high school and in fields geared towards various professions at vocational schools. For example, prospective industrial

managers take courses on identifying value streams and indicators; prospective wholesalers take courses on controlling buying processes. Therefore, most business and economics teachers in Germany need PCK that enables them to teach diverse, practically relevant, action-oriented content for a large number of commercial professions.

We focused on the content areas of sales, buying processes, and principles of economics because they cover central topics from both business administration and economics, and they are relevant to most professions that rely on business and economic knowledge (Brötz et al. 2011). In addition, these content areas are targeted in national and international assessments (e.g., Winther and Achtenhagen 2009; Walstad et al. 2013) and cover internationally recognized core competencies students should acquire in business and economics classes, such as understanding and solving various problems and social conflicts, making decisions, and seeing things critically and from multiple perspectives (e.g., Council for Economic Education [CEE] 2010; Davies 2012).

Research hypotheses

Test content

One of the five major criteria indicating the validity of an assessment, as defined by the Standards of Educational and Psychological Testing, refers to the connection between the theoretical construct and the test content (AERA et al. 2014). Analyses of content validity aim to determine the extent to which both single items and the compiled test are representative of and relevant to the construct being assessed (Haynes et al. 1995). Items must be derived systematically from a previously defined construct to ensure validity (Kane 2006).

The representativeness of the test items and overall test was deduced primarily from conceptual considerations. Our structural model served to describe the test-specific definition of PCK of business and economics and delineate the item domain, that is, the range of potential items to-be-included (McDonald 1999).

Items were judged relevant if they represented PCK necessary for teaching business and economics. Document analyses had provided a sound basis of relevant content from business and economics curricula and requirements for the teaching profession. To ensure practical relevance of items, we interviewed experts in the areas of both university- and school-based business and economics teacher education during item construction and revision.

Regarding representativeness and relevance of test items, Hypotheses I and II were specified as follows:

1. Test items cover the content and cognitive structures of the construct PCK of business and economics as specified in the test definition.
2. Test items cover aspects of PCK relevant to teaching business and economics.

Internal structure

Internal structure refers to the dimensionality of the test, which should correspond to the theoretically modeled structure of the construct (AERA et al. 2014). We used confirmatory factor analysis and models from item response theory (IRT; Embretson and Reise 2000) to test the relationship between the theoretically modeled structure of PCK

of business and economics and responses to the test items. Since our operationalized test definition of PCK of business and economics focused on case-related PCK only and because research on the dimensionality of PCK in other domains is inconclusive (e.g., Hill et al. 2004; Kleickmann et al. 2012), we assumed a 1D latent construct. Although our definition included knowledge of both teaching and learning processes, analyses of teaching practices have shown that these types of knowledge are mutually dependent. Therefore, our conceptual differentiation between content associated with lesson planning and reactions to student statements might not be reflected in the empirical data. Similarly, we assumed that the cognitive process of creating is very similar to the cognitive processes of analyzing and applying and might not be clearly distinguishable in the empirical data. We tested the internal structure of the data set with a 1D measurement model and with two 2D models for comparison (one 2D model for content and cognitive structure, respectively).

We used differential item analysis to examine whether the test assessed the same latent variable across groups of test takers (AERA et al. 2014). Differential item functioning (DIF) indicates, for similar overall ability, a systematic overestimation or underestimation of particular item parameters for certain subgroups (Camilli 1993). DIF should be considered an item bias compromising estimation of the person ability only if the difference in the item parameters can be explained by a meaningful, construct-irrelevant difference between subgroups. If differences in subsets of items can be explained by construct-relevant factors, differential item analysis indicates relative strengths and weaknesses of subgroups with respect to the assessed construct (Camilli 1993). We expected groups with greater teaching experience to have an edge in completing PCK items. Furthermore, we assumed that female and male participants would perform equally well on the test items.

Accordingly, we specified the following hypotheses:

3. PCK of business and economics, as defined in our test definition, is a 1D construct.
4. There is no gender bias of the item format.
5. Subgroups with greater practical experience have an edge in completing PCK items.

Results

Item construction and test content

Interviews with eight experts, including university- and school-based educators of pre-service business and economics teachers as well as experienced business and economics teachers, and the consultation of practical materials (e.g., teaching materials and exam tasks from university classes, teaching materials from seminars, lesson plans for demonstration lessons) provided descriptions of everyday situations in business and economics teaching that covered all four areas of the item construction matrix (lesson planning and reacting to student statements, crossed with applying/analyzing and creating). Some examples that address particular domain-specific student difficulties in the three key content areas are: difficulties in differentiating between guarantee and warranty, difficulties in understanding the role of the state in national economy (see also Leiser, 1983, p. 308f), difficulties in applying abc analysis, for example to differentiate between value share and quantity share (see also Weber, 1994, p. 154ff). These descriptions were used

to develop situation-based item contexts referring to business and economics teaching with a strong orientation towards practical requirements (Carlson 1990).

To represent the theoretically modeled cognitive processes, we selected suitable response formats (Schmeiser and Welch 2006). The processes of analyzing and applying were judged to be elicited adequately with forced-choice questions, while an open-response format was deemed necessary for the process of creating.

All items consisted of four parts: a meta-context, identical for all items in a content area and providing general curricular background information; a sub-context; a question; and an answer format. Figure 2 shows an example item from the content area of buying processes. In this item, the meta-context provides specific curricular details, and test takers are instructed to assume the role of a teacher at a vocational school teaching a class of prospective wholesalers and foreign trade operators.

The sub-context describes the students' previous knowledge and the targeted competency. In this item, the task proposed in the sub-context does not match the lesson goal. As this item is part of the lesson planning set, test takers are instructed to revise the task, taking into account the lesson goal and target group. To respond correctly to the item, test takers need to remember and understand relevant subject-specific teaching aspects (e.g., it is a class of prospective wholesalers) and subject-matter aspects (e.g., the structure of decision-making processes). Hence, to respond correctly, test takers need subject-specific teaching knowledge and only little subject-matter knowledge because additional information has been provided in the item description (e.g., specifying price as decision criterion rather than simply a quantitative criterion). Test takers need to analyze the relevant aspects given in this item and apply them to the specific learning situation, for example by analyzing what 'taking uncertain factors into consideration' means for prospective wholesalers and by creating their own solutions. The open-ended item prompts test takers to suggest two ways of altering the task. Possible alterations include adding another offer to the task in order to create a real decision situation; adding qualitative aspects to the task such as adherence to delivery schedules in accordance with the qualitative comparison of offers; or adjusting the task to the students' background and life experience, for example, by phrasing the task specifically for trainees of wholesaling and foreign trade.

[Sub-context:] You teach an advanced class. Your students of wholesaling and foreign trade have learned how to choose suppliers based on a quantitative comparison of the offers (price as decision criterion). Now, your aim is to ensure that students can make decisions by considering uncertain factors. Your starting point is the following task for your students.

You are a salesperson for cell phones and receive an offer by a supplier for cell phone cases for a total amount of 500 Euros. The supplier grants you a discount of 15%. Furthermore, if the payment results within 8 days, an additional cash discount of 2% will be granted.

Work in teams and calculate and evaluate the purchase price!

[Question:] How would you alter the task to achieve your aim? Give 2 specific options. (In bullet points, please.)

[Response format: open-ended]

- 1.
- 2.

Fig. 2 Sample PCK item (Kuhn 2014, p. 316)

The initial item pool consisted of 34 items covering the four cells of the matrix. The relevance and other aspects of the initial items were evaluated qualitatively by the aforementioned experts. Think-aloud interviews (Ericsson and Simon 1993; Leighton et al. 2011) were conducted with five pre-service business and economics teachers to detect formal deficiencies in the items, such as inaccurate language or construct-irrelevant demands (Kuhn and Brückner 2013). The revision process, which was conducted in close cooperation with the eight experts, resulted in a paper–pencil test with 17 items, representing the test definition as shown in Fig. 3 with regard to content and cognition (hypothesis I). A standardized rating by 13 experts, six of whom had not been involved in the test construction, additionally confirmed the relevance of the item content for business and economics teaching (hypothesis II). Confirmation means that the experts indicated on a 4-point scale either ‘I strongly agree’ or ‘I agree.’ For all items, relevance of content was confirmed by at least 80 % of the experts; for 11 of the 17 items, at least 90 % of the experts confirmed relevance of content. The experts also confirmed that the underlying business and economics content areas and core competencies are relevant to teaching business and economics.

Administration of the test and internal structure

Survey and sample

In 2011, we administered the test to the target group (Table 1) and collected demographic data on the participants (e.g., gender, completion of content-specific courses). The test took 45 min to complete.

Case-related pedagogical content knowledge in B&E		
	Applying & Analyzing	Creating
Knowledge of teaching objectives, content, and methods	5 items (forced-choice)	6 items (open-ended)
Knowledge of students’ learning process	3 items (forced-choice)	3 items (open-ended)
in the core content areas of sales, buying processes and principles of economics targeted to core competencies of students		

Fig. 3 Item construction according to 2-by-2-matrix based on the test definition (Kuhn 2014, p. 142)

Table 1 Descriptive data of the sample for the empirical study

Target group N = 338	N	Age (SD) (years)	Gender (female) (%)	PCK test scores	
				Mean	SD
B&E pre-service teachers university phase (bachelor level)	125	23.5 (3.0)	58	8.74	3.80
B&E pre-service teachers university phase (master level)	51	26.3 (1.9)	49	11.35	4.78
B&E pre-service teachers practical training phase (after university)	109	31.0 (5.0)	57	13.58	3.96
B&E in-service teachers	53	42.3 (9.8)	55	13.42	3.77

SD standard deviation, PCK pedagogical content knowledge, B&E business and economics

The test was scored with a maximum of 26 points. In accordance with the theoretically described cognitive requirements, the nine open-ended items requiring creating one's own solutions had double weighting and, thus, were awarded 0, 1, or 2 points; the eight forced-choice items requiring only applying and analyzing were awarded 0 or 1 point. To ensure objectivity in the scoring of the open-ended items, a coding manual was developed in cooperation with the above-mentioned experts. Two trained coders independently scored 62 % of the open-ended responses. Intercoder reliability, as indicated by Krippendorff's alpha for ordinal data, showed a good average value of 0.88 and ranged between 0.70 and 0.96 for all items.

Analysis of factor structure

First, we tested the dimensionality of the construct by analyzing the factor structure assuming a less restrictive, congeneric model. To this end, we conducted a confirmatory factor analysis using the software Mplus version 6 (Muthén and Muthén 1998–2010). The model fit of the one-factor model was compared against the fit of the two-factor models. The results were interpreted with regard to the reliability measure Cronbach's alpha.

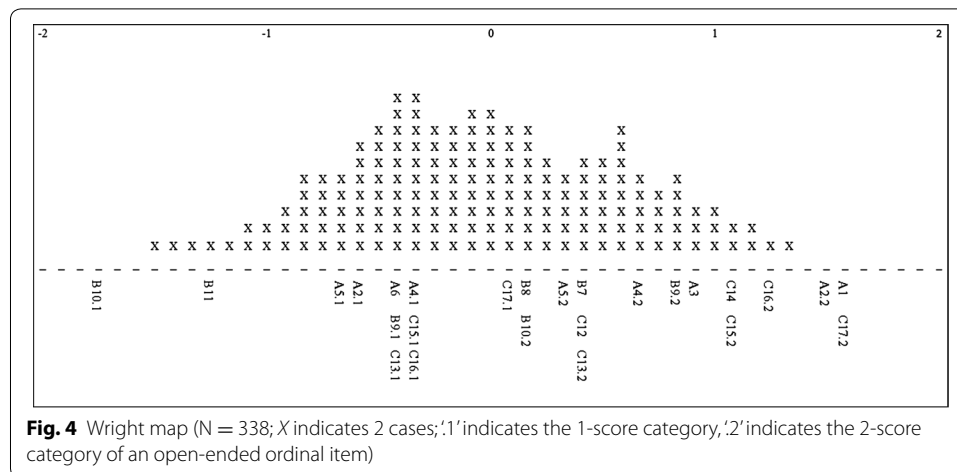
The fit measures for the one-factor model indicated a good fit and, thus, confirmed a one-factor structure of PCK of business and economics as specified in the test definition (WLSMV estimator, $N = 338$: RMSEA = 0.03; CFI = 0.94; TLI = 0.93; WRMR = 0.90). Three of the 17 items had positive, but non-significant factor loadings in the unstandardized solution ($p > 0.05$). In the standardized solution, only one item had a factor loading with $p > 0.05$. The analysis indicates a factorially homogeneous group of items, which can be considered representative of a single dimension (hypothesis III).

We tested the one-factor model against both two-factor models. In Alternative Model 1, we divided the items between the two factors of lesson planning and reacting to student statements. In Alternative Model 2, the items were attributed either to applying and analyzing or to creating. The fit of these two models¹ was almost the same as the fit of the one-factor model. However, Alternative Model 1 had a latent correlation of 1.06, which means the model estimate could not be interpreted at all, and Alternative Model 2 had a latent correlation of 0.96, which does not support the assumption of a two-factor structure. To further interpret Alternative Model 1, we conducted an analysis based on a multi-dimensional IRT model in ConQuest (Wu et al. 2007). We calculated a latent correlation of 0.93 between the two factors, lending further support to the hypothesis of a 1D construct.

Testing for a one-factor structure under the more restrictive condition of an essentially tau-equivalent model, we found, as expected, that this restrictive model was not supported.² The empirical results showed that factor loadings were not identical while homogeneity of items was confirmed; hence, the calculated value of Cronbach's alpha of 0.68 was interpreted as being close to the lower bound of the acceptable reliability range (Krohne and Hock 2007; McDonald 1999). For the purpose of this project and in view of the lack of suitable alternative tests, this reliability was deemed acceptable (Hair et al. 1998).

¹ Alternative Model 1: RMSEA = 0.032, CFI = 0.939, TLI = 0.930, WRMR = 0.903; Alternative Model 2: RMSEA = 0.032, CFI = 0.938, TLI = 0.929, WRMR = 0.903, both models based on WLSMV estimator, $N = 338$.

² One-factor model with factor loadings of all 17 indicators restricted to 1, WLSMV estimator, $N = 338$; RMSEA = 0.064; CFI = 0.724; TLI = 0.722; WRMR = 1.520.



Rasch scaling

Despite the differences in the factor loadings, which seemed to suggest a two-parameter (2PL) model, the items met the common criterion of fit to the Rasch model for ordinal data (hypothesis III) as calculated using ACER ConQuest 2.0: partial credit model (Masters 1982), $N = 338$, mean of latent ability distribution = 0. As shown in the Wright map (Wilson 2005; Fig. 4), the thresholds spread reasonably well over the person ability distribution. As expected, the thresholds of the 2-category were in the higher ability range, while thresholds of the 1-category were in the lower ability range. The precision of the person ability estimates was represented by the expected a posteriori (EAP) estimation of 0.71, which showed, as expected, a good match with Cronbach's alpha. However, the mean threshold of 0.27 meant that there were not enough easy items for precise estimations in the lower ability range.

The item fit measure weighted mean square (WMNSQ) showed acceptable values (between 0.91 and 1.14); only one item showed a significant deviation from the model with a t value of 2.4, which was beyond the acceptable interval of $[-1.96; 1.96]$. The point-biserial correlation values of the score categories increased in order of increasing category scores, with negative correlation values for category 0 (Wu et al. 2007).

Differential item analyses provided evidence as to whether the test assessed the same latent variable across groups of test takers. Three forced-choice items showed moderate effects ($-0.56; 0.47; 0.56$), indicating differences in the item functioning between female participants ($N = 150$) and male participants ($N = 188$) (Wilson 2005). However, since differences in item functioning were not observed systematically across all three items (one item was systematically easier for males, while two items were systematically easier for females), the analysis did not indicate a gender bias of the item format. As intended, there was no significant differences in the mean test performances of the female and male participant groups (hypothesis IV).³

In the differential analysis of those with teaching experience (pre-service teachers in their practical phase of teacher education after university, as well as experienced

³ In line with the theoretical framework, the group of pre-service teachers in the practical phase and in-service teachers showed a significantly higher mean test performance than the group of pre-service teachers in the university phase (Kuhn 2014).

in-service business and economics teachers; N = 162) and those without teaching experience (pre-service teachers in the university phase of teacher education; N = 176), two items showed moderate effects (0.46; 0.52), and 10 items showed large effects (0.65–1.24), indicating that the items functioned differently for the two subgroups. All these items were systematically easier for in-service teachers and pre-service teachers who had already entered the practical phase than for pre-service teachers who were still in the university phase. Therefore, these effects indicated that individuals with greater practical experience scored better. This construct-relevant explanation supported hypothesis V, emphasizing the importance of practical experience (Camilli 1993; Roussos and Stout 1996).

The results indicated that there was not only a factorially homogenous group of items, but also a Rasch homogenous group of items representing the same latent variable. Thus, we considered the manifest sum score of the test a sufficient statistic of PCK of business and economics as specified in our test definition (Masters 1982).

Response processes and relationships to other variables

Comprehensive validation of our test on PCK of business and economics involved assessment of the above-mentioned measures and other validity criteria (AERA et al. 2014). Thus, to present a sound argument for the validity of our test, we highlight key findings from these investigations. The entire validation process of the international Standards for Educational and Psychological Testing is illustrated in Fig. 5.

Cognitive interviews using the think-aloud method with pre-service business and economics teachers provided preliminary empirical evidence that the theoretically predicted thought processes of analyzing, applying, and creating are, indeed, elicited during item response processes, as indicated by verbalized responses (evidence based on

Aspects of validity			Methods	
"test content"	I)	Test items cover representative share of the test definition.	Item construction process (2-by-2-matrix)	✓
	II)	Test items have relevant content.	Item construction process, standardized expert interviews (N=13)	✓
	III)	Test items are adequate.	Item construction process, standardized expert interviews (N=13)	✓
"response processes"	IV)	Thinking processes converge with verbalized response processes.	Think aloud method (N=6)	✓
"internal structure"	V)	The construct is one-dimensional.	Confirmatory factor analysis (N = 338, WLSMV estimator: RMSEA=0.03, CFI=0.94, TLI=0.93) Rasch scaling (N=338, Partial Credit-Model: WMNSQ [0.91; 1.14], one t-value outside the range [-1.96; 1.96])	✓
"relationship to other variables"	VI)	Subject matter knowledge can be differentiated from PCK.	Correlation of sum scores (N=97)	✓
	VII)	Level of PCK corresponds with study model (contrast group vs. target group).	Known groups method (N=480)	✓
	VIII)	Target group's level of PCK corresponds with stages of teacher education.	Known groups method (N=338)	✓

Fig. 5 Validation process of the Standards for Educational and Psychological Testing (see Kuhn 2014, p. 129)

response processes; Kuhn and Brückner 2013). Analyses of relationships to other variables provided evidence of discriminant and criterion validity (evidence based on relationships to other variables). For example, comparative group analyses confirmed that the level of PCK among different groups of in-service teachers in the field of business and economics increased as additional classes on subject-specific teaching methodology were introduced and as the number of practical teaching phases increased within their study model. Additionally, as expected, the level of PCK increased with teaching expertise over different stages of teacher education. In-service teachers at the stage of professional, fully autonomous teaching at schools scored higher than pre-service teachers at university (Kuhn 2014).

Discussion

Although the results reported above provide evidence of reliability and validity of our test on PCK of business and economics, we situate this work within the larger landscape of debates about whether knowledge for teaching can be developed and assessed outside of the complex classroom context. Our instrument is consistent with other efforts to develop paper-and-pencil assessments of PCK (e.g., Park et al. 2008). In addition, other researchers have attempted to document teachers' PCK in written form using structured templates (Loughran et al. 2001; van der Valk and Broekman 1999). However, concerns have been raised about whether these methods really elicit PCK or something else (Alonzo 2007, Hambleton 2006) and whether they capture the complexity of classroom practices (Baxter and Lederman 1999). While it is possible to provide contextual information in written form and, thus, assess part of teachers' case knowledge (Rohaani et al. 2009), the fact remains that teachers are engaged in tasks that differ from actual teaching. Consistent with earlier claims (van Driel et al. 1998), recent research (Alonzo et al. 2012) points to the importance of a more spontaneous, flexible, and adaptive form of PCK that allows teachers to listen and respond to students' ideas.

With greater clarity about the construct of PCK of business and economics established in this study, our next step is to explore the use of video to assess pre-service and in-service teachers' strategic knowledge in our three-year follow-up study (Heinze et al. 2016). Kagan (1990) criticized earlier efforts to elicit teachers' cognitions as relying upon materials that are 'so artificially neat that the data they generate are irrelevant to the problems encountered in classrooms' (p. 456). Videos preserve some of the authenticity and complexity of classrooms situations (Alonzo et al. 2012) and can capture transient moments of classroom instruction. Since we cannot access teachers' reasoning and use of knowledge while they are engaged in teaching, this is as close as we can come to exploring PCK in action (Janík and Miková 2006) and, thus, to honoring even more the context-specific nature of this knowledge (Grossman et al. 1989). Work to elicit teachers' PCK using video has demonstrated that this approach may capture different knowledge from that revealed by paper-and-pencil measures of PCK (Kersting et al. 2010). Previous studies have aimed to determine PCK through teachers' reflections on videos of their own instruction (Alonzo and Kim in press), of others' instruction (Kersting et al. 2010), and of interviews with students (Alonzo and Kim in press). We see particular promise in eliciting teachers' strategic knowledge through direct responses to video-recorded displays of student thinking (Lindmeier 2011).

The paper–pencil test we presented in this paper combined with our newly developed video-/performance-based instrument will serve as a valuable tool to foster professional teaching and teacher education in the subject area of business and economics. In the future, a valid and comprehensive measure of teachers' PCK of business and economics could provide a better understanding of the role PCK plays for student learning and of how teachers acquire PCK of business and economics during teacher education and professional practice. To this end, formative diagnostic assessment could be employed during the different phases of teacher education and practical training. Although the test instrument was created for use in Germany, it is based on internationally agreed upon core concepts of business and economics and adheres to content standards and, therefore, could be adapted for use to support business and economics education in other countries.

Authors' contributions

All authors contributed substantially to this work. CK and OZ raised funding for the project and designed the study. CK developed the theoretical framework in consultation with OT. Data analysis for this paper was carried out by CK in consultation with AA. All authors discussed together the manuscript at all stages. All authors read and approved the final manuscript.

Authors' information

Christiane Kuhn (CK) is a research associate and lecturer at the Department of Business and Economics Education at Johannes Gutenberg University, Mainz, Germany. Her research interests lie in the development and validation of tools to assess and foster pre- and in-service teachers' practice-oriented competencies in the domain of business and economics.

Alicia C. Alonzo (AA) is an associate professor in the Department of Teacher Education at Michigan State University. Her research focuses on knowledge and tools to support assessment practices (at classroom and large-scale levels) that can transform learning opportunities for both students and their teachers, particularly in K-12 science education.

Olga Zlatkin-Troitschanskaia (OZ) is a Chair of Business and Economics Education at Johannes Gutenberg University, Mainz, Germany. She conducts national and cross-nationally comparative empirical research in vocational and higher education with a major focus on modeling and measurement of competencies in higher education.

Author details

¹ Department of Business and Economics Education, Johannes Gutenberg University, Jakob Welder-Weg 9, 55099 Mainz, Germany. ² Department of Teacher Education, Michigan State University, 620 Farm Lane Blvd, Erickson Hall, Room 307, East Lansing, MI 48824, USA.

Acknowledgements

This work was supported by the Federal Ministry of Education and Research under Grant 01JG0928.

Competing interests

The authors declare that they have no competing interests.

Received: 22 February 2016 Accepted: 20 May 2016

Published online: 14 June 2016

References

- Alonzo AC (2007) Challenges of simultaneously defining and measuring knowledge for teaching. *Meas Interdiscip Res Perspect* 5:131–137
- Alonzo AC, Kim J (in press) Declarative and dynamic pedagogical content knowledge as elicited through two video-based interview methods. *J Res Sci Teach*. doi:10.1002/tea.21271
- Alonzo AC, Kobarg M, Seidel T (2012) Pedagogical content knowledge as reflected in teacher-student interactions. Analysis of two video cases. *J Res Sci Teach* 49(10):1211–1239
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014) Standards for educational and psychological testing. American Educational Research Association, Washington
- Anderson LW, Krathwohl DR (2001) A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives, Longman, New York
- Baumert J, Kunter M, Blum W, Brunner M, Voss T, Jordan A, Klusmann U, Krauss S, Neubrand M, Tsai YM (2010) Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *Am Educ Res J* 47(1):133–180. doi:10.3102/0002831209345157
- Baxter JA, Lederman NG (1999) Assessment and measurement of pedagogical content knowledge. In: Gess-Newsome J, Lederman NG (eds) Examining pedagogical content knowledge. The construct and its implications for science education. Kluwer Academic Publishers, Dordrecht, pp 147–161

- Berry A, Loughran J, van Driel JH (2008) Revisiting the roots of pedagogical content knowledge. *Int J Sci Educ* 30:1271–1279
- Blinder A (2015) What did we learn from the financial crisis, the great recession, and the pathetic recovery? *J Econ Educ* 46(2):135–149. doi:10.1080/00220485.2015.1015190
- Blömeke S, Suhl U, Kaiser U (2011) Teacher education effectiveness: quality and equity of future primary teachers' mathematics and mathematics pedagogical content knowledge. *J Teacher Educ* 62:154–171
- Bromme R (1995) What exactly is "pedagogical content knowledge"?—critical remarks regarding a fruitful research program. In: Hopmann S, Riquarts K (eds) *Didaktik and/or curriculum*. IPN, Kiel, pp 205–216
- Brötz R, Kaiser F, Brings C, Peppinghaus B, Warmbold-Jaquinet V, Krieger A, Noack I, Nies N, Schaal T (2011) *Gemeinsamkeiten und Unterschiede kaufmännisch-betriebswirtschaftlicher Aus- und Fortbildungsberufe* [Similarities and differences between commercial professions based on vocational or advanced training]. BIBB, Bonn
- Brückner S, Förster M, Zlatkin-Troitschanskaia O, Walstad WB (2015) Effects of prior economic education, native language, and gender on economic knowledge of first-year students in higher education. A comparative study between Germany and the USA. *Stud High Educ* 40(3):437–453. doi:10.1080/03075079.2015.1004235
- Camilli G (1993) The case against item bias detection techniques based on internal criteria. Do item bias procedures obscure test fairness issues. In: Holland PW, Wainer H (eds) *Differential item functioning*. Lawrence Erlbaum Associates, Hillsdale, pp 397–413
- Carlson R (1990) Assessing teachers' pedagogical content knowledge. *Item development issues*. *J Pers Eval Educ* 4:157–173
- Carlton ST (2012) Using content specialists for detecting item and test unfairness. In: Secolsky C, Denison DB (eds) *Handbook on measurement, assessment, and evaluation in higher education*. Routledge, New York, pp 310–320
- Cochran-Smith M, Cannady M, McEachern K, Mitchell K, Piazza P, Power C, Ryan A (2012) Teachers' education and outcomes. Mapping the research terrain. *Teach Coll Rec* 114(10):1–49
- Council for Economic Education (CEE) (2010) Voluntary national content standards in economics, 2nd edn. <http://www.councilforeconed.org/wp/wp-content/uploads/2012/03/voluntary-national-content-standards-2010.pdf>
- Council for Economic Education (CEE) (2011) Survey of the States. Economic and personal finance education in our nation's schools. <http://www.councilforeconed.org/wp/wp-content/uploads/2011/11/2011-Survey-of-the-States.pdf>
- Darling-Hammond L, Newton SP, Wei RC (2013) Developing and assessing beginning teacher effectiveness. The potential of performance assessments. *Educ Assess Eval Account* 25(3):179–204. doi:10.1007/s11092-013-9163-0
- Davies P (2012) Threshold concepts in economic education. In: Hoyt GM, McGoldrick K (eds) *International handbook on teaching and learning economics*. Mass: Edward Elgar, Cheltenham, pp 250–258
- De Jong O, van Driel J (2004) Exploring the development of student teachers' PCK of the multiple meanings of chemistry topics. *Int J Sci Math Educ* 2:477–491. doi:10.1007/s10763-004-4197-x
- Depaepe F, Verschaffel L, Kelchtermans G (2013) Pedagogical content knowledge. A systematic review of the way in which the concept has pervaded mathematics educational research. *Teach Teach Educ* 34:12–25
- Desimone LM (2009) Improving impact studies of teachers' professional development. Toward better conceptualizations and measures. *Educ Res* 38(3):181–199. doi:10.3102/0013189X08331140
- De Wever B, Vanderlinde R, Tuytens M, Aelterman A (eds) (2016) *Professional learning in education: challenges for teacher educators, teachers and student teachers*. Academia Press, Gingko-imprint
- Educational Testing Service (2011) *The Praxis Series. Business education. Content knowledge (0101). Test at a glance*. <https://www.ets.org/s/praxis/pdf/0101.pdf>
- Embretson SE, Reise S (2000) *Psychometric methods. Item response theory for psychologists*. Lawrence Erlbaum Associates, Mahwah
- Ericsson KA, Simon HA (1993) *Protocol analysis. Verbal reports as data*. MIT Press, Cambridge
- European Commission (2007) *Improving the quality of teacher education. Communication from the commission to the council and the European parliament*. http://www.ec.europa.eu/education/com392_en.pdf
- Fenstermacher GD (1994) The knower and the known: the nature of knowledge in research on teaching. *Rev Res Educ* 20:3–56
- Förster M, Brückner S, Zlatkin-Troitschanskaia O (2015) Assessing the financial knowledge of university students in Germany. *Empir Res Vocat Educ Train* 7(6):1–20
- Fritsch S, Berger S, Seifried J, Bouley F, Wuttke E, Schnick-Vollmer K, Schmitz B (2015) The impact of university teacher training on prospective teachers' CK and PCK—a comparison between Austria and Germany. *Empir Res Vocat Educ Train* 7(4):1–20
- German Federal Statistical Office (2014) *Studierende an Hochschulen—2013/2014* [Students at Universities—2013/2014]. Wiesbaden
- Gill A, Gratton-Lavoie C (2011) Retention of High School Economics Knowledge and the Effect of the California State Mandate. *J Econ Educ* 42(4):319–337
- Gold B, Holodyski M (2015) Development and construct validation of a situational judgment test of strategic knowledge of classroom management in elementary schools. *Educ Assess* 20(3):226–248
- Grossman PL, Wilson SM, Shulman LS (1989) Teachers of substance. Subject matter knowledge for teaching. In: Reynolds M (ed) *Knowledge base for the beginning teacher*. Pergamon Press, Oxford, pp 23–36
- Hair J, Anderson R, Tatham R, Black W (1998) *Multivariate Data Analysis*, 5th edn. Prentice-Hall Inc, New Jersey
- Hambleton RK (2006) Psychometric models, test designs and item types for the next generation of educational and psychological tests. In: Bartram D, Hambleton RK (eds) *Computer-based testing and the internet. Issues and advances*. Wiley, New York, pp 77–90
- Haynes SN, Richard DCS, Kubany ES (1995) Content validity in psychological assessment. A functional approach to concepts and methods. *Psychol Assess* 7(3):238–247
- Heinze A, Kuhn C, Lindmeier A, Zlatkin-Troitschanskaia O (2016) ELMaWi—assessing subject-specific competencies in teacher education in mathematics and economics. A quasi-experimental validation study with a focus on domain-specificity. In: Zlatkin-Troitschanskaia O, Pant HA, Lautenbach C, Toepfer M (eds) *Modeling and measuring*

- competencies in higher education—validation and methodological innovations (KoKoHs). (KoKoHs Working Papers, 11). Humboldt University & Johannes Gutenberg University, Berlin & Mainz
- Hill HC, Schilling SG, Ball DL (2004) Developing measures of teachers' mathematics knowledge for teaching. *Elem Sch J* 105:11–30. doi:10.1086/428763
- Hill HC, Rowan B, Ball DL (2005) Effects of teachers' mathematical knowledge for teaching on student achievement. *Am Educ Res J* 42(2):371–406
- Hopmann S, Riquarts K (1995) Didaktik and/or curriculum. Basic problems on comparative Didaktik. In: Hopmann S, Riquarts K (eds) *Didaktik and/or curriculum*. IPN, Kiel, pp 9–40
- Janík T, Miková M (2006) The dynamic nature of pedagogical content knowledge. Video case study from physics instruction at Czech lower secondary school. Paper presented at the European Conference on Educational Research, Geneva
- Kagan DM (1990) Ways of evaluating teacher cognition: inferences concerning the Goldilocks principle. *Rev Educ Res* 60:419–469
- Kane MT (2006) Validation. In: Brennan RL (ed) *Educational Measurement*, 4th edn. Praeger, Westport, pp 17–64
- Kersting NB, Givvin KB, Sotelo FL, Stigler JW (2010) Teachers' analyses of classroom video predict student learning of mathematics. Further explorations of a novel measure of teacher knowledge. *J Teach Educ* 61:172–181
- Kleickmann T, Richter D, Kunter M, Elsner J, Besser M, Krauss S, Baumert J (2012) Teachers' content knowledge and pedagogical content knowledge. The role of structural differences in teacher education. *J Teach Educ* 20(10):1–17. doi:10.1177/0022487112460398
- Knupp T, Harris DJ (2012) Building content and statistical test specifications. In: Secolsky C, Denison DB (eds) *Handbook on measurement, assessment, and evaluation in higher education*. Routledge, New York, pp 239–256
- König J, Blömeke S, Paine L, Schmidt WH, Hsieh FJ (2011) General pedagogical knowledge of future middle school teachers. On the complex ecology of teacher education in the United States, Germany, and Taiwan. *J Teach Educ* 62:188–201. doi:10.1177/0022487110388664
- Krohne HW, Hock M (2007) *Psychologische Diagnostik [Psychological diagnostics]*. Kohlhammer, Stuttgart
- Kuhn C (2014) *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung [Pedagogical content knowledge of teachers in business and economics. Theoretical modeling, test development, and validation]*. Verlag Empirische Pädagogik, Landau
- Kuhn C, Brückner S (2013) Analyse des fachdidaktischen Wissens von (angehenden) Lehrkräften in der kaufmännisch-verwaltenden Bildung mit der Methode des lauten Denkens [Analyzing pedagogical content knowledge of (future) teachers of business administration using the thinking-aloud method] *Zeitschrift bwp@Berufs- und Wirtschaftspädagogik—online* 24:1–20. http://www.bwpat.de/ausgabe24/kuhn_brueckner_bwpat24.pdf. Accessed 25 Jun 2013
- Kuhn C, Happ R, Zlaktin-Troitschanskaia O, Beck K, Förster M, Preuß D (2014) Kompetenzentwicklung angehender Lehrkräfte im kaufmännisch-verwaltenden Bereich—Erfassung und Zusammenhänge von Fachwissen und fachdidaktischem Wissen [Competence development among future teachers in the field of business administration – assessment and relationship of content knowledge and pedagogical content knowledge]. *Zeitschrift für Erziehungswissenschaft [J Educ Sci]* 17(1):149–167. doi:10.1007/s11618-013-0456-3
- Leighton JP, Heffernan C, Cor MK, Gokiert RJ, Cui Y (2011) An experimental test of student verbal reports and teacher evaluations as a source of validity evidence for test development. *Appl Measur Educ* 24(4):324–348
- Leiser D (1983) Children's conceptions of economics. The constitution of a cognitive domain. *J Econ Psychol* 4:297–317
- Lindmeier A (2011) Modeling and measuring knowledge and competencies of teachers. A threefold domain-specific structure model for mathematics. Waxmann, Münster
- Loughran J, Milroy P, Berry A, Gunstone R, Mulhall P (2001) Documenting science teachers' pedagogical content knowledge through PaP-eRs. *Res Sci Educ* 31:289–307
- Masters GN (1982) A Rasch model for partial credit scoring. *Psychometrika* 47:149–174
- McDonald RP (1999) *Test theory. A unified treatment*. Erlbaum, Mahwah
- Mislevy R, Haertel G (2006) *Implications of Evidence-Centered Design for Educational Testing (Draft PADI Technical Report 17)*. SRI International, Menlo Park
- Muthén LK, Muthén BO (1998–2010) *Mplus User's Guide*, 6th edn. Muthén & Muthén, Los Angeles
- National Council for Accreditation of Teacher Education (2008) *Professional standards for the accreditation of teacher preparation institutions*. Washington. <http://www.ncate.org/documents/standards/NCATE%20Standards%202008.pdf>
- Organisation for Economic Co-operation and Development (2014). Revised scoping paper for an AHELO main study. 19–20 November 2013. (unpublished document)
- Park S, Chen YC, Jang J (2008) Developing measures of teachers' pedagogical content knowledge for teaching high school biology. Paper presented at the annual meeting of the Association for Science Teacher Education, St. Louis
- Pellegrino JW, Chudowsky N, Glaser R (eds) (2001) *Knowing what students know. The science and design of educational assessment*. The National Academies Press, Washington
- Rohaan EJ, Taconis R, Jochems WM (2009) Measuring teachers' pedagogical content knowledge in primary technology education. *Res Sci Technol Educ* 27:327–338
- Roussos L, Stout W (1996) A multidimensionality-based DIF analysis paradigm. *Appl Psychol Meas* 20(4):355–371
- Schmeiser CB, Welch CJ (2006) Test development. In: Brennan RL (ed) *Educational measurement*, 4th edn. Praeger, Westport, pp 307–353
- Schmelzing S, van Driel JH, Jüttner M, Brandenbusch S, Sandmann A, Neuhaus BJ (2013) Development, evaluation, and validation of a paper-and-pencil test for measuring two components of biology teachers' pedagogical content knowledge concerning the 'cardiovascular system'. *Int J Sci Math Educ* 11:1369–1390
- Shavelson RJ (2013) An approach to testing and modeling competence. In: Blömeke S, Zlaktin-Troitschanskaia O, Kuhn C, Fege J (eds) *Modeling and measuring competencies in higher education. Tasks and Challenges*. Sense Publishers, Rotterdam, pp 29–43
- Shulman LS (1986) Those who understand: knowledge growth in teaching. *Educ Res* 15(2):4–14
- Shulman LS (1987) Knowledge and teaching. Foundations of the new reform. *Harv Educ Rev* 57(1):1–22

- van der Valk AE, Broekman HGB (1999) The lesson preparation method. A way of investigating pre-service teachers' pedagogical content knowledge. *Eur J Teach Educ* 22:11–22
- van Driel JH, Verloop N, de Vos W (1998) Developing science teachers' pedagogical content knowledge. *J Res Sci Teach* 35(6):673–695
- Veal WR, Tippins DJ, Bell J (1999) The evolution of pedagogical content knowledge in prospective secondary physics teachers. Retrieved from ERIC database. (ED 443719)
- Vidal Uribe R (2013) Measurement of learning outcomes in higher education. The case of Ceneval in Mexico. In: Blömeke S, Zlatkin-Troitschanskaia O, Kuhn C, Fege J (eds) Modeling and measuring competencies in higher education. Tasks and challenges. Sense Publishers, Rotterdam, pp 137–146
- Walstad WB (2013) Economic understanding in US high school courses. *Am Econ Rev* 102(3):659–663
- Walstad WB, Watts M, Rebeck K (2007) Test of understanding in college economics Examiner's manual, 4th edn. Council on Economic Education, New York
- Walstad WB, Rebeck K, Butters RB (2013) Test of economic literacy. Examiner's manual. Council for Economic Education, New York
- Weber S (1994) Vorwissen in der betriebswirtschaftlichen Ausbildung. Eine struktur- und inhaltsanalytische Studie. Dissertation, Universität Göttingen. Gabler, Wiesbaden
- Wilson M (2005) Constructing measures. An item response modeling approach. Erlbaum, Mahwah
- Winther E, Achtenhagen F (2009) Measurement of vocational competencies—a contribution to an international large-scale assessment on vocational education and training. *Empir Res Vocat Educ Train* 1:88–106
- Wu ML, Adams RJ, Wilson MR, Haldane SA (2007) ACER ConQuest version 2.0. Generalised item response modeling software. ACER Press, Camberwell
- Zlatkin-Troitschanskaia O, Shavelson RJ, Kuhn C (2015) The International state of research on measurement of competency in higher education. *Stud High Educ* 40(3):393–411. doi:10.1080/03075079.2015.1004241

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

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

Submit your next manuscript at ▶ springeropen.com
