Zitiervorschlag: Holtsch, D., Hartig, J., & Shavelson, R. J. (2019). Do practical and academic preparation paths lead to differential commercial teacher “quality”? *Vocations and Learning*, 12(1), 23–46.
<https://doi.org/10.1007/s12186-018-9208-0>

Zur Verfügung gestellt auf PHIQ:

PHIQ-DOI: <https://doi.org/10.18747/PHSG-coll3/id/727>

Original-DOI: <https://doi.org/10.1007/s12186-018-9208-0>

Dokumentart: Journal Article

Version: accepted version

Copyright-Hinweis: This is a post-peer-review, pre-copyedit version of an article published in *Vocations and Learning* (1874-785X). The final authenticated version is available online at: <https://doi.org/10.1007/s12186-018-9208-0>

Lizenz: Alle Rechte vorbehalten

Do Practical and Academic Preparation Paths Lead to Differential Commercial Teacher “Quality”?

Dr. Doreen Holtsch¹

Prof. Dr. Johannes Hartig²

Prof. Dr. Richard Shavelson³

¹ University of Zurich

Institute of Education

Kantonsschulstrasse 3

Zurich, CH 8001

doreen.holtsch@uzh.ch

Phone: +41 (0)44 634 6602

ORCID: 0000-0003-1577-1898

² German Institute for International Educational Research (DIPF)

Centre for Educational Quality and Evaluation, Educational Measurement

Schloßstraße 29

Frankfurt am Main, DE 60486

hartig@dipf.de

³ Stanford University

Margaret Jacks Professor of Education (Emeritus)

332 School of Education Building

Stanford, CA, USA 94305

richs@stanford.edu

Do Practical and Academic Preparation Paths Lead to Differential Commercial Teacher “Quality”?

1	Introduction.....	1
2	Theoretical and Conceptual Framework.....	3
2.1	Pathways in VET Teacher Education.....	3
2.2	Cognition Vertex: <i>Economics and Society</i> Teachers' Content Knowledge and Pedagogical Content Knowledge	5
2.3	Research Questions	10
3	Methods.....	11
3.1	Participants	11
	Teachers' Socio-demographic Characteristics	11
	Teacher Pathways	12
3.2	Data Collection.....	16
	Measure of Content Knowledge.....	16
	Measure of Pedagogical Content Knowledge	17
	Scoring of the PCK Items.....	19
3.3	Statistical Procedures	20
	Research Question 1: Validity and Reliability of CK and PCK Measures	20
	Research Question 2: Pathway Dependent Teacher Quality.....	21
4	Results.....	22
4.1	Interpretation Vertex: Validity and Reliability	22
	Content Knowledge Structure and Reliability	22
	Pedagogical Content Knowledge Structure and Reliability.....	23
	Relationship to Other Variables	25
	Gender	26
4.2	RQ 2: Pathways to Teaching and Their Association with CK and PCK	26

5	Discussion and Limitations	31
5.1	RQ 1: Validity and Reliability	31
5.2	RQ 2: Association of Teachers' Pedagogical Content Knowledge and Content Knowledge with Different Pathways	33
5.3	Limitations	34
5.4	Future Research	36
6	Appendix	

1 Introduction

Teachers matter and teaching quality is essential to a successful education system. This statement holds true for Switzerland, perhaps more so than elsewhere because of its intertwined workforce and vocational education and training (VET) system (cf. Hoeckel, Field, & Grubb, 2009); quality, then, depends on both school education and work training. The VET system attracts two-thirds of all Swiss students each year, combining part-time school education with on-the-job training in a host company (State Secretariat for Education, Research and Innovation (SERI), 2016). In this way, Switzerland produces a stable workforce that contributes to one of the world's most stable economies (Hoffman & Schwartz, 2015). Therefore, teaching quality is paramount to a successful VET system from both a policy perspective and a student outcome perspective because high-quality teaching produces high-performing students (Author(s) (2007) - removed for blind review, Baumert et al., 2010).

The Swiss educational system offers various pathways to teaching at VET schools that are perceived as more or less rigorous by policy makers and the public. This variation in teacher training has led the Swiss government to question pathway quality and request reforms (Criblez, 2016). Two main pathways exist to becoming a teacher at a commercial VET school, namely, through universities of applied sciences and through universities. Compared to genuinely research oriented universities, universities of applied sciences are more practically oriented (Swiss Coordination Centre for Research in Education (SCCRE), 2014, p. 166). Consequently, teachers from different pathways may not have had equal opportunities to learn during teacher education and training (cf.

Blömeke & Delaney, 2012; Brunner et al., 2006). Research on school teaching has shown that teachers' content knowledge (CK) and pedagogical content knowledge (PCK) developed during teacher education and training positively impact student development (Baumert et al., 2010). Teachers having followed different teacher education and training pathways may thus vary substantially with regard to teaching quality and student outcomes.

This paper focuses on teachers at commercial VET schools because the commercial sector is the most popular professional field among students entering the Swiss VET system (SERI, 2016). More specifically, we focus on teachers of *Economics and Society (E&S)*, considered the core subject in commercial VET because of its high numbers of lessons in the school curriculum and the weight given to it in final exams (Staatssekretariat für Bildung, Forschung und Innovation (SBFI), 2017). *E&S* includes both commercial education for activities in the workplace and economic-civic education for economic and social citizenship. The *E&S* syllabus includes *Economics, Business Administration, Finance and Accounting, Law and History and Political Science* (SBFI, 2017).

To date, little is known about commercial VET school teachers' CK and PCK as indicators of teaching quality across the multiple entry pathways. We are aware that teacher and teaching quality encompass far more than CK and PCK (e.g., Shulman, 1987). Hence, we speak of indicators of quality and use "quality" in this reserved sense

throughout the paper. Therefore, we ask whether alternative paths produce differentially qualified teachers, as indicated by graduates' CK and PCK.

To address this question we needed measures of CK and PCK. While a test measuring CK seemed to be adaptable in an international context, there were no adaptable tests measuring Swiss teachers' PCK in particular. Therefore, we ask the subsequent research questions:

1. Are the modified measure of CK and the newly developed measure of PCK used in this study reliable and valid?
2. Are alternative paths to teacher certification associated with differences in teacher "quality"?

Answering these questions presupposes the following: (a) reliable and valid measures of Swiss teachers' CK and PCK, and (b) substantial differences in paths to becoming a commercial teacher in a vocational school. To this end, we begin by describing pathways and then turn our attention to the two measures of quality.

2 Theoretical and Conceptual Framework

2.1 Pathways in VET Teacher Education

Most Swiss *E&S* teachers are educated in a multiphase program. *In the first phase*, teachers complete either a practically oriented pathway with bachelor's degree at a university of applied sciences, or an academically oriented pathway with a master's

degree at a university in *Economics* or *Business Administration* (SBFI, 2015; cf. institutions of higher education SERI, 2017).

The second phase consists of a one-year teaching diploma program including professional practice and pedagogical content training. An *E&S* teaching diploma is offered by both universities and universities of teacher education; universities of applied sciences do not offer the diploma. However, third parties may offer *E&S* teaching diploma courses completed with recognized certificates (SCCRE, 2014).

Another criteria relevant for becoming VET teachers is to gain practical experience, e.g., in a company in the commercial field. Therefore, both the practically and academically oriented pathway include at least a half year of practical experience (SBFI, 2015).

Although the Swiss government defined basic requirements of VET teacher education and training (SBFI, 2015) each of the two different pathways covers a different (institutional-based) level of education in the first phase. Moreover, each of the two pathways involves variations resulting from options in the second phase. These variations may be reflected in teachers' CK and PCK because of differences in students' backgrounds and opportunities to learn at universities of applied sciences or universities.

Analyses of mathematics and VET school pre-service teachers' opportunities to learn at universities revealed widely varying positive correlations with their CK and PCK (cf.

Blömeke et al., 2010; Fritsch et al., 2015; Kuhn et al., 2014). This variation might be due to different conceptualizations and operationalization of both opportunities to learn and knowledge (cf. König, Ligtoet, Klemenz, & Rothland, 2017). Nevertheless, in studying pathways to teaching, in-service teachers' opportunities to learn should be investigated.

Retrospective approaches are challenging because the quantity and the quality of opportunities to learn are difficult to reconstruct in detail. Another approach, the one we took, is to consider differences between types of institutions that teachers attended to teach *E&S* at a VET school. Different combinations of institutions of teacher education and training would result in chronologically constructed pathways for Swiss *E&S* teachers. With the aid of these alternative pathways, namely, practically and academically oriented, we then explore variations in teachers' CK and PCK (cf. Brunner et al., 2006).

2.2 Cognition Vertex: *Economics and Society* Teachers' Content Knowledge and Pedagogical Content Knowledge

If we are to address our two research questions regarding “quality” produced by alternative education and training paths, we first need reliable and valid measures of CK and PCK. In what follows, we briefly describe the German version of the U.S. *Test of Understanding in College Economics* (TUCE) (Walstad, Watts, & Rebeck, 2007) as a measure of CK. We then turn attention to the development of a Swiss PCK measure.

Zlatkin-Troitschanskaia and colleagues (Brückner, Förster, Zlatkin-Troitschanskaia, Happ, et al., 2015; Brückner, Förster, Zlatkin-Troitschanskaia, & Walstad, 2015; Zlatkin-

Troitschanskaia, Förster, Brückner, & Happ, 2014; Zlatkin-Troitschanskaia, Förster, Schmidt, Brückner, & Beck, 2015) translated and adapted the U.S. TUCE-4 (Walstad et al., 2007) for use in Germany, thus providing a reliable and valid CK measure that we used in this study.¹ We investigate whether we can replicate its psychometric properties in Swiss *E&S* teachers.

Unlike the measure of CK, we found no adaptable tests to measure Swiss teachers' PCK (for Germany see Kuhn, 2014; Kuhn, Alonzo, & Zlatkin-Troitschanskaia, 2016). We make this claim based on recent research indicating that cross-national test adaptations for commercial VET are generally challenging. The reason is that minor contextual differences, e.g., regarding instruction at VET school, may cause major adaptation work and adaptation proved to be expensive and time-consuming (Author(s), 2016b – removed for blind review). Consequently, we developed a PCK measure from scratch based on prior research.

To develop the PCK test we followed the assessment triangle (National Research Council, 2001, pp. 44–51). The triangle has vertices of *cognition*—definition of the construct to be measured, in this case PCK; *observation*—the tasks derived from the definition used to produce observable responses; and *interpretation*—the reliability and validity of PCK score interpretation (Author(s), 2010 – removed for blind review; National Research Council, 2001, pp. 44–51).

¹ The *TUCE* was successfully translated, adapted, and validated in the project *Modeling and measuring competencies in business and economics among students and graduates by adapting and further developing existing American and Latin-American measuring instruments* (EGEL/TUCE), funded by the German Federal Ministry of Education and Research. For more information, see: <http://www.wiwi-kompetenz.de/eng/index.php> (Accessed 4 May 2018).

Researchers commonly define PCK as a combination of subject-specific content knowledge and pedagogical knowledge (cf. Depaepe, Verschaffel, & Kelchtermans, 2013, p. 15) consistent with Shulman's (1986) concept of PCK (cf. analyses for mathematics, Depaepe et al., 2013). However, Shulman himself further developed his ideas about PCK (cf. Carlsen, 1999; Gess-Newsome, 1999, pp. 3–4; Shulman, 1986, 1987). Thus, a simple reference to Shulman comes with complexities. Therefore, we analyzed three widely cited approaches to defining and measuring PCK: *Learning Mathematics for Teaching Project* (LMT) (Ball, Thames, & Phelps, 2008; Hill, Ball, & Schilling, 2008), *Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy* (COACTIV) (Baumert & Kunter, 2013; Baumert et al., 2010; Krauss, Baumert, & Blum, 2008; Krauss, Brunner, et al., 2008) and *Teacher Education and Development Study: Learning to Teach Mathematics* (TEDS-M) (Blömeke & Kaiser, 2014; Tatto et al., 2008). Moreover, we studied two approaches focusing on E&S teachers in Germany: *Modelling and Measurement of Content Knowledge and Pedagogical Content Knowledge in Business and Economic Education* (KoMeWP) (Bouley et al., 2015; Fritsch et al., 2015) and Kuhn's (2014) dissertation project (Kuhn et al., 2016). We found an interdisciplinary overlap for defining PCK as comprising knowledge of (a) how to represent and explain content and (b) handling students' misconceptions in a particular subject (cf. Depaepe et al., 2013, p. 22; Krauss, Baumert, et al., 2008). These two facets relate to the interactive PCK facets in TEDS-M and are likewise included in Kuhn et al.'s (2016) approach.

In addition to a construct definition, our study is concerned with whether these research programs provide evidence that the PCK construct can be measured reliably and separately from CK. Although all the conceptualizations varied in breadth and depth, the capacity to measure the construct PCK and distinguish it from CK was confirmed to a greater or lesser degree (e.g., Hill et al., 2008; Krauss, Brunner, et al., 2008; Senk et al., 2012). Because of partly equivocal empirical results, LMT revealed that *knowledge of content and of students* (KCS) was multidimensional (Ball et al., 2008, pp. 396, 403; Hill et al., 2008, pp. 385–388, 395) while both COACTIV and KoMeWP confirmed their PCK measures to have three facets: (a) representing and explaining content, (b) dealing with students' misconceptions, and (c) dealing with instructional tasks (Bouley et al., 2015; Krauss, Brunner, et al., 2008).

Accounting for these definitions, we focused on three PCK facets: (1) explaining content, (2) addressing students' misconceptions and (3) cognitively activating students through tasks. Moreover, these facets directly relate to students' understanding and learning in the classroom (e.g., Author(s), 2007 – removed for blind review; Baumert et al., 2010). Our conceptualization of PCK thus most closely follows that of COACTIV (Krauss, Baumert, et al., 2008; Krauss, Brunner, et al., 2008) and KoMeWP (Bouley et al., 2015; Fritsch et al., 2015; Mindnich, Berger, & Fritsch, 2013).

Thus far, we have compared existing PCK definitions and empirical findings. As a next step, *E&S teachers' PCK* must be specified because its facets unfold in the subject-

specific domain. Although *E&S* teachers are trained to teach the entire *E&S* curriculum, some teachers specialize in content areas within *E&S*, namely, *Economics*, *Business Administration*, *Accounting*, or *Law*. *E&S*-PCK covers a heterogeneous content domain and each content area might require slightly different PCK. *Accounting*, for example, includes more mathematical processes and *Business Administration* includes more managerial processes. However, an underlying commonality exists: *Economics* includes basic concepts that are relevant to all content areas. Consequently, we focus on the PCK that is embedded in the content area of *Economics*, specifically, the *Economics PCK*.

In summary, we define teachers' PCK in the field of *Economics* as having three facets:

- (1) Explaining economic content, including representing and illustrating economic content and concepts, namely, using authentic tables; diagrams; and statistics dealing with, for example, unemployment.
- (2) Anticipating and addressing students' economics misconceptions, e.g., misinterpreting economic principles.
- (3) Cognitively activating students, which involves not only applying and developing cognitively activating tasks but also asking questions to encourage students' thinking and understanding of economic concepts, e.g., the equilibrium of supply and demand.

2.3 Research Questions

Because we had already chosen the German TUCE for measuring teachers' CK, we focused on developing a test to measure the *Economics* PCK of *E&S* teachers. Our PCK definition served as the basis for test development. After PCK test development, we addressed the following research questions (RQ):

RQ1: Are the measures of CK and PCK used in this study reliable and valid?

AERA et al. (2014) suggested multiple sources of validity evidence. We aimed to provide evidence for our validity claim with respect to: (a) test content, (b) response processes, (c) internal test structure (including reliability and dimensions), and (d) relationships with conceptually related constructs (discriminant evidence). Evidence based on (a) and (b) was in part embedded particularly in the PCK test development process (e.g., content sampling and interviews). Evidence for (c) and (d) is provided empirically via test application in the target group. The combination of these four sources of evidence provides a basis for examining reliability and validity (AERA et al., 2014, pp. 21–22) of the measurement of teachers' *Economics* CK and PCK.²

The Swiss governments' *educational policy question* was a driving factor for this study: Are alternative pathways to teacher certification equivalent? Consequently, we ask:

² Compare also Kuhn et al. (2016) for applying validity criteria while developing a test.

RQ2: Do alternative paths to teacher certification lead to differences in teacher “quality”?

We address this question by disentangling different paths to certification. Once disentangled, we examine mean CK and PCK performance among paths.

3 Methods

3.1 Participants

Teachers’ Socio-demographic Characteristics

The Swiss *E&S* teacher population cannot be defined precisely because of teachers’ flexible individual teaching loads. Therefore, we recruited *E&S* teachers in two waves. In the first wave, 95 *E&S* teachers of 85³ classes were randomly selected from the population of 357 *E&S* commercial VET classes in German-speaking Switzerland in August 2012. Of the 95 teachers, 88 took part in our study’s first wave, which involved responding to a survey on instructional quality in 2014. In the study’s second wave in 2015, we recruited an additional 86 *E&S* teachers to increase the overall sample size for the main study of teachers’ CK and PCK. For the instructional survey and the study of CK and PCK 69 teachers participated both waves; 19 teachers dropped out between 2014 and 2015 (dropouts did not differ in characteristics from those remaining). Overall, then, in the main study reported here we had 69 teachers from the first wave and 86 teachers from the second wave totaling 155 teachers. At the school level, a total of 36 commercial VET schools from a population of 49 commercial VET schools in German-

³ Several classes were taught by two teachers.

speaking Switzerland participated in both studies (for details on sampling see Author(s), 2018b – removed for blind review).

A comparison of the randomly selected teachers in the first and the purposively recruited teachers in the second wave revealed significant differences, on average, in age, years of teaching experience, and teaching load per week (Table 1).

Table 1 Characteristics of the teacher sample

	<i>Number¹</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>Median</i>	<i>Minimum</i>	<i>Maximum</i>
Age in years						
Combined sample	154	46.5	9.7	46	28	66
Random sample ²	87	48.4	9.4	48	30	66
Convenience sample	86	44.9	10.0	46	28	65
Teaching experience in years						
Combined sample	155	14.9	9.8	13	1	42
Random sample ²	88	17.2	9.6	16	1	38
Convenience sample	86	13.2	10.1	11	1	42
Average teaching load per week						
Combined sample	152	80.7	21.5	90.0	15	125 ³
Random sample ²	67	85.5	18.0	90.0	45	125 ³
Convenience sample	85	77.0	23.3	85.0	15	100

Notes. ¹Sample size varies depending on missing answers. ²The random sample of the year 2014 includes 19 teachers who did not participate in the year 2015. ³In exceptional cases.

Teacher Pathways

We constructed alternative pathways to certification using educational and professional pathways (i.e., prior commercial VET program, academic pathway, *E&S* teaching

diploma) (Table 3).⁴ Contrary to the subsample differences, the samples did not differ significantly as to whether they completed a commercial VET program, a university or university of applied sciences. However, teachers from the subsamples differed significantly regarding the proportion of completed *E&S* teaching diplomas. For further pathway analysis, we focused on pathways with an *E&S* teaching diploma. Therefore, we combined the random and the convenience samples for further analyses and included all 155 *E&S* teachers who participated in the main study (for details on sampling see Author(s), 2018b – removed for blind review).

Table 2 Pathways in teacher professional development

	<i>Number</i> ¹	Percent
Completed commercial VET program		
Combined sample	150	36.0
Random sample ²	85	35.3
Convenience sample	83	37.3
Completed university track		
Combined sample	149	78.5
Random sample ²	82	84.1
Convenience sample	82	73.2
Completed <i>E&S</i> teaching diploma		
Combined sample	153	84.3
Random sample ²	83	91.6
Convenience sample	85	77.6

Notes. ¹Sample size varies depending on missing data. ²The random sample of the year 2014 includes 19 teachers who did not participate in the year 2015.

We found that approximately 79 % of the *E&S* teachers completed the academically oriented track and 21 % a practically oriented track. Eight teachers did not provide information about their subject degree or did not finish a higher education degree. 36 % of *E&S* teachers finished a commercial VET program before starting their studies, and

⁴ XXX (removed for blind review) participated in developing the biographical questionnaire.

nearly 84 % of *E&S* teachers completed a teaching diploma in *E&S* from a university of teacher education or university (see for detailed pathway analysis Author(s), 2017b – removed for blind review).

Teachers from universities of applied sciences more frequently completed a commercial VET program than teachers from universities (Table 3). Rather, teachers without a teaching diploma frequented a university pathway.

Table 3 Percentages of teachers from university of applied sciences and university who completed either a commercial VET program or a teaching diploma in *E&S*

	University of applied sciences (<i>N</i> = 32)	University (<i>N</i> = 115)
Completed commercial VET program ¹	84	20
Completed teaching diploma in <i>E&S</i> ²	88	84

Notes. ¹Three cases missing. ²Two cases missing.

The academically and practically oriented pathways are considered the two main teacher training pathways in Switzerland. However, we found additional variation within these two pathways and subsequently identified the most frequently chosen academic (*A*₁ and *A*₂) and practically oriented (*B*) pathways (Fig. 1).

Fig. 1 Three typical training and education pathways of Swiss VET teachers (cf. Author(s), 2017b – removed for blind review)

Note. Work experience means working in the commercial field.

Pathways A_1 and A_2 are considered academically oriented pathways including a master's degree from a university whereas pathway B is considered practically oriented including a bachelor's degree from a university of applied sciences. Work experience indicates experience in the commercial field. Teachers from pathways A_1 and A_2 completed the *E&S* teaching diploma considerably more often at a university than teachers from pathway B (Table 6). Although 85 teachers covered the three main pathways, 70 teachers completed other pathways, which resulted in subsamples of very few teachers. For example, teachers completed a pathway similar to B, except for one of the two stages of work experience, or teachers completed a PhD before enrolling in the *E&S* teaching diploma program.

3.2 Data Collection

We investigated *E&S* teachers' *Economics* PCK and CK in spring 2015. Highly trained test administrators conducted and monitored standardized surveys at commercial VET schools. The CK and the PCK tests had fixed durations of 50 and 60 minutes, respectively. To motivate teachers to participate, we provided individual written feedback about their results compared to the entire sample. The 29 pages of feedback also included details about the teacher sample, example test items and suggestions for guided self-reflection.

Measure of Content Knowledge

The teachers' CK measure had to meet the government's minimum requirements that is defined as knowledge at the bachelor level. The German TUCE (Brückner, Förster, Zlatkin-Troitschanskaia, Happ, et al., 2015; Brückner, Förster, Zlatkin-Troitschanskaia, & Walstad, 2015; Zlatkin-Troitschanskaia et al., 2014) consists of 60 four-alternative multiple choice items measuring bachelor students' microeconomic and macroeconomic knowledge. Both *Microeconomics* and *Macroeconomics* are represented by six content categories (Walstad et al., 2007). For our study⁵, we selected 18 macroeconomic and 17 microeconomic⁶ items of medium difficulty from the German TUCE such that the twelve content categories were each represented by three items. We did so based on an analysis of the Swiss economics bachelor's curricula at three universities and a university of applied sciences that had the highest enrollment of *Economics* and *Business Administration* students in German-speaking Switzerland. Curricula analysis

⁵ Used with permission. WiwiKom-Test. Copyright© 2014 JGU Mainz, FB 03, Wirtschaftspädagogik 1, Mainz. All rights reserved. For more information, see http://www.wipaed.uni-mainz.de/lis/1085_DEU_HTML.php

⁶ Only two items were available for one of the six microeconomic content categories.

revealed considerable content overlap in compulsory economics modules (cf. Author(s), 2016a – removed for blind review). Moreover, the majority of TUCE's content categories were covered by Swiss universities' curricula.

Measure of Pedagogical Content Knowledge

After defining *E&S* teachers' PCK in *Economics*, we developed test items corresponding to the *Economics* PCK construct – the *observation* vertex. To address test content representativeness, we developed the PCK test items in four ways: (1) conceptually and systematically based on the facets of the PCK definition, (2) concretely based on observations of representative *E&S* teaching situations in a video study with teachers in *Economics*, (3) external review and (4) item-piloting with pre-service teachers. Each provided pieces of evidence on content representativeness; we outline them briefly in the following (cf. test development Author(s), 2018a; Author(s), 2017a – removed for blind review).

The items were developed to include five elements (cf. Kuhn et al., 2016, p. 9)⁷:

- (1) context (i.e., situations in *Economics* lessons),
- (2) PCK facet (e.g., dealing with students' misconceptions),
- (3) *Economics* curricular content dimension (e.g., conceptual knowledge of the Gross Domestic Product),
- (4) cognitive process in teaching *Economics* (e.g., analyzing or creating explanations of supply and demand), and
- (5) item format (i.e., constructed-response vs. complex multiple choice).

⁷ Kuhn et al. (2016, p. 9) applied four parts for constructing PCK items.

First, we developed PCK items conceptually to link the construct definition with test content specification as well as items, material and item format (AERA, 2014, p. 14). We developed two item types: (1) constructed-response items (Fig 3 and Fig 4) and (2) selected-response items, namely, complex multiple-choice (Fig 5) and forced-choice items. Complex multiple choice items required test takers to select one of two options (e.g., *apply* or *does not apply*) regarding four statements (cf. Findeisen, 2017; Organisation for Economic Co-operation and Development (OECD), 2014, p. 37).

Second, to obtain evidence for the extent to which the test content embodies typical teaching activities (cf. AERA, 2014, p. 14), we observed the *E&S* teachers teaching *Economics*. Criterion teaching situations constitute authentic and reliable sources for a valid representation of teaching economics content (cf. Author(s), 2010 – removed for blind review) and were sampled from a previous video study with nine *E&S* teachers in commercial VET schools teaching *Economics*. We used the observations to modify the conceptually developed items and materials (e.g., an economy textbook figure), and questions and statements (e.g., a student's misinterpretation of the economic cycle). We ultimately developed of 31 so-called criterion items (15 constructed-response items, 15 complex multiple choice items, and 1 forced-choice item), which were based on the aforementioned criterion situations.

Third, we used external interviews to provide evidence that the teachers' observed answers and performance required thinking that is considered to represent PCK (cf.

AERA et al., 2014, pp. 14–15). A first group of eight interviewees ($N = 5$ teacher education faculty staff, $N = 1$ experienced *E&S* teacher, and $N = 2$ novice *E&S* teachers) first solved and commented on eight to twelve criterion items and the corresponding material. They then evaluated each solved item with regard to, for example, authenticity in *E&S* teaching. This analysis disclosed ambiguous items and material. A second group of 15 *E&S* pre-service teachers worked through 13 of the 31 items in a pilot study at a university in German-speaking Switzerland. Their responses showed that some items were relatively easy to answer (e.g., 14 of 15 respondents answered all items correctly). Such items had to be adjusted or rejected for the main study as they would not enable us to discriminate between teachers. With the aid of a rotated design, items were reviewed during the external interviews and piloted at least twice (Appendix: Table 7).

This series of pilot studies reduced the initial 31 items to 22 items for the main study (Appendix: Table 8). The test booklet included an introduction and information that told test takers that all the items addressed the topic of *Economics* in *E&S* lessons for commercial apprentices in VET schools. The paper-and-pencil test included 13 constructed-response items (Example in Appendix: Fig. 4), one forced choice item, and eight complex multiple choice items (Example in Appendix: Fig. 5).

Scoring of the PCK Items

The PCK test included complex multiple choice items and constructed-response items. The complex multiple choice items were scored on a three-point scale—four correct

responses counted for two points; three correct responses one point and two or fewer correct responses zero points. These item scores were modeled as partial-credit items.

The teachers' responses to the constructed-response items were coded in two stages. In the first stage, two trained graduate-student teachers independently coded the correctness of item responses (0, 1), reaching 82 % agreement and, according to Landis and Koch (1977, p. 165), an acceptable *Cohen's Kappa* of 0.60. In the second stage, they reached consensus when there was disagreement.

3.3 Statistical Procedures

Research Question 1: Validity and Reliability of CK and PCK Measures

Are the CK and PCK tests scores reliable and valid measures? To address this question we used the responses of the 155 teachers. In the CK and the PCK test, 2 % and 4 % of the responses were missing, respectively. In the CK test, five teachers did not answer the final question whereas in the PCK test, six teachers left the final question unanswered. The missing responses were coded as incorrect answers since we assumed that the time to work on the test was sufficient and that a missing response indicated the respondent's lack of knowledge.

We began by examining the internal structure of both measures to determine whether we could retrieve the conceptual structure of CK (macro- and microeconomics items) and PCK (explaining economic content, economic misconceptions, and student cognitive activation). Once we settled on the structure of the tests, we calculated internal-consistency reliabilities for each subscale.

To investigate the question of score-interpretation validity, we examined the data for evidence that the CK and PCK tests correlated with each other as expected (AERA et al., 2014). To provide discriminant evidence, we analyzed latent correlations between CK and PCK in a multidimensional confirmatory factor analysis (CFA). Correlations between PCK and CK dimensions should be markedly lower than 1 and lower than the latent correlation between microeconomic CK (MiCK) and macroeconomic CK (MaCK). Furthermore, we investigated performance differences between female and male teachers in the PCK test by including gender as an observed predictor in a structural equation model. For the CFA models, we used criteria for global fit indices based on recommendations by Hu and Bentler (1999): the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) should be close to .95 or higher, the standardized root mean squared residual (SRMR) should be .09 or lower, and the root mean squared error of approximation (RMSEA) should be .06 or lower.

Research Question 2: Pathway Dependent Teacher Quality

To address the educational policy question, we analyzed the *E&S* teachers' pathways with the aid of descriptive analysis. We then used multivariate general linear models (GLM) to predict test scores for PCK, MiCK, and MaCK. We chose the multivariate approach to account for dependencies of PCK, MiCK, and MaCK. First, we applied multivariate analyses of variance (MANOVAs) with three dichotomous predictors, namely, completed commercial VET program, academic pathway, and completed *E&S* teaching diploma as independent variables. Second, we identified three main pathways, which include the aforementioned predictors. More specifically, we investigated the

relationship between these pathways and the results of the PCK, MiCK and MaCK tests. For this purpose, we applied another MANOVA with the pathway membership as a categorical independent variable. Third, we included gender as another factor and repeated every analysis. To obtain a better understanding of the multivariate results, univariate tests for all the dependent variables were conducted within each MANOVA.

All the descriptive analyses and linear models were computed with SPSS version 21. CFA was performed in Mplus version 7.2 (Muthén & Muthén, 1998–2015).

4 Results

In the following sections, we address our two main research questions: (1) can we measure CK and PCK reliably and validly? And if so, (2) are different paths to teacher certification associated with variation in teacher quality, as measured by CK and PCK?

4.1 Interpretation Vertex: Validity and Reliability

Content Knowledge Structure and Reliability

Given the large number of items and the small sample of teachers, we used item parcels to analyze the dimensional structure of CK. We constructed six content-wise parcels for MiCK and for MaCK. A two-dimensional model with the expected dimensions for microeconomics and macroeconomics fitted the data very well ($\chi^2 = 55.1$; $df = 53$; $p = .397$; RMSEA = .016; CFI = .99; TLI = .99), with a latent correlation of .80 between dimensions. A unidimensional model fitted significantly worse ($\Delta\chi^2 = 8.8$; $\Delta df = 1$; $p = .003$). We decided to use the two-dimensional model for CK for subsequent analyses. The CK test yielded reliabilities of $\alpha = .62$ for the MiCK scale and $\alpha = .72$ for the MaCK

scale. These results were in line with the German results (Zlatkin-Troitschanskaia et al., 2015, pp. 125–126).

Pedagogical Content Knowledge Structure and Reliability

We first examined the PCK test items using classical test theory based on its conceptual structure. Even though the mean item difficulty of all 22 items was $p = .53$, the item discriminations, the average item intercorrelation, and the Coefficient Alpha (α) for all the items (unidimensional) and the PCK facets (three-dimensional) were low. The CFA for both the unidimensional and three-dimensional PCK model proved to fit the data unsatisfactorily. However, in these analyses, poorly fitting items became apparent. We decided to remove items from the scale based on both classical item discrimination and non-significant factor loadings from the unidimensional CFA model. Seven complex multiple choice items, the forced-choice item and two constructed-response items were removed. This removal resulted in a reduced scale of twelve items (11 constructed-response and one complex multiple choice) with difficulties ranging from .13 to .73 and a mean of .51. The remaining items still covered all three PCK facets, with two items for explaining, two for misconceptions, and eight for cognitive activation. However, the reliability of the reduced PCK scale was still low, at $\alpha = .56$ (based on WLE .55 and on PV .56). The item discriminations were also low, ranging from .12 to .32.

For the remaining twelve items, we then ran a unidimensional CFA. All the items had significant factor loadings ranging from .24 to .52. However, the model fit remained unacceptable ($\chi^2 = 82.7$; $df = 54$; $p = .007$; RMSEA = .059; CFI = .72; TLI = .66). An inspection of residuals showed that this finding was a result of dependencies between

the items referring to the same item stem (testlets). Consequently, the unidimensional model was extended by two testlet factors – one testlet of two items and one of four items. This bifactor model yielded an acceptable fit ($\chi^2 = 57.4$; $df = 49$; $p = .191$; RMSEA = .033; CFI = .92; TLI = .89). Based on this result, we kept the number of twelve items for our PCK measure and used this measure for further analyses.

Because the PCK test was conceptually developed as representing three PCK facets (section 2.2), and given the small sample of teachers, we decided to make use of item parceling (Little, Cunningham, Shahar, & Widaman, 2002) to model PCK in the more complex models including other dimensions and variables. We constructed three indicators by calculating the means of the twelve items for the PCK facets explaining (2), misconceptions (2), and cognitive activation (8), respectively (Little, 2013, p. 21). Please note that the items having loadings on the testlet factors in the bi-factor model are aggregated in the same parcel. The three indicators were treated as indicators of the general PCK dimension (Little, 2013, p. 20). With three indicators measuring one dimension, a CFA model is just-identified and its fit is trivial ($df = 0$). To examine whether the three indicators represent the PCK facets equally, a CFA with parallel measures (equal loadings and equal residual variances for all indicators) was tested. The model fitted the data very well ($\chi^2 = 1.8$; $df = 4$; $\chi^2/df = 0.448$; CFI = 1.000; TLI = 1.12; RMSEA = .000; SRMR = .053). Hence, the model with parallel measurement was used as a measurement model for PCK within the subsequent, more complex models.

Relationship to Other Variables

To examine the overall structure of teachers' knowledge, including the relationships between PCK and CK, we analyzed a CFA model⁸ with three dimensions: PCK, MiCK, and MaCK. The content-wise CK item parcels (micro- and macroeconomics) and facet-wise PCK item parcels (explaining, misconceptions, and cognitive activation) and measurement models tested in the previous separate models represent the latent variables.⁹

Fig. 2 Standardized solution of the joint CFA for CK and PCK (Author(s), 2018a – removed for blind review)

Notes. PCK = Pedagogical content knowledge; MiCK = Microeconomic content knowledge; MaCK = Macroeconomic content knowledge; Expl = Explaining; Miscon = Misconceptions; Cogact = Cognitive Activation.

The model fitted the data well ($\chi^2 = 99.7$; $df = 91$; $p = .249$; RMSEA = .025; CFI = .97; TLI = .96). The standardized coefficients of the overall model are displayed in Fig. 2.

⁸ For MiCK and MaCK, we replicated the aforementioned two-dimensional structure of Zlatkin-Troitschanskaia et al. (2015) in a CFA manner. By including PCK, the CFA had an exploratory factor analysis (EFA) character.

⁹ The COACTIV project used a similar procedure with PCK subscales and CK parcels (Krauss, Brunner, et al., 2008).

The latent correlation between PCK and MiCK turned out to be higher (.57) than that with MaCK (.38).

Gender

Gender differences in PCK, MiCK, and MaCK were analyzed by extending the three-dimensional CFA to include gender as an observed predictor variable (female = 0, male = 1); the model fit was acceptable. Regression coefficients were standardized using the variances of the latent variables only ('Y-standardization'; Muthén & Muthén, 1998–2015); they can be interpreted as the difference in standard deviations of the knowledge dimensions between female and male teachers. There was no significant gender difference in PCK ($\beta = 0.02$; $p = .937$). For CK, male teachers had an advantage in both MiCK ($\beta = 0.52$; $p = .018$) and MaCK ($\beta = 0.54$; $p = .008$).

4.2 RQ 2: Pathways to Teaching and Their Association with CK and PCK

For years, the Swiss government has been concerned with the quality of VET teacher preparation across various entry-to-teaching pathways. Our focus was whether the various pathways produced equally qualified teachers as measured by PCK, MiCK and MaCK or whether the pathway outputs differ. The pathways varied with regard to whether teachers had (1) completed a prior commercial VET program or not, (2) graduated from a university of applied sciences or a university, (3) completed an *E&S* teaching diploma or not, and (4) a combination of these predictors.

Table 4 provides an overview of the descriptive statistics for PCK, MiCK and MaCK depending on the predictors.

Table 4 Descriptive statistics: Association of pathway with gender, CK and PCK (values are percentages of gender, PCK, MiCK and MaCK of correct answers) (Author(s), 2018a – removed for blind review).

	Total sample	Commercial VET program		Academic track		E&S teaching diploma	
		Not completed	Completed	University of applied sciences	University	No	Yes
<i>N</i>	155	96	54	32	117	24	129
Female	28.4	32.3	20.4	25.0	29.9	37.5	26.4
PCK							
<i>M</i>	52.1	51.5	53.8	57.5	50.7	49.7	52.8
<i>SD</i>	16.6	16.6	17.2	15.0	16.8	17.0	16.6
MiCK							
<i>M</i>	60.9	62.8	58.6	57.9	62.0	53.7	62.4
<i>SD</i>	17.7	18.6	15.5	13.3	18.5	20.5	16.9
MaCK							
<i>M</i>	73.3	74.2	71.9	75.9	72.9	68.5	74.5
<i>SD</i>	17.0	17.6	16.3	14.0	17.6	21.5	16.0

Notes. PCK = Pedagogical content knowledge; MiCK = Microeconomic content knowledge; MaCK = Macroeconomic content knowledge.

Overall, although differing in preparation, the teachers' performances on the PCK, MiCK and MaCK tests were quite similar. Nevertheless, some specific differences can be found between teachers from universities of applied sciences and universities, particularly with regard to PCK. Moreover, teachers with a teaching diploma generally outperformed teachers without a teaching diploma on every test. Teachers with a

university degree outperformed teachers with a degree from a university of applied sciences only on the MiCK test.

To check the interpretation of observed differences, three MANOVAs were estimated with each of the three predictor variables as single predictors, namely, teachers who had completed a commercial VET program or not (model 1), graduated from a university of applied sciences or a university (model 2), and completed a teaching diploma in *E&S* or not (model 3). Moreover, a fourth MANOVA (model 4) was estimated including all the predictors simultaneously (Table 5).

None of the models showed significant multivariate effects for prior completion of a commercial VET program and teaching diploma (Table 5). The multivariate effect of academic track was significant both when used as a single predictor (model 2; Wilk's $\Lambda = .935$; $p = .021$) or with the other predictors (model 4; Wilk's $\Lambda = .936$; $p = .031$). Except for PCK ($p = .047$) in model 2 favoring teachers from universities of applied sciences, none of the other univariate tests for the three dependent variables reached significance. Even though there was no multivariate effect of teaching diploma, we found a univariate effect on MiCK ($p = .027$) in model 3. Nevertheless, this result indicates the overall profiles differ in the three dependent variables (the mean vectors) of teachers from different tracks differ, but not the means in all the individual variables. This effect can occur if the multivariate difference is due to a combination of variables (Field, 2009, p. 608 et seq.; Pituch & Stevens, 2016, p. 156).

All four models were repeated, including gender as a covariate (Table 5). The results essentially confirmed the analyses without gender. None of the effects for the prior completion of a VET program and teaching diploma were significant, and the multivariate effect of academic track was significant when gender was included in model 2 (Wilk's $\Lambda = .933$; $p = .019$) with a univariate effect for the individual dependent variable PCK ($p = .045$). When gender was included in model 4, the effect for academic track was significant (Wilk's $\Lambda = .936$; $p = .031$). With one exception, no univariate effects were found for the individual dependent variables PCK and MiCK; the exception was MaCK ($p = .037$) for VET program in model 4.

Table 5 Alternative pathways (models): Results of MANOVA and MANCOVA

Model		Without gender		With gender	
	Predictor	Multivariate effects	Univariate effects	Multivariate effects	Univariate effects
1	Completed commercial VET program or not	n.s.	n.s.	n.s.	n.s.
2	University or university of applied sciences	$p = .021$	PCK $p = .047$	$p = .019$	PCK $p = .045$
3	Teaching diploma in E&S or not	n.s.	MiCK $p = .027$	n.s.	MiCK $p = .039$
4	Combination of all predictors simultaneously				
	Commercial VET program	n.s.	n.s.	n.s.	MaCK $p = .037$
	University	$p = .031$	n.s.	$p = .031$	n.s.
	Teaching diploma in E&S	n.s.	n.s.	n.s.	n.s.

Although the practically and academically oriented pathways are considered the two main teacher-training pathways in Switzerland we found variation within these two pathways and subsequently identified the three most frequent pathways (Fig 2): A1 academic with work experience, A2 academic, and B universities of applied sciences (practically oriented). Note that the analyses were generated from a subsample of 85 teachers covering pathway A1, A2, and B. Descriptive results comparing these alternative paths are provided in Table 6.

Table 6 Descriptive statistics for three typical pathways to certification (values are percentages of gender, PCK, MiCK and MaCK of correct answers)

	Total sample	Subsample	A ₁ Academic with work experience	A ₂ Academic	B Practically oriented
<i>N</i>	155	85	40	30	15
Female	28.4	31.8	40.0	30.0	13.3
Teaching diploma at a university	57.4	68.2	74.4	89.3	26.7
PCK					
<i>M</i>	52.1	53.2	53.5	49.2	60.3
<i>SD</i>	16.6	16.5	15.3	17.3	16.4
MiCK					
<i>M</i>	60.9	62.8	61.6	66.7	58.0
<i>SD</i>	17.7	17.8	19.7	18.4	8.3
MaCK					
<i>M</i>	73.3	74.4	72.2	78.3	72.2
<i>SD</i>	17.0	16.1	16.6	16.3	13.5

Notes. PCK = Pedagogical content knowledge; MiCK = Microeconomic content knowledge; MaCK = Macroeconomic content knowledge.

Teachers from universities of applied sciences (B) appeared to outperform teachers from universities (A₁ and A₂) on the PCK test (Table 6). However, when we performed a MANOVA on PCK, MiCK, and MaCK with pathway as an independent variable, we found no significant effect (Wilk's $\Lambda = .909$, $p = .119$). Moreover, we considered gender as an additional factor and applied MANCOVA. Again, no multivariate effect for pathway was found (Wilk's $\Lambda = .899$, $p = .074$) (note: female teachers are over-represented in pathway A₁ and under-represented in pathway B).

5 Discussion and Limitations

The Swiss government is interested in the quality of E&S teachers from multiple training and education pathways. PCK and CK are possible measures of teachers' quality. As a CK measure, we used a German version of the TUCE with known reliability and validity, which we replicated in the Swiss sample. However, because no PCK measure was adaptable to the Swiss context, we developed a PCK test. We then asked whether we could measure teachers' *Economics* CK and PCK validly and reliably (RQ 1). With the measures of PCK and CK in hand, we could then, in a very preliminary way, address the *educational policy question* (RQ 2) of whether the different paths to certification produced teachers who systematically varied with regard to these measures of "quality".

5.1 RQ 1: Validity and Reliability

Whereas the reliability of the MiCK and MaCK was adequate, the newly developed PCK test provided scores with rather low reliability ($\alpha = .56$) even after excluding poorly performing items (especially complex multiple choice). This fact raised the question of whether the remaining items produced a sufficient measure for PCK in *Economics*. To

address this question, we note that the test covers the three PCK facets: explaining, misconceptions, and cognitive activation. Future revisions of the test with additional items covering the three PCK facets hold promise. Moreover, comparing PCK tests' reliabilities reported by KoMeWP (Schnick-Vollmer et al., 2015, p. 28, EAP reliability .64) and Kuhn (2014, pp. 194–196, Cronbach's $\alpha = .675$) in the field of *E&S* were higher reliabilities but still less than desirable. Krauss, Neubrand, et al. (2008, p. 240) also reported rather low reliabilities for each PCK facet ($.51 \leq \alpha \leq .63$) and a moderate reliability for overall PCK ($\alpha = .77$).

We interpret our results along two lines. First, teachers may have had difficulties with the complex multiple choice item format. Unlike the complex multiple choice items, a constructed-response item did not require a single right answer but a self-generated answer, systematized and defined in a coding manual. We learned from teacher interviews that initially less appropriate answers might still prove correct with appropriate argumentation. Second, we may have encountered the same challenges with PCK test development as Hill et al. (2008). They reported low reliabilities for teachers' knowledge of content and of students, which they explained with their multidimensional PCK construct that might have occurred in the Economics PCK test as well.

Information about the PCK test in terms of *interpretation*, particularly discriminant evidence of validity, was provided by the correlation analysis and the CFA with CK. The CK dimensions MiCK and MaCK correlated more highly with each other than either did

with PCK separately. Medium-sized correlations indicated that the PCK test appeared to be separate from the CK dimensions MiCK and MaCK. Although the PCK test involved fairly balanced microeconomic and macroeconomic content, we found a higher correlation between MiCK and PCK. We note that it is more challenging for *E&S* teachers to make macroeconomic content accessible and relevant to students' everyday private and professional lives than microeconomic content.

5.2 RQ 2: Association of Teachers' Pedagogical Content Knowledge and Content Knowledge with Different Pathways

Our lessons learned from the answers to the measurement question provide a basis for the *educational policy question*. As a first step, we analyzed the PCK, MiCK and MaCK test results depending on three typical pathway characteristics, namely, the prior completion of a commercial VET program, academic track and completed teaching diploma in *E&S*. Multivariate analyses of variance revealed a significant difference between teachers from a university of applied sciences and teachers from a university. Essentially, teachers from a university of applied sciences scored significantly better in the PCK test than teachers from a university.

Even though a previously completed VET program was an important positive predictor for future teachers' PCK and/or CK or *Economics* students in other studies (Brückner, Förster, Zlatkin-Troitschanskaia, & Walstad, 2015; Fritsch et al., 2015), it did not predict PCK, MiCK or MaCK in this study. Consequently, our finding suggests that the effect of a completed VET program may not hold true for more experienced in-service teachers because of their opportunities to learn during teaching. We did not find multivariate

significant differences between teachers with and without a completed teaching diploma regarding their PCK, MiCK and MaCK test performance.

As a second step, we accounted for differing training and education pathways. We identified three typical pathways into teaching *Economics*—combinations of prior completion of a commercial VET program, academic track and completed *E&S* teaching diploma—and repeated the MANOVA for these pathways. Surprisingly, this MANOVA yielded no significant difference among the three pathways indicating that they were associated with similar results regarding teachers' PCK, MiCK and MaCK. Despite the apparent PCK mean differences between academically oriented pathways (A₁ and A₂) and the practically oriented pathway (B), there was no significant multivariate effect, not even with gender as an additional factor. Accounting for all three aforementioned predictors simultaneously would require larger subsample sizes.

5.3 Limitations

This study has three major limitations. First, achieving a fairly representative and large sample of *E&S* teachers in Switzerland is challenging because no statistical data are available on the population of *E&S* teachers. Moreover, many commercial VET teachers work part-time (see Table 1), leading to a varying teacher population year-to-year. Nevertheless, we sampled 155 teachers in two waves, which is encouraging considering 357 commercial VET classes are in the population. Whereas the first wave randomly sampled *E&S* classes, the second wave was conveniently sampled. However, teachers' CK and PCK test performance between the randomly sampled and

conveniently sampled teachers did not differ. Hence, sample size constraints meant we could not analyze differential item functioning (e.g., gender) or measurement invariance for subsamples (e.g., gender, pathways) that would be interesting, albeit it would require a larger sample, perhaps larger than the population of *E&S* teachers in Switzerland.

Second, a multidimensional *Economics* PCK construct with low reliability might result from a PCK test based on real teaching situations. On the one hand, situation-based items should provide representative and authentic measures for PCK in *Economics*. On the other hand, however, situation-specificity represents content validity and could reduce the reliability of the overall PCK factor because different situation-based items measure situation-specific knowledge. This issue can be regarded as a typical validity-reliability dilemma (e.g., Rost, 2018; Slomp & Fuite, 2004) that might apply for the PCK test with the rather low reliability presented here. The question arises regarding how many such items are needed to measure the multidimensional construct reliably and whether this test is practical.

The third limitation of the study is that the results of the PCK test are based on the same sample as the validity evidence. Ideally, the results should be replicated by an independent sample. Therefore, future research should aim at collecting more validity evidence for measuring PCK, e.g., answers and response processes from different samples as *Economics* students and novice teachers (cf. AERA et al., 2014).

5.4 Future Research

To conclude, the question of whether practical and academic preparation paths lead to differential commercial teacher “quality” leads to complex answers. Overall, future research should include indicators for measuring students’ perception of instructional quality and students’ performance.

Furthermore, the PCK test requires in-depth research because teachers’ PCK and its correlation with instructional quality, for example cognitive activation, was one of the main predictors of student achievement in COACTIV (Baumert et al., 2010, pp. 163–164). Even if the definition of this PCK facet differs in our study, it is a promising prospect to link teachers’ data with apprentices’ perceptions of the cognitive activation of *E&S* lessons and apprentices’ competence development in *Economics* (Author(s), 2018c – removed for blind review). Future research on the PCK test itself should involve further item development covering all the PCK facets to split the prospective modified PCK test into reliable subtests for each facet (cf. Rost, 2018), as noted above.

In order to address the policy question additional analysis are conceivable. First, in-depth analysis of opportunities to learn, namely, comparing contents of courses dealing with teachers’ CK and PCK provides further insights into differences among and within pathways (cf. Blömeke & Delaney, 2012). Second, pathways to certification are more complex than those represented above. Further and alternative segmentation based on additional variables (e.g., years of teaching, further education) might prove informative. Moreover, alternative modeling approaches, such as latent profile analysis, to arrive at

pathways might be carried out that might reveal larger clusters of similar teachers' characteristics and pathways.

From a more comprehensive perspective, PCK and CK are just two aspects of teacher "quality". Most empirical studies on teachers' professional competence, including this one, sought to measure situation-based knowledge. However, some have argued that knowledge and professional action are only analytically distinguishable and merge in teaching situations (cf. Author(s), 2015 – removed for blind review; Dreyfus & Dreyfus, 1980; Neuweg, 2014; Oser, 2013, pp. 46–47; Reusser & Messner, 2002). Therefore, in-depth research on the assemblage of teacher knowledge, beliefs, behavior in concrete teaching situations in the classroom, and reflection on teaching situations are needed to inform teacher education and training.

6 Appendix

Table 7 Example item development

(1) and (2) conceptually developed original item based on observations in the video study	<i>You would like to illustrate the development of unemployment within a year in Switzerland. Mention three aspects in keywords, which you must include when applying this table in the E&S lesson.</i>	
(3) Input during expert interviews and (4) Piloting	Interviews: <i>'I know, did not know now exactly, ..., what do you mean now with 'aspects'? ... But I think, the unemployment development should probably be the crucial point [of the table]' (Interview 1).</i>	Item piloting with future teachers (examples): <ul style="list-style-type: none"> • <i>Change of unemployment</i> • <i>Calculation of unemployment</i> • <i>Change of population</i> • <i>Total number and percentage</i>
Revised item for the main study	<i>You would like to illustrate the development of unemployment within a year in Switzerland. Name two reasons in keywords to state why the table is not adequate to illustrate this.</i>	

Table 8 Distribution of PCK items with respect to construct facets and item formats

Facet	Response format	Stems	Items	Points	Σ Points
Explaining	CMC ¹	4	4	7	13
	CR	3	3	6	
Students' misconceptions	CMC	3	3	6	9
	CR	2	2	3	
Cognitive activation	CMC	2	2	4	22
	CR	4	8	18	
Σ		18	22	44	44

Note. ¹One forced choice item.

Your colleague taught the international division of labor. As shown below, he used a blackboard to compare the absolute and comparative advantage.

Absolute advantage	Comparative advantage
Cheese and clothing are produced in Switzerland and France. Both countries benefit from the product exchange if the production of cheese is cheaper in Switzerland and the production of clothing cheaper in France.	Cheese and clothing are produced in Switzerland and France. The two countries still benefit from the product exchange even if both the production of cheese and clothing is cheaper in France.

The students react skeptically to the explanation of comparative advantage. Which alternative and student-oriented example do you recommend to your colleague to explain the comparative advantage?

Answer in keywords:

My example to illustrate the comparative advantage:

_____
_____

Fig. 3 Constructed-response item example for explanation

You are teaching the topic 'market economy' and you explain with the help of the following figure that the demand curve is affected by the budget.

Source: Own representation (English translation by authors)

A student raises his hand and says:

This is illogical. It would mean that after having finished my apprenticeship and having a significant higher income I would have to drink significantly more coffee. But I can only drink as much coffee as I already drink.'

What is your response to the student? Please note your response in keywords:

_____
_____

Fig. 4 Constructed-response item example for a student's misconception

At the end of a lesson, you try to deepen the stages of the economic cycle for students with the following picture:

Source: Figure based on Rüfenacht, H., Saxer, U., & Tobler, T. (2014). Brennpunkt Wirtschaft und Gesellschaft. Band 3. Ausgabe für Lehrpersonen. Zürich: Verlag SKV. Section 2, p. 2. (English translation by authors, translation and reprint with permission)

Which of the following teacher questions prompts a cognitive conflict for a student regarding the economic cycle? Please check one box in each row.

Teachers' question	Cognitive conflict	No cognitive conflict
A Why do some economic participants influence the economic development of Switzerland more strongly than others?	☐	☐
B Under which conditions do the GDP growth and the rate of unemployment rise simultaneously in Switzerland?	☐	☐
C How can the growth trend be positive if the economic situation declines?	☐	☐
D Why does the GDP growth of Switzerland develop negatively with stable private consumption at the same time?	☐	☐

Fig. 5 Complex multiple choice item example for cognitive activation

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington: AERA.
- Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching. Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator, Fall 2005*, 14–46.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407. doi:10.1177/0022487108324554
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (pp. 25–48). New York: Springer.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. doi:10.3102/0002831209345157
- Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM – The International Journal on Mathematics Education*, 44(3), 223–247. doi:10.1007/s11858-012-0429-7
- Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & H. W. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn. TEDS-M results* (pp. 19–47). Dordrecht: Springer.
- Blömeke, S., Suhl, U., Kaiser, G., Felbrich, A., Schmotz, C., & Lehmann, R. (2010). Lerngelegenheiten und Kompetenzerwerb angehender Mathematiklehrkräfte im internationalen Vergleich. *Unterrichtswissenschaft*, 38(1), 29–50.
- Bouley, F., Wuttke, E., Schnick-Vollmer, K., Schmitz, B., Berger, S., Fritsch, S., & Seifried, J. (2015). Professional competence of prospective teachers in business and economics education: Evaluation of a competence model using structural equation modeling. *Peabody Journal of Education*, 90(4), 491–502. doi:10.1080/0161956X.2015.1068076
- Brückner, S., Förster, M., Zlatkin-Troitschanskaia, O., Happ, R., Walstad, W. B., Yamaoka, M., & Asano, T. (2015). Gender effects in assessment of economic knowledge and understanding: Differences among undergraduate business and economics students in Germany, Japan, and the United States. *Peabody Journal of Education*, 90(4), 503–518. doi:10.1080/0161956X.2015.1068079
- Brückner, S., Förster, M., Zlatkin-Troitschanskaia, O., & Walstad, W. B. (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. *Studies in Higher Education*, 40(3), 437–453. doi:10.1080/03075079.2015.1004235
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T., . . . Neubrand, M. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9(4), 521–544. doi:10.1007/s11618-006-0166-1
- Carlsen, W. S. (1999). Domains of teacher knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge. The construct and its implications for science education* (pp. 133–144). Dordrecht: Springer.

- Criblez, L. (2016). Switzerland: Teacher education. In T. Sprague (Ed.), *Education in non-EU countries in Western and Southern Europe. Education around the world* (pp. 99–121). London: Bloomsbury Academic.
- 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. *Teaching and Teacher Education*, 34, 12–25. doi:10.1016/j.tate.2013.03.001
- Dreyfus, H. L., & Dreyfus, S. E. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Berkeley: University of California.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Findeisen, S. (2017). *Fachdidaktische Kompetenzen angehender Lehrpersonen. Eine Untersuchung zum Erklären im Rechnungswesen*. Wiesbaden: Springer.
- 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. *Empirical Research in Vocational Education and Training*, 7(4), 1–20. doi:10.1186/s40461-015-0014-8
- Gess-Newsome, J. (1999). Pedagogical content knowledge: An introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge. The construct and its implications for science education* (pp. 3–17). Dordrecht: Springer.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372–400.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Hoeckel, K., Field, S., & Grubb, W. N. (2009). *Learning for jobs. OECD reviews of vocational education and training. Switzerland*. Paris: Organisation for Economic Co-operation and Development (OECD).
- Hoffman, N., & Schwartz, R. (2015). *Gold standard: The Swiss vocational education and training system. International comparative study of vocational education systems*. Washington, DC: National Center on Education and the Economy.
- Hu, L. & 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. doi:10.1080/10705519909540118
- König, J., Ligtoet, R., Klemenz, S., & Rothland, M. (2017). Effects of opportunities to learn in teacher preparation on future teachers' general pedagogical knowledge: Analyzing program characteristics and outcomes. *Studies in Educational Evaluation*, 53, 122–133. doi:<http://dx.doi.org/10.1016/j.stueduc.2017.03.001>
- Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM – The International Journal on Mathematics Education*, 40(5), 873–892. doi:10.1007/s11858-008-0141-9
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716–725. doi:10.1037/0022-0663.100.3.716
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M., & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *Journal für Mathematik-Didaktik*, 29(3/4), 223–258.

- Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung*. Landau: Empirische Pädagogik.
- Kuhn, C., Alonzo, A. C., & Zlatkin-Troitschanskaia, O. (2016). 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. *Empirical Research in Vocational Education and Training*, 8(5), 1–18. doi:10.1186/s40461-016-0031-2
- Kuhn, C., Happ, R., Zlatkin-Troitschanskaia, O., Beck, K., Förster, M., & Preuße, D. (2014). Kompetenzentwicklung angehender Lehrkräfte im kaufmännisch-verwaltenden Bereich – Erfassung und Zusammenhänge von Fachwissen und fachdidaktischem Wissen. In E. Winther & M. Prenzel (Eds.), *Perspektiven der empirischen Berufsbildungsforschung. Kompetenz und Professionalisierung. Zeitschrift für Erziehungswissenschaft. Sonderheft 22* (pp. 149–167). Wiesbaden: Springer.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi:10.2307/2529310
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 151–173. doi:10.1207/S15328007SEM0902_1
- Mindnich, A., Berger, S., & Fritsch, S. (2013). Modellierung des fachlichen und fachdidaktischen Wissens von Lehrkräften im Rechnungswesen – Überlegungen zur Konstruktion eines Testinstruments. In U. Faßhauer, B. Fürstenau, & E. Wuttke (Eds.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2013. Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)* (pp. 61–72). Opladen: Barbara Budrich.
- Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus. Statistical analysis with latent variables. User's guide*. Los Angeles, CA: Muthén & Muthén.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment. Committee on the Foundations of Assessment. Pellegrino, J. W.; Chudowsky, N.; Glaser, R., (Eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academy Press.
- Neuweg, G. H. (2014). Das Wissen der Wissensvermittler. Problemstellungen, Befunde und Perspektiven der Forschung zum Lehrerwissen. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (pp. 583–614). Münster: Waxmann.
- Organisation for Economic Co-operation and Development (OECD). (2014). *PISA 2012 technical report*. Paris: OECD.
- Oser, F. (2013). „I Know How to Do It, But I Can't Do It“: Modeling competence profiles for future teachers and trainers. In S. Blömeke, O. Zlatkin-Troitschanskaia, C. Kuhn, & J. Fege (Eds.), *Modeling and measuring competencies in higher education: Tasks and challenges* (pp. 45–59). Rotterdam: Sense.
- Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*. New York: Routledge.
- Reusser, K., & Messner, H. (2002). Das Curriculum der Lehrerinnen- und Lehrerbildung – ein vernachlässigtes Thema. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 20(3), 282–299.
- Rost, J. (2018). Reliabilitäts-Validitätsdilemma. In M. A. Wirtz (Ed.), *Dorsch – Lexikon der Psychologie*. Bern: Hogrefe. <https://m.portal.hogrefe.com/dorsch/reliabilitaets-validitaetsdilemma/>. Accessed 6 May 2018.

- Schnick-Vollmer, K., Berger, S., Bouley, F., Fritsch, S., Schmitz, B., Seifried, J., & Wuttke, E. (2015). Modeling the competencies of prospective business and economics teachers: Professional knowledge in accounting. *Zeitschrift für Psychologie*, 223(1), 24–30. doi:10.1027/2151-2604/a000196
- Senk, S. L., Tatto, M. T., Reckase, M., Rowley, G., Peck, R., & Bankov, K. (2012). Knowledge of future primary teachers for teaching mathematics: An international comparative study. *ZDM – The International Journal on Mathematics Education*, 44(3), 307–324. doi:10.1007/s11858-012-0400-7
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Slomp, D. H., & Fuite, J. (2004). Following Phaedrus: Alternate choices in surmounting the reliability/validity dilemma. *Assessing Writing*, 9(3), 190–207. doi:<https://doi.org/10.1016/j.asw.2004.10.001>
- Staatssekretariat für Bildung, Forschung und Innovation (SBFI). (2015). *Rahmenlehrpläne – Berufsbildungsverantwortliche vom 01. Februar 2011, (Stand 1.1.2015)*. Bern: SBFI.
- Staatssekretariat für Bildung, Forschung und Innovation (SBFI). (2017). *Verordnung des SBFI er die berufliche Grundbildung Kauffrau/Kaufmann mit eidgen schem Fähigkeitszeugnis (EFZ) vom 26. September 2011 (Stand am 1. Mai 2017) i.V.m. Leistungszielkataloge der SKAAB (Stand 1. Januar 2017)* Bern: SBFI.
- State Secretariat for Education, Research and Innovation (SERI). (2016). *Vocational and professional education and training in Switzerland. Facts and figures 2016*. Bern: SERI.
- State Secretariat for Education, Research and Innovation (SERI). (2017). *Higher education and research in Switzerland*. Bern: SERI.
- Swiss Coordination Centre for Research in Education (SCCRE). (2014). *Swiss education report 2014*. Aarau: SCCRE.
- Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M). Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Walstad, W. B., Watts, M., & Rebeck, K. (2007). *Test of understanding in college Economics: Examiner's manual (Forth Edition)*. New York: National Council on Economic Education.
- Zlatkin-Troitschanskaia, O., Förster, M., Brückner, S., & Happ, R. (2014). Insights from a German assessment of business and economics competence. In H. Coates (Ed.), *Higher education learning outcomes assessment. International perspectives* (pp. 175–197). Frankfurt am Main: Peter Lang.
- Zlatkin-Troitschanskaia, O., Förster, M., Schmidt, S., Brückner, S., & Beck, K. (2015). Erwerb wirtschaftswissenschaftlicher Fachkompetenz im Studium. Eine mehrbenenanalytische Betrachtung von hochschulischen und individuellen Einflussfaktoren. In S. Blömeke & O. Zlatkin-Troitschanskaia (Eds.), *Kompetenzen von Studierenden. Zeitschrift für Pädagogik. Beiheft 61* (pp. 116–135). Weinheim: Beltz Juventa.

Author(s) (2007). Removed for blind review.
Author(s) (2010). Removed for blind review.
Author(s) (2013). Removed for blind review.
Author(s) (2015). Removed for blind review.
Author(s) (2016a). Removed for blind review.
Author(s) (2016b). Removed for blind review.
Author(s) (2017a). Removed for blind review.
Author(s) (2017b). Removed for blind review.
Author(s) (2018a). Removed for blind review.
Author(s) (2018b). Removed for blind review.
Author(s) (2018c). Removed for blind review.

Supplemental Material

Summary of Definitions of Pedagogical Content Knowledge (PCK)

Project	Content Area	Facets of PCK—Teachers Knowledge of:
Learning Mathematics for Teaching (LMT) ¹	Mathematics	1. Content and students, i.e. understanding students' learning processes
Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy (COACTIV) ²	Mathematics	1. Instruction, i.e. how to present and make content comprehensible 2. Students' cognitions (e.g., misconceptions, errors) 3. Tasks' potential and alternative solution paths
Teacher Education and Development Study: Learning to Teach Mathematics (TEDS-M) ³	Mathematics	1. Curriculum 2. Planning teaching and learning 3. Enacting for teaching and learning (e.g., explaining or presenting mathematical content)
Modelling and Measurement of Content Knowledge and Pedagogical Content Knowledge in Business and Economic Education (KoMeWP) ⁴	Accounting	1. Representing and explaining content 2. Students' cognitions and errors 3. Tasks and instructional tools
Kuhn's ⁵ dissertation	Sales, Buying Processes and Economic Principles	1. Teaching objectives, content, and methods 2. Students' learning processes

Notes. ¹Ball, Hill, and Bass (2005), Ball, Thames, and Phelps (2008), Hill, Rowan, and Ball (2005), Hill, Ball, and Schilling (2008). For more information, see: <http://www.umich.edu/~lmtweb/> (Accessed 4 May 2018). ²Baumert et al. (2010), Baumert and Kunter (2013), Krauss, Baumert, and Blum (2008), Krauss, Brunner, et al. (2008). For more information, see: <https://www.mpib-berlin.mpg.de/coactiv/en/index.php> (Accessed 4 May 2018). ³Blömeke and Delaney (2012), Blömeke and Kaiser (2014), Senk et al. (2012), Tatto et al. (2008). For more information, see: <http://www.iea.nl/other-iea-studies> (Accessed 4 May 2018). ⁴Bouley et al. (2015, p. 493), Fritsch et al. (2015), Mindnich, Berger, and Fritsch (2013). For more information, see: http://www.kompetenzen-im-hochschulsektor.de/235_ENG_HTML.php (Accessed 4 May 2018). ⁵Kuhn (2014), Kuhn, Alonzo, and Zlatkin-Troitschanskaia (2016).

References

- Ball, D. L., Hill, H. C., & Bass, H. (2005). Knowing mathematics for teaching. Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator, Fall 2005*, 14–46.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407. doi:10.1177/0022487108324554
- Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV project* (pp. 25–48). New York: Springer.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1), 133–180. doi:10.3102/0002831209345157
- Blömeke, S., & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM – The International Journal on Mathematics Education*, 44(3), 223–247. doi:10.1007/s11858-012-0429-7
- Blömeke, S., & Kaiser, G. (2014). Theoretical framework, study design and main results of TEDS-M. In S. Blömeke, F.-J. Hsieh, G. Kaiser, & H. W. Schmidt (Eds.), *International perspectives on teacher knowledge, beliefs and opportunities to learn. TEDS-M results* (pp. 19–47). Dordrecht: Springer.
- Bouley, F., Wuttke, E., Schnick-Vollmer, K., Schmitz, B., Berger, S., Fritsch, S., & Seifried, J. (2015). Professional competence of prospective teachers in business and economics education: Evaluation of a competence model using structural equation modeling. *Peabody Journal of Education*, 90(4), 491–502. doi:10.1080/0161956X.2015.1068076
- 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. *Empirical Research in Vocational Education and Training*, 7(4), 1–20. doi:10.1186/s40461-015-0014-8
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372–400.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371–406.
- Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers' pedagogical content knowledge and content knowledge: Validation of the COACTIV constructs. *ZDM – The International Journal on Mathematics Education*, 40(5), 873–892. doi:10.1007/s11858-008-0141-9
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary

- mathematics teachers. *Journal of Educational Psychology*, 100(3), 716–725. doi:10.1037/0022-0663.100.3.716
- Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung*. Landau: Empirische Pädagogik.
- Kuhn, C., Alonzo, A. C., & Zlatkin-Troitschanskaia, O. (2016). 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. *Empirical Research in Vocational Education and Training*, 8(5), 1–18. doi:10.1186/s40461-016-0031-2
- Mindnich, A., Berger, S., & Fritsch, S. (2013). Modellierung des fachlichen und fachdidaktischen Wissens von Lehrkräften im Rechnungswesen – Überlegungen zur Konstruktion eines Testinstruments. In U. Faßhauer, B. Fürstenau, & E. Wuttke (Eds.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2013. Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE)* (pp. 61–72). Opladen: Barbara Budrich.
- Senk, S. L., Tatto, M. T., Reckase, M., Rowley, G., Peck, R., & Bankov, K. (2012). Knowledge of future primary teachers for teaching mathematics: An international comparative study. *ZDM – The International Journal on Mathematics Education*, 44(3), 307–324. doi:10.1007/s11858-012-0400-7
- Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M). Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

Doreen Holtsch completed a commercial apprenticeship before she obtained a diploma degree in Business Administration and Business Education at the Humboldt-University in Berlin in 2002. She holds a doctorate degree in Economics from the University of Rostock in 2007. She pursued a teaching diploma for commercial VET schools in 2009. She worked as a senior research assistant at the Georg-August-University Gottingen and has worked at the University of Zurich since 2011, where she investigates learning and teaching processes at commercial VET schools.

Johannes Hartig has studied psychology at the Goethe University Frankfurt and completed a doctorate degree in 2003 in Frankfurt. From 1998 to 2003 he worked as a research assistant in the Methods Division at the Institute of Psychology of Goethe University Frankfurt. From 2003 to 2008 he worked as a research assistant in the Unit Educational Quality and Evaluation of the DIPF. After a professorship for Methods of Empirical Educational Research at the University of Erfurt he has been Professor for Educational Measurement and Head of the Unit Educational Measurement in the Department Educational Quality and Evaluation in the DIPF since 2010.

Richard Shavelson earned his PhD in Educational Psychology in 1971 at the Stanford University. He held professorships at UCLA, the UCSB, and Stanford. He has been Professor at the Graduate School of Education since 1995 and between 1995 and 2000 he was the Dean of the Graduate School of Education at the Stanford University. The Margaret Jacks Professor of Education & I. James Quillen Dean (Emeritus) focuses his work, amongst others, on measurement, psychometrics, and cognitive psychology.