

**Research Bank**

Journal article

**Measuring Chinese teacher professional competence : Adapting and validating a German framework in China**

**Yang, Xinrong, Kaiser, Gabriele, König, Johannes and Blömeke, Sigrid**

This is an Accepted Manuscript version of the following article, accepted for publication in *Journal of Curriculum Studies*.

Yang, X., Kaiser, G., König, J. and Blömeke, S. (2018). Measuring Chinese teacher professional competence : Adapting and validating a German framework in China. *Journal of Curriculum Studies*, 50(5), pp. 638-653.  
<https://doi.org/10.1080/00220272.2018.1502810>.

It is deposited under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivatives License](#), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

## **Measuring Chinese Teacher Professional Competence: Adapting and Validating a German Framework in China**

Xinrong Yang (Southwest University), Gabriele Kaiser (University of Hamburg, Australian Catholic University), Johannes König (University of Cologne), Sigrid Blömeke (University of Oslo)

**Abstract:** The study presented in the paper has two main goals. The first is to review and compare teacher competence frameworks developed in Eastern (Chinese) and Western (German) contexts, exemplified for the domain of mathematics. Major similarities of the two contexts could be reconstructed in the conceptualization of teacher competence as a multidimensional construct comprising knowledge, teaching-related skills, and beliefs. Distinct differences could be identified as well, with the Chinese frameworks including a wider range of teacher-competence facets and emphasizing more teaching-related competencies than the Western (German) frameworks. The second purpose is to adapt and validate a German framework of the measurement of mathematics teacher competence in a Chinese context. This adaptation and validation uses exemplarily mathematics teacher, in detail follow-up-studies of the international Teacher Education and Development Study: Learning to Teach Mathematics (TEDS-M). With the integration of a qualitative approach (e.g. elemental validity) and a quantitative approach (e.g. construct validity) to validate the framework, the results of both approaches suggest a satisfactory validity for the adaptation. Overall, the results point out that the examined teacher competence framework and its instruments can be used for comparative analyses in Germany and China.

**Key words:** teacher competence; adaptation and validation; comparative study; mathematics

### **1. Introduction**

Over the past decades, international large-scale comparative studies of student achievement have consistently reported East Asian students outperforming their Western counterparts despite East Asian students’ accompanying negative affective outcomes such as low interest and confidence in learning (Leung, 2001, 2014). Such differences in mathematics learning outcomes have exerted a tremendous impact on the reforms of educational policy and systems around the world (Deng &

Gopinathanb, 2016). At the turn of this century, many Eastern countries and Western countries launched curriculum reforms aimed at improving the effectiveness of education. These reforms can be well observed in the domain of mathematics education, but are not restricted to it.

In these reformed curriculum standards, the description of the nature of the subject and the required methods for its learning and teaching have been fundamentally revised (Ni, Li, Li, & Zhang, 2011). However, to directly link a reformed curriculum with students’ learning is an arduous task since many factors influence and restrict the effectiveness of the implementation of curriculum (Krupa & Confrey, 2017). The successful implementation of reform approaches in classrooms requires especially teachers to play a substantially different role (Remillard, 2005). Teachers’ professional competence, such as relevant knowledge and skills required by the reformed curriculum, undoubtedly, is another critical factor that will influence an effective and successful implementation of education reform initiatives (Lloyd, 2008; Remillard, 2005). Therefore, a deep understanding of the theoretical construct of teachers’ professional competence – and possible opportunities to improve it – are of practical importance in teacher education worldwide (Guerriero, 2017; Schleicher, 2016; Townsend & Bates, 2007). Besides this practical relevance, modelling and comparing the structure of teacher competencies in various contexts are regarded as necessary to theoretically advance research in this field (Seidel & Stürmer, 2014). Meeting this goal requires the development of measurement frameworks and instruments that will allow researchers to examine teachers’ competencies in various contexts.

In the field of education, however, the concept of teacher competence is ‘difficult to capture’ (Johansson, Myrberg, & Rosén, 2014, p. 564), because essentially, teacher competence is situationally and societally determined (Blömeke & Kaiser, 2017). Nevertheless, if we are to reach a clearer understanding of its theoretical construct, we must work to find a common ground

of teacher competence ‘beyond different cultural traditions’ (Caena, 2014, p. 311). In addition, teacher competence has been described as a continuum and a dynamic process rather than a stable state of knowledge or skills (Blömeke, Gustafsson, & Shavelson, 2015). Such a characterization makes it more complicated to define and measure teacher competence. Even so, there has been an increasing interest in theoretical discussions on the construct of teacher competence. In recent years, several theoretically derived models of teachers’ professional competence have been developed from different perspectives accompanied by research endeavours to develop instruments to measure teachers’ professional competence (e.g. Guerriero, 2017; Kaiser, Blömeke, König, Busse, Döhrmann, & Hoth, 2017; Kunter, Klusmann, Baumert, Richter, Voss, & Hachfeld, 2013). However, so far, most frameworks and measurement instruments regarding teacher competence have been developed by Western researchers, especially those with a strong focus on German traditions. To our knowledge no such instrument has been developed in China – or anywhere in East Asia - to measure teacher competence.

Finding a common ground of teacher competence beyond diverse cultures and comparing teacher competence across cultural traditions are necessary if we are to reach a clearer theoretical understanding of the concept. An in-depth understanding of any similarities and differences in teacher competence between Western and Eastern cultures has the potential, for example, to explain the differences in mathematics learning outcomes that have been observed in international comparative studies (Leung, 2001, 2014). To meet this aim, the first step should be the development of instruments with high reliability and validity based on sound theoretical frameworks. In this paper, we contribute to this goal by analyzing the possibility of adapting a Western framework of teacher competence and its measurement instruments to an Eastern context.

## 2. Literature Review

### 2.1 Teacher-competence framework in the Western (German) context

Over the last two decades, there has been an increasing research interest in the professional competence of teachers in Western contexts. This wave of interest was largely fuelled by a criticism about the inefficiency of teacher education when it came to adequately developing teachers’ professionalism. This critique led to the implementation of the IEA’s *Teacher Education and Development Study: Learning to Teach Mathematics (TEDS-M)* in 2007 (Tatto, Schwille, Senk, Ingvarson, Peck, & Rowley, 2008), which has since become particularly important within the German discussion. Current approaches to investigate competence in the field of teacher education research have their theoretical underpinnings in the work of Weinert (2001). His view on professional competencies includes an individual’s knowledge, ability, and willingness to cope with demands in a specific situation successfully; thus, competence includes ‘all of the mental resources of individuals that are used to master tasks, acquire knowledge and achieve a good performance (Mulder, Weigel, & Collins, 2007, p. 69). Thus, teacher competence is described as a multidimensional construct consisting of cognitive facets (e.g. knowledge) and affective-motivational facets (e.g. beliefs) (Kaiser et al., 2017).

Defined in this way, teacher competencies are latent-domain-related dispositions and will be manifested within teaching performance (Klieme, Hartig, & Rauch, 2008). During the last decade, specifically in the field of mathematics education, but also in other domains such as science and language education, a cognitive approach on the professionalism of teachers focusing on their knowledge facets has dominated the frameworks developed in various Western contexts. For example, the Mathematical Knowledge for Teaching (MKT) framework of the *Mathematics Teaching and Learning-to-Teach Project (MTLT)* in the United States (Ball et al. 2008) or the

Knowledge for Teaching English framework (Freeman, Katz, Gomez, & Burns, 2015) mainly include knowledge facets. Similarly, TEDS-M, a large-scale international comparative study on teacher education, is based on the construct of competence as proposed by Weinert (2001) in that it includes pre-service teachers’ professional knowledge as well as their affective-motivational characteristics. Referring to the theoretical work by Shulman (1986), knowledge for teachers includes in TEDS-M the following sub-domains: 1) content knowledge (MCK) about the main mathematics areas relevant for teachers, 2) pedagogical content knowledge (PCK), including curricular knowledge, knowledge of planning for mathematics teaching and learning of enacting mathematics, and 3) general pedagogical knowledge (GPK) covering knowledge about teaching and learning and about the evaluation of students’ achievements (Döhrmann, Kaiser, & Blömeke, 2012; Tatto et al., 2008). In the affective-motivational facet, based on the work of Richardson (1996), epistemological beliefs about the nature of the subject and beliefs about the learning and teaching of the subject are distinguished. In addition, affective dispositions (e.g. motivation) and metacognitive abilities (e.g. self-regulation) were also included. In follow-up-studies TEDS-M was adapted to German and English teaching with the same theoretical framework using similarly structured instruments, which took into account the different nature of language education, such as different distinctions of the sub-domains, different linkages between the subject and its teaching methods (e.g. Blömeke et al., 2011).

Similarly, in the *Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students’ Mathematical Literacy Project (COACTIV)* mathematics teachers’ professional competence was also defined from a cognitive approach referring to Shulman (1986), which is composed of professional knowledge as well as beliefs, motivational orientation, and self-regulation (Kunter et al., 2013). This work was enriched and continued by the inclusion of other

subjects, namely German, English, Latin, physics and religious education. Despite the great variety of the subjects the same theoretical framework and similar test instruments and testing procedures could be used, which points to a strong commonalities of professional competencies across the different school subjects (Krauss et al., 2017).

A common fundamental characteristic shared by these teacher competence frameworks is that teacher competence is conceptualized as cognitive and motivational resources, which underlie teachers’ classroom performance. However, in practice, such resources will not be simply and directly transformed into performance but instead be mediated by other cognitive skills that are more closely related to teachers’ performance in complex classroom situations (Blömeke et al. 2015). Therefore, situational and societal characteristics of teacher competence should be taken into account concerning the development of teacher competence frameworks (Blömeke & Kaiser 2017).

Indeed, the present professional competence research has already shifted to use situated professionalism as the main approach; that is, competence only gets meaning in a certain context (Mulder, 2014). Researchers have argued that traditional conceptual dichotomies were misleading as they ignored either the stable dispositional or the more variable situational competence facets. They suggest that instead of thinking of competence as a dichotomous characteristic, it should be regarded as a continuum with many steps in between the dispositional facets (i.e. cognition and affect) versus situated performance in the classroom (Blömeke et al., 2015).

An example of such a study aiming to gain a more complete understanding of teacher competence, based on the integration of cognitive and situated approaches for the investigation of teachers’ professional competence was carried out as German follow-up study to TEDS-M (TEDS-FU). The theoretical framework of the original study was extended to include teachers’

professional noticing as situation-specific skills. In order to capture professional noticing three sub-facets were distinguished, (a) perception, i.e. perceiving particular events in an instructional setting, (b) interpretation, i.e. interpreting the perceived activities in the instructional setting, and (c) decision-making: either as anticipating responses to students’ activities or as proposing alternative instructional strategies (so-called PID model). It is generally believed that ‘processes such as the perception and interpretation of a specific job situation together with decision-making may mediate between disposition and performance’ (Blömeke et al., 2015, p.7). As such the PID model comprises situated and perceptually oriented facets of teacher competence to enrich the cognitive perspective of the teacher competence framework developed amongst others in TEDS-M (Kaiser et al., 2017).

## **2.2. Teacher-competence framework in the Chinese context**

It has been widely accepted by Chinese researchers that teacher competence or teacher capacity is a major factor in the educational process that will directly influence or even determine teaching effectiveness for many years (e.g. Ning, 2010; Ye, 1998). However, only a few studies investigated the construct of teacher competence in China before the year of 2000. Moreover, as pointed out by Wei et al. (2017), those few studies mainly focused on the theoretically assumed construct of teacher competence from the perspective of general pedagogy or psychology without considering the characteristics of a specific schooling subject (similarly Wang et al., 2012). For example, Ye (1998) theoretically argued that the competence needed for teachers to cope with future challenges has to include adequate educational beliefs, a complete professional knowledge structure, and practical abilities. However, such theoretically driven proposals provide scant information for the development of empirical instruments with which to examine teacher competence in practice, especially for subject teachers (Wang et al., 2012).



After the most recent curriculum reform in China in 2000, there has been an increasing interest in discussing teacher competence, including how it relates to the new requirements of the reformed curriculum and the development of the construct and its features at both the academic and policy levels (Ning, 2010). For example, in the National Professional Standards for Teachers issued by the Ministry of Education China in 2012, professional requirements for teachers at different school levels were clearly described (see Table 1 for a brief description).

**Table 1**

*Overview on the Professional Standards for Primary and Secondary School Teachers*

Dimensions	Main Aspects
Professional ethics and morality	The understanding of teaching as a profession; Attitudes and behaviour to students; and Attitudes and behaviour toward education and teaching.
Professional knowledge	Pedagogical knowledge; Subject matter knowledge; Subject matter pedagogical content knowledge; and Knowledge of educational context.
Professional ability	Teaching design; Implementation of teaching; Classroom management and education; Assessment and evaluation; and Reflection and self-development.

As shown in Table 1, qualified teachers in China should not only possess professional knowledge and abilities related to classroom teaching, they also must have further skills, such as the ability to self-evaluate their own teaching. Similar requirements have been discussed by researchers in the academic field. For example, Du’ s (2011) theoretically based competence framework for primary and secondary school teachers includes the following seven core abilities: 1) to design teaching; 2) to develop and use curriculum resources; 3) to express themselves clearly and demonstrate topics clearly, such as through lucid demonstration with oral language, body

language and text; 4) to interact, such as through questioning; 5) to manage teaching and learning processes; 6) to assess student learning; and 7) to conduct educational research.

Most recently, Wei et al. (2017) developed a teaching-competence model based on interviews with expert teachers in China in which respondents described the abilities most important for teachers. The resulting model contains the following six aspects: 1) subject literacy, such as subject-related knowledge, skills, methods, and thinking; 2) basic teaching skills, such as using language, body language, and text to express ideas clearly to facilitate students’ understanding, along with ICT usage skills; 3) general pedagogical knowledge and subject pedagogical content knowledge; 4) ability to design, implement, and evaluate teaching; 5) regard for teaching ethics, self-efficacy and other aspects; 6) ability to reflect on teaching and conduct educational research. Similar aspects were identified in Wang and Hu’ s (2016) empirical investigation conducted in China.

Concerning the professional competence of mathematics teachers, several frameworks exist that meet the requirements of mathematics teaching after the latest mathematics curriculum reform. For example, Fu and Liu (2005) discussed the theoretical construct of mathematics teachers’ competence, highlighting the four main concepts shown in Table 2.

**Table 2**

*Construct of (Mathematics) Teacher Ability*

First-level ability	Second-level ability
Basic ability	Language expression, interaction, ICT use, and life-long learning.
Mathematical ability	Spatial imagination, abstract generalization, reasoning and proving, operating, and problem solving.
Mathematics teaching ability	Designing, implementing, monitoring teaching and reflecting on it.

Similar ideas could be found in most recent frameworks developed by mathematics-education researchers as well (e.g. Zhang & Xu, 2016). The reviewed teacher competence or teaching ability frameworks developed in China cover a range of different aspects of teacher competence. Overall, common aspects could be identified within the frameworks, namely: 1) knowledge; 2) teaching-related skills such as designing and implementing teaching, assessing students’ learning, and evaluating and reflecting on teaching; 3) conducting educational research; and 4) professional ethics and moral values. Given the repeated appearance of these aspects across studies, it seems that we now have a relatively clear sense of the content of professional competence as it relates to (mathematics) teachers. However, until now, there have been no instruments available that could be used to measure those aspects of teacher competence in China.

### **2.3 Similarities and differences of teacher competence between the two contexts**

Important similarities among the theoretical constructs of teacher competence can be identified even though social and cultural differences exist between Eastern and Western contexts. The first similarity is that teacher competence is conceptualized as a multidimensional construct in both settings. In the important teacher-competence frameworks in China, a variety of aspects such as knowledge, teaching ability, and the ability to conduct educational research were included; the frameworks developed in Western contexts also highlighted multiple aspects such as knowledge and beliefs. The second main similarity is that teacher knowledge is the core component of teacher competence. Although sometimes different terminologies are used to describe various facets of knowledge, and different classifications are employed in the two contexts, the essential meanings are more or less the same. For example, in China, the ability to design and implement teaching is highlighted in every framework of teacher competence; similar aspects such as knowledge of

planning for domain-specific teaching and learning, and knowledge of enacting the content of the subject for teaching and learning, are included in the German TEDS-M and TEDS-FU frameworks.

Moreover, items similar to the situational aspects of teacher competence highlighted in the TEDS-FU framework can also be identified in the Chinese frameworks. For example, Chinese teachers are required to demonstrate the ability to implement and monitor teaching flexibly according to students’ understanding of the teaching content (e.g. Du, 2011; Fu & Liu, 2005). This is similar to teachers’ professional noticing from TEDS-FU, which encompasses teachers’ perception of students’ difficulties and problems, their fast mental interpretation of those issues, and their ability to adjust their teaching plans accordingly (Kaiser et al., 2017).

However, despite these similarities, distinct differences can also be identified between these two cultures. The biggest difference is that, compared with their Western counterparts, frameworks in the Chinese context cover a wider range of teacher competencies, including the ability to conduct educational research and evaluate teaching (Yang, 2014). Secondly, although some common aspects are described in the frameworks in the two contexts, the degree to which they are emphasized is different. For example, the aspect of knowledge is not as heavily emphasised in Chinese frameworks as it is in German frameworks. In contrast, the ability to design, implement, and monitor teaching is more accentuated in Chinese frameworks than in German ones.

In addition, unlike the systematic research in the field of teacher competence in a Western (German) context, the Eastern (Chinese) research on teacher competence is less theory-driven and not very systematically conducted. Consequently, unlike the availability of evaluation instruments in the field of teacher competence in the Western (German) context, especially in the field of mathematics education, so far, no instruments have been developed specifically for application in a Chinese context. In view of this, the second aim of this paper is to analyze the possibility of adapting evaluation instruments developed in a Western context to examine in-service

mathematics teacher competence, namely those of TEDS-FU from Germany, by testing their validity in an Eastern context (China). As the comparison results above have shown, almost all the aspects contained in the teacher-competence framework developed in Western studies were also reflected in the Eastern (Chinese) context. Therefore, it can be assumed that the instruments developed within Western contexts can be adapted to the Chinese context to close the gap of the missing assessment instruments for mathematics teacher competence.

### **3. Methods**

#### **3.1 The development of teacher-competence assessment instruments in Germany**

##### ***3.1.1 The development of assessment instruments of teachers' professional knowledge***

The tests used in TEDS-FU to assess MCK, MPCK, and GPK were originally developed in the TEDS-M study. To avoid cultural bias, the TEDS-M items were developed by expert groups from all participating countries, and the item pool was reviewed by international experts and then reviewed again within each participating country. All the items using multiple-choice, open-response, and constructed-response format were translated and re-translated in all necessary languages, the translated items were tested for adequacy, correctness, and clarity of wording by experts in mathematics and mathematics pedagogy from each participating country. To ensure high psychometric quality, measures such as construct validity, internal consistency, score reliability, and measurement invariance were taken into consideration (Tatto et al., 2008).

In the German follow-up studies of TEDS-M, TEDS-FU, and TEDS-Instruct, shortened versions of the MCK and GPK tests were used to reduce the test-burden for the participants without loss of reliability. For MCK, the abbreviated version only contains 25 items, which cover the three included sub-domains of mathematics: number, algebra, and geometry. For MPCK, a slightly

extended version was used, comprising items that cover knowledge related to curriculum and planning as well as knowledge about enacting the teaching of mathematics. For GPK, the shortened version included 39 items, which were fairly equally distributed across aspects such as lesson planning, motivation, classroom management, and assessment.

### ***3.1.2 The development of instruments of teacher noticing***

In order to examine teachers’ professional noticing, three video-vignettes were developed in TEDS-FU, which cover a range of central mathematical topics usually taught in German schools in grades 8 to 10. The three video-vignettes do not consist of episodes taken from real classroom teaching; instead, they are scripted plots meant to represent different types of schools in Germany. In addition, each of the video-vignettes focused on different phases of teaching – such as the introduction of a mathematical task, or work on a task followed by a whole-class discussion of the results – to investigate teachers’ professional noticing comprehensively. Each video-vignette lasts around four minutes and provides an overview about the whole lesson.

After observing each of the three video-vignettes, the participating mathematics teachers were asked to answer closed and constructed-response items referring to the observed classroom scenes. The items distinguished between P\_PID (Perception, Interpretation and Decision-making from a pedagogical perspective) and M\_PID (Perception, Interpretation and Decision-making from a mathematics pedagogical perspective). The first kind of items mainly focused on teachers’ pedagogy-related noticing, such as lesson structure, classroom management and assessment, and the second kind focused on teachers’ mathematics instruction-related noticing, such as different ways of explaining a mathematical concept and identifying task types and mathematical ideas. Generally, the items required teachers to notice mathematics classroom teaching holistically. In total, there were 38 closed items (22 P\_PID and 16 M\_PID) using Likert-scales and 36 constructed-

response items (18 P\_PID and 18 M\_PID). An expert rating meeting was implemented to generate the rating scales to the closed items. The coding manual was developed through a systematic analysis of relevant theories and literature and extensive discussion among experienced mathematics teachers and mathematics education researchers. It was extensively piloted before it was used in TEDS-FU to improve its validity. Various approaches, like curricular analyses of the mathematical content and comprehensive expert workshops, were employed to ensure the content validity of the instruments (Hoth et al. 2016).

### **3.2 Translation and cross-cultural adaptation of the measurement instruments.**

The translation and cross-cultural adaptation of the instrument developed in Germany to the Chinese context was carried out in accordance with the International Test Commission (2017) guidelines. Normally, multistep processes such as forward translation, synthesis of the translations, back translation, expert reviews, and pre-testing are utilized to adapt an instrument to another cultural context (Beaton et al., 2000; International Test Commission, 2017). However, as the knowledge part of the instrument consisted of the released parts of the tests on GPK, MCK, and MPCK already used in Taiwan within the TEDS-M-study – they were directly implemented to test teachers’ cognitive competence in Chongqing, where the present study was conducted. Although subtle cultural differences between Mainland China and Taiwan exist, the deep-rooted cultural values on the teaching and learning of mathematics, shaped by the Confucian heritage culture (CHC), are very similar across most East Asian regions, including Taiwan and Chongqing, even up to the present time (Leung, 2001). Therefore, only one modification was made to the knowledge part of the instrument: the traditional Chinese writing used in Taiwan was converted to simplified Chinese to meet the mathematics education standards in Chongqing. In detail, the knowledge part of the instrument used in China included 22 dichotomously coded items used to assess MCK

(open-response answers were required for eight items, and the other 14 items were multiple-choice), 24 dichotomously coded items used to assess MPC (including 4 open-response items and 20 multiple-choice items), and 22 dichotomously coded items used to assess GPK.

Overall, the translation and cross-cultural adaptation mainly occurred within the professional noticing part of the instrument, which aimed to investigate Chinese mathematics teachers’ situational competence. In the first step, a forward translation from German to English was performed by a native German-speaking mathematics-education postgraduate student with a minor degree in English language. The accuracy of the translation was then checked by the translator, the second author, and the original main developer of the instrument, who made necessary corrections to ensure the original intentions (Banville, Desrosiers, & Genet-Volet, 2000). The first author reviewed relevant Chinese literature in order to find Chinese terminologies that were as closely related to the meaning of the English (or German) terms in order to secure the semantic, idiomatic, experiential, and, particularly, conceptual equivalence between the original German item and its Chinese translation (Beaton et al., 2000; Hambleton, 2000). The first author translated the instruments from English to Chinese, and the Chinese translation was checked and corrected by a Chinese professor from mathematics education with many years of experience in international studies and research. To ensure the content validity, the Chinese version of the instruments was double-checked by a native Chinese speaking PhD student majoring in the German language. Slight modifications were made to the video-script and some expressions used within the instrument. After this, two secondary school mathematics teachers in China were invited to answer the whole instrument and to be interviewed about how they understood every item. Modifications were further made according to their comments and suggestions.

### **3.3 Approaches for validation**



The adaptation of an instrument to a new context should be followed ideally by the test of psychometric properties, i.e. evidence for the validity of the instrument should be provided (Beaton et al., 2000; International Test Commission, 2017). The concept of validity, however, has changed over the years as the field of educational and psychological measurement has developed, and so far, there is no general agreement in regard to what validity is and how it should be measured (e.g. Kane, 2001; Lissitz & Samuelson, 2007; Newton & Shaw, 2016). One common view within contemporary discussions of validity is that in addition to the traditional sources of validity, such as content validity and internal structure validity, evidence based on response processes and consequences are accepted as important sources of validity evidence that should be included in validation. Support for this view appears in the most recent edition of the Standards for Educational and Psychological Testing (AERA, APA, and NCME, 2014).

To validate the instruments in this study, qualitative methods were also employed. More specifically, we examined the ‘elemental validity’ of the instrument for the professional noticing part of the instrument. ‘Elemental validity’ was proposed by Schilling and Hill (2007) to validate instruments of teacher knowledge based on Kane’s (2001, 2006) ‘argument-based approach’ to validity. The elemental assumption relates to individual items in a measurement instrument and how well the items reflect teachers’ mathematical knowledge independent of the influence of less relevant factors on some teachers’ responses, such as the ability to master test-taking strategies; furthermore, for inference purposes, teachers’ reasoning for a particular item should be consistent with the answer they gave (Schilling & Hill, 2007). To analyze elemental validity, cognitive interviews are necessary to examine, for each item, whether teachers’ thinking is consistent with the answer they provided and whether the correct answer was accompanied by

correct thinking or argumentation (Hill, Dean, & Goffney, 2007). Furthermore, cognitive interviews can provide information on whether a teacher drew on the type of knowledge that items intended to tap (Hill et al., 2007). This approach has been employed by previous research to validate instruments of teacher knowledge developed in the United States in contexts like Ireland (Delaney, 2012) or Ghana (Cole, 2012). Similarly, in the present study, elemental validity of the instrument for the video-based part of the measurement instrument was examined to make sure that for each item, the same or similar knowledge and skills of teachers was examined in both relevant cultural contexts.

In addition, content validity and construct validity for the whole instrument were examined in the Chinese context. Content validity mainly studies whether the content of a measurement instrument covers the content it is supposed to measure. In the present study, content validity also aimed to discover whether the content and skills being examined in the German context also match the Chinese contexts. Content validity consists of content-related argumentations and is usually evaluated through expert opinions. Construct validity is defined as the extent to which the relations among items, domains, and concepts support a priori hypotheses about the logical relations that should exist with other measures. Construct validity is commonly accepted as the central component in validation work.

### **3.4 Participants and data collection**

As described above, three approaches (content validity, elemental validity, and construct validity) were used successively to validate the German instruments in the Chinese context. Participants and data collection methods for each approach were as follows:

Concerning content validity, two mathematics education researchers and six junior secondary school mathematics teachers in China discussed the Chinese versions of all test instruments.

For elemental validity, we carried out a study with eight mathematics teachers from two junior secondary schools with two to 29 years of teaching experience. The teachers answered all items of the professional noticing part at the same time in a meeting room of their school. Afterwards, a cognitive interview was conducted with each teacher, in which teachers were asked to explain their answers – especially the reasoning or thinking processes they used to reach their answers. For the unanswered items, they were asked to explain the reasons why they were unable to answer. The interviews lasted from 55 to 76 minutes. All the interviews were audiotaped and transcribed by the first author for analysis.

For the examination of construct validity, we carried out a study consisting of 203 junior secondary school mathematics teachers in Chongqing from different academic backgrounds, school locations (rural and urban), and with varying teaching experiences. The teachers were chosen from 19 junior secondary schools in Chongqing, the biggest metropolitan city in Western China, 64 (31.5%) of them were female. During the time of the assessment 69 (34.0%) taught in rural schools. Their teaching experience ranged from two to 36 years, with an average of 14.6 years. The teachers were assigned randomly to one of the five lecture rooms, where they completed a three-hour timed paper-and-pencil test that included the display of the three video-vignettes.

## **4. Results**

### **4.1 Content validity results**

Like the approaches employed in TEDS-FU to ensure content validity, the expert team in China was required to examine all the test items from various perspectives, such as performing curriculum analyses of the mathematics content and taking into consideration the reality and tradition of Chinese mathematics teaching and learning. For the knowledge part, the expert team only gave a few suggestions, requesting the modification of some phrasing to make them closer to the common use in Mainland China. For the video part of the instrument, the major suggestions made by the expert team were to delete items that were closely related to the German mathematics curriculum but were not included in Chinese curriculum, i.e. items on the heterogeneity or multicultural backgrounds of students. In addition, suggestions were made to make a few expressions more closely connected to the common language used by Chinese teachers.

#### **4.2 Elemental validity results**

In order to analyze the consistence between the eight teachers’ answers and their thinking, and to determine whether they had used the knowledge and skills the items aimed to examine, the 37 closed items and 27 constructed-response items were examined separately. For the closed items, their answers were checked according to the coding manual developed and used in TEDS-FU and TEDS-Instruct. For the correct answers, their explanations were also checked to examine whether or not the teachers actually attended to the events being tested. The reasons for the wrong answers and non-answered items were also checked.

For the closed items, answers by eight teachers on 37 items provide 296 data points for analysis. Results of this part show that 238 (80.40%) data points were correct, 33 (11.15%) data points were wrong, and 25 (8.45%) data points were unanswered. For the correct answers, all the examined teachers were able to provide reasonable explanations instead of using test-taking strategies or guessing. That is, they either could provide the evidence that they noticed the

respective events when they watched the three video-vignettes, or they knew how to use relevant knowledge to justify their given answers. Regarding the wrong answers, seven data points related mathematical topics and the teachers clearly realized that they made a mistake when they answered the items. For the other wrong and unanswered items, mainly relating to classroom management, the teachers had not attended to the events or forgotten.

For the constructed-response items, the teachers’ answers were first coded according to the coding manual developed in TEDS-FU and TEDS-Instruct. For the correct answers, their explanations were further analyzed according to the descriptions given in the manual; if their explanations matched one of the descriptions, that answer and the underlying reasoning and thinking were counted as consistent with the type of knowledge or skills the item supposed to test. Otherwise, the answer was counted as inconsistent. In addition, the wrong answers and unanswered items were also analyzed with reference to the reasons why the teachers gave a wrong answer or did not answer the item. If their explanations were wrong or they clearly pointed out that they had not known how to solve such items, the answer was also counted as consistent between the answer and the reasoning.

With eight teachers and 27 items, a total of 216 data points were analyzed. Results of this part show that 127 (58.80%) data points were correct, 42 (19.44%) data points were wrong, and 47 (21.56%) data points were unanswered. For the 127 correct answer points, all eight teachers’ explanations were found to match or be similar to at least one aspect of the descriptions in the coding manual. This suggests that the examined teachers reasoned correctly or applied the knowledge and skills that the items intended to examine. For the unanswered items, there were 24 data points that the teachers clearly expressed that they had not known how to fully answer. However, for the other 23 unanswered points, the teachers either said that they had not had the

time for answering or had not fully understood the question. When teachers mentioned that they had not understood a question, information like how they understood the item or possible suggestions to modify the questions was collected and was further used to modify the expressions of these items.

19 of the 42 wrongly answered data points contained wrong explanations by the teachers. Another twelve referred to teachers’ lacking the knowledge to answer the item, even with guessed answers. However, all of the remaining eleven wrong answers related to the same task, which required the teachers to recall a short interaction between the teacher and a female student shown in the video. In the sequence, a female student asks the teacher during a group-work phase whether he has already marked the tests. He answers that she has been successful, although the test was not easy. The girl says that she only had been lucky. The underlying theories of gender differences in the self-attribution of success has been discussed for decades in Western mathematics education and has served as partial explanation for gender differences especially in mathematics (Wolfe et al.,1980). Therefore, this item worked in TEDS-FU and TEDS-Instruct satisfactorily when posed to German teachers, but the Chinese teachers reacted to this item very differently. Six of the eight teachers evaluated the student’ s behaviour as unacceptable as her question was not related to the teaching content. There are several possible explanations for this strong difference between the Chinese and the German teachers: amongst others, the strong focus of Chinese teachers on content-related issues in teaching in contrast to societal aspects like gender issues. This item reveals how important it is to carefully examine the cultural influence on how test items may be interpreted.

### **4.3 Construct validity**

For the data collected from the 203 teachers for the study of construct validity, the open response items were coded according to the coding rubrics developed in the studies TEDS-M, TEDS-FU, and TEDS-Instruct. Independent raters coded 25 percent of the questionnaire; good values of Cohen’s Kappa were reached ( $k > 0.76$  and  $K_{\text{average}} = 0.84$ ). For all the open response items, items with no response or an incorrect response were rated with 0, and each correct answer was rated with 1 (for items with several sub-items, the sum of the correct answers was used). The relative item difficulties for a one-parameter (Rasch model) item response theory (IRT) model were calculated separately on the five dimensions of GPK, MCK, MPCK, P\_PID, and M\_PID. Items with extreme difficulty were removed for the final analysis in order to avoid distortions. The internal consistency of the remaining items in the sub-dimensions of the five dimensions was estimated using Cronbach’s alpha reliability coefficient. The reliability scores for each aspect ranged from 0.56 to 0.88, indicating acceptable or good reliability for a statistical analysis.

The second step comprised the evaluation of the psychometrical properties of the five assessment domains. Confirmatory factor analyses were carried out using *Mplus* to validate the structure of GPK, MCK, MPCK, P\_PID, and M\_PID separately. Due to the relatively small number of participants in the present study, parcels of items were used, and the resulting item mean scores of each of the parcels were used as manifest indicators to produce more stable factor solutions (Little et al., 2002). Items with the same task stem within the knowledge test and items testing a similar topic (e.g. classroom management, teachers’ behaviour, and students’ errors) in the video-based part of the test were assigned to one parcel to account for potential task-specific variance. The model fit of the five models is reported in Table 3. The following fit indices were used to evaluate the fit of each of the five models: the ratio of chi-square values to the degrees of freedom ( $\chi^2/df$ ), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and the comparative fit index (CFI). The following values for the criteria

were used: For the  $\chi^2/df$ , values close to 5 were seen as adequate fit (Schumacker, & Lomax, 2004), while values less than or equal to 3 were judged as reflecting a good fit (Ullman, 2001); for RMSEA and SRMR, values less than 0.05 indicate a very good fit and values less than 0.08 indicate a reasonable fit (Browne & Cudeck, 1993; Hu & Bentler, 1999); for CFI, values higher than .95 represent a good fit of the model to the data, and values above .90 show an adequate fit (Hu & Bentler, 1999).

According to these criteria, as shown in Table 3, confirmatory factor analyses revealed that the five models are a good or an acceptable fit to the data, which confirm the structure of the five constructs GPK, MCK, MPCK, P\_PID, and M\_PID as separate domains for Chinese teachers’ professional competencies. Therefore, it is reasonable to say that the measurement instruments of teachers’ knowledge and teachers’ noticing have acceptable validity in the context of Chinese mathematics education.

**Table 3**

*Model Fit Indices for the Five Assessed Domains*

Assessment	$\chi^2/df$	p value	CFI	RMSEA	SRMR
GPK	4.81	p<0.001	0.944	0.038	0.061
MCK	2.63	p<0.001	0.973	0.016	0.060
MPCK	4.78	p<0.001	0.984	0.020	0.046
P_PID	4.31	p<0.001	0.989	0.014	0.033
M_PID	2.74	p<0.001	0.961	0.023	0.047

## 5. Discussion and conclusions

Summarizing, the present addressed two aims: first, to compare (mathematics) teacher-competence frameworks developed in Eastern (Chinese) and Western (German) contexts to examine possible similarities and differences between these two contexts; second, to adapt a mathematics teacher competence measurement framework developed in Germany to a Chinese



context and validate it. The adaption and validation of this measurement framework builds a meaningful foundation for cross-country comparative studies in the field of teacher competence given that very little cross-country evidence is available about the similarities and differences of teacher competence (Blömeke, 2017).

As reviewed and discussed above, even though the educational, societal, and cultural contexts between Eastern and Western countries are quite different, important similarities between the theoretical conceptualization of teacher competence and between the frameworks of teacher competence can be identified. The main similarity between the frameworks of (mathematics) teacher competence between the two contexts is that fundamentally, teacher competence is conceptualized as a multidimensional construct comprising various types of knowledge, teaching-related skills, beliefs, and moral and ethical aspects, with knowledge and teaching-related skills as the core competencies in both contexts. Although different terminologies were used to describe various facets of teacher competence, and teacher competence was classified differently in different frameworks, the essential meaning is fundamentally the same. For instance, the ability to design and implement teaching was emphasized in all the frameworks of teacher competence in China (as protagonist for Eastern countries), such aspects were also included as necessary component in the frameworks from Germany (as protagonist for Western countries). In addition, situationally oriented frameworks of teacher competence have been proposed to enrich the traditionally more cognitively oriented frameworks of teacher competence in the field of teacher education research (Kaiser et al. 2015). Although such aspects were not clearly specified in the frameworks developed in the Chinese context, similar aspects could be identified either implicitly or explicitly. For example, Chinese teachers were required to demonstrate the ability to implement teaching and monitor their teaching flexibly according to their students’ understanding of the content being relayed (e.g. Du 2011; Fu & Liu, 2005). Therefore, such similarities further suggest

that the integration of cognitive and situational approaches to investigate teacher competence, which is now a prominent approach in the field of teacher competence (e.g. Kaiser et al., 2015, 2017), also describes the core of the discussion on teacher competence research in China. Overall, the comparative findings provide a solid theoretical foundation for the second purpose for the study – that is, the adaptation and validation of a teacher-competence model developed with strong German tradition to a Chinese context.

Distinct differences were identified between the frameworks developed in the two contexts. The first main difference is that the Chinese researchers employed a broader view to define teacher competence in their frameworks, including items – such as the ability to research and evaluate teaching – that are not mentioned in the Western (German) frameworks. In addition, for some aspects of teacher competence, the extent to which they are emphasized in the two contexts is different. As discussed above, the Chinese frameworks stressed more heavily the aspect of teaching-related competence, whereas the frameworks developed in Western (German) contexts put more emphasis on teachers’ professional knowledge. Such differences on the one hand suggest that teacher competence is indeed a societal and cultural construct (Blömeke & Kaiser, 2017). On the other hand, such differences also make measurement and, especially, cross-country comparative study a challenging endeavour. However, the existence of societal and cultural differences may further influence the development and growth of teacher competence in different contexts. Teachers from different contexts may possess different weaknesses and strengths in terms of teacher competence. Therefore, in the field of teacher competence research, it is necessary and meaningful to conduct cross-country comparative studies aimed at reforming teacher education and improving students’ mathematics learning (Blömeke, 2017).

The second purpose of the present study was to adapt and validate a German framework of the measurement of mathematics teacher competence to China. As described above, the German framework integrates a cognitive and a situational approach of teacher competence, i.e. from the cognitive perspective, knowledge central to mathematics teachers’ work (e.g. GPK, MPCK, and MCK), and professional noticing as the situational aspect. As reported above, the released items testing GPK, MPCK, and MCK developed in TEDS-M had been validated already in East Asia within the TEDS-M-study. The approaches of content validity and construct validity were employed to examine validity of the instrument in the new Chinese context. Findings of the study suggest that for content validity, according to the experts we assembled, all the items match the teaching and learning reality in Chongqing well. In addition, the analyses confirm the structure of the three constructs GPK, MCK, MPCK as distinguished in TEDS-M. Therefore, it is reasonable to conclude that the three knowledge measurement instruments could be adapted to test Chinese mathematics teachers’ cognitive part of professional competence.

For the situational part of the professional competence, as described above, professional noticing was included in the framework of teacher competence developed in TEDS-FU. As this part of the instrument has not been used in an East Asian context like China, an integration of quantitative and qualitative methods was used to examine the validity of this instrument in the context of Chongqing. The integration of qualitative approaches (e.g. elemental validity) and quantitative approaches (e.g. analysis of psychometric properties) has been described by previous researchers as a high-quality adaptation process while extending the use of a measurement to another country (Fauskanger et al., 2012). Concerning the content validity in the present study, all the expert members commented that the majority of the items developed in the German context

could be adapted in Chongqing apart of a few items strongly related to the German mathematics educational discussion.

To examine the consistency between teachers’ answers and their thinking, and especially to examine whether or not the Chinese teachers referred to the same knowledge or similar skills as their German counterparts did while answering the items, elemental validity as proposed by Hill et al. (2007) was examined. Elemental validity results suggest that for the majority of the items, the Chinese teachers’ answers and their reasoning and thinking was consistent with expectations and prior results. Similar to their German counterparts, for most of the items the Chinese teachers also used the expected knowledge and similar skills to answer the items even though for quite a few items, societal and cultural bias were identified. The elemental validity results suggest that although the situational part of the instrument was originally developed in Germany, it is possible to adapt it to another, quite different, context as the topics used in the three video-vignettes and the skills that were supposed to be examined are in common. However, societal and cultural differences between the two contexts should not be ignored, which led to the exclusion of a few items strongly related to German mathematics education. The further quantitative results also confirmed that the construct of P\_PID and M\_PID, which are the two main components of the situation-specific aspects of teacher competence, can be separated and distinguished. Therefore, it is also reasonable to say that the two measurement parts focusing on P\_PID and M\_PID could be adapted from the German context to investigate situational facets of Chinese mathematics teachers’ professional competence.

Validating measures of teachers’ competence is a challenging work – especially validating a framework in a quite different context (Delaney, 2012). The present study intended to integrate a qualitative approach and a quantitative approach to validate a framework from the

German context within the Chinese context. The results from both approaches suggest a satisfactory validity for the adaptation of the framework from the German to the Chinese context. To summarise, the present study has the potential to serve as a meaningful foundation for cross-country comparisons of teachers’ professional competence.

However, the study contains shortcomings, which have to be pointed out. First, all the data collected for the validity studies were collected in one district in Chongqing. Data collected from other regions in China may be needed to enhance the examination of the validity of the framework for China as a whole. In addition, only a small number of teachers were involved in the qualitative and the quantitative approaches. Therefore, for elemental validity, it might be possible that new inconsistencies may be identified, if more teachers would be involved. For construct validity, due to the constraint of the small number of participants, parcel scores were developed, which might influence the internal structure of the five constructs. Moreover, for the cognitive aspect of validity, most of the items were released items of TEDS-M that might also influence the internal structure of the measurement of the domains GPK, MPCK, and MCK. However, as validation is an ongoing process, follow-up studies are planned that will enrich the cognitive part of the instrument. Furthermore, we are aiming to involve a larger number of teachers from various places in China in our next study, which has the potential of further improving the validity of the framework in Germany.

## References

- AERA, APA, NCME (2014). *Standards for Educational and Psychological Testing*. Thousand Oaks: Sage.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59, 389 – 407.
- Banville, D., Desrosiers, P., & Genetvolet, Y. (2000). Translating questionnaires and inventories using a cross-cultural translation technique. *Journal of Teaching in Physical Education*, 19(3), 374-387.

- Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186-3191.
- Blömeke, S. (2017). Modelling teachers' professional competence as a multi-dimensional construct. In S. Guerriero (Ed.), *Pedagogical Knowledge and the Changing Nature of the Teaching Profession* (pp. 119-135). Paris: OECD Publishing.
- Blömeke, S., Bremerich-Vos, A., Haudeck, H., Kaiser, G., Nold, G., Schwippert, K., & Willenberg, H. (2011) (Eds.). *Kompetenzen von Lehramtsstudierenden in gering strukturierten Domänen. Erste Ergebnisse aus TEDS-LT*. Münster: Waxmann.
- Blömeke, S., Gustafsson, J.-E. & Shavelson, R. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie*, 223, 3-13.
- Blömeke, S. & Kaiser, G. (2017). Understanding the development of teachers' professional competencies as personally, situationally, and societally determined. In D. J. Clandinin, & J. Husu (Eds.), *International handbook of research on teacher education* (pp. 783 - 802). Thousand Oakes, CA: Sage.
- Bornstein, R. F. (2011). Toward a process-focused model of test score validity: Improving psychological assessment in science and practice. *Psychological Assessment*, 23(2), 532-544.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061 - 1071.
- Bottcher, W. (2013). Curriculum, Evaluation, and Control in Germany. In W.F. Pinar (Ed.), *International Handbook of Curriculum Research* (2nd ed.) (pp. 201-207). New York: Taylor & Francis.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (eds.), *Testing structural equation models* (pp. 136 - 162). Newbury Park, CA: Sage.
- Caena, F. (2014). Teacher Competence Frameworks in Europe: policy - as - discourse and policy - as - practice. *European Journal of Education*, 49(3), 311-331.
- Cole, Y. (2012). Assessing elemental validity: the transfer and use of mathematical knowledge for teaching measures in Ghana. *ZDM Mathematics Education*, 44(3), 415-426.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.
- Delaney, S. (2012). A validation study of the use of mathematical knowledge for teaching measures in Ireland. *ZDM Mathematics Education*, 44(3), 427-441.
- Deng, Z., & Gopinathan, S. (2016). PISA and high-performing education systems: Explaining Singapore' s education success. *Comparative Education*, 52(4), 449-472.
- Döhrmann, M., Kaiser, G., & Blömeke, S. (2012). The conceptualization of mathematics competencies in the international teacher education study TEDS-M. *ZDM Mathematics Education*, 3(44), 325 - 340.
- Du, P. (2011). Study on the contents of the criterion of primary and secondary school teachers' basic teaching ability. *Curriculum, Teaching Material and Method*, 31(8), 95-100.
- Fauskanger, J., Jakobsen, A., Mosvold, R., & Bjuland, R. (2012). Analysis of psychometric properties as part of an iterative adaptation process of MKT items for use in other countries. *ZDM Mathematics Education*, 44(3), 387-399.
- Freeman, D., Katz, A., Gomez, P. A., & Burns, A. (2015). English-for-Teaching: rethinking teacher proficiency in the classroom. *ELT Journal*, 69, 129-139.
- Fu, M., & Liu, Y. (2005). On the ability structure of modern mathematics teachers. *Curriculum, Teaching Material and Method*, 25(4), 78-82.

- Guerriero, S. (Ed.) (2017). *Pedagogical Knowledge and the Changing Nature of the Teaching Profession*. Paris, France: OECD.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for crosscultural assessment* (pp. 3 - 38). Mahwah, NJ: Lawrence Erlbaum.
- Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement*, 5(2-3), 81-92.
- Hoth, J., Schwarz, B., Kaiser, G., Busse, A., König, J., & Blömeke, S. (2016). Uncovering predictors of disagreement: ensuring the quality of expert ratings. *ZDM Mathematics Education*, 48(1-2), 83-95.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, 6, 1 - 55.
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests (second edition)*. [www.InTestCom.org](http://www.InTestCom.org).
- Johansson, S., Strietholt, R., Rosén, M., & Myrberg, E. (2014). Valid inferences of teachers' judgements of pupils' reading literacy: does formal teacher competence matter? *School Effectiveness and School Improvement*, 25(3), 394-407.
- Kaiser, G. (1999). Comparative studies on teaching mathematics in England and Germany. In G. Kaiser, E. Luna & I. Huntley (Eds.), *International comparisons in mathematics education* (pp. 140-150). London: Falmer Press.
- Kaiser, G., & Blömeke, S. (2013). Learning from the Eastern and the Western debate: the case of mathematics teacher education. *ZDM Mathematics Education*, 45(1), 7-19.
- Kaiser, G., Blömeke, S., König, J., Busse, A., Döhrmann, M., & Hoth, J. (2017). Professional competencies of (prospective) mathematics teachers—cognitive versus situated approaches. *Educational Studies in Mathematics*, 94(2), 161-182.
- Kaiser, G., Busse, A., Hoth, J., König, J., & Blömeke, S. (2015). About the complexities of video-based assessments: Theoretical and methodological approaches to overcoming shortcomings of research on teachers' competence. *International Journal of Science and Mathematics Education*, 13(2), 369 - 387.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of educational measurement*, 38(4), 319-342.
- Kane, M. T. (2006). Content-related validity evidence in test development. In M. S. Downing, & M. T. Haladyna (Eds.), *Handbook of test development* (pp. 131 - 153). Mahwah: Erlbaum.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3 - 22). Göttingen, Germany: Hogrefe & Huber.
- Krauss, S., Lindl, A., Schilcher, A., Fricke, M., Göhring, A., Hofmann, B., Kirchhoff, P., & Mulder, R. H. (Eds.). (2017). *FALKO: Fachspezifische Lehrerkompetenzen. Konzeption von Professionswissenstests in den Fächern Deutsch, Englisch, Latein, Physik, Musik, Evangelische Religion und Pädagogik*. Münster: Waxmann.
- Kristjansson, E., Desrochers, A., & Zumbo, B. (2003). Translating and Adapting Measurement Instruments for Cross-Linguistic and Cross-Cultural Research: A Guide for Practitioners. *Canadian Journal of Nursing Research*, 35(2), 127 - 142.

- Krupa, E. E., & Confrey, J. (2017). Effects of a reform high school mathematics curriculum on student achievement: whom does it benefit? *Journal of Curriculum Studies*, 49(2), 191-215.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805.
- Leung, F. K. S. (2001). In search of an East Asian identity in mathematics education. *Educational Studies in Mathematics*, 47(1), 35-51.
- Leung, F. K. S. (2014). What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal*, 26(3), 579-605.
- 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*, 9, 151 - 173.
- Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, 36(8), 437-448.
- Lloyd, G. M. (2008). Curriculum use while learning to teach: One student teacher's appropriation of mathematics curriculum materials. *Journal for Research in Mathematics Education*, 39(1), 63-94.
- Mulder, M. (2014). Conceptions of Professional Competence. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International Handbook on Research into Professional and Practice-based Learning* (pp. 107 - 137). Dordrecht: Springer.
- Mulder, M., Weigel, T., & Collins, K. (2007). The concept of competence in the development of vocational education and training in selected EU member states: a critical analysis. *Journal of Vocational Education & Training*, 59(1), 67-88.
- Newton, P., & Shaw, S. (2016) Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, 23(2), 178-197
- Ni, Y., Li, Q., Li, X., & Zhang, Z. H. (2011). Influence of curriculum reform: An analysis of student mathematics achievement in Mainland China. *International Journal of Educational Research*, 50(2), 100-116.
- Ning, H (2010). The theoretical model on the standard for teacher competence (in Chinese). *Journal of Educational Research*, 370(11), 77-82.
- Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211-246.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of Research on Teacher Education* (vol. 2) (pp. 102-119). New York, NY: Macmillan.
- Schilling, S. G., & Hill, H. C. (2007). Assessing measures of mathematical knowledge for teaching: A validity argument approach. *Measurement*, 5(2-3), 70-80.
- Schleicher, A. (2016). *Teaching excellence through professional learning and policy reform*. Paris, France: OECD.
- Secolsky, C., Wentland, E., & Denison, B. (2011). The need for documenting validation transactions: a qualitative component of the testing validation process. *Quality & Quantity*, 45(6), 1303-1311.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Erlbaum.
- Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional vision in preservice teachers. *American Educational Research Journal*, 51(4), 739-771.



- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4 - 14.
- 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*. East Lansing: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Townsend, T., & Bates, R. (2007). *Handbook of teacher education: Globalisation, standards and professionalism in times of change*. Springer.
- Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (pp. 653 - 771). Needham Heights, MA: Allyn & Bacon.
- Wang, B., & Hu, W. (2016). Science teacher instruction competence model construction: A vision of the experts (in Chinese). *Teacher Education Research*, 28(6), 65-74.
- Wang, L., Lin, H., Ma, C., & Hu, W. (2012). The research situation and trend analysis of teacher ability in the last thirty years (in Chinese). *Theory and Practice of Education*, 2012(32), 38-42.
- Wei, S., Wang, Q., Jia, Y., & Chen, F. (2017). A model of subject teaching competences (in Chinese). *Teacher Education Research*, 29(4), 84-91.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45 - 66). Göttingen: Hogrefe.
- Wolfe, P.L., Pedro, J.D., DeVaney Becker, A., & Fennema, E. (1980). Sex Differences in High School Students' Causal Attributions of Performance in Mathematics. *Journal for Research in Mathematics Education*, 11(5), 356-366.
- Yang, X. (2014). *Conception and Characteristics of Expert Mathematics Teachers in China*. Wiesbaden: Springer Spektrum.
- Ye, L. (1998). On teachers' professional competences for new century (in Chinese). *Educational Research and Experiment*, 1998(1), 41-46.
- Zhang, Q., & Xu, W. (2016). On teachers' professional literacy and its construct: Based on considerations of teacher professional standard and mathematics curriculum standard (in Chinese). *Journal of Mathematics Education*, 25(4), 69-73.