

The original source of publication of this article is:

Blömeke, S. & Delaney, S. (2012). Assessment of teacher knowledge across countries: A review of the state of research. *ZDM* 44 (3), 223-247.

The published article is available electronically and The final publication is available at www.springerlink.com at <http://www.springerlink.com/content/w7214t0124j75835/>. The version below may differ in minor ways from the published version.

Assessment of teacher knowledge across countries: A review of the state of research

Sigrid Blömeke & Séan Delaney

Abstract

This review presents an overview of research on the assessment of mathematics teachers' knowledge as one of the most important parameters of school quality. Its focus is on comparative and international studies that allow for analyzing the cultural dimensions of teacher knowledge. In a first step, important conceptual frameworks underlying comparative studies of mathematics teachers' knowledge are summarized. Then, key instruments designed to assess the content knowledge and pedagogical content knowledge of future and practicing mathematics teachers in different countries are described. Core results from comparative and international studies are documented, including what we know about factors influencing the development of teacher knowledge and how the knowledge is related to teacher performance and student achievement. Finally, we discuss the challenges connected to cross-country assessments of teacher knowledge and we point to future research prospects.

Keywords: Mathematics teacher education, comparative study, mathematics content knowledge (MCK), mathematics pedagogical content knowledge (MPCK), large scale assessment

This special ZDM issue on the “Assessment of teacher knowledge across countries” is edited by Deborah Loewenberg Ball, Sigrid Blömeke, Seán Delaney, and Gabriele Kaiser. It brings together two research fields that have developed largely independently from each other: comparative studies on prospective mathematics teachers’ content knowledge and pedagogical content knowledge at the end of teacher preparation (Blömeke and Kaiser), and research on the mathematical knowledge of practicing teachers in several countries (Ball and Delaney). The combination of these two research fields provides the reader with an overview of what is going on in mathematics education research across countries on the knowledge of teachers.

In our first review of this topic for this journal, we had to summarize the state of research as follows: “Teacher-education research lacks a common theoretical basis, which prevents a convincing development of instruments and makes it difficult to connect studies to each other” (Blömeke, Felbrich, Müller, Kaiser & Lehmann, 2008). Since then, not only the research on prospective but also the research on practicing mathematics teachers’ knowledge has continued to develop. Two research groups were particularly productive by assessing teacher knowledge with direct measures: one from Michigan State University in the context of “Mathematics Teaching in the 21st Century (MT21; see e.g. Schmidt, Blömeke & Tatto, 2011)” and the “Teacher Education and Development Study: Learning to Teach Mathematics (TEDS-M; see e.g. Tatto, Schille, Senk, Rodriguez, Bankov, & Reckase, in press; Blömeke, Kaiser & Lehmann, 2010)”, the second one from the University of Michigan in the context of “Learning mathematics for teaching” (LMT; see e.g. Delaney, Ball, Hill, Schilling & Zopf, 2008; Hill, Ball & Schilling, 2008)”. This pioneering work has paved the way for the present special issue.

Our introductory paper summarizes what we have learned in editing this issue, and it presents an overview of research on the assessment of mathematics teachers’ knowledge going on in other research groups. Our focus is on comparative and international studies because these are innovative areas covering long-lasting research gaps. During the past two decades, the interest in international comparative studies on mathematics teachers has increased (Cochran-Smith & Zeichner, 2005; Darling-Hammond, 2000). Mathematics teachers will play a central role in the preparation of future generations’ K-12 students. An examination of mathematics teachers’ knowledge is therefore an important parameter of school quality. It is of high relevance to ascertain whether and how teacher training contributes to the development of teacher knowledge.

At the same time, international comparisons allow for analyzing cultural dimensions of teacher knowledge. By developing international studies, many matters of course are questioned which may remain unquestioned in national studies. The structure and the content of mathematics teacher training depend on a deeper rationale which is a result of

factors which may be at least partly cultural. Like the water in the fish's tank, such cultural givens are too often invisible (Blömeke & Paine, 2008) – and international comparisons provide the chance to move beyond the familiar, and to see with a kind of “peripheral vision” (Bateson, 1994).

The results of comparative studies also provide benchmarks of what level and quality of teacher knowledge can be achieved and they point at country-specific strengths and weaknesses. Efforts to fill existing research gaps have been made since the late 1990s. Several comparative small-scale studies on mathematics teachers and mathematics teacher training are available (e.g., An, Kulm and Wu, 2004; Ma, 1999; Burghes, 2008). Much of the teacher research, however, neglected the content domain, focused on other subdomains of mathematics teachers' competencies like beliefs (Bramald, Hardman, & Leat, 1995; Calderhead, 1996) or intended to capture knowledge by self reports. Studies including direct measures of teacher knowledge and cross-country studies are still needed (Brouwer, 2010; Wilson, Floden, & Ferrini-Mundy, 2001).

In many countries, the results of comparative studies on K-12 student achievement have led to fundamental reforms of the school system. The publication of the PISA 2000 results in Germany, for example, one of the first international studies the country took part in, and the realization that Germany performed at a mediocre level only – in contrast to the country's self-image – virtually led to a “shock”. Heated debates and soul-searching among policy makers, researchers, and lay people finally resulted in changes. Thus, comparative studies of student knowledge provided the chance to understand educational phenomena in a new way. We hope that teacher and teacher training research across countries will produce similar effects.

Our paper is organized as follows: First, we summarize important theoretical models underlying comparative and international studies of mathematics teachers' knowledge. Second, we develop a model of factors assumed to influence the development of teacher knowledge during teacher training and the relationship of teacher knowledge and student achievement. Third, we describe the study design and key instruments developed to assess the content knowledge and pedagogical content knowledge of future and practicing mathematics teachers. Fourth, core results from comparative and international studies from several countries on the structure and the level of this knowledge and how it is associated with teacher training, teacher performance and student achievement are documented. Fifth, we discuss the challenges connected to cross-country assessments of teacher knowledge. Each section is subdivided into research on future teachers and practicing teachers.

1 Modeling mathematics teacher knowledge: Conceptual frameworks

In his recent book *How we think*, Alan Schoenfeld (2010, p. 187) describes in-the-moment decision making as follows: “People’s decision making in well practiced, knowledge-intensive domains can be fully characterized as a function of their orientations, resources, and goals.” Teaching is such a knowledge-intensive domain, and Schoenfeld points out that mathematics knowledge is the most important resource of mathematics teachers. He regards it an important enterprise (p. 203) to develop analytical distinctions of knowledge facets and to clarify why particular knowledge facets are accessed in some classroom situations by teachers and others are not. Affective-motivational facets like orientations and goals or self regulation are supposed to be decisive in these processes as well. Only together the full range of teacher competencies underlying classroom performance is understood. In the following, we present two conceptual frameworks that model the cognitive and the affective-motivational dispositions of mathematics teachers: TEDS-M and LMT.

1.1 Conceptual framework of TEDS-M as a comparative study on prospective teachers

In 2008, a comparative study was carried out that focused on the outcomes of teacher training with standardized testing. Nationally representative samples of primary and lower secondary mathematics teachers in their final year of teacher training from 17 countries in Africa, the Americas, Asia and Europe were examined, as well as representative samples of teacher educators and training institutions (Tatto et al., 2008). TEDS-M was carried out under the supervision of the International Association for the Evaluation of Educational Achievement (IEA).¹ The study looked at how teachers of mathematics were trained and what kinds of knowledge and beliefs they had at the end of their training. More than 24,000 prospective teachers were surveyed.

The professional knowledge of teachers can be divided into several facets: content knowledge, pedagogical content knowledge, curricular knowledge and generic pedagogical knowledge (Shulman, 1985). In the context of TEDS-M as a study about prospective mathematics teachers, the content knowledge was the knowledge of mathematics. Pedagogical content knowledge referred to knowledge about the teaching and learning of mathematics. In the TEDS-M framework, it included curricular knowledge, too. Pedagogical knowledge, finally, was the knowledge typically acquired in a teacher training program that is not subject-matter related (Blömeke & Paine, 2008).

Mathematics content knowledge (MCK), in this framework, includes not only basic factual knowledge of mathematics but also the conceptual knowledge of structuring and

¹ TEDS-M was funded by the IEA, the National Science Foundation (REC 0514431) and the participating countries. In Germany, the German Research Foundation funded TEDS-M (DFG, BL 548/3-1). The instruments are copyrighted by the International Study Center at Michigan State University (ISC). The views expressed in this paper are those of the authors and do not necessarily reflect the views of the IEA, the ISC, the participating countries or the funding agencies.

organizing principles of mathematics as a discipline (Shulman, 1987): why a specific approach is important and where it is placed in the universe of approaches to mathematics. Mathematics pedagogical content knowledge (MPCK) includes the subject-related knowledge for teaching. Shulman (1987, p. 9) characterizes it as an “amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding”. A mathematics teacher has to know about typical preconditions of students and how to represent a topic in the best possible way. Curricular knowledge is part of it and includes teaching materials and curricula.

Table 1. Core situations which mathematics teachers are expected to manage (Tatto et al., 2008)

Mathematical curricular knowledge	<ul style="list-style-type: none"> Establishing appropriate learning goals Knowing about different assessment formats Selecting possible pathways and seeing connections within the curriculum Identifying the key ideas in learning programs Knowledge of the mathematics curriculum
Knowledge of planning for mathematics teaching and learning [pre-active]	<ul style="list-style-type: none"> Planning or selecting appropriate activities Choosing assessment formats Predicting typical student responses, including misconceptions Planning appropriate methods for representing mathematical ideas Linking didactical methods and instructional designs Identifying different approaches for solving mathematical problems Planning mathematics lessons
Enacting mathematics for teaching and learning [interactive]	<ul style="list-style-type: none"> Analyzing or evaluating students' mathematical solutions or arguments Analyzing the content of students' questions Diagnosing typical student responses, including misconceptions Explaining or representing mathematical concepts or procedures Generating fruitful questions Responding to unexpected mathematical issues Providing appropriate feedback

An important implication of the TEDS-M framework that modeled teacher knowledge as a facet of teacher “competencies” (Weinert, 2001) was its link to classroom situations. Since competencies were assumed to represent a latent trait that underlies performance, teacher knowledge was regarded as situated and applied by nature (Blumer, 1969). The classroom situations a teacher has to deal with have to be set by constitutive features of the teaching profession. To determine which features were to be regarded as constitutive, TEDS-M referred to existing standards for the national teacher training programs in its participating

countries (e.g., KMK, 2004; NCTM, 2000). Table 1 documents the problems which mathematics teachers were expected to solve in TEDS-M, based on these standards.

In this applied perspective, it is important to distinguish between three different types of knowledge (Shulman, 1986): knowledge in teaching, that is *propositional knowledge* of, e.g., student errors or misconceptions without being related to a specific classroom context; *case-based knowledge* that includes prototypes, borderline cases and analogies based on individual experiences; and *strategic knowledge* or “practical wisdom” for situations when a teacher is overwhelmed by the multidimensionality and speed of what is going on in the classroom. As far as possible in a paper-and-pencil test, TEDS-M tried to cover the first two types of knowledge.

TEDS-M looked also at the professional beliefs held by the future mathematics teachers. Beliefs were defined by Richardson (1996, 103) as “psychologically held understandings, premises, or propositions about the world that are felt to be true”. As Schoenfeld (2010) pointed out, teacher beliefs are crucial for the perception of classroom situations and for decisions on how to act (Leder, Pekhonen, & Törner, 2002; Leinhardt & Greeno, 1986). Therefore, they connect knowledge and action. In this sense, they are also an indicator of the type of instruction that mathematics teachers will use in their future teaching (Brown & Rose, 1995). If beliefs are operationalized specifically to both the content being taught and the challenges a specific classroom situation presents, empirical evidence exists for a link between teacher beliefs and student achievement (Bromme, 1994). In TEDS-M, several beliefs facets were distinguished, in particular epistemological beliefs about the nature of mathematics and beliefs about the teaching and learning of mathematics (Thompson, 1992).

By distinguishing between knowledge and beliefs and thus to include cognitive and affective-motivational teacher dispositions and by stressing the situative and applied nature of teacher knowledge, the TEDS-M framework can be connected to a measurement tradition prominent in educational psychology: the measurement of competencies. Competencies as defined in general by Weinert (2001) and specifically with regard to teaching by Bromme (1997) means to have the cognitive ability to develop effective solutions for job-related problems and the motivational, volitional and social willingness to successfully and responsibly apply these solutions in various situations (see Figure 1).

Besides beliefs as one facet of the latter affective-motivational dimension, a teacher's professional motivation and self regulation are important. A teacher that regulates her behavior is able to define her professional objectives, to decide on appropriate strategies in order to achieve her objectives and to apply them in various situations. Furthermore, she monitors and evaluates her behavior systematically guided by metacognition (Butler & Winne, 1995; Perry, Phillips, & Hutchinson, 2006; Boekaerts & Corno, 2005). Strengths and

weaknesses are identified and – if necessary – either behavior is adjusted or professional development activities are taken. Self regulation capacities are a necessary precondition in order to be successful as a teacher in the long run.

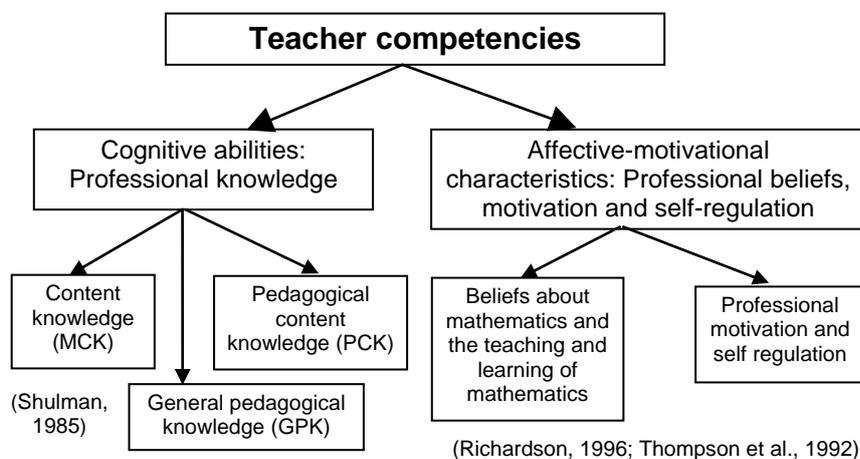


Figure 1. Conceptual framework of teacher competencies

Such a multidimensional notion of teacher competencies that underlie classroom performance did not only lead the item development in TEDS-M but also comparative studies like MT21 or the follow-up studies to TEDS-M (TEDS-FU) carried out in Germany, Taiwan and the USA. TEDS-FU looks currently into the transition of the samples tested in TEDS-M from teacher training into the job (see the forthcoming special IJSME issue, edited by the three national TEDS-M coordinators in these countries, Blömeke, Schmidt & Hsieh).

1.2 Conceptual framework of studies on practicing teachers

The studies on practicing teachers that are considered here differ from the TEDS-M studies in at least two ways. First, the studies focus mainly on elementary teachers. Although, measures of secondary teachers' mathematical knowledge, which can be used at scale, have been developed by the German COACTIV (Cognitive Activation in the Classroom) research group (e.g. Krauss, Baumert, & Blum, 2008), none of the studies reported here have adapted these measures for use in other countries. Instead the measures used were developed for use with elementary teachers by the U.S. LMT (Learning Mathematics for Teaching) research group (e.g. Hill & Ball, 2004). Although the LMT group has subsequently developed measures for studying teachers of middle school students, the team's original focus was on knowledge held by teachers of elementary students.

Second, unlike the conceptual framework which was developed for the purposes of comparing prospective teachers participating in the TEDS-M study, the conceptual

framework for mathematical knowledge for teaching (MKT) was developed by researchers at the University of Michigan (Ball & Bass, 2003, p. 399) in order to better understand teacher knowledge in the United States. The conceptual framework of MKT was inspired by Shulman's (1986) idea of pedagogical content knowledge and seeks to categorize the domains of knowledge needed to do the work of teaching mathematics (Ball, Thames & Phelps, 2008). It consists of two broad categories – subject matter knowledge and pedagogical content knowledge, each of which has been further subdivided into the domains of common content knowledge, specialized content knowledge, and horizon content knowledge on one hand; and knowledge of content and students, knowledge of content and teaching, and knowledge of content and curriculum on the other. Each domain refers to a hypothesized type of mathematical knowledge that is needed by teachers.

Common content knowledge (CCK) refers to mathematical knowledge "used in settings other than teaching" (Ball, Thames, & Phelps, 2008, p. 399) and an example would be recognizing and naming a two-dimensional shape such as a rectangle or a pentagon. Specialized content knowledge (SCK) is mathematical knowledge and skill that is "not typically needed for purposes other than teaching" (Ball et al., 2008, p. 400), such as knowing a range of definitions of shapes that are both comprehensible to students of different age levels, and mathematically accurate and complete. Knowledge of content and students (KCS) "combines knowing about students and knowing about mathematics" (Ball et al., 2008, p. 401) and would involve knowing for example that a square remains a square even if it is rotated 45 degrees. Knowledge of content and teaching (KCT) refers to knowledge of mathematics combined with knowledge of teaching and would include knowing how to select a poster to support the teaching of shapes by using non-examples and non-stereotypical examples. A provisional domain is horizon content knowledge (HCK) which is an "awareness of how mathematical topics are related over the span of mathematics included in the curriculum" (Ball et al., 2008, p. 403). It includes knowledge of the wider discipline of mathematics insofar as its content and practices can inform the work of teaching. Knowledge of content and curriculum (KCC) is another provisional category. Measures based on the framework have been developed for all domains except HCK and KCC and it is these measures which have appealed to researchers outside the United States.

An important distinction to notice between the TEDS-M study and studies using measures based on MKT is that the TEDS-M study was conducted in conjunction with the IEA, whereas the theory of MKT and the measures based on the theory were developed by the U.S.-based Learning Mathematics for Teaching (LMT) project. Although both frameworks are related to classroom situations, the teaching which informed the development of MKT was specifically U.S. teaching (Ball et al., 2008). Consequently, when measures based on

the theory are used outside the United States, researchers need to evaluate the suitability of using U.S. measures to study the mathematical knowledge held by teachers in other countries. This is because the theory is based on the practice of teaching and if the practice of teaching is a cultural activity as some would argue (Stigler & Hiebert, 1999), the theory of MKT may also be culture-specific. Nevertheless, the availability of high quality measures of MKT that could be used at scale (Hill, Schilling, & Ball, 2004) has made it worthwhile to evaluate their suitability for use in several countries, despite the challenges involved in doing so.

One challenge is that a coordinated study of practicing teachers' mathematical knowledge across countries has not yet been funded. This situation contrasts with the substantial funding of several large-scale comparative studies – TIMSS and PISA – of students' mathematical knowledge. The MKT studies that are reported here were conducted locally, are small-scale in nature, and are situated in specific countries. Despite being small in scale, the lessons learned in such studies can inform comparative studies of teacher knowledge, should they take place in the future. The range of settings for the studies – involving countries in Europe, Africa and Asia – provides a robust and diverse test for the U.S. measures. Such diversity is likely to ensure that the studies are well placed to evaluate as well as to contribute to the conceptual framework of MKT.

2 Modeling the link from mathematics teacher education to student achievement

The topic of this special issue and our review, the knowledge of prospective and practicing mathematics teachers, is a crucial link between mathematics *teacher education* and *student achievement* in mathematics. In the first perspective, teacher knowledge represents a core criterion for effective teacher training and is thus a dependant variable. In the second perspective, teacher knowledge represents an important predictor of student achievement and is thus an independent variable. Only both perspectives together provide an appropriate view on the relationship of mathematics teacher training and what is accomplished in schools – multiply mediated by complex context factors.

2.1 Mathematics teacher education and teachers' professional knowledge

In order to examine which factors may influence the development of teacher knowledge – the focus of our review and this special issue – during teacher training, potentially influential factors were divided into three categories in TEDS-M:

- 1) the individual characteristics of future teachers
- 2) the institutional characteristics of teacher training and

3) the national country context.

In studies of school effectiveness, K-12 students' background is almost always a powerful predictor of achievement. Specifically with respect to mathematics, gender (Hyde, Lindberg, Linn, Ellis, & Williams, 2008), socio-economic status (Mueller & Parcel, 1981) and language background (Walter & Taskinen, 2007) as well as prior knowledge (Simmons, 1995) and motivation (Singh, Granville, & Dika, 2002) play an important role (Scheerens & Bosker, 1997). Such a network of individual predictors may apply to the knowledge acquisition during mathematics teacher training as well.

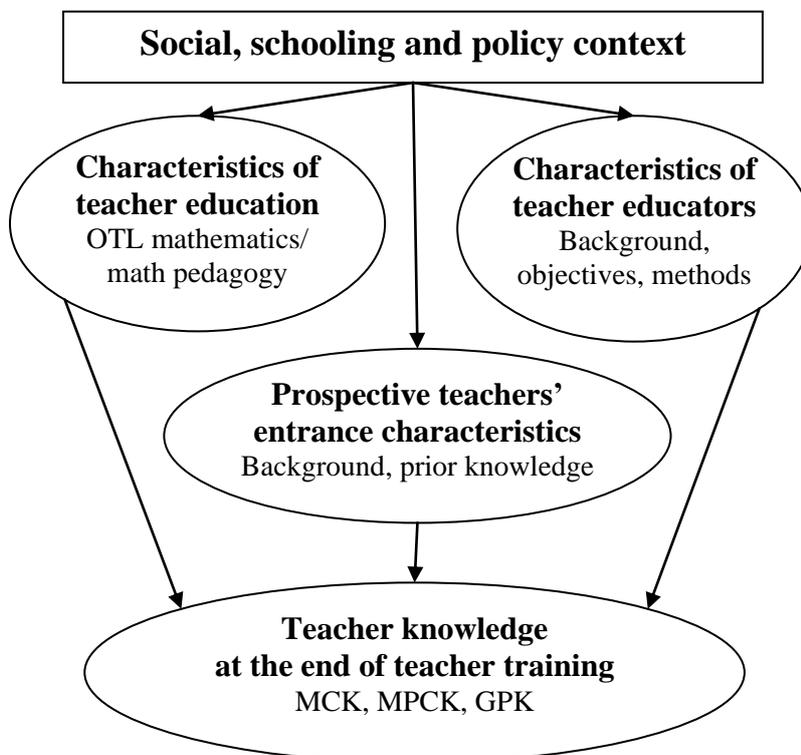


Figure 2. Conceptual framework of factors assumed to influence teacher knowledge (see also Tatto et al., 2008)

With respect to institutional characteristics, TEDS-M followed the tradition of the IEA in connecting educational opportunity and educational achievement. As it was done in the “Third International Mathematics and Science Study (TIMSS)”, opportunities to learn (OTL) were framed as content coverage, specifically as “the content of what is being taught, the relative importance given to various aspects of mathematics and the student achievement relative to these priorities and content” (Travers & Westbury, 1989). OTL were in this sense defined as future mathematics teachers' encountering occasions to learn about particular topics during teacher training, including the characteristics of their educators like background, teaching objectives and teaching methods used. Since subject matter specificity is the defining element of an

educational opportunity (Schmidt, McKnight, Valverde, Houang, & Wiley, 1997), the particular topics reflected the areas of mathematics and mathematics pedagogy.

OTL in teacher training can be regarded as having been intentionally developed by educational policymakers and teacher training institutions (Stark & Lattuca, 1997; Schmidt et al., 2008). They give characteristic shape and direction to instruction. Every choice provides some OTL at the expense of others. National program choices in this sense reflect particular visions of what mathematics teachers are supposed to know and be able to do in class and how teacher training should be organized in order to provide the knowledge and skills necessary for successful accomplishment of their professional tasks.

Thus, teacher knowledge at the end of teacher training is assumed to depend on individual background characteristics of the prospective mathematics teachers, their differential learning experiences during teacher training, opportunities to learn provided by their training institutions and the social, schooling and policy context they are living in (see Figure 2).

2.2 *Mathematics teachers' professional knowledge and student achievement*

If the characteristics of prospective mathematics teachers depend on their entering characteristics, the learning experiences received from interactions with teacher educators and their teacher education program, and from their social, schooling and policy context, additional factors such as experience (e.g. Hill, 2007) and professional development (Bell, Wilson, Higgins & McCoach, 2010) may influence the knowledge held by practicing teachers. Our interest, however, in teachers' knowledge is not ultimately an end in itself but as a means of improving student achievement. For many years researchers have attempted to study the link between teacher mathematical knowledge and student achievement in what are often referred to as educational production function studies (Monk, 1989). Begle conducted an early educational production function study in 1972 (Begle, 1972; Eisenberg, 1977). He subsequently used his own research and studies by others to conclude that beyond a certain level, mathematical knowledge matters little for student achievement (Begle, 1979). In the studies referred to by Begle mathematical knowledge was measured by tests of general mathematical knowledge administered to teachers (e.g. Begle, 1972). Begle's studies and the studies he reviewed provided evidence that proxy measures of teacher knowledge (e.g. math courses studied) and performance on generic mathematics test items are not good predictors of student learning, suggesting that more sophisticated means of studying teachers' mathematical knowledge were needed.

More recent studies that attempt to link teacher knowledge and student achievement have used "pencil-and-paper" tests. Their goal is to test teachers' performance on mathematics problems that occur in teaching rather than on general mathematical

knowledge tasks. They focus on specific mathematical domains, and they are developed by teams with expertise in mathematics, mathematics education and psychometrics (Hill, Sleep, Lewis, & Ball, 2007). An example of such a study is one by Hill, Rowan and Ball (2005) where it was found that every standard deviation of difference in teacher knowledge was worth the equivalent of two to three weeks additional instruction time in predicting the gains made by first and third grade students on standardized math test scores. In third grade the effect size of teacher knowledge “rivalled that of SES and students' ethnicity and gender, while in the first grade models the effect size was not far off” (p. 396). Although such findings in a single study need to be treated with caution, they confirm the importance of pursuing teacher knowledge as a key variable in student achievement.

A German study conducted by Baumert and colleagues (2010) studied the effects of one domain of mathematical knowledge, pedagogical content knowledge (PCK), on student achievement at the end of grade ten. The study found a substantial effect size of PCK on student achievement, an effect which is distinguishable from the effect of general mathematical knowledge and other factors such as track membership. Although this finding too is promising, it is limited to studying teacher knowledge in a single country.

Comparative studies of teacher knowledge have tended to be on a smaller scale and consequently have not established “direct and statistical relationships between teaching-related factors and student mathematics performance” (Wang & Lin, 2005). Although studies such as Ma's (1999) demonstrated that some Chinese teachers have deeper mathematical knowledge than some U.S. teachers, another study found that whether U.S. or Chinese students perform better on mathematics problems varies by problem type (Cai, 2000). More importantly for our interest, factors other than teacher knowledge, such as classroom instruction or placement of topics on the school curriculum (Cai, 2000; Wang & Lin, 2005), may account for differences in student performance between the two countries. Studying the relationship between teacher knowledge and student achievement across several countries could yield important insights in mathematics education. Consequently, insights from country-specific studies of the teacher knowledge/student achievement relationship and from studies of teacher knowledge across countries will inform how challenges inherent in such research can be addressed.

3 Study design and instruments to assess mathematics teachers' knowledge

Most comparative studies have to deal with the same methodological challenges. In order to reduce the complexity, this section focuses again on selected studies as examples. The largest study on prospective mathematics teachers was TEDS-M. Its study design and instruments are

therefore described and evaluated in detail. An instrument for the assessment of practicing teachers' mathematical knowledge applied in many countries is the LMT test which is also documented.

3.1 *Design of TEDS-M as a study on prospective teachers*

3.1.1 *Sampling*

The target population of TEDS-M consisted of students in their final year of teacher training who were on track to receive a license to teach mathematics in primary or lower secondary schools (Tatto et al., 2008). A teacher training program was included if it prepared primary teachers for one of the grades 1 through 4 as the common denominator of level 1 education in the "International Standard Classification of Education" (primary or basic education, cycle 1; UNESCO, 1997), or if it prepared lower secondary teachers for grade 8 as the common denominator of level 2 education (lower secondary or basic education, cycle 2).

In a two-stage process, random samples were drawn from the target populations in each participating country. The samples were stratified according to important teacher training features like "route" (consecutive vs. concurrent programs), "type" of program (grade span the license includes, e.g., grades 5 through 9 vs. 5 through 12), "focus" of opportunities to learn (with or without extensive mathematics) and "region" (e.g., federal state) in order to reflect accurately the distribution of prospective primary and lower secondary teachers' characteristics at the end of training.

Table 2. Countries participating in TEDS-M

Botswana	Chile	Germany	Georgia
Canada	Malaysia	Norway	Oman
Philippines	Poland	Russia	Switzerland
Singapore	Spain	Taiwan	Thailand
U.S.A.			

In 2008, approximately 14,000 prospective primary and 9,000 lower secondary mathematics teachers from 17 countries (see Table 2) were tested on their knowledge of mathematics and mathematics pedagogy by a standardized paper-and-pencil assessment. All countries had to meet the IEA's quality requirements, as set out in TIMSS or PIRLS. This included controlling the translation processes, monitoring test situations and meeting the required participation rates. The aim of these standards was to ensure that bias resulting from non-response was kept within acceptable limits.²

² The participation rates in four countries on the primary level (Chile, Norway, Poland and the USA) and five countries on the lower secondary level (Chile, Georgia, Norway, Poland and the USA) did not fully meet the required benchmarks. Their results are therefore reported in an annotated way. In Poland, Switzerland and the

3.1.2 Instruments

TEDS-M sought to measure prospective teachers' MCK and MPCK at the end of their training. For this purpose, a 60-minute paper-and-pencil assessment was completed during a standardized and monitored test session. In order to capture the desired breadth and depth of teacher knowledge, a matrix design was applied. Five primary and three lower secondary test booklets were developed with rotating blocks of items using a balanced-incomplete-block design (Tatto et al., 2008). The 74 (primary study) or 76 (secondary study) items of the MCK tests covered number (e.g., fractions and decimals, irrational numbers), algebra (e.g., equations/formulae and functions) and geometry (e.g., geometric shapes, location and movement) with approximately equal weight and, to a lesser extent, data. In addition, three cognitive dimensions were covered by the items: knowing (e.g., recalling and computing); applying (e.g., representing and implementing) and reasoning (e.g., analyzing and justifying). A third heuristic emphasized different levels of expected difficulty (novice, intermediate and expert). Sample items are given in Figures 3 and 4 (the full sets of released items are available from tedsm@msu.edu).

Three students have drawn the following Venn diagrams showing the relationships between four quadrilaterals: rectangles (RE), parallelograms (PA), rhombuses (RH) and squares (SQ).

[Tian] [Rini] [Mia]

Which student's diagram is correct? Check one box.

A. [Tian] o₁

B. [Rini] o₂

C. [Mia] o₃

Figure 3. Sample item from the TEDS-M primary test of MCK

Prove the following statement:
If the graphs of linear functions

$$f(x) = ax + b$$

and

$$g(x) = cx + d$$

U.S.A. the coverage of the target population was reduced and in Norway the sample composition did not fully meet the TEDS-M definition of the target population. In the U.S.A., a substantial proportion of missing values was observed. The results of these countries are reported in an annotated way as well. Canada had to be excluded from the study because the country missed the benchmarks to a serious extent.

$(f + g)(x)$ must also go through P .

Figure 4. Sample item from the TEDS-M lower secondary test of MCK

The 32 (primary study) or 27 (lower secondary study) items of the MPCK test covered two facets: Knowledge of curricula and planning, which is necessary before a teacher enters the classroom (e.g., establishing appropriate learning goals, knowing about different assessment formats or linking didactic methods) as well as interactive knowledge about how to enact mathematics for teaching and learning (e.g., diagnosing typical student responses, including misconceptions, explaining or representing mathematical concepts or procedures, providing appropriate feedback). In line with the MCK test, three levels of expected difficulty and four content areas were distinguished. An example is given in Figure 5.

The item development was mainly informed by the MT21 study (Schmidt, Blömeke, & Tatto, 2011), as well as the two Michigan studies entitled “Knowing mathematics for teaching algebra” (KAT; Ferrini-Mundy, Floden, McCrory, Burrill, & Sandow, 2005) and “Learning mathematics for teaching” (LMT; Hill, Ball, & Schilling, 2008). Three item formats were used: multiple choice, complex multiple choice and open constructed response.

When teaching children about length measurement for the first time, Mrs. [Ho] prefers to begin by having the children measure the width of their book using paper clips, and then again using pencils.

Give **TWO** reasons she could have for preferring to do this rather than simply teaching the children how to use a ruler?

Figure 5: Sample item from the TEDS-M primary test of MPCK

3.1.3 Limitations of the TEDS-M design

If one is to evaluate the nature of the TEDS-M tests as done by Döhrmann, Kaiser and Blömeke (in this issue), one can summarize that the MCK and MPCK of prospective teachers was successfully conceptualized and efficiently surveyed through the TEDS-M assessments. The items covered the domains a priori defined and they were suitable for identifying different proficiency levels of prospective teachers from various countries. The authors therefore confirm the overall reliability and validity of the tests from an international point of

view. However, they also point to limits of the assessment. Due to cultural differences between the participating countries, the items did not cover the entire range of the knowledge teachers should acquire during teacher training. Some teacher tasks are only relevant in certain countries but not in others. Thus, the corresponding knowledge was not assessed.

In addition, Döhrmann, Kaiser and Blömeke (in this issue) characterize the orientation of the conceptual framework and the item pool as slightly biased towards a pragmatic conception of teaching and learning, predominantly in place in English-speaking countries. Facets common in continental Europe were taken into account to some extent only. In this European tradition, the mathematics and the mathematics pedagogy tests would have had to include more argumentation and proof items, for example. Also, fundamental ideas of central mathematical concepts such as number or percentage and different ways to introduce them in class would have had to have a higher priority.

Hsieh, Lin and Wang (in this issue) argue along the same line. Confirming the overall validity of the TEDS-M tests, they point out that the tests have more items testing MCK than MPCK which may demonstrate an unbalanced focus of interest. In the TEDS-M design, more information can be derived about MCK than about MPCK although – based on conceptual considerations or empirical evidence – many colleagues would make a plea for MPCK as the crucial facet of teacher knowledge (see e.g. Baumert et al., 2010). More data on MPCK would be urgently needed then. Furthermore, Hsieh, Lin and Wang (in this issue) would prefer an assessment of teacher knowledge as conceptualized by Niss (2003) who emphasized the fundamental characteristics of mathematical thinking across the content domains of mathematics rather than to assess their knowledge in these specific domains (see also Hsieh, 2010; Hsieh, Wang, Hsieh, Tang, Chao, & Law et al., 2010).

An open question is to what extent the TEDS-M tests were measurement invariant across countries. The number of countries was only 15, with even smaller numbers of country groups from similar educational traditions or with substantial proportions of teachers using different languages within the countries. Thus, a potential cultural bias or a potential language bias could only be examined to a limited extent. There is no commonly agreed upon threshold, either, above which a lack of measurement invariance would affect results from cross-country comparisons. However, first evidence suggests that the MCK and MPCK assessments may not have been *completely* equivalent in all TEDS-M countries (Blömeke, Houang & Suhl, 2011).

Although rigorous quality control had taken place (as it always does in IEA studies), language and cultural differences between and within countries seem to be related to *how well* these traits were measured in the TEDS-M countries. The language problem seemed to be larger with respect to MCK than to MPCK. Blömeke, Houang and Suhl (2011) attributed this result to a long history of schooling in a different language than used at home in the case

of MCK. Its acquisition had probably already suffered from language disadvantages before the prospective teachers entered university. In addition, Blömeke, Houang and Suhl (2011) found a potential cultural influence on the measurement properties, too. The factor loadings were surprisingly high in the two Eastern European countries Poland and Russia. Although these countries had not been particularly strongly involved in the test development, it seemed as if the TEDS-M tests were more closely connected to mathematics and mathematics pedagogy traditions in these countries.

3.1.3 *Scaling options*

Scaled scores in TEDS-M were created separately for MCK and MPCK in one-dimensional models using item response theory. Such models stress the conceptual difference between the two knowledge facets. For dichotomous items, the standard Rasch model and for polychotomous items the partial credit model were used (see Tatto et al., in press). Both item types were analyzed simultaneously with ACER Conquest software (Wu, Adams, Wilson, & Haldane, 2007). The resulting achievement estimates were transformed into a scale with an international mean of 500 and a standard deviation of 100 test points.

An alternative way of scaling the data was used by Blömeke, Houang and Suhl (2011). They applied two-dimensional models that can take the conceptual overlap between the latent traits examined, i.e. MCK and MPCK, into account (Reckase, 2009). Whereas traits such as reading literacy or mathematics literacy, typically found in PIRLS or TIMSS, are relatively well-defined, such conceptual clarity does not exist with respect to teacher knowledge. Researchers are still struggling to separate its facets (Graeber & Tirosh, 2008). Studies by Schilling, Blunk and Hill (2007) or Krauss, Brunner, Kunter, Baumert, Blum, Neubrand et al. (2008) demonstrated that MCK and MPCK were highly correlated.

Therefore, Blömeke, Houang and Suhl (2011) used a multidimensional IRT approach in which MCK and MPCK were represented with a general and a nested factor ("within-item multidimensionality," Adams et al., 1997). The model represented Shulman's idea that the nested factor MPCK was a mixture of different abilities and that mathematics pedagogy items measured this mix. According to this idea, solving the mathematics pedagogy items required MCK as a general ability but also specific MPCK. The results supported, in fact, the contention that the nature of teacher knowledge is multidimensional (Blömeke, Houang & Suhl, 2011).

The model isolated the specific MPCK trait from MCK. It represented an elaborated model of the interaction between teachers and items. Thus, the model yielded distinctive profiles of strengths and weaknesses in the mathematics teachers' knowledge in the different countries. This was particularly evident in the case of MPCK. Only in this model, the importance of OTL in mathematics pedagogy was revealed, too. The more a country had

focused on mathematics pedagogy in relation to mathematics during teacher training, the more likely it would be to have a high MPCK mean.

3.1.4 Data analysis

Due to the complex sampling design, standard errors were estimated using balanced repeated replication (BRR) (Tatto et al., in press). Weights were determined by Statistics Canada according to the sampling design and adjusted for non-participation and non-response. Parameter estimations were determined using the International Database Analyzer provided by the IEA. Therefore, the TEDS-M results gave a sound picture of the professional knowledge of prospective mathematics teachers who in 2008 were in their final year of teacher training.

The data were analysed on two levels of aggregation because it was important to distinguish between an evaluation of the national teacher training systems and an evaluation of specific teacher training programs within countries. Both approaches have their benefits and their limitations:

(1) Due to the traditional policy orientation of IEA's large-scale assessments, TEDS-M focused on the national level on the one hand. This approach stressed the overall educational effectiveness of a nation, regardless of the structure of its education system. In this perspective, with regard to international competitiveness, it considered what a nation accomplishes as a whole.

(2) Additional information was gained by looking into program types. Thus, it was possible to learn about pathways to success within countries, i.e. without confounding variables like cultural or societal features. Note that the relatively small sample sizes in the case of teachers (compared to students) became even smaller when types of programs were examined and that the precision of estimates was probably lower because the sampling target was mainly on the national level. This approach has therefore to be used with caution.

3.2 Use of MKT measures to study on practicing teachers

Because the studies that used the MKT measures to study teachers' knowledge are country-specific, they differ in terms of how samples of respondents were selected, in terms of the specific measures used, and in terms of how data were analyzed. This is because studies conducted to date have not been coordinated centrally and the purpose and the resources available differed from one setting to another. What they have in common is that they all accessed the database of multiple choice measures of MKT that were developed at the University of Michigan. The measures available related to each the four hypothesized domains of MKT.

Figure 6 is an example of an MKT item which is considered to tap into teachers' common content knowledge (CCK). A teacher responding to the item needs to consider various strategies that were used to check if a number is prime, where only one strategy is valid. Answer (a) is incorrect because it is possible that 371 could be divisible by two numbers higher than 9 and not by a number less than 9. The number 19 is the square root of 371 making 19 a factor (although in this case 7, one of the numbers to be checked, is also a factor). Response option (b) is incorrect because the 3 in the hundreds place is separated from the tens and units digits and claimed to be prime. However, the 3 represents 300 which is not a prime number. Response (c) considers every prime number less than 20 as a possible factor and because 20^2 is greater than 371, if it has more than exactly two factors, one of them must be less than 20. Response (d) is incorrect for reasons similar to response (b). This item taps into general mathematical knowledge of prime numbers and although it is information a teacher would use, it is not specific to the work of teaching and it requires no knowledge of students or teaching.

Ms. Chambreaux's students are working on the following problem:

Is 371 a prime number?

As she walks around the room looking at their papers, she sees many different ways to solve this problem. Which solution method is correct? (Mark ONE answer.)

- a) Check to see whether 371 is divisible by 2, 3, 4, 5, 6, 7, 8, or 9.
- b) Break 371 into 3 and 71; they are both prime, so 371 must also be prime.
- c) Check to see whether 371 is divisible by any prime number less than 20.
- d) Break 371 into 37 and 1; they are both prime, so 371 must also be prime.

Figure 6.

Sample CCK item from the LMT database of MKT multiple choice items. Released items are available from http://sitemaker.umich.edu/lmt/files/LMT_sample_items.pdf.

Figure 7 is an example of an MKT measure that taps into specialized knowledge of mathematics. A respondent is expected to analyze four different potential representations of a calculation involving multiplication of fractions. In response (a) two unit area models are shown, with $1\frac{1}{2}$ shaded in grey and $\frac{2}{3}$ of the $1\frac{1}{2}$ is indicated with oblique lines. Response (b) is similar but the units are partitioned into sixths rather than halves. The third

model is considered to be unsuitable because one unit is a rectangle and the second is a circle. Response (d) uses a number line model to show a line measuring $1 \frac{1}{2}$ partitioned in three equal parts, each measuring $\frac{1}{2}$. Two of these parts (or $\frac{2}{3}$ of the line measuring $1 \frac{1}{2}$ units) are shaded indicating the product of $1 \frac{1}{2}$ and $\frac{2}{3}$. The knowledge required here is purely mathematical but it is mathematical knowledge that is specific to teaching.

At a professional development workshop, teachers were learning about different ways to represent multiplication of fractions problems. The leader also helped them to become aware of examples that do not represent multiplication of fractions appropriately.

Which model below cannot be used to show that $1 \frac{1}{2} \times \frac{2}{3} = 1$? (Mark ONE answer.)

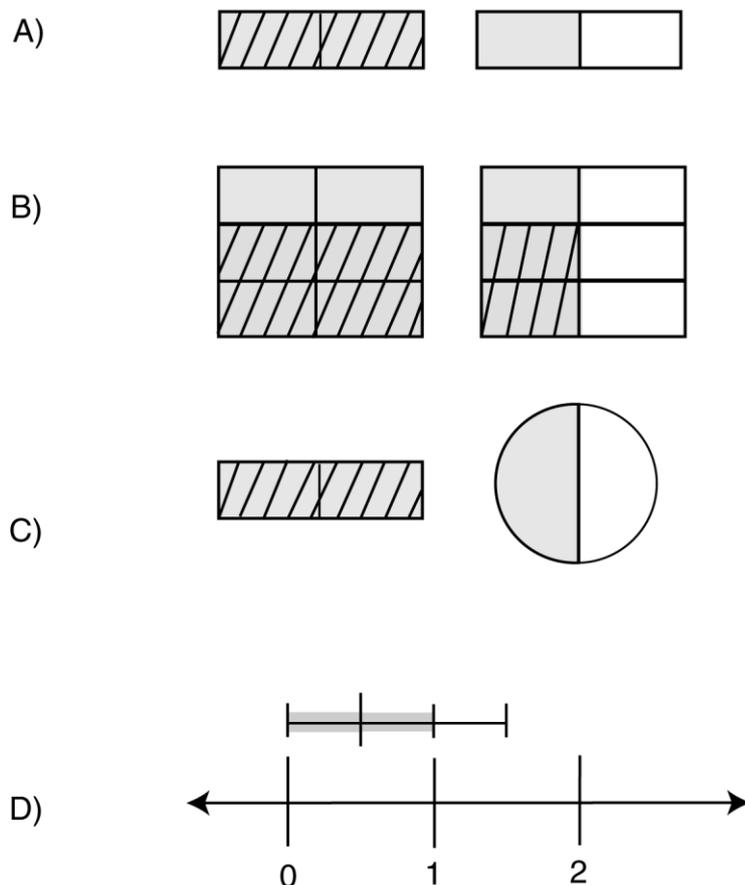


Figure 7.

Sample SCK item from the LMT database of MKT multiple choice items. Released items are available from http://sitemaker.umich.edu/lmt/files/LMT_sample_items.pdf.

Figure 8 is a sample KCS item and it draws on the teacher's knowledge of mathematics and students combined. In three samples of students' work, children have subtracted incorrectly. However, the errors made differ. In sample I, the student exchanged one hundred for ten units and in sample II the student exchanged one thousand for ten units. A teacher who has knowledge of students and mathematics will recognize that both problems relate to a misunderstanding of renaming a number using principles of place value. In sample III the student exchanged one thousand for nine hundreds, eight tens and ten units, which is also incorrect but the error made is more sophisticated than a direct swap of hundreds or thousands for ten units as is the case in I or II, so the correct answer is (a).

Mrs. Jackson is planning mini-lessons for students around particular difficulties that they are having with subtracting from large whole numbers. To target her instruction more effectively, she wants to work with groups of students who are making the same kind of error, so she looks at a recent quiz to see what they tend to do. She sees the following three student mistakes:

I	II	III
$\begin{array}{r} 4 \ 12 \\ \cancel{502} \\ - 6 \\ \hline 406 \end{array}$	$\begin{array}{r} 4 \ 15 \\ \cancel{35005} \\ - 6 \\ \hline 34009 \end{array}$	$\begin{array}{r} 6 \ 9 \ 8 \ 15 \\ \cancel{7005} \\ - 7 \\ \hline 6988 \end{array}$

Which have the same kind of error? (Mark ONE answer.)

- a) I and II
- b) I and III
- c) II and III
- d) I, II, and III

Figure 8.

Sample KCS item from the LMT database of MKT multiple choice items. Released items are available from http://sitemaker.umich.edu/lmt/files/LMT_sample_items.pdf.

The LMT research team in the United States has set a standard for using such items and for analyzing responses. A substantial database of items has been developed and many of these have appeared on various survey forms used in studies by the research group. The purposes for which measures may be used and how results are presented are restricted and

users of the measures are required to participate in a training workshop before using the measures. They may not be used for high stakes purposes such as decisions related to appointments or tenure and raw frequencies cannot be discussed publicly or compared with other groups (<http://sitemaker.umich.edu/lmt/terms>). The measures are suitable for measuring growth in teacher knowledge, studying how teacher knowledge relates to student achievement, studying the mathematical knowledge that teachers need, and for looking at how that knowledge is organized (Hill et al., 2004).

Factor analyses of responses to the items can be compared to the hypothesized domains of teacher knowledge (Hill et al., 2004). Item response theory (IRT) models can be used to score the teachers' responses in standard deviations where the mean is 0 (Hill, 2007). The difficulty of the items and their ability to discriminate among teachers are also estimated using IRT models. Studies were conducted in the United States to evaluate the validity of using the MKT measures. Although more work remains to be done on validation, and in understanding MKT and its measurement, even within the United States, Hill, Ball, Blunk, Goffney and Rowan (2007) found that teachers' scores on the measures could predict mathematical features of the teachers' instruction and student achievement.

One early study by the group looked at how the MKT measures might be adapted for use outside the United States (Delaney, Ball, Hill, Schilling and Zopf, (2008). This study made four categories of changes to items: to the general cultural context, the school cultural context, the mathematical substance and other changes. As other researchers have used the measures in various countries, including non-English speaking countries, they have proposed additional or alternative categories of changes to be considered (see Mosvold, Fauskanger, Jakobsen, & Melhus, 2009; and Kwon, Thames & Pang, this issue) when measures developed for use in one country are adapted for use elsewhere.

4 Core results on the professional knowledge of mathematics teachers

4.1 Prospective teachers

4.1.1 MCK and MPCK by countries

Prospective primary teachers from Taiwan achieved the most favourable MCK result of all of the countries participating in the TEDS-M primary study (Blömeke et al., 2011; see Table 3). The difference from the international mean of 500 test points was large – more than one standard deviation, which is according to Cohen (1988) a highly relevant difference. The achievement of primary teachers from the U.S. was slightly above the international mean and roughly on the same level as the achievement of teachers in Germany and Norway. Their

difference from the international mean was significant but of low practical relevance. These groups of teachers also reached significantly lower performance levels than Swiss and Thai teachers. If we take into account the Human Development Index used by the U.N. in order to indicate the social, economic and educational developmental state of a country, the high performance of teachers from Russia and Thailand was striking.

Regarding MPCK, the achievement of prospective primary teachers from the U.S. was roughly on the same level as the achievement of teachers in Norway, which was significantly above the international mean. In this case, the difference from the international mean was of practical relevance. Teachers from Singapore and Taiwan outperformed the U.S. teachers. Whereas Singapore was behind Taiwan in the case of MCK, these countries were on the same level in the case of MPCK. Regarding MPCK, Norway and the U.S. were only half of a standard deviation behind the two East Asian countries, whereas this difference reached one standard deviation regarding MCK.

Table 3. Knowledge of prospective primary teachers by country

Country	MCK		Country	MPCK	
	Mean	SE		Mean	SE
Taiwan	623	4.2	Singapore	593	3.4
Singapore	590	3.1	Taiwan	592	2.3
Switzerland*	543	1.9	Norway ^{1 n}	545	2.4
Russia	535	9.9	U.S.A. ^{*** 1 3}	544	2.5
Thailand	528	2.3	Switzerland*	537	1.6
Norway ^{1 n}	519	2.6	Russia	512	8.1
U.S.A. ^{*** 1 3}	518	4.1	Thailand	506	2.3
Germany	510	2.7	Malaysia	503	3.1
International	500	1.2	Germany	502	4.0
Poland ¹	490	2.2	International	500	1.3
Malaysia	488	1.8	Spain	492	2.2
Spain	481	2.6	Poland ¹	478	1.8
Botswana	441	5.9	Philippines	457	9.7
Philippines	440	7.7	Botswana	448	8.8
Chile ¹	413	2.1	Chile ¹	425	3.7
Georgia	345	3.9	Georgia	345	4.9

¹ Combined participation rate < 75%

³ High proportion of missing values

* Colleges of education in German-speaking regions

** Institutions with concurrent programs

*** Public universities

ⁿ The results for Norway are reported by combining the two datasets available.

Prospective lower secondary teachers from Taiwan, Russia, Singapore, Poland and Switzerland significantly outperformed teachers from the other countries regarding MCK (Blömeke et al., 2012; see tables 4 and 5). If we take into account the Human Development Index used by the U.N., the performance of lower secondary mathematics teachers from Russia and Poland was remarkable. Regarding MPCK, the achievement of Taiwanese and

Russian teachers was outstanding. The achievement of teachers from Singapore, Switzerland and Russia was also well above the international mean.

The ranking of countries in TEDS-M was very similar to the ranking of countries in TIMSS (Mullis, Martin & Foy, 2008) which allows a preliminary tentative conclusion that we are talking about a cyclic relationship – with the option to improve student achievement by increasing mathematics teachers' professional knowledge.

Tables 4 & 5. Knowledge of future secondary mathematics teachers by country

MCK of future lower secondary mathematics teachers		MPCK of future lower secondary mathematics teachers	
Country	Mean	Country	Mean
Taiwan	667 (3.9)	Taiwan	649 (5.2)
Russia	594 (12.8)	Russia	566 (10.1)
Singapore	570 (2.8)	Singapore	553 (4.7)
Poland** 1	540 (3.1)	Switzerland*	549 (5.9)
Switzerland*	531 (3.7)	Germany	540 (5.1)
Germany	519 (3.6)	Poland** 1	524 (4.2)
U.S.A.*** 1 3	505 (9.7)	U.S.A.*** 1 3	502 (8.7)
International	500 (1.5)	International	500 (1.6)
Malaysia	493 (2.4)	Thailand	476 (2.5)
Thailand	479 (1.6)	Oman	474 (3.8)
Oman	472 (2.4)	Malaysia	472 (3.3)
Norway ^{2 n}	444 (2.3)	Norway ^{2 n}	463 (3.4)
Philippines	442 (4.6)	Philippines	450 (4.7)
Botswana	441 (5.3)	Georgia ¹	443 (9.6)
Georgia ¹	424 (8.9)	Botswana	425 (8.2)
Chile ¹	354 (2.5)	Chile ¹	394 (3.8)

- * Colleges of education in German-speaking regions
- ** Institutions with concurrent programs
- *** Public universities
- ¹ Combined participation rate < 75%
- ² Combined participation rate < 60%
- ³ High proportion of missing values
- ⁿ The results for Norway are reported by combining the two datasets available.

Hsieh, Lin and Wang (in this issue) examined the TEDS-M data in detail based on the notion of thought-oriented mathematical knowledge (Niss, 2003). They found that, although Taiwan outperformed Singapore, these two East Asian countries shared the same structural pattern in their responses (see also Hsieh, Law, Shy, Wang, Hsieh, & Tang, 2011). The pattern suggested that, when compared with other countries, Taiwanese and Singaporean teachers performed relatively better with respect to mathematical language, including representing mathematical entities and handling mathematical symbols or formalisms, than with respect to modeling and reasoning.

As another relative weakness of prospective teachers in Taiwan, the authors discovered that – when compared with other countries – Taiwan performed worse in diagnosing student achievement. This result confirmed findings of domestic studies in Taiwan that teachers demonstrated an incomplete understanding of student learning. The study of Hsieh, Lin and Wang (in this issue) also shows that – when Taiwanese prospective teachers were incapable of providing a formal proof – they did not try a heuristic approach. Overall, these results suggest a specific cultural pattern in teacher performance in Taiwan as discussed in the literature (Leung, 2001; Leung, Graf, & Lopez-Real, 2006) and probably related to basic cultural features of the society (Hofstede, 1983, 1993).

Senk, Tatto, Reckase, Rowley, Peck and Bankov (in this issue) analyze the TEDS-M data by comparing types of teacher training programs. They point out large structural variation in how teachers were trained to teach mathematics across countries. The authors group teacher training programs into four groups. Primary teachers trained as mathematics specialists tend to have higher MCK and MPCK than those trained as generalists. However, within each group of teacher training programs differences of about 100 to 200 test points in MCK, or one to two standard deviations of the population, occur between the highest and the lowest achieving countries. Differences in MPCK range from about 100 to 150 score points. The authors infer from these results that the relative performance within countries may vary greatly, especially if more than one teacher training program exists.

4.1.3 Factors related to prospective mathematics teachers' professional knowledge

According to the TEDS-M results, countries differed with respect to the demographic background of their prospective teachers and the opportunities to learn they received during teacher training (Tatto et al., in press). Both the individual as well as the institutional characteristics level influenced the acquisition of teachers' professional knowledge (Blömeke et al., 2012; Schmidt, Houang & Cogan, in this issue; Wong, Boey, Lim-Teo & Dindyal, in this issue).

Influence of individual characteristics on teacher knowledge

Gender effects: With respect to prospective primary teachers from the 15 countries that participated in TEDS-M, we have to note significant achievement differences in favor of male compared with female teachers in most countries (Blömeke et al., 2011). Pronounced gender gaps existed at the end of teacher training in particular with respect to MCK. The gender effect did not apply to the same extent to MPCK; in Malaysia, female teachers even outperformed male teachers. The MCK differences between male and female teachers were the largest in Poland, whereas only in a few countries (Malaysia, the Philippines, Botswana, and Germany) no significant differences occurred. The comparative study MT21 provided first evidence that gender-related achievement differences in MCK might apply to lower secondary teachers as well (Blömeke & Kaiser, 2010).

The MCK gender effect may be a cumulative effect. MCK has been built over a long period of schooling and teacher training, whereas MPCK was taught only at university. The well-known K-12 disadvantages in the mathematics achievement of girls compared with boys in Western countries (Hyde et al., 2008) may result in differences in prior knowledge as well as in motivational differences and thus play out in teachers' MCK.

That several countries were able to avoid gender inequalities indicates that teacher achievement may reflect cultural patterns. With respect to MPCK, its pedagogical nature may

reduce women's disadvantages, too. Evidence exists that female future teachers tend to support pedagogical motives more strongly than male future teachers do, specifically in comparison to subject-specific motives (Eberle & Pollak, 2006).

Language effects: Another background characteristic associated with prospective teachers' professional knowledge in some countries was their *language* background (Blömeke et al., 2011). In Germany, the United States, and Thailand, differences of high practical relevance occurred in MCK as well as in MPCK. The differences were always in favor of those teachers whose first language matched the official language of instruction in teacher training. Thus, students with a minority background, speaking a different language background at home than the language of instruction was, were at a disadvantage.

This difference may result from selection effects during schooling. The language background is an important predictor of K-12 achievement (Coleman et al., 1966; Thomas & Collier, 1997). Students with a different language background than the one used in instruction, may have been filtered out during schooling or at the beginning of teacher training. An indicator for this interpretation is that the proportion of teachers with a different language background is lower in the prospective teacher force than on average in the K-12 student population. Again, several countries (e.g. Spain) were successful in avoiding differential language effects. It is worthwhile to examine in detail how these countries accomplished language equity.

Prior knowledge: Characteristics strongly associated with prospective teachers' MCK and MPCK, not only in a few countries but more or less universally, were the perceived high-school achievement as well as the number of mathematics classes at school (Blömeke et al., 2012). Effect sizes were large in both cases. Assuming that both predictors were appropriate to indicate prior knowledge, these results are in accordance with the general state of research (see e.g. Anderson & Lebière, 1998; Simmons, 1995). Higher prior knowledge facilitates the acquisition of new knowledge, e.g. by supporting the integration of new information into existing schemata, the modification of knowledge structures or the compilation and chunking of knowledge.

Motivation effects: A final set of individual characteristics associated with MCK and MPCK was motivation. Subject-related motives were positively related whereas extrinsic motives were negatively related to teacher training outcomes if other individual predictors were controlled (Blömeke et al., 2012). It seems as if the persistence to overcome mathematics-related learning difficulties or to invest time and energy in the learning of mathematics decreases if somebody wants to become a teacher primarily because she wants the long-term security of the job but increases if she is interested in the subject (Wigfield & Eccles, 2000).

Several countries concluded from these kinds of results that an active recruitment policy is necessary in order to ensure a high-quality pool of applicants from which teacher training can draw. This, however, raises questions of feasibility with respect to salary, prestige and motivation as the teaching profession does not always compare favorably with other professions available to highly mathematically literate college graduates. In Singapore, all students selected into teacher training receive full salaries as if they were already practicing. Tuition fees are paid by the state (Wong, Boey, Lim-Teo & Dindyal, in this issue). Schmidt, Houang and Cogan (in this issue) describe with respect to the U.S. a different measure to accomplish the same objective. If the U.S. were to successfully implement the Common Core, the new federal mathematics curriculum, the achievement of students in eighth grade might look different and resemble more that of high scoring countries. This might then result in the U.S. being more competitive internationally even if teachers were drawn from the same part of the distribution as is currently done.

Institutional characteristics

Prior to TEDS-M, the state of research had indicated large cultural diversity in the curricula of teacher training across countries (Bishop, 1988). However, a study by Adler, Ball, Krainer, Lin and Novotna (2005) analyzed 160 papers about mathematics teacher training and pointed out that most studies were restricted to small-scale qualitative research and English-speaking countries. In quantitative studies, only the type of license or the number of courses taken was used to define OTL. These measures reflected the amount of content coverage without taking into account which content was offered.

TEDS-M was the first study that provided in-depth information about OTL. Blömeke and Kaiser (in this issue) summarize with respect to primary teacher training that, in fact, a comprehensive core curriculum accepted in all TEDS-M countries neither existed in mathematics pedagogy nor in mathematics. At the same time, the authors were able to conclude that the heterogeneity may be less pronounced than usually discussed. In mathematics (number theory and probability) and mathematics pedagogy (teaching methods) topics existed which were taken by most prospective primary teachers.

Furthermore, it was sufficient to distinguish between a few profiles of OTL in mathematics (“advanced university mathematics”, “basic university mathematics” and a restriction to “school mathematics”) as well as in mathematics pedagogy (a “broad mathematics pedagogy curriculum”, a “functional mathematics pedagogy curriculum” and “teaching methods”) to describe appropriately the mathematics and mathematics pedagogy curriculum across the participating countries. The dominating philosophy across the TEDS-M countries was to provide OTL in basic university mathematics and a broad mathematics pedagogy curriculum.

The common topics and the low number of profiles may reflect shared visions of what primary teachers are supposed to know before they enter the profession. This result confirmed a prior result from a comparative study on lower-secondary mathematics teacher training in six countries (Blömeke, Kaiser & Lehmann, 2008; Schmidt, Blömeke & Tatto, 2011). Its data indicated that specific OTL profiles may exist and that these may have been influenced by culture: In five countries, the multiple institutions where teacher training took place tended to cluster together with respect to the OTL offered, suggesting country level agreement reflecting a cultural effect (Schmidt et al., 2008).

However, a closer examination of the OTL revealed also patterns of high-achieving vs. low-achieving TEDS-M countries. The level of mathematics and the emphasis of mathematics teaching practice were higher in the first cases (Hsieh, Yang & Shy, in press). Schmidt, Houang and Cogan (in this issue) confirmed these aggregated country-level results through multi-level modeling. They found within-country evidence of a significant relationship between teacher training and both MCK and PCK. OTL in mathematics and mathematics pedagogy predicted an increase in MCK of about half of a standard deviation which is a substantial effect size. The largest single effect was produced by OTL closely related to experiences with mathematics instruction. In addition, the more OTL in linear algebra and calculus the prospective teachers had had, the better they did on MCK. OTL in mathematics had not only a strong direct influence on MCK though but also on MPCK, and they probably mediated the effects of OTL in mathematics pedagogy (Blömeke et al., 2012).

Wong, Boey, Lim-Teo & Dindyal (in this issue) summarized in this sense the evidence for Singapore that teacher training counts. Even well-qualified recruits require proper training to realize their potentials. Therefore, their programs are regularly revised in Singapore to ensure that they are responsive to both external changes such as recruitment numbers and education initiatives launched by the state and within-institution research and feedback from the student teachers about their training.

The TEDS-M results are in line with recent evidence from large national longitudinal studies. These had revealed that pure structural features, such as program or degree type, do not have significant effects on short-term outcomes of teacher training, such as teacher knowledge, or long-term outcomes, such as teacher retention or student achievement (Goldhaber & Liddle, 2011). In contrast, especially in the case of mathematics teachers the evidence increasingly suggests that the quality of programs does have an impact on teacher outcomes (Boyd et al. 2009; Constantine et al. 2009). Content courses in mathematics are an important part of these quality features as they provide the background knowledge and the conceptual and factual knowledge necessary to present mathematics topics to learners in a meaningful way and to connect the topics to one another as well as to the learner's prior knowledge and future learning objectives (Cochran-Smith and Zeichner 2005; Wilson et al.

2001). In this context, the curriculum sequence and delivery seems to have an important influence on graduates' subject matter knowledge (e.g., Tatto, Lerman, & Novotna, 2010).

Knowing the content, however, provides only a foundation for mathematics teaching. Student achievement is higher if a strong subject-matter background is combined with strong educational credentials (Clotfelter et al. 2007). The importance of professional preparation, specifically the understanding of how learners acquire mathematical knowledge, how to teach racially, ethnically and linguistically diverse students and using a wide array of instructional strategies, represents another robust finding of teacher education research across various studies (Constantine et al. 2009; NRC 2010). A third robust finding on the impact of OTL on the outcomes of teacher education is the quality of the teaching methods experienced, in particular, the opportunity to engage in actual teaching practices, such as planning a lesson or analyzing student work, rather than only listening to lectures (Boyd et al. 2009).

4.2 *Practicing mathematics teachers*

MKT is a practice-based theory in the sense that MKT is identified with reference to the practice of teaching. Because the practice of teaching can vary from country to country, the content of MKT could vary accordingly. Furthermore, the MKT measures differ from other instruments for measuring teacher knowledge because they are not criterion referenced (Hill, Sleep, et al., 2007). Criteria stating what mathematical knowledge teachers should hold have not been specified. Instead the measures were designed so that teachers could be ordered "relative to one another and to the underlying trait being assessed" (p. 131). Therefore the studies of the MKT held by practicing teachers in various countries could not be compared as the knowledge of prospective teachers could be in the TEDS-M study. In some cases, the researchers' goal was not so much to measure teachers' knowledge at present but to develop guidelines to enable such study in the future.

For example, the study by Kwon, Thames and Pang (this issue) considers changes made to the multiple-choice items in order to use them with Korean teachers. Despite initially proposing a more elaborate scheme of potential changes to items and despite making various changes, such as replacing terminology relating to base ten materials with an alternative term and subsequently with a diagram, the authors conclude by advocating a conservative approach to adaptation in order to maintain the integrity of the items. They identify potential risks that are inherent in adaptation such as reducing or increasing the mathematical demand of the items or losing the validity of an item. Such an argument is helpful for others who seek to adapt the items for use outside the United States.

The study by Fauskanger, Jakobsen, Mosvold and Bjuland (this issue) describe an iterative process that can be used to determine how well items – adapted or not – work when

administered to a group of teachers outside the United States. Point biserial correlations (how teachers' performances on a given item correlate with their performances on all other items) and the relative ordering of item difficulties (measured in standard deviations where an item that a teacher with average ability has a 0.5 chance of responding correctly has a difficulty of 0) for U.S. and Norwegian teachers was studied in order to identify items which could be probed further in focus group interviews. Although different causes of the problems were hypothesized and the precise problems ultimately remain somewhat uncertain, possible reasons include differences in defining mathematical objects, the use of unfamiliar teaching contexts and problematic aspects of translation. When combined with the approach of Kwon and her colleagues (this issue), the findings of the Norwegian study provide a means of identifying problematic items, but mindful of the risks inherent in any adaptation, changes should only be made if a compelling reason exists to do so.

Ng (this issue) studied Indonesian teachers' performances on geometry items. Having studied the point biserial correlations, the item difficulties and the reliability of the measures and compared them to similar statistics among U.S. teachers, he concludes that the geometry measures "may not be a good set of assessments to evaluate Indonesian teachers' mathematical knowledge for teaching geometry." He believes that a reason for this is attributable to national differences between the United States and Indonesia with regard to differences in how shapes are classified. Nevertheless, he believes that the process of studying the performance outside the United States of measures based on MKT can contribute to the overall development of our understanding of MKT.

Both Cole (this issue) and Delaney (this issue) evaluate the validity of using the MKT measures in Ghana and Ireland respectively. Cole looks at the consistency between three teachers' scores on individual items and the mathematical reasoning for choosing the responses they chose. Cole finds that most items could be used validly in Ghana despite evidence of "cultural incongruence." Sources for the incongruence came from possible differences in teaching practices, a specific question format, the length of time taken to do the test, and taking the test in English where despite English being the language of schooling, none of the teachers in the study were native English speakers.

Delaney (this issue) attempted to validate the use of the measures to study Irish teachers' MKT. Like Cole (this issue), he found that in general teachers' thinking was consistent with their responses to the measures. The factors found among teachers' responses to the items in Ireland are similar to the factors found among U.S. teachers' responses. However, the organization of the factors in both countries differs from the hypothesized domains. Finally, Delaney (this issue) found that only in the case of five of ten teachers did their MKT score predict the mathematical quality of instruction to be found in their teaching. This contrasted with findings in a similar study in the United States (Hill, Ball,

et al., 2007). He concludes by identifying challenges to validating the use of the MKT measures in settings outside the United States.

5 Challenges of comparative large-scale assessments of teacher knowledge

It is a methodological challenge to assess the development of knowledge among prospective teachers in the context of a differentiated tertiary education system. Not only do a variety of institutions, teacher training programs, and job requirements exist, but also an outcome that is hard to define and even harder to measure. Another challenge is the comparative perspective of the studies presented in this review. Research perspectives have to be adjusted across borders and educational traditions.

5.1 Generalizability

In most studies reported above, a definition of teacher knowledge as a context-specific disposition was applied. The knowledge can be acquired and it is needed to cope successfully with domain-specific classroom situations and tasks. However, several controversies are unsolved. What is the role of attitudes and beliefs in this context? In future research, in addition to MCK and MPCK as subject-specific facets of mathematics teachers' knowledge, other cognitive criteria like general pedagogical knowledge or affective characteristics like teacher beliefs should be included (as it was done in TEDS-M) in order to develop a full model. Such an approach would increase the validity of studies on teacher knowledge.

With respect to theory development and generalizability, it seems important not to focus solely on individual universities and their teacher training programs but to include research across (sometimes) idiosyncratic features and thus to generate meaningful theories. Shavelson (2012) unpacks competency as a complex ability construct closely related to real-life-situation performance. How to make it amenable to measurement is exemplified by research from business, military and education sectors. The generalizability theory, a statistical theory for modeling and evaluating the dependability of competency scores, is applied to several of these examples. The paper then pulls the threads together into a general competency measurement model.

Shavelson points out that there are limitations to measuring competency in terms of resources, costs and time on various levels. Performance assessment is an issue that has been discussed for a long time (Kane, 1992). It is difficult to generalize results from one situation to another, i.e. problems with reliability (Brennan & Johnson, 1995), and it is difficult to validate the measures (Kane, Crooks & Cohen, 1999). Although these limitations apply to

teacher research as well, it is worthwhile to examine the phenomenon of dependability and generalizability in more detail – also with new instruments beyond the limitations of paper-and-pencil tests.

In this context, the slight difference in spelling between “competency” and “competence” in the English language may be relevant for evaluating the generalizability of results. Sadler (2012) points out that a conceptual distinction can be made between the two terms which in turn leads to distinct measurement approaches. A “competency” often means an identifiable practice. “Competence,” in contrast, often consists of a large number of discrete competencies which could be tested independently. Competence involves being able to select from and then orchestrate a set of competencies to achieve a particular end within a particular context. The competent person makes multi-criterion judgments that are consistently appropriate and situation-sensitive. What is more, the range of situations faced by many professional practitioners is potentially infinite. Decomposing competence into manageable components in order to facilitate judgments has value in certain contexts, but the act of decomposition can obscure how a practitioner would juggle the various bits together to form a coherent whole. It may be worthwhile to follow this assumption with research projects that compare the results of more integrative and holistic judgments with the results of more analytic approaches.

Also, it is necessary to consider the knowledge development of prospective and practicing teachers along a continuum of lifelong learning (Musset, 2010). Such an approach would also allow including classroom observations of teacher performance and possibly even student achievement in order to examine the construct validity of measures.

5.2 Benefits and limits of comparative research

Like everyone else, researchers are embedded in their own culture, and so they often overlook matters of culture. This is particularly the case for teacher training, given the unique way in which it incorporates or touches upon many different levels of education and stands at the intersection of education and other social, economic and political forces (Blömeke & Paine, 2008). This embedded character of the system of teacher training in any one country makes looking beyond that country's experience mandatory in order to recognize the assumptions which drive it, which are all too often taken for granted. The investigation of another teacher training system in a foreign country, for example, and the discovery that it is possible to organize the training differently sheds new light on domestic systems. The recognition of this cultural boundedness of teacher training is an argument for approaching a comparative study in ways that maximize opportunities for cross-cultural communication and the direct examination of concepts (LeTendre, 1999).

As such, language problems become important in comparative studies and are far more demanding to resolve than a “simple” translation of instruments or responses (National Research Council, 2003). Of course, at one level, this is a common, familiar and well-studied aspect of cross-cultural studies, for which there are now widely-used conventions of translation, back translation and so on (Hambleton, 2002). In teacher training, however, more language-related challenges exist that require attention. They are a problem of cultural boundaries. In some countries, Ghana for instance, the language of schooling may vary from the language of the home for many students. Many terms from native languages cannot be translated because adequate English terms are missing and vice versa. It is even difficult to name the process by which future teachers learn their profession: is it teacher education, is it teacher training or is it perhaps teacher preparation?

These questions relate to deeper and often tacit assumptions about schooling, teaching and learning to teach. As these terms connect to broadly shared cultural beliefs, the uniqueness of their meaning often is not explicit and can easily escape scrutiny unless outsiders to the cultural community stumble over them and begin to enquire about them (Blömeke & Paine, 2008). Behind the apparently simple choice of whether to refer to the practice as teacher education, teacher training, teacher preparation, or something else, lie other aspects of history, policy, social values and cultural norms. These are worth examining in detail.

These conceptual challenges of comparative research are extended by methodological challenges. Owing to the low number of countries, in the TEDS-M analyses a “one size fits all approach” (van Ewijk & Sleegers, 2010) has to be used though with parameter estimates the same for all countries in multi-level analyses (see e.g. Blömeke et al., 2012). Thus, a risk exists that country-specific variation in the effect sizes of some predictors is overlooked. At least for the larger countries in the TEDS-M sample, it seems therefore worthwhile to estimate country-specific models in addition to cross-country models.

Future research should also examine in more detail the question of measurement invariance in TEDS-M (Blömeke, Houang & Suhl, 2011). Van de Vijver (1998, 43) points to a serious threat: “An instrument is biased if its scores do not have the same psychological meaning across the cultural groups involved; more precisely, an instrument is biased if statements about (similarities and differences of) its scores do not apply in the psychological domain of the scores.” Equivalence is thus the objective to be achieved. It consists of several dimensions (Vandenberg & Lance, 2000): conceptual equivalence of the latent trait in each group, equivalent associations between operationalizations in each group, and the extent to which they are influenced to the same degree by the same factors.

Hierarchical IRT and multiple-group confirmatory factor analysis provide the tools to determine important properties such as configural invariance, metric invariance, and scalar

invariance (Fox, 2005; Vandenberg & Lance, 2000). Even if full invariance – which is rarely accomplished in cross-cultural research – cannot be determined for TEDS-M, such studies would reveal the extent to which partial invariance is supported. Approaches could then be taken to appropriately deal with the problems. Using hierarchical IRT, for example, de Jong, Steenkamp, and Fox (2007) were able to relax all invariance requirements across groups while retaining the possibility to make substantive comparisons. Such studies would be relevant not only with respect to the TEDS-M assessment data but also, and perhaps more importantly, with respect to the OTL and beliefs data, given the likelihood of self-reported data being even more vulnerable to bias (Blömeke et al., 2010a, 2010b).

5.3 Research on teacher knowledge as a tool to improve teacher training

Wong, Boey, Lim-Teo and Dindyal (in this issue) discuss the value of studies like TEDS-M for teacher training. They point out that the released MCK and MPCK items can be used as a training resource. In fact, the Singaporean TEDS-M team is preparing a book consisting of these released items, the scoring guides, the Singapore results against international benchmarks, and samples of constructed responses. Teacher educators can use these materials in their lessons with future cohorts of prospective teachers then by, e.g., exploring strategies to remedy misconceptions, by designing classroom activities that mirror the scenarios described in the TEDS-M items, by linking the assessment items to the TEDS-M framework and thus to analyze conceptions of teacher knowledge. Thus, although the TEDS-M items were originally created as a summative assessment of teacher knowledge at the end of teacher training, it can be used as a formative assessment of teacher knowledge.

Teacher educators may also want to compare the outcomes of different programs and different institutions in their country. Within almost all countries, huge between-program disparity existed. This means that within the same cultural context some institutions are doing better than others. They may represent a benchmark and provide important information about features of teacher education which can be more easily adapted than features from other countries. Especially the structure and content of the mathematics and the mathematics pedagogy curriculum should be put to the test.

From those countries achieving high scores in TEDS-M, we may want to learn about promising ideas on how to organize teacher training programs. Again the Singapore example may serve as a role model (Wong, Boey, Lim-Teo and Dindyal, in this issue). Here, mathematicians teach the content courses and mathematics educators teach the pedagogical content courses, but they belong to the same department. Under this organization, there are many opportunities for mathematicians and mathematics educators to work in committees and projects that draw on their separate expertise to achieve the common goal of training competent mathematics teachers. They can also share information

about the prospective teachers. Furthermore, all mathematicians learn to supervise practice of prospective teachers at secondary schools through a process of mentoring, and this requirement provides an important opportunity for them to observe first-hand school mathematics teaching and to share their views as a subject specialist with the prospective teachers.

For achieving an increase of teacher education effectiveness, the TEDS-M study points to two further measures, each with separate effects. Providing OTL in mathematics as well as increasing entrance selectivity may have positive consequences for the outcomes of teacher training and thus in the long run for student achievement in mathematics. Mathematics is one of the most important school subjects and a gatekeeper to academic and professional success. Investments in the training of teachers should therefore pay off quickly. Entrance selectivity is a sensitive issue, however. Not everywhere is teaching such a popular and rewarding job that enough applicants for teacher education are available. Higher selectivity, however, may increase the reputation of the profession in the long run so that institutions can recruit from a larger pool.

In addition to such reforms, policymakers have to be aware of the continuing problem of societal inequalities in teacher education outcomes. Special support of female teachers when it comes to the acquisition of MCK in order to overcome cumulative disadvantages of a long history of K-12 schooling seems to be a meaningful measure in many TEDSM countries.

5.4 Adaptation

Because MKT is a practice-based theory (Ball & Bass, 2003), and teaching practices may be cultural in nature (Stigler & Hiebert, 1999), differences may exist in relation to the form MKT takes in different countries. In order that teachers in a given country can respond to the items without being distracted by names or contexts that would be unusual or non-existent for them, some adaptation is necessary. Indeed all of the studies of MKT adapted the measures they used in their research. However, such adaptations must not avoid distorting the mathematical content or the mathematical demands of the measures. Guidelines for adaptation exist for studies such as TIMSS (Johansone & Malak, 2008) and PISA and similar guidelines will be needed for measures of MKT if the measures are to be used in diverse countries around the world. Recommendations from the studies in this issue could be useful in developing such guidelines.

5.5 Validation

Despite the challenges found by Cole (this issue) and Delaney (this issue), and despite the general absence of validity analyses in educational assessments (Hill, Sleep, et al., 2007), validation of the use of the measures needs to be an important part of studies of MKT in any

countries in which the measures are used. Attending to this will help to clarify the organization of the sub-domains of MKT. It will also ensure that the items are tapping knowledge that is needed to teach mathematics. Above all, it will ensure that the measures are valid for the uses to which they will be put.

6 Concluding remarks

This review presented an overview of research on the assessment of mathematics teachers' knowledge as one of the most important parameters of school quality. Its focus was on comparative and international studies that allow for analyzing the cultural dimensions of teacher knowledge. We presented in detail the conceptual frameworks underlying TEDS-M and MKT/LMT, the instruments designed to assess the content knowledge and pedagogical content knowledge of future and practicing mathematics teachers in different countries and core results of its level and structure.

Although cross-national and comparative surveys of student knowledge have a longer track record and attract more sustained funding and attention than studies of teacher knowledge, the current issue acknowledges the importance of learning about teacher knowledge by studying it beyond individual countries. More needs to be learned about adapting and validating measures for use in other countries. However, studying across countries has the potential to offer insights into the original frameworks and contribute to a better and clearer conception of the frameworks underlying TEDS-M and MKT/LMT.

It appears as if – not only in the context of TEDS-M and MKT/LMT – the research on mathematics teacher knowledge has made important progress. Shulman's model of teacher knowledge leads many studies so that the traits examined intend to represent the same. Our summary in section 1 demonstrates this intention with respect to TEDS-M and LMT. The similarities in turn make it easier to compare the instruments and to connect the studies' results to each other than it was in prior research.

Still, many challenges exist. Cross-country equivalence of meaning and predictive validity are the most important ones. The many studies connected to LMT had their focus on this perspective. LMT started as a one-nation enterprise but is expanding to many countries all over the world. In contrast, TEDS-M started as a comparative study but is now complemented by many national studies that go into more details. We are starting to get ideas about how teacher knowledge develops and how it is connected to teacher education and student achievement. It turns out, as assumed, that teacher knowledge is the crucial link between mathematics teacher education and student achievement in mathematics. How much and in which quality opportunities to learn are provided significantly influences the knowledge achieved during teacher training.

In turn, teacher knowledge represents an important predictor of student achievement because a mathematics teachers' decision making in class is a function, among others, of her mathematical knowledge (Schoenfeld, 2010). Thus, the two perspectives together – research on prospective and practicing teachers – provide an appropriate view on the relationship between mathematics teacher education and what is accomplished in schools – although multiply mediated by complex context factors. Here, a lot of further research is needed so that we will be able to understand the nature of teacher competencies underlying classroom performance.

References

- Adams, R. J., Wilson, M., & Wang, W.C. (1997). The multidimensional random coefficients multinomial logit. *Applied Psychological Measurement, 21*, 1-24.
- Adler, J., Ball, D., Krainer, K., Lin, F. L., & Novotna, J. (2005). Reflections on an emerging field: Research on mathematics teacher education. *Educational Studies in Mathematics, 60*, 359-381.
- An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school mathematics teachers in China and the U.S. *Journal of Mathematics Teacher Education, 7*, 145-172
- Anderson, J. R., & Lebière, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum.
- Ball, D. L., & Bass, H. (2003). *Toward a practice-based theory of mathematical knowledge for teaching*. Paper presented at the Annual meeting of the Canadian Mathematics Education Study Group, Edmonton, AB.
- 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.
- 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.
- Begle, E. G. (1972). Teacher knowledge and student achievement in algebra *School Mathematics Study Group Reports, No. 9*. Washington, D.C.: National Science Foundation.
- Begle, E. G. (1979). *Critical variables in mathematics education*. Washington D.C.: Mathematical Association of America and the National Council of Teachers of Mathematics.
- Bell, C.A., Wilson, S.M., Higgins, T. & McCoach, D.B. (2010). Measuring the effects of professional development on teacher knowledge: The case of developing mathematical ideas. *Journal for Research in Mathematics Education, 41* (5), 479-512.
- Cai, J. (2000). Mathematical thinking involved in U.S. and Chinese students' solving of process-constrained and process-open problems. *Mathematical Thinking and Learning, 2*(4), 309-340.

- Delaney, S., Ball, D. L., Hill, H. C., Schilling, S. G., & Zopf, D. (2008). "Mathematical knowledge for teaching": Adapting U.S. measures for use in Ireland. *Journal of Mathematics Teacher Education*, 11(3), 171-197.
- Eisenberg, T. A. (1977). Begle revisited: Teacher knowledge and student achievement in algebra. *Journal for Research in Mathematics Education*, 8(3), 216-222.
- Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis*, 29(2), 95-114.
- Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35, No. 5, 330-351.
- Hill, H. C., Ball, D. L., Blunk, M. L., Goffney, I. M., & Rowan, B. (2007). Validating the ecological assumption: The relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspectives*, 5(2 & 3), 107-118.
- 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.
- Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' knowledge for teaching. *The Elementary School Journal*, 105, No. 1, 11-30.
- Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge. In F. K. Lester (Ed.), *Handbook for research on mathematics education (2nd ed)* (pp. 111-155). Charlotte, NC: Information Age.
- Johansone, I., & Malak, B. (2008). Translation and national adaptations of the TIMSS 2007 assessment and questionnaires. In J. F. Olson, M. O. Martin & I. V. S. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 63-75).
- 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, 873-892.
- Ma, L. (1999). *Knowing and teaching elementary mathematics*. Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.
- Monk, D. H. (1989). The education production function: Its evolving role in policy analysis. *Educational Evaluation and Policy Analysis*, 11(1), 31-45.
- Mosvold, R., Fauskanger, J., Jakobsen, A., & Melhus, K. (2009). Translating test items into Norwegian - without getting lost in translation. *Nordic Studies in Mathematics Education*, 14(4), 101-123.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.

- Wang, J., & Lin, E. (2005). Comparative studies on US and Chinese mathematics learning and the implications for standards-based mathematics teaching reform. *Educational Researcher*, 34, No. 5, 3-13.
- Weinert, F. E. (2001). Concepts of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salgnik (Eds.), *Defining and selecting key competencies* (pp. 45-66). Göttingen, Germany: Hogrefe.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Wilson, S. M., Floden, R. E. & Ferrini-Mundy, J. (2001). *Teacher preparation research. Current knowledge, gaps, and recommendations*. Washington: Center for the Study of Teaching and Policy.
- Wu, M. L., Adams, R., Wilson, M., & Haldane, S. (2007). *ACER Conquest: Generalised Item response Modelling Software (Version 2.0)*. Melbourne ACER.