

Family Background, Entry Selectivity and Opportunities to Learn: What Matters in Primary Teacher Education? An International Comparison of Fifteen Countries

Sigrid Blömeke, Ute Suhl, Gabriele Kaiser, and Martina Döhrmann

Abstract First findings of IEA’s “Teacher Education and Development Study in Mathematics (TEDS-M)” had revealed differences in the demographic background, prior knowledge, opportunities to learn (OTL), and outcomes of primary teacher education between future teachers from different countries. In this chapter, two hypotheses are examined: (1) OTL and teacher background are significant predictors of mathematics content knowledge (MCK) and mathematics pedagogical content knowledge (MPCK) as teacher education outcomes. (2) OTL effects are partly mediated by differential student teacher intake. Data from multilevel models reveal that effects on MCK are in general larger than on MPCK. Gender, prior knowledge and OTL in mathematics are significantly related to both types of outcomes whereas other background characteristics affect MPCK only. Motivation mediates the effects of prior knowledge and the OTL effects are partly mediated by teacher intake. Consequences for educational policy are discussed based on these results. Policymakers have on the one hand to be aware of the continuing problem of societal inequalities. Providing OTL in mathematics as well as increasing entrance selectivity may, on the other hand, have positive consequences for the outcomes of primary teacher education.

Keywords Effectiveness (of teacher education) · Teacher knowledge · MCK · MPCK · Opportunities to learn · OTL · Value added · Comparative study · Pre-service teacher education

Adjusted version of Blömeke et al. (2012).

S. Blömeke (✉) · U. Suhl
Department of Education, Humboldt University of Berlin, 10099 Berlin, Germany
e-mail: sigrid.bloemeke@staff.hu-berlin.de

G. Kaiser
Department of Education, University of Hamburg, 20146 Hamburg, Germany

M. Döhrmann
Department of Sciences and Mathematics, University of Vechta, 49337 Vechta, Germany

1 Introduction

With the publication of comparative studies on K-12 *student* achievement, the competencies of their *teachers* have become areas of considerable interest. This interest is reflected by the “Teacher Education and Development Study: Learning to Teach Mathematics (TEDS-M)” which examined the competencies of mathematics teachers in fifteen countries at the end of their training (Tatto et al. 2008, 2013).¹ Mathematics teachers have a central role in the preparation of future generations of K-12 students. Mathematics not only belongs to the core academic subjects worldwide (Mullis et al. 2008) but is also essential for meeting everyday occupational requirements (Freudenthal 1983).

An examination of mathematics teachers’ competencies and to ascertain whether and how teacher education contributes to their development is therefore one of the most important parameters of school quality. Efforts to fill corresponding research gaps have already been made since the 1990s (Cochran-Smith and Zeichner 2005; Darling-Hammond 2000). Most of the research, however, focused on future teacher *beliefs* as one subdomain of teacher competencies (see e.g. Bramald et al. 1995; Calderhead 1991; Tamir 1988). Large-scale assessments or studies including direct measures of teacher *knowledge* as another subdomain of teacher competencies are still widely lacking (Brouwer 2010; Wilson et al. 2002).

TEDS-M offers a unique chance to examine the relationship of teacher education and future teachers’ knowledge in detail. It was the first comparative large-scale assessment of higher education in which graduates from fifteen countries were tested. The first descriptive results revealed significant mean differences in the future teachers’ background, their opportunities to learn (OTL) during teacher education, and outcomes in terms of mathematics content knowledge (MCK) and mathematics pedagogical content knowledge (MPCK) between countries (Babcock et al. 2010; Blömeke et al. 2011, 2012; Hsieh et al. 2010; Oser et al. 2010). It had to remain an open question though to what extent teacher background and OTL influenced the outcomes. This relationship, examined with respect to future primary teachers, is the focus of the present chapter.

In examining the effects of program characteristics on teacher education outcomes, the chapter contributes to effectiveness research on the level of the teacher education system. It transfers an approach frequently used in K-12 research where effectiveness is defined as “the degree to which schools achieve their goals, in comparison with other schools that are ‘equalized’, in terms of student-intake” (Scheerens 2000, p. 20). The advantage of an approach on the system level is a high precision of estimates due to large sample sizes. At the same time our study is a first approximation of a value-added model because the effects of OTL in teacher

¹The international costs of TEDS-M were funded by the IEA, the US National Science Foundation (REC 0514431) and the participating countries. The instruments are copyrighted by the TEDS-M International Study Center at MSU (ISC). This publication was made possible by grant number BL 548/3-1 from the German Research Foundation (DFG). Any views expressed in this paper are those of the authors and do not necessarily reflect the views of the funders or the ISC.

education are not distorted by teacher background (McCaffrey et al. 2003). The advantage of such an approach is that it filters out characteristics which are not under the control of teacher education.

2 Theoretical Framework

2.1 *Opportunities to Learn*

TEDS-M followed the tradition of the IEA in connecting educational *opportunity* and educational *achievement*. As it was done in TIMSS, 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 and Westbury 1989). OTL were in this sense defined as future primary teachers’ encountering occasions to learn about particular topics during teacher education. Since subject matter specificity is the defining element of an educational opportunity (Schmidt et al. 1997), in the case of TEDS-M as a study about “learning to teach *mathematics*” the particular topics reflected the areas of mathematics and mathematics pedagogy. Teaching mathematics represents a small but important part of primary teachers’ responsibilities since they usually work as class teachers and teach most subjects.

OTL in teacher education can be regarded as having been intentionally developed by educational policymakers and teacher education institutions (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 primary teachers are supposed to know and be able to do in class and how teacher education should be organized in order to provide the knowledge and skills necessary for successful accomplishment of their professional tasks.

In expansion of TIMSS, TEDS-M also examined the *quality* of OTL, e.g. the teaching methods experienced during teacher education (McDonnell 1995). The idea of teacher education as a model for future teaching in class has always played an important role in pedagogical discourse; see e.g. the theory of “signature pedagogies” developed by Shulman (2005). In this paper we take OTL quality into account by including research-based learning approaches.

2.2 *Teacher Background*

In studies of school effectiveness, not only OTL but also K-12 student background is almost always a powerful predictor of achievement. Specifically with respect to mathematics, gender, socio-economic status and language background as well as generic and domain-specific prior knowledge play an important role (Scheerens and

Bosker 1997). Equity with respect to these characteristics is rarely accomplished. It is reasonable to assume that the same applies to teacher education.

Mathematics has been regarded as a male-dominated subject for a long time (Burton 2001). Longitudinal and trend studies reveal that even though differential mathematics achievement by *gender* has decreased over the past decades, females still show lower achievement compared with their male counterparts in mathematics tests in higher school grades and college (Fan et al. 1997; Hyde et al. 2008). The reasons for such inequity mainly point to socio-psychological aspects: the females had received less support and encouragement from teachers and parents and they had had fewer opportunities to learn mathematics (Henrion 1997). One of the few studies on gender effects in teacher education, the comparative study “Mathematics Teaching in the 21st Century (MT21)”, carried out with future lower-secondary mathematics teachers (Schmidt et al. 2011), provides evidence that gender-related achievement differences in mathematics also apply to teacher education. Male lower secondary teachers from Germany significantly outperformed their female counterparts in mathematics tests (Blömeke and Kaiser 2010).

Differences in *language background* are a characteristic known to be associated with K-12 student achievement as well. In many countries, those students whose OTL occur in their second language perform significantly worse than first-language learners (Walter and Taskinen 2007). Classroom discourse plays a major role in this context as Schütte and Kaiser (2011) show with respect to German primary students. The magnitude of language disadvantages is usually increased by the difference between language skills sufficient for communication at home or with peers and the language proficiency necessary for school success (Council of Chief State School Officers 1990; Cummins 1983). Correspondingly, Thomas and Collier (1997) found cumulative effects in the sense that language effects increase in higher grades of K-12 schooling.

Students' *socio-economic status* is generally significantly associated with achievement as well (Coleman et al. 1966). The higher the SES, the better students perform in tests. SES represents in this context access to resources important for learning like wealth or education (Mueller and Parcel 1981). These resources are actively used, or implicitly play out, as support for student progress.

Prior *generic and domain-specific knowledge* has to be included in a study about teacher education effects not only because it has frequently proven to be associated with K-12 student achievement in a strongly positive way (Simmons 1995) but also because not correcting for it could result in an overestimation of other background or institutional effects (Goldhaber and Brewer 1997; Thomas and Mortimore 1996). Prior knowledge has to be regarded as probably having been affected by these characteristics in the past.

Motivation is often positively related to learning outcomes, especially if the learning tasks are complex (Benware and Deci 1984; Grolnick and Ryan 1987) and if motivation is modeled as intrinsic motivation (Singh et al. 2002). With respect to teachers, intrinsic reasons to decide on this profession can be distinguished into *altruistic-pedagogical* and *subject-related motives* (Brookhart and Freeman 1992; Watt and Richardson 2007). How these affect cognitive outcomes of primary teacher

education is an open question. The effects of *extrinsic motivation* on achievement—with respect to teachers it is related to job security and job benefits (Brookhart and Freeman 1992)—are generally mixed (Ryan and Deci 2000). A controversy exists about the extent to which motivation can be regarded as a background characteristic at all. Some researchers argue that including motivation in effectiveness studies would require the inclusion of a variable that may mediate real background effects like socio-economic status and therefore represents an explanation of how these effects play out. Thus, motives will only be included stepwise in the present study about primary teacher education.

2.3 Outcomes of Teacher Education

TEDS-M is based on the notion of professional competencies as they are defined in general by Weinert (2001) and specifically with regard to teaching by Taconis et al. (2004). Competencies in this tradition mean the cognitive and affective-motivational wherewithal to solve job-related problems successfully. In the case of TEDS-M, cognitive abilities have been categorized into three facets which are frequently discussed in the literature: mathematics content knowledge (MCK), mathematics pedagogical content knowledge (MPCK), and—due to feasibility reasons in only three countries: Germany, Taiwan, and the US—general pedagogical knowledge (GPK) (Blömeke 2002; Shulman 1985). The job-related problems to be dealt with by future primary teachers were defined according to existing standards (see e.g. NCTM 1991).

In the present study, MCK and MPCK were used as indicators of the outcomes of primary teacher education. Since we had GPK data from only three countries, we had to leave out this component of teacher competencies. But by including two subdimensions of teachers' professional competencies we lowered the risk of a "mono-operation bias" (De Maeyer et al. 2010): Evidence exists that teachers need to draw on MCK *and* MPCK in order to foster student achievement in mathematics (Baumert et al. 2010). If we used only one of these as outcome indicator, we would miss the breadth of teacher competencies. Although school effectiveness research has established a certain degree of consistency across cognitive outcome measures (Scheerens and Bosker 1997; Thomas et al. 1997), we do not have the same kind of information in teacher education research.

3 Hypotheses

In line with the results from school effectiveness research, we hypothesize that OTL matter for teacher education outcomes (H1). More specifically, we expect that across the fifteen TEDS-M countries OTL in mathematics and mathematics pedagogy as well as research-based learning during primary teacher education significantly predict outcomes in terms of MCK and MPCK. The strengths of the relationships may

vary though. OTL in mathematics should have a stronger impact on MCK than on MPCK, and OTL in mathematics pedagogy should have a stronger impact on MPCK than on MCK because the respective predictors and outcomes correspond more closely to each other. Still, we expect cross-effects, especially an influence of OTL in mathematics on MPCK, because MPCK requires by definition MCK and the two latent traits correlate. Research-based learning should have a stronger influence on MPCK because—in the way it was defined in TEDS-M (including videos of mathematics instruction, for example)—it is much more prominent in the field of mathematics pedagogy than in mathematics.

At the same time, and again in line with the results from school effectiveness research, we suppose that background matters for teacher education outcomes (H2). In particular, we hypothesize significant effects of gender (in favor of males), socioeconomic status (in favor of higher SES) and language background (in favor of first-language learners), prior generic and domain-specific knowledge (in favor of those primary teachers with higher perceived high-school achievement), and motivation (in favor of those with higher altruistic-pedagogical and subject-related motives and lower extrinsic motives) on the acquisition of MCK and MPCK.

Finally, we hypothesize that OTL effects are partly mediated by differential teacher intake (H3). The first descriptive results of TEDS-M had revealed that the composition of future teachers differed in many countries by teacher education program (Tatto et al. 2013). This applied especially to prior knowledge in the sense that teachers who reported better high-school achievement were more often selected—either formally by the institutions or by self-selection—for teacher education programs with more OTL in mathematics and mathematics pedagogy.

4 Study Design

4.1 Sample

The target population of the present study was defined as future teachers in their final year of teacher education who would receive a license to teach mathematics in primary schools (Tatto et al. 2008). This definition included primary teachers who would work as class teachers. A teacher education program was identified as preparing primary teachers if the license covered one of the grades 1 through 4 as the common denominator of education level 1 in the “International Standard Classification of Education” (primary or basic education, cycle 1; UNESCO 1997).

In a two-stage process, random samples were drawn from this target population in each participating country. The samples were stratified according to important teacher education features like “route” (consecutive vs. concurrent programs), “type” of program (grade span the license included, e.g. grades 1 through 4 vs. 1

Table 1 Participating countries in the TEDS-M primary study

Botswana	Chile ^e	Germany	Georgia
Malaysia	Norway ^{d,e}	Philippines	Poland ^{b,e}
Russia	Spain	Switzerland ^a	Singapore
Taiwan	Thailand	USA ^{c,e,f}	

^aColleges of Education in German-speaking regions

^bInstitutions with concurrent programs

^cPublic Universities

^dResults for Norway are reported by combining the two data sets available to cover the entire population of primary future teachers

^eCombined Participation Rate <75 %

^fSubstantial proportion of missing values

through 10) or “focus” of opportunities to learn (with or without extensive opportunities to learn mathematics) in order to reflect accurately the distribution of primary teachers’ characteristics at the end of their training.

In 2008, about 14 000 future primary teachers from more than 500 teacher education programs in fifteen countries (see Table 1) were tested on their MCK and MPCK in a standardized paper-and-pencil assessment. All countries had to meet the quality requirements of the “International Association for the Evaluation of Educational Achievement (IEA)” as known from studies like the “Third International Mathematics and Science Study (TIMSS)”. These included controlling of translation processes, monitoring of test situations, and meeting participation rates. If a country missed the participation benchmark only slightly, its results are reported briefly (“Combined Participation Rate <75 %”). This applies to Chile, Norway, Poland, and the US. In the US, about a quarter of the primary sample had to use a shortened version of the survey instrument for administrative reasons. Therefore, the basic proportion of missing values is higher than in other countries.

In most countries, TEDS-M covered the full target population. Only Switzerland, Poland and the US had to limit their study for economic or other reasons. In Poland, due to difficulties identifying the target population, it was not feasible to include about 10 % of the teacher education institutions where teachers were trained in consecutive programs only. In the US, it was not feasible immediately to include private universities where about one third of the teachers in the target population were trained. They were examined in a separate step; the results did not differ systematically from those at public universities. In Switzerland, only the German speaking regions agreed to participate in the study. Particularly complex is the composition of the Norwegian sample. Data from two different primary programs are available for this country. Although these sub-populations are not completely disjunct because students had the chance to change to the other programs, the present chapter combines them in order to cover the entire population of primary future teachers in Norway.

4.2 Instruments

The *gender* variable was dichotomous with two values (0: female, 1: male). Across the fifteen TEDS-M countries, on average 81 % of the primary teachers in their final year of training were female (range: 59 % in Botswana through 100 % in Georgia).

The *language* spoken at home in contrast to the official language of instruction in teacher education was captured with a four-point Likert scale (0: “never” through 3: “always”). A distinct difference between two groups of countries existed. In Botswana, Malaysia, and the Philippines, future teachers were tested in English although this was the language always or almost always spoken at home by less than 13 %. In Singapore, Thailand and Taiwan, between 30 and 40 % of the teachers always or almost always spoke a different language at home. In the other nine countries, between 86 and 99 % of the future teachers always or almost always spoke the official language of instruction at home.

Measuring *socio-economic status* (SES) is complex. Owing to its multidimensionality, SES can be indicated by different aspects or be a composite of parental education, home resources, parental occupation, and/or parental income (Sirin 2005; van Ewijk and Sleegers 2010). These subdimensions are commonly associated with each other but represent different aspects of societal inequality. Based on their meta-analysis, van Ewijk and Sleegers (2010) recommend either the use of a composite or the use of one single indicator as continuous variables. Dichotomies have to be regarded as unreliable measures of the underlying continuous construct. Including several SES indicators may lead to ambiguity in the interpretation and the true effect would probably be underestimated. Therefore, in the present study *parent education* was used as an indicator of SES. It was separately measured for future teachers’ fathers and mothers on scales covering the seven most important ISCED levels (1 = “primary” through 7 = “beyond ISCED 5A”). One variable was created to represent the parents’ highest education level. On average, almost 40 % of the primary teachers had parents with a university degree (range: 12 % in Botswana through 52 % in Norway).

Perceived high-school achievement was used as a proxy for *generic prior knowledge*. It was measured across school subjects with a five-point Likert scale representing the perceived high-school achievement compared with a future teacher’s age cohort (1: “generally below average” through 5: “always at the top”). Across the TEDS-M countries, about 38 % of the primary teachers reported high-school achievement at or near the top (range: 14 % in Germany through 58 % in Malaysia).

Domain-specific prior knowledge was surveyed through the number of mathematics classes taken during K-12 schooling as a proxy (five-point Likert scale from 1: “below year 10” through 5: “year 12 (advanced level)”). Across the fifteen countries, 68 % of the primary teachers reported at least twelve years of mathematics at school with a minimum of 0 % in Russia where high school ends after grade 11 and a maximum of 100 % in Taiwan and Poland where twelve years of mathematics are mandatory.

The *motives* to become a teacher were captured in three subdimensions: *altruistic-pedagogical*, *subject-related* and *extrinsic motivation*. Four, two or three

statements respectively had to be rated on four-point Likert scales (1: “not a reason” through 4: “a major reason”). An indicator of altruistic-pedagogical motives was e.g. “I like working with young people.” An indicator of subject-related motives was “I love mathematics” and an indicator of extrinsic motives was “I seek the long-term security associated with being a teacher.” On average, altruistic-pedagogical motives dominated the decision to become a primary teacher much more ($M = 3.18$, $SD = 0.65$) than extrinsic motives ($M = 2.05$, $SD = 0.69$) but also more than subject-related motives ($M = 2.04$, $SD = 0.79$). In an international context, the reliability of the pedagogical scale was sufficient (Cronbach’s $\alpha = 0.73$) whereas the reliability of the other two scales was only at or slightly above the critical limit ($\alpha = 0.50$ or 0.60 respectively). Thus, the number of items turned out to be too low. If we had used more items, we still would have achieved a higher reliability though. In any case, we have to be wary of drawing conclusions in the context of this study if we do not find significant correlations.

The OTL index for *mathematics* was based on the future primary teachers’ responses to what extent content was covered in 15 domains across three key areas: (1) continuity and functions, e.g. beginning calculus or multivariate calculus, (2) discrete structures and logic, e.g. linear algebra or number theory, and (3) geometry, e.g. axiomatic geometry or differential geometry. Opportunities to learn probability and statistics were ignored in this paper because the corresponding knowledge is only poorly represented in the mathematics test. The index represents a regression score ($M = 0$, $SD = 1$) with a minimum of -0.75 in Germany ($SD = 0.94$) and a maximum of 1.56 in Thailand ($SD = 0.46$) from a factor analysis with the three OTL indices which explained 68 % of the variance.

The OTL index for *mathematics pedagogy* was based on eight domains, including foundations like the development of mathematics ability and thinking, and instructional applications like developing teaching plans. The index once again represents a regression score ($M = 0$, $SD = 1$) based on a factor analysis with the two counts which explains 71 % of the variance. The minimum was -1.05 in Germany ($SD = 1.11$) and the maximum was 0.75 in Malaysia ($SD = 0.73$).

In TEDS-M, *teaching methods* were captured in several subdomains. For the purpose of this paper, the scale “research-based learning” was chosen which was the only one that corresponds with subject-specific OTL and points to their academic nature of teacher education. Its reliability was good ($\alpha = 0.83$). Four statements covered the reading of research papers as well as active research strategies like analyzing videos. They had to be rated on four-point Likert scales (1: “never” through 4: “often”). Across the fifteen countries, primary teachers reported a medium level of research-based learning during teacher education ($M = 2.36$, $SD = 0.81$) with the lowest level in Germany ($M = 1.65$, $SD = 0.67$) and the highest in Russia ($M = 2.76$, $SD = 0.70$).

TEDS-M sought to measure future teachers’ *MCK* and *MPCK* as outcomes at the end of primary teacher education. For this purpose, a 60-minute paper-and-pencil assessment had to be completed during a standardized and monitored test session. The items were supposed to depict classroom performance of mathematics teachers in grades 1 through 4 as closely as possible. A matrix design with five test booklets of

the type “Balanced Incomplete Block Design” was applied. Scaled scores were created separately for MCK and MPCK in 1-dimensional models using item response theory. For dichotomous items, the standard Rasch model and for polychotomous items the partial credit model were used (see Tatto et al. 2013). Both item types were analyzed simultaneously with ACER Conquest software (Wu et al. 2007). The resulting achievement estimates were transformed into a scale with an international mean of 500 and a standard deviation of 100 test points.

The 74 items of the mathematics test covered number (25 items), algebra (23) and geometry (21) but only to a small extent data (5). Three cognitive dimensions were covered: knowing (33), applying (29) and reasoning (12). About a quarter of the TEDS-M items have been released by the IEA and are available at: teds@msu.edu.

The 32 items of the mathematics pedagogy test covered two subdimensions: pre-active curricular and planning knowledge (16 items) which is necessary before a teacher enters the classroom (e.g. establishing appropriate learning goals, knowing different assessment formats or linking pedagogical methods and instructional designs, identifying different approaches for solving mathematical problems) and interactive knowledge about how to enact mathematics for teaching and learning (16 items; e.g. diagnosing typical students’ responses including misconceptions, explaining or representing mathematical concepts or procedures, providing appropriate feedback).

4.3 Validity of the TEDS-M Measures

As an IEA study, TEDS-M had to meet the benchmarks set by prior large-scale assessments like TIMSS in order to prove validity of its instruments. First of all, the item development had to follow a conceptual framework (Tatto et al. 2008) and it had to be connected to previous research. These precautions provided strong validity-related evidence regarding the content of the scales as well as their meaningfulness and appropriateness. To avoid cultural bias, items had to be sent in from all participating countries. The item pool was reviewed by large groups of experts, and this on the international level and within the participating countries. Translation processes had to follow strict rules and they were controlled by the IEA headquarter. All national research coordinators had to approve the final version of the different instruments in order to satisfy ethical aspects of the research.

In addition to this conceptual validity, measures were taken to ensure high psychometric quality, including the provision of internal-consistency evidence, score reliability evidence, and particularly evidence of measurement invariance (see Tatto et al. 2013). Based on data from an extensive pilot study, initial exploratory factor analyses were carried out. These were followed by confirmatory factor analyses based on data from the main study and referring to the conceptual framework in order to assess the fit of each scale to the data. The structure of the scales was similar to the pilot findings and there was strong consistency between the primary future teacher and secondary future teacher studies. These results again provided validity-related evidence regarding the construct definitions.

To assess the degree to which these factor structures were invariant across countries, Multiple Group Confirmatory Factor Analysis (MCFA) was used. The results provided evidence of the fit of the given factor structure in each country—an important test to defend the meaningfulness of each scale within and across countries.

OTL was measured by asking the future teachers what they perceived had been covered. Such self-reported data always includes certain kinds of risks. Therefore, evidence was collected to prove the validity of these data by correlating the future teacher data to curriculum data (Blömeke and Kaiser 2010).

4.4 Data Analysis

The analyses took the multi-level structure of the TEDS-M data into account. The international sampling plan used a stratified multi-stage probability sampling design (Tatto et al. 2013). The future teachers (individual level) were randomly selected from a list of future teachers for each of the randomly selected teacher education institutions in a country. Teachers from all teacher education programs (level 2) offered by an institution selected were considered in scope if the license formally allowed for the teaching of mathematics in one of the grades 1 through 4 (including in a class teacher's role) and if they were in their final year of teacher education. Countries represented the third level in our multi-level analyses.

Explicitly modeling the cluster structure has several advantages. First, we obtain statistically efficient estimates of regression coefficients and correct standard errors (Hox 2002). Second, and this is important in the context of this paper, we can use covariates at any level of the hierarchy which enables us to examine the extent to which differences in achievement are accountable for by OTL or teacher background. One measure in this context is to adjust for intake differences.

The influence of individual level characteristics (teacher background) on MCK and MPCK was examined first. The background variables were introduced by group centering in order to separate level-1 effects from higher-level effects accurately. When level-2 effects were examined (OTL and teacher intake), the individual-level variables were controlled and therefore introduced by grand mean centering them. In order to determine the mediating effect of teacher intake, it is not only important to state separate significant effects of the predictors as well as of the mediator on outcomes of teacher education but also a significant relationship between the predictors and the mediator (Baron and Kenny 1986). Therefore, an additional multi-level model will be estimated in which this relationship is examined.

In order to check how justified it was to aggregate the OTL and teacher-intake data (self-reported high-school achievement) which were collected on the individual level, we estimated the $ICC(K)$ and the $r_{wg(J)}$ indices as indicators of reliability across our clusters and agreement within these (McGraw and Wong 1996; James et al. 1993). Overall, the results indicated that it was justified to aggregate these measures (see Table 2). Based on the $ICC(K)$ index, we can conclude that all four

Table 2 Indices of reliability and agreement of future primary teachers with respect to self-reported OTL and high-school achievement

	$ICC(K)$	$ICC(K)$ range	$r_{wg(J)}$	$r_{wg(J)}$ range
OTL in mathematics	0.85	(0.53–1.00)	0.90	(0.83–0.97)
OTL in mathematics pedagogy	0.88	(0.55–0.99)	0.84	(0.77–0.90)
Research-based learning	0.88	(0.63–0.99)	0.67	(0.50–0.78)
Teacher intake	0.83	(0.50–0.97)	0.59	(0.38–0.79)

measures were stable enough across programs in the 15 TEDS-M countries to use them as composites (LeBreton and Senter 2008). The average reliability was very good and none of the scales dropped below 0.50 in any of the countries.

Based on the $r_{wg(J)}$ index, the within-group agreement was sufficient with respect to both OTL measures and research-based learning as well. The perceived high-school achievement showed only moderate agreement though (ibid.). However, the average reliability across teacher education programs was high. This result pointed to a lack of consensus *within* programs—may be because of an insufficient supply of applicants so that institutions had to fill their slots with a wide range of future teachers—but still to a relative high consistency across programs which is the more important feature in the context of our analyses.

Within countries, it can reasonably be assumed that effects of predictors play out in the same way. Thus, slopes were defined to be the same across programs in our multi-level analyses. In contrast, due to possible cultural differences between countries the strength of effects like gender could vary on this level. If the number of countries were large enough, random slopes should be estimated. However, due to the already relatively low number of countries this procedure was not feasible in our case and the strength of predictor effects was defined to be the same across countries as well.

One question was whether the model for the examination of OTL effects had to include these variables—introduced on the aggregated level—on the individual level as well. In many studies of composition effects this is a common practice and it is recommended in technical handbooks (see e.g. Snijders and Bosker 1999) because peer effects would be overestimated otherwise. In fact, we followed this recommendation when we examined the role of teacher intake. The focus was different when OTL were concerned, however. Here, we were not interested in separating individual and composition effects. The variables represented OTL offered by programs according to their specific requirements but may have been used with some variation by the future teachers. The mixture of level-1 and level-2 effects is therefore precisely what we would like to obtain.

By including two indicators of teacher education outcomes—MCK and MPCK—we increased the construct validity of our study. At the same time, however, we “bought” an increased risk of type 1 errors because our dependant variables were correlated to each other (Hox 2002). The range varied from a strong manifest correlation in Poland ($r = 0.68$) to a low correlation in Botswana ($r = 0.28$). A multi-

variate multi-level model would have taken care of this problem but it was not feasible. We already had three levels to consider—future teachers, teacher education programs and countries—so that adding another level would have led to unstable results and difficulties in interpreting the results. Given the obvious fact that the risk of missing important effects is negligible (De Maeyer et al. 2010), we applied two univariate three-level models.

Weights were incorporated in order to reflect non-response rates so that robust population estimates could be obtained. Teacher education programs with fewer than four future teachers in an institution were excluded from the analyses in order to insure stable estimates. This measure reduced the original data set of 13 871 primary teachers in their final year of teacher education to 13 829 (=99.7 %) nested in 527 teacher education programs and fifteen countries.

Given this large sample size, statistical significance is not sufficient to distinguish between practically relevant results and results less relevant. Therefore, each effect will be discussed with respect to its practical relevance based on its proportion of one standard deviation. All analyses were done with HLM for Windows Version 6.08.

5 Results

5.1 Variance in the Outcomes of Primary Teacher Education

The unconditioned models revealed that the country level explained a large proportion of variance in the outcomes of primary teacher education. About one-third of the MCK as well as of the MPCK variance was explained by this level (see the footnotes below Tables 4 and 5). This result reflects the huge disparity in the country means (see Table 3). Systematic variance also existed between teacher education programs within countries. The proportion of variance in the future teacher level was higher in the case of MPCK than MCK.

5.2 Effects of Background Characteristics on Teacher Education Outcomes

Our data generally supported H2 in that background matters for outcomes of primary teacher education. We have to be careful, however. There was large variation depending on whether we examined MCK or MPCK and whether we examined demographics, prior knowledge or motivation.

With respect to MCK (see Table 4), *gender* turned out to be the most important individual characteristic across the participating TEDS-M countries. On average, a difference between one-fifth—if gender was introduced separately—or even a quarter of a standard deviation—if the other background characteristics

Table 3 Means and standard errors (SE) of future primary teachers' MCK and MPCK

Mathematics content knowledge		Mathematics pedagogical content knowledge	
Country	Mean (SE)	Country	Mean (SE)
Taiwan	623 (4.2)	Singapore	593 (3.4)
Singapore	590 (3.1)	Taiwan	592 (2.3)
Switzerland ^a	543 (1.9)	Norway ^{d,e}	545 (2.4)
Russia	535 (9.9)	USA ^{c,e,f}	544 (2.5)
Thailand	528 (2.3)	Switzerland ^a	537 (1.6)
Norway ^{d,e}	519 (2.6)	Russia	512 (8.1)
USA ^{c,e,f}	518 (4.1)	Thailand	506 (2.3)
Germany	510 (2.7)	Malaysia	503 (3.1)
International	500 (1.2)	Germany	502 (4.0)
Poland ^{b,e}	490 (2.2)	International	500 (1.3)
Malaysia	488 (1.8)	Spain	492 (2.2)
Spain	481 (2.6)	Poland ^{b,e}	478 (1.8)
Botswana	441 (5.9)	Philippines	457 (9.7)
Philippines	440 (7.7)	Botswana	448 (8.8)
Chile ^e	413 (2.1)	Chile ^e	425 (3.7)
Georgia	345 (3.9)	Georgia	345 (4.9)

^aColleges of Education in German-speaking regions

^bInstitutions with concurrent programs

^cPublic Universities

^dResults for Norway are reported by combining the two data sets available in order to approximate a country mean

^eCombined Participation Rate <75 %

^fHigh proportion of missing values

were controlled—between male and female teachers existed in favor of the males. This is a highly substantial effect. In contrast, future teachers' *language background* and their *parents' education* were influential but the effect sizes were small. The effect of language even disappeared when the teachers' motivation was controlled.

Important for the acquisition of MCK were both proxies of prior knowledge, the *perceived high-school achievement* as well as the *number of mathematics classes*. Those future primary teachers within a program who perceived themselves as good students compared to their peers and reported more years of mathematics during schooling performed better on average in our MCK test. One more year of mathematics and a one-point difference on the perceived high-school achievement scale led to a difference of about twelve test points. Once motivation was introduced, the effect sizes of perceived high-school achievement and number of math-

Table 4 Three-level modeling of future primary teachers' MCK regressed on background characteristics

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)
<i>Individual predictors</i>										
Gender	20.7***					23.7***				24.7***
	(3.2)					(2.9)				(2.8)
Parent education		2.2**				1.8*				2.0*
		(0.8)				(0.9)				(0.9)
Language background			2.9*			3.3**				ns
			(1.5)			(1.2)				
High-school achievement				11.7***		12.3***				9.8***
				(1.0)		(1.0)				(0.8)
Mathematics classes					13.8***	11.8***				9.5***
					(1.5)	(1.3)				(1.4)
Pedagogical motives							ns			-4.4***
										(1.2)
Subject-rel. motives								13.8***		13.2***
								(2.0)		(1.6)
Extrinsic motives									-2.2**	-6.9***
									(0.8)	(1.6)

Variance components in the unconditioned model: Country 33.6 %, Teacher education program 18.6 %, Future teacher 47.9 %. Predictors have been centered around their group means (programs within institutions) in order to obtain pure level-1 effects. ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ematics classes taken decreased slightly. This result may indicate a mediating effect.

Motivation itself had a varying influence on the acquisition of MCK—depending on which subdimension was concerned. Across the fifteen TEDS-M countries, the correlation of *subject-matter related motives* to *subject-matter knowledge* was positive and especially strong, even stronger than prior knowledge. The correlation of *extrinsic motivation* to MCK was significant as well but negative. *Altruistic-pedagogical motives* neither supported nor limited substantially the acquisition of MCK if this characteristic was introduced separately. If all background characteristics were controlled, a small negative effect emerged.

With respect to MPCK (see Table 5), fewer or less substantial effects of background characteristics existed across the fifteen TEDS-M countries. With respect to demographics, only *gender* had on average a small significant effect in favor of male primary teachers and this effect even disappeared if the other background variables were controlled. Neither which language a future teacher spoke at home nor his/her *parents' educational background* was significantly correlated to the acquisition of MPCK.

Table 5 Three-level modeling of future primary teachers' MPCK regressed on background characteristics

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)
<i>Individual predictors</i>										
Gender	4.8*					ns				ns
	(2.2)									
Parent education		ns				ns				ns
Language background			ns			ns				ns
High-school achievement				10.4***		10.1***				8.8***
				(0.8)		(0.9)				(0.9)
Mathematics classes					9.0***	7.1***				5.5***
					(1.1)	(0.7)				(0.6)
Pedagogical motives							3.4**			ns
							(1.3)			
Subject-rel. motives								10.9***		9.1***
								(1.2)		(1.5)
Extrinsic motives									-3.0***	-7.2***
									(0.9)	(1.6)

Variance components in the unconditioned model: Country 31.6 %, Teacher education program 11.5 %, Future teacher 56.9 %. Predictors have been centered around their group means (programs within institutions) in order to obtain pure level-1 effects. ns: not significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In contrast, both proxies of *prior knowledge* turned out to be significantly influential in relation to the acquisition of MPCK. On average, those future primary teachers who indicated better school achievement by one point (e.g. the difference between “generally about average for my year level” and “generally above average for my year level”) performed better by ten test points. One more year of mathematics at school added another seven test points. Also in this case, the effect sizes of perceived high-school achievement and the number of mathematics classes decreased slightly once motivation was introduced.

Motivation itself significantly influenced the acquisition of MPCK. Across the TEDS-M countries, the correlation of *subject-matter related motives* to this type of outcome had about the same positive effect size as perceived high-school achievement. If introduced separately, *altruistic-pedagogical motives* had a positive effect on the acquisition of MPCK as well. The effect size was small, however, and disappeared altogether if all background characteristics were controlled. It is important to note that extrinsic motives were generally significantly negatively correlated to the acquisition of MPCK. A one-point difference on the four-point Likert scale was associated with a loss of seven points in the MPCK test.

5.3 Effects of Opportunities to Learn on Teacher Education Outcomes

With respect to the acquisition of MCK (see Table 6), two of the program features were of high relevance across the TEDS-M countries: OTL taken in mathematics and OTL taken in mathematics pedagogy. Thus, the data strongly supported H1. Both factors led to differences in MCK of one-third or almost a quarter of a standard deviation in favor of those future teachers in a program where they had had one standard deviation more of the respective OTL during teacher education. In particular, OTL taken in mathematics explained a substantial proportion of variance in the outcomes of primary education between programs whereas the proportion was relatively low in the case of OTL taken in mathematics pedagogy. Correspondingly, the data revealed that—if the OTL were mutually controlled—a substantial proportion of the mathematics pedagogy effect on MCK was mediated by OTL in mathematics. Across the TEDS-M countries, the effect size was almost halved.

The research-based learning during primary teacher education generally did not have a significant effect. The acquisition of MCK was neither supported nor limited significantly by reading research papers or using active research strategies like analyzing videos. In this respect, H1 has to be rejected.

Some of the results for MPCK correspond to the MCK results (see Table 7). Similarly to MCK, OTL in mathematics were important for the acquisition of MPCK. In addition to background characteristics, this type of OTL explained a substantial proportion of variance between teacher education programs in the fifteen TEDS-M countries. No matter whether OTL in mathematics were introduced separately or whether other OTL characteristics were controlled, a difference of a quarter of a standard deviation in MPCK existed in favor of those future teachers whose program had offered one standard deviation more of OTL in mathematics during teacher education.

Slightly less additional variance in MPCK was explained by OTL taken in mathematics pedagogy. If introduced separately, the data revealed that future primary teachers who had taken more of these topics performed better in our test, and this by one-fifth of a standard deviation. Similarly to MCK but against our hypothesis, the relevance of OTL in mathematics pedagogy decreased if the OTL in mathematics were controlled. Across the TEDS-M countries, the effect size was more than halved then.

An interesting deviance from the MCK results was the relevance of research-based learning for the acquisition of MPCK. Even though the substantial positive effect disappeared if the other two OTL variables were controlled, the separate effect may point to an important feature of primary teacher education. The proportion of variance explained across the TEDS-M countries by using active and passive research strategies and the average gain in test points corresponded to the effect size of OTL in mathematics pedagogy.

Table 6 Three-level modeling of future primary teachers' MCK regressed on OTL in teacher education and teacher intake (controlling for background characteristics)

	M0	M1	M2	M3	M4	M5	M6
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)
<i>Individual predictors</i>							
Gender	25.5*** (3.2)	25.0*** (2.8)	25.4*** (3.1)	25.4*** (3.1)	25.0*** (2.8)	25.1*** (3.0)	24.7*** (2.8)
Parent education	2.2* (0.9)	2.3* (1.0)	2.2* (0.9)	2.2* (0.9)	2.3* (1.0)	2.2* (0.9)	2.2* (0.9)
Language background	ns	ns	ns	ns	ns	ns	ns
High-school achievement	10.3*** (0.9)	10.3*** (0.9)	10.2*** (0.9)	10.3*** (0.9)	10.3*** (0.9)	9.7*** (0.8)	9.8*** (0.8)
Mathematics classes	9.8*** (1.3)	9.7*** (1.2)	9.8*** (1.4)	9.8*** (1.3)	9.7*** (1.3)	9.7*** (1.4)	9.6*** (1.3)
Pedagogical motives	-4.7*** (1.0)	-4.4*** (1.2)	-4.8*** (1.0)	-4.8*** (1.0)	-4.4*** (1.2)	-4.7*** (1.0)	-4.4*** (1.2)
Subject-rel. motives	14.4*** (1.9)	13.6*** (1.5)	14.3*** (1.9)	14.3*** (1.9)	13.6*** (1.5)	14.1*** (1.8)	13.5*** (1.5)
Extrinsic motives	-7.5*** (1.8)	-7.4*** (1.7)	-7.5*** (1.8)	-7.5*** (1.8)	-7.4*** (1.7)	-7.2*** (1.7)	-7.2*** (1.6)
<i>Program predictors</i>							
OTL in mathematics		34.1* (14.9)			31.5* (14.6)		25.9* (11.8)
OTL in math pedagogy			23.7* (11.9)		13.7 [†] (7.6)		ns
Research-based learning				ns	ns		ns
Teacher intake						44.7** (13.8)	31.9*** (3.7)
<i>R</i> ²	24.6 %	38.1 %	29.5 %	27.5 %	40.7 %	37.8 %	46.3 %

*R*²: Proportion of null-model variance on the program level explained by the respective model. Individual predictors have been centered around the grand mean in order to control for level-1 effects and thus obtain pure level-2 effects. OTL predictors have been centered around their group means (countries) in order to obtain pure level-2 effects. ns: not significant, [†] *p* < 0.10, * *p* < 0.05, ** *p* < 0.01, *** *p* < 0.001

5.4 The Role of Teacher Intake

Entry selection according to perceived high-school achievement seemed to play a major role in the acquisition of MCK and MPCK across the fifteen TEDS-M countries although program effects in terms of OTL in mathematics were still substantial even after controlling for teacher intake and background effects (see Tables 6 and 7, M6). The data revealed that across the TEDS-M countries a difference of about two-

Table 7 Three-level modeling of future primary teachers' MPCK regressed on OTL in teacher education and teacher intake (controlling for background characteristics)

	M0	M1	M2	M3	M4	M5	M6
	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>b</i>
	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)	(SE)
<i>Individual predictors</i>							
Gender	ns	ns	ns	ns	ns	ns	ns
Parent education	ns	ns	ns	ns	ns	ns	ns
Language background	ns	ns	ns	ns	ns	ns	ns
High-school achievement	9.3*** (1.0)	9.3*** (1.0)	9.3*** (1.0)	9.3*** (1.0)	9.3*** (0.9)	8.6*** (0.8)	8.8*** (0.9)
Mathematics classes	6.0*** (0.6)	5.8*** (0.6)	6.0*** (0.6)	5.8*** (0.6)	5.8*** (0.6)	5.8*** (0.6)	5.7*** (0.6)
Pedagogical motives	ns	ns	ns	ns	ns	ns	ns
Subject-rel. motives	10.3*** (2.3)	9.4*** (1.6)	10.1*** (2.2)	10.2*** (2.3)	9.3*** (1.6)	9.9*** (2.0)	9.2*** (1.6)
Extrinsic motives	-7.8*** (1.8)	-7.7*** (1.7)	-7.8*** (1.8)	-7.8*** (1.8)	-7.7*** (1.7)	-7.4*** (1.6)	-7.4*** (1.6)
<i>Program predictors</i>							
OTL in mathematics		27.0* (12.8)			24.6* (12.3)		19.9* (9.9)
OTL in math pedagogy			20.5* (9.6)		9.7 [†] (5.1)		ns
Research-based learning				22.2* (10.5)	ns		ns
Teacher intake						37.6** (12.2)	26.5*** (3.5)
R^2	20.1 %	32.2 %	25.9 %	25.3 %	36.7 %	32.9 %	41.7 %

R^2 : Proportion of null-model variance on the program level explained by the respective model. Individual predictors have been centered around the grand mean in order to control for level-1 effects and thus to obtain pure level-2 effects. OTL predictors have been centered around their group means (countries) in order to obtain pure level-2 effects. ns: not significant, [†] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

fifths of a standard deviation in MCK as well as in MPCK existed between teacher education programs in favor of those programs where the primary teachers reported a one-point higher mean school achievement level if this indicator of teacher intake was introduced separately. These are highly substantial effects. The corresponding school achievement effect on the individual level decreased only slightly after the composite was introduced.

If the composition characteristic was introduced in addition to the OTL characteristics, the intake effect and the effects of OTL in mathematics on MCK and

Table 8 Two-level modeling of teacher intake regressed on OTL in primary teacher education

	M1 <i>b</i> (SE)	M2 <i>b</i> (SE)	M3 <i>b</i> (SE)
<i>Predictors</i>			
OTL in mathematics	0.15 [†] (0.08)		ns
OTL in mathematics pedagogy		0.19** (0.07)	0.15* (0.08)

Variance components in the unconditioned model: Country 46.0 %, Teacher education program 54.0 %. Predictors have been centered around their group means (countries) in order to obtain pure level-2 effects. ns: not significant, [†] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

MPCK decreased by 13 or 11 and by 6 or 5 test points respectively and the effects of OTL in mathematics pedagogy disappeared completely. These results pointed to a mediating effect in the sense that primary teachers with a better perceived school achievement were selected or selected themselves to a higher extent for programs with more OTL to learn mathematics and mathematics pedagogy so that the entrance differences mediated the OTL effects.

In order to support this hypothesis, it is necessary to show in addition to our previous results that OTL as predictors significantly influenced teacher intake as the assumed mediator as well (Baron and Kenny 1986). For this purpose an additional two-level model was examined with programs as level 1 and countries as level 2. This model allowed us to use composition of programs according to perceived high-school achievement as the dependant variable and OTL in mathematics and mathematics pedagogy as predictors.

In fact, if introduced separately both OTL characteristics showed a systematic relationship with the mean level of perceived high-school achievement in teacher education programs (see Table 8). In particular, the effect of OTL in mathematics pedagogy was significant which fits well to our prior results. It seems as if OTL in mathematics pedagogy are an especially important feature of primary teacher education programs that drove the (self-)selection process—and thus have an indirect effect on MCK and MPCK. If examined separately without taking teacher intake into account (Tables 6 and 7, M2 and M4), there is a significant relationship of OTL in mathematics pedagogy to MCK and MPCK. This relationship disappears if one controls for teacher intake. In contrast, OTL in mathematics does not have a significant relationship to teacher intake. So, the effects remain in M6 compared to M4 (see Tables 6 and 7).

6 Discussion

Data from the comparative TEDS-M study revealed that the mathematics content knowledge (MCK) and the mathematics pedagogical content knowledge (MPCK) of primary teachers differed significantly at the end of teacher education between the participating countries and between teacher education programs within countries. In this chapter, we examined to what extent teacher background, prior knowl-

edge, motivation, opportunities to learn (OTL) during teacher education and teacher intake influenced the knowledge acquisition across countries on average in order to contribute to a global theory of teacher education effectiveness.

Our hypothesis that teacher background generally influenced the outcomes of teacher education (H2) was only partly supported by the data. *Gender* turned out to be an important individual characteristic but only with respect to the acquisition of MCK and not with respect to MPCK. In the first case, university training may have suffered from cumulative effects during a long history of gender inequity in K-12 schooling (Hyde et al. 2008). The acquisition of MPCK started only after that, which may have reduced the disadvantages of females.

Against our hypothesis, the *language background* of the teachers and their *parents' education* were relevant neither for MCK nor for MPCK. Given that these are important predictors on the school level (Coleman et al. 1966; Thomas and Collier 1997), this result is surprising. It seems as if the many selection processes during schooling had filtered out those students who were at a disadvantage because of their background.

In contrast, our data strongly confirmed our hypotheses that the *perceived high-school achievement* as well as the *number of mathematics classes at school* significantly correlate with MCK and MPCK. Effect sizes were large in both cases. Assuming that both predictors are appropriate to indicate prior knowledge, these results are in accordance with the general state of research (see e.g. Anderson and Lebière 1998; Simmons 1995). A possible explanation may be that 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.

With respect to motivation, it is important to distinguish between subdimensions because it had either no practically relevant (*altruistic-pedagogical motives*) or contradictory effects (positive: subject-related motives, negative: extrinsic motives) on the outcomes of primary teacher education. 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 s/he wants the long-term security of the job but increases if s/he is interested in the subject (Wigfield and Eccles 2000). Some evidence surfaced that motivation was one of the channels through which prior knowledge played out. Further research is needed at this point but such a result would support the critical evaluation laid out at the beginning of the paper that motivation should not be regarded purely as a background characteristic.

With respect to program characteristics, the data supported our hypotheses that OTL and teacher intake are highly relevant to teacher education outcomes (H1 and H3). Both features were introduced as aggregated variables on the program level in order to increase the reliability of the measures. In fact, the *ICC(K)* estimates revealed strong agreement within programs.

OTL in mathematics were of outstanding relevance for the outcome of primary teacher education. They had not only a strong direct influence on MCK but also on MPCK and they probably mediated the effects of OTL in mathematics pedagogy.

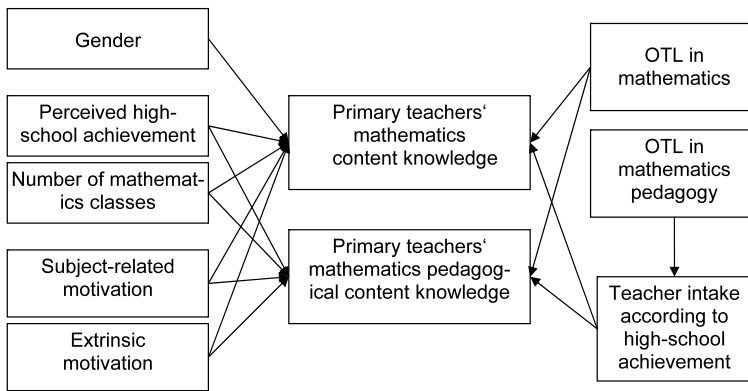


Fig. 1 Hypothetical model of the effects of teacher background, opportunities to learn and teacher intake on outcomes of primary teacher education

These in turn probably mediated the effect of research-based learning although further research is needed about the specifics of these processes.

Besides the relevance of OTL, the relevance of entry selection at the beginning of primary teacher education—either carried out officially by an institution or program or implicitly happening as self-selection by the future teachers—became apparent as well. OTL in mathematics pedagogy were an important feature here and thus had an indirect effect on MCK and MPCK. This result probably reflects the widespread nature of primary teacher education programs as trainings of generalists with broader coverage of mathematics pedagogy than of mathematics. The larger this coverage is, the more it attracts students with higher self-perceived high-school achievement who in turn show higher MCK and MPCK at the end of teacher education.

In addition, the composition effect significantly mediated the effects of opportunities to learn. It is important to note, however, that OTL in mathematics were still substantial even after controlling for teacher intake and background effects.

These results lead to a first hypothetical model of the effectiveness of primary teacher education from a global perspective, which is summarized in Fig. 1.

Before conclusions are drawn, we have to point out some methodological limitations of our study. TEDS-M was a cross-sectional study with a retrospective self-report about school achievement. Longitudinal data and a better measure of prior knowledge are needed for far-reaching conclusions. Furthermore, owing to the low number of countries we had to use a “one size fits all approach” (van Ewijk and Slegers 2010) with parameter estimates the same for all countries. Thus, a risk exists that country-specific variation in the effects sizes of some predictors was overlooked (with respect to variation in gender and language effects by country see Blömeke et al. 2011). At least for the larger countries in the TEDS-M sample, it seems therefore worthwhile to estimate country-specific models.

In future research, in addition to MCK and MPCK as subject-specific criteria of teacher education outcomes, other cognitive criteria like general pedagogical knowledge or affective characteristics like teacher beliefs should be included in order to

develop a full model. Such an approach would increase the validity of a study of teacher education effectiveness. In this context the increased risk of type 1 errors owing to correlation between different criteria should be addressed as well, e.g. by multi-level structural equation modeling.

With respect to effects of single variables, we have to point out that the SES effect may have been underestimated because a single indicator instead of a composite was used (van Ewijk and Slegers 2010). To create a composite, in our study data about parental occupation were missing. In addition, the reliability of the scale measuring extrinsic motivation was at a critical limit. Since we discovered a significant effect in any case, we can assume that its size was underestimated as well.

7 Conclusions

If school effectiveness can be defined as “the degree to which schools achieve their goals, in comparison with other schools that are ‘equalized’, in terms of student-intake” (Scheerens 2000, p. 20), we examined in this chapter the effectiveness of teacher education in 527 programs from fifteen countries with respect to MCK and MPCK as cognitive outcomes after equalizing their teacher intake. Future research should continue this line of research but aim at improving some of the methodological weaknesses discussed above. Also, it seems necessary to include classroom observations of teacher performance and possibly even K-12 student achievement to examine the construct validity of our outcome measures. With respect to OTL, it may be beneficial to go into more detail instead of examining broad constructs like “OTL in mathematics” to gain more insight into the relationship between program characteristics and knowledge acquisition. Subdomains like number or algebra or indicators like types of practical experience are worth examination.

Policymakers have to be aware of the continuing problem of societal inequalities even 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 TEDS-M countries.

For achieving an increase of teacher education effectiveness, our study points to two potential measures, each with separate effects. Providing OTL in mathematics as well as increasing entrance selectivity may have positive consequences for the outcomes of primary teacher education 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 at primary schools 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.

Teacher educators may 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.

References

- Anderson, J. R., & Lebière, C. (1998). *The atomic components of thought*. Mahwah: Erlbaum.
- Babcock, J., Babcock, Ph., Buhler, J., Cady, J., Cogan, L., Houang, R., Wight, K., et al. (2010). *Breaking the cycle: an international comparison of US mathematics teacher preparation*. East Lansing: Michigan State University Press.
- Baron, R. M., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual strategy and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., et al. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom and student progress. *American Educational Research Journal*, *47*(1), 133–180.
- Benware, C., & Deci, E. L. (1984). Quality of learning with an active versus passive motivational set. *American Educational Research Journal*, *21*, 755–765.
- Blömeke, S. (2002). *Universität und Lehrerbildung*, Bad Heilbrunn: Klinkhardt.
- Blömeke, S., & Kaiser, G. (2010). Mathematics teacher education and gender effects. In H. Forgasz, K.-H. Lee, J. Rossi Becker, & O. Bjorg Steinsthordottir (Eds.), *International perspectives on gender and mathematics education* (pp. 263–283). Charlotte: Information Age.
- Blömeke, S., Suhl, U., & Kaiser, G. (2011). Teacher education effectiveness: Quality and equity of future primary teachers' mathematics and mathematics pedagogical content knowledge. *Journal of Teacher Education*, *62*(2), 154–171.
- Blömeke, S., Suhl, U., Kaiser, G., & Döhrmann, M. (2012). Family background, entry selectivity and opportunities to learn: What matters in primary teacher education? An international comparison of fifteen countries. *Teaching and Teacher Education*, *28*, 44–55.
- Bramald, R., Hardman, F., & Leat, D. (1995). Initial teacher trainees and their views of teaching and learning. *Teaching and Teacher Education*, *11*(1), 23–31.
- Brookhart, S. M., & Freeman, D. J. (1992). Characteristics of entering teacher candidates. *Review of Educational Research*, *62*, 37–60.
- Brouwer, C. N. (2010). Determining long term effects of teacher education. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education* (Vol. 7, pp. 503–510). Amsterdam: Elsevier.
- Burton, L. (2001). Fables: the tortoise? The hare? The mathematically underachieving male? In B. Atweh, H. Forgasz, & B. Nebres (Eds.), *Sociocultural research on mathematics education—an international perspective* (pp. 379–392). Mahwah: Erlbaum.
- Calderhead, J. (1991). The nature and growth of knowledge in student teaching. *Teaching and Teacher Education*, *7*(56), 531–535.
- Cochran-Smith, M. & Zeichner, K. M. (Eds.) (2005). *Studying teacher education. The report of the AERA Panel on Research and Teacher Education*. Mahwah: Erlbaum.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., York, R. L., et al. (1966). *Equality of educational opportunity*. Washington: US Gov Printing Office.

- Council of Chief State School Officers (1990). *School success for limited English proficient students. The challenge and state response*. Washington: Council of Chief State School Officers.
- Cummins, J. (1983). *Heritage language education. A literature review*. Toronto: Ministry of Education.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8(1), 1–50.
- De Maeyer, S., van den Bergh, H., Rymenans, R., & Van Petegem, P. (2010). Effectiveness criteria in school effectiveness studies: Further research on the choice for a multivariate model. *Educational Research Review*, 5, 81–96.
- Fan, X., Chen, M., & Matsumoto, A. R. (1997). Gender differences in mathematics achievement. Findings from the National Education Longitudinal Study of 1988. *Journal of Experimental Education*, 65, 229–242.
- Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.
- Goldhaber, D., & Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources*, 32(3), 505–523.
- Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality and Social Psychology*, 52, 890–898.
- Henrion, C. (1997). *Women in mathematics. The addition of difference*. Bloomington: Indiana University Press.
- Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah: Erlbaum.
- Hsieh, F.-J., Wang, T.-Y., Hsieh, Ch.-J., Tang, S.-J., Chao, G., Law, Ch.-K., Shy, H. Y., et al. (2010). *A milestone of an international study in Taiwan teacher education—an international comparison of Taiwan mathematics teacher preparation (Taiwan TEDS-M 2008)*. Taiwan: National Research Center of Taiwan Teacher Education and Development Study in Mathematics.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Diversity: Gender similarities characterize math performance. *Science*, 321(5888), 494–495.
- James, L. R., Demaree, R. G., & Wolf, G. (1993). rWG: an assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78, 306–309.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815–882.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica: RAND.
- McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, 17(3), 305–322.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46.
- Mueller, Ch. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development*, 52, 13–30.
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 International Mathematics Report. Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill: Boston College.
- NCTM (National Council of Teachers of Mathematics) (1991). *Professional standards for teaching mathematics (ENC-001451)*. Reston: NCTM.
- Oser, F., Biedermann, H., Brühwiler, Ch., Kopp, M., Krattenmacher, S., & Steinmann, S. (2010). *Deutschschweizer Lehrerausbildung auf dem Prüfstand. Wie gut werden unsere angehenden Lehrpersonen ausgebildet? Ein internationaler Vergleich*. Fribourg: University of Fribourg.
- Ryan, R. M., & Deci, E. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54–67.
- Scheerens, J. (2000). *Improving school effectiveness*. Paris: UNESCO/IIIEP.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Schmidt, W., Houang, R., Cogan, L., Blömeke, S., Tatto, M. T., Hsieh, F. J., Paine, L., et al. (2008). Opportunity to learn in the preparation of mathematics teachers: its structure and how it varies

- across six countries. *ZDM—The International Journal on Mathematics Education*, 40(5), 735–747.
- Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., & Wiley, D. E. (1997). *Many visions, many aims: A cross-national investigation of curricular intentions in school mathematics*. Dordrecht: Kluwer Academic.
- Schmidt, W. H., Blömeke, S., & Tatto, M. T. (2011). *Teacher Education Matters. A Study of The Mathematics Teacher Preparation from Six Countries*. New York: Teacher College Press.
- Schütte, M., & Kaiser, G. (2011) Equity and the quality of the language used in mathematics education. In B. Atweh, M. Graven, W. Secada, & P. Valero (Eds.) *Mapping Equity and Quality in Mathematics Education*. New York: Springer.
- Shulman, L. (1985). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 3–36). New York: Macmillan Co.
- Shulman, L. (2005). Signature pedagogies. <http://www.hub.mspnet.org/index.cfm/11172>.
- Simmons, P. (1995). Gender, ethnic and socio-ethnic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years. *British Educational Research Journal*, 21, 465–483.
- Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *Journal of Educational Research*, 95(6), 323–332.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research 1990–2000. *Review of Educational Research*, 75(3), 417–453.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis*. London: Sage.
- Taconis, R., van der Plas, P., & van der Sanden, J. (2004). The development of professional competencies by educational assistants in school-based teacher education. *European Journal of Teacher Education*, 27, 215–240.
- Tamir, P. (1988). Subject matter and related pedagogical knowledge in teacher education. *Teaching and Teacher Education*, 4(2), 99–110.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing: Michigan State University Press.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2013, forthcoming). *Teacher education and development study in mathematics (TEDS-M): Technical handbook*. Amsterdam: IEA.
- Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary school effectiveness. *Research Papers in Education*, 11, 5–33.
- Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Stability and consistency in secondary schools' effects on students' GSCE outcomes over 3 years. *School Effectiveness and Improvement*, 8(2), 169–197.
- Thomas, W. P., & Collier, V. P. (1997). *School effectiveness for language minority students*. Washington: National Clearinghouse for English Language Acquisition.
- Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula* (Vol. 1). Oxford: Pergamon.
- UNESCO (1997). *International standard classification of education—ISCED*. Paris: UNESCO.
- van Ewijk, R., & Slegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5, 134–150.
- Walter, O., & Taskinen, P. (2007). Kompetenzen und bildungsrelevante Einstellungen von Jugendlichen mit Migrationshintergrund in Deutschland. Ein Vergleich mit ausgewählten OECD-Staaten. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme, R. Pekrun et al. (Eds.), *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 337–366). Munich: Waxmann.

- Watt, H. M. G., & Richardson, P. W. (2007). Motivational factors influencing teaching as a career choice: Development and validation of the FIT-Choice Scale. *Journal of Experimental Education*, 75(3), 167–202.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies*, Göttingen: Hogrefe & Huber (pp. 45–66).
- Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, 6, 49–78.
- Wilson, S., Floden, R., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education*, 53(3), 190–204.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. (2007). ACER ConQuest 2.0: General item response modelling software [computer program manual]. Camberwell, Australia: Australian Council for Educational Research.