



Mathematical content knowledge and knowledge for teaching: exploring their distinguishability and contribution to student learning

Charalambos Y. Charalambous¹ · Heather C. Hill² · Mark J. Chin^{2,3} · Daniel McGinn^{2,4}

Published online: 12 August 2019
© Springer Nature B.V. 2019

Abstract

During the last three decades, scholars have proposed several conceptual structures to represent teacher knowledge. A common denominator in this work is the assumption that disciplinary knowledge and the knowledge needed for teaching are distinct. However, empirical findings on the distinguishability of these two knowledge components, and their relationship with student outcomes, are mixed. In this replication and extension study, we explore these issues, drawing on evidence from a multi-year study of over 200 fourth- and fifth-grade US teachers. Exploratory and confirmatory factor analyses of these data suggested a single dimension for teacher knowledge. Value-added models predicting student test outcomes on both state tests and a test with cognitively challenging tasks revealed that teacher knowledge positively predicts student achievement gains. We consider the implications of these findings for teacher selection and education.

Keywords Content knowledge · Dimensionality · Teacher knowledge · Mathematics · Student learning · Teaching-specific knowledge

Introduction

There is wide consensus among scholars, teacher educators, and policymakers that teacher knowledge is an important asset in teachers' toolkits. Despite this consensus, the composition of this knowledge and its role in supporting student learning has long been debated.

✉ Charalambos Y. Charalambous
cycharal@ucy.ac.cy

¹ Department of Education, University of Cyprus, Theophanides Building, Rm 508, 11-13 Dramas St, 1077 Nicosia, Cyprus

² Harvard Graduate School of Education, Gutman Library, Room 445, 6 Appian Way, Cambridge, MA 02138, USA

³ Harvard University, Cambridge, MA, USA

⁴ Center for Education Policy Research at Harvard University, 4th Floor, 50 Church St, Cambridge, MA 02138, USA

Drawing on Shulman and colleagues' work (Shulman 1986, 1987; Wilson et al. 1987), scholars in mathematics education have proposed different conceptual structures for this knowledge (Ball et al. 2008; Grossman 1990; McCrory et al. 2012; Rowland et al. 2005). These conceptual structures often separate pure content knowledge (CK) from the knowledge that facilitates teaching. However, some scholars (Bernarz and Proulx 2009; Huillet 2009) doubt whether CK is separable from the knowledge needed for teaching, and empirical work on the distinguishability of these two components provides mixed findings. In fact, a systematic review of 60 research papers focusing on pedagogical content knowledge (PCK, Depaepe et al. 2013) surfaced disagreements among scholars about whether CK and PCK are distinct or intertwined. Perhaps equally inconclusive is the empirical evidence on the predictive validity of teacher knowledge on student learning. Despite the accumulated body of research examining this issue, the magnitude of the contribution of teacher knowledge to student learning is still unclear.

To contribute to the ongoing discussions about the structure of teacher knowledge and to help illuminate the role of teacher knowledge in student learning, this paper asks: Is teachers' CK distinguishable from the knowledge needed for teaching, or do these two types of knowledge compose a unidimensional construct? Further, how do these two knowledge components contribute to student learning, as measured by student achievement gains on standardized tests?

Answering these questions will shed light on two important issues. From a theoretical perspective, more empirical evidence is needed to understand the nature of teacher knowledge. Conceptualizations of teacher knowledge advanced thus far—from Shulman's work (1986) to more recent ones—suggest multidimensionality, yet empirical evidence supporting this dimensionality is inconsistent at best. From a practical standpoint, determining the contribution of teacher knowledge to student learning might inform ongoing efforts to improve teacher selection and education [cf. Committee on the Study of Teacher Preparation Program in the United States 2010; National Mathematics Advisory Panel (NMAP), 2008].

We structure the remainder of this paper into five sections. In the second section, we review theoretical work delineating the structure of teacher knowledge in mathematics, the focal subject of this paper, then review empirical attempts to examine this structure, and examine studies exploring the association of teacher knowledge to student learning. In this section, we also outline the ways in which the present study complements and extends prior work. After outlining our research questions ([Theoretical perspectives on the knowledge needed for teaching](#)), we then detail the methods we pursued to empirically explore the dimensionality of teacher knowledge and its association to student learning ([Empirical findings on content knowledge and knowledge needed for teaching](#)). In the last two sections, we present the findings of this exploration and discuss their implications for teacher selection and education.

Background

Theoretical perspectives on the knowledge needed for teaching

Interest in the structure of teacher knowledge is not new. Almost two and a half thousand years ago, Aristotle distinguished between a *knower*, who simply learns and possesses the subject matter, and a *master*, who not only knows the content, but can also teach it.

However, only in the mid-1980s did the distinction between content knowledge and the knowledge needed for teaching come to the forefront of scholarly research. In his seminal work, Shulman (1986) defined CK in terms of both substance and syntax of the discipline itself. As he described, the teacher

need not only understand that something is so [i.e., substance]; the teacher must further understand why it is so, on what grounds its warrant can be asserted, and under what circumstances our belief in its justification can be weakened and even denied [i.e., syntax]. (p. 9)

Shulman also hypothesized that CK alone would not suffice for the work of teaching. Instead, he and others (Ball and McDiarmid 1990; Grossman 1990; Shulman 1986, 1987; Wilson et al. 1987; Wilson and Wineburg 1988) held that an accomplished teacher also holds general pedagogical knowledge (e.g., classroom management techniques and strategies), knowledge of learners and their characteristics, knowledge of educational contexts, curriculum knowledge, and pedagogical content knowledge (PCK). Defined as the “special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding” (Shulman 1986, p. 8), PCK includes an understanding of different forms of representations, analogies, illustrations, examples, explanations, and demonstrations that the teacher can employ while teaching. It also encompasses knowledge of why specific topics are easy or difficult to learn, and what (mis)conceptions students might hold.

Shulman’s work catalyzed the thinking of many scholars who, in turn, created a robust field of inquiry around what teachers know and how they think about specific content (see, e.g., Lannin et al. 2013; Mitchell et al. 2014; Remillard and Kim 2017; Sleep 2012; Steele et al. 2013). In accordance with Shulman’s views, most authors make a clear distinction between CK and additional knowledge needed for teaching. For example, scholars working on the Teacher Education and Development Study in Mathematics (TEDS-M) make a distinction between mathematics content knowledge and mathematics PCK, with the former pertaining to fundamental definitions, concepts, and procedures within mathematics, and the latter including knowledge of how to represent content to students, and specifically, taking into account their prior knowledge and difficulties (Blömeke et al. 2011). In Rowland and colleagues’ *Knowledge Quartet* (Rowland et al. 2005; Rowland et al. 2009), CK (named “overt subject knowledge”) appears as one of 17 elements composing the four dimensions of this framework. While this component pertains to the “foundation” dimension, three other dimensions—transformation, connections, and contingency knowledge—capture knowledge needed for teaching.

A clear distinction between these two types of knowledge is also reflected in the *Knowledge for Algebra for Teaching* framework (McCrary et al. 2012). Here, two types of CK are proposed: CK equivalent to the content taught in middle and high school (knowledge of school algebra), and more advanced CK, related to calculus and abstract algebra. The elements related to the knowledge needed for teaching in this framework include knowledge of typical student errors, canonical uses of school mathematics, and knowledge of curriculum trajectories. Knowledge for teaching is also reflected in the latter two components of Tchoshanov’s (2011) work, which includes knowledge of facts and procedures (which corresponds to CK), as well as knowledge of concepts and connections and knowledge of models and generalizations.

In Ball and colleagues’ work on *Mathematical Knowledge for Teaching* (MKT) (Ball et al. 2008), common content knowledge (CCK)—or the mathematical knowledge commonly found in settings other than teaching—is one of six separate domains of teacher knowledge.

The other five domains include specialized content knowledge (SCK; wholly mathematical, but specialized to the work of teaching), horizon content knowledge (HCK; knowledge of topics on the academic horizon), knowledge of how students learn the content (knowledge of content and students; KCS), knowledge of how best to teach specific materials (knowledge of content and teaching; KCT), and knowledge of content and the curriculum (KCC). As explained below, in this article we largely draw on a specific type of CCK—advanced common content knowledge—and SCK and KCT, as knowledge needed for teaching.

Although many academics view CK and the knowledge needed for teaching as theoretically discrete elements, not all agree with this distinction. Drawing on Chevallard's anthropological theory of didactics, Huillet (2009), for example, argues that the distinction between CK and PCK is artificial and that these two components cannot be separated from each other in practice. Similarly, Bernarz and Proulx (2009) argue that teachers' decisions at any given point in time are concurrently informed by mathematical, didactical, and pedagogical considerations, thus making it hard to distinguish between disciplinary-based content knowledge and the content knowledge needed for teaching. Thus, for these scholars, "at the heart of a teacher's practice [one] can find a very specific knowledge, composed of *intertwined* [...] dimensions" (p. 14, emphasis added).

In sum, with some notable exceptions, most theoretical models of teacher knowledge suggest that it is multidimensional. We now turn to empirical investigations into this issue.

Empirical findings on content knowledge and knowledge needed for teaching

A number of studies have empirically evaluated claims that teachers' knowledge contains multiple, separable elements. In most cases, authors investigated such claims by developing survey instruments measuring CK and some aspect of the knowledge needed for teaching, administering the instrument to a large number of teachers and then subjecting the resulting data to factor analyses, structural equation modeling, or other analytic techniques designed to determine dimensionality. In our review of these studies, we first start with those that capitalized on the distinction between CK and PCK proposed by Shulman and then continue with other studies based on the MKT conceptualization, the conceptualization considered here.¹

Krauss and colleagues' (2008, 2013) work represents perhaps the earliest and most systematic attempt to explore the relationship between CK and PCK. This group administered a test including CK and PCK items to a representative sample of 10th-grade mathematics teachers in Germany ($N=198$). A subsequent structural equation model showed that the data were best fit by a model representing the CK and PCK factors separately; these factors were, however, highly correlated ($r=0.79$). Interestingly, when the same analysis was run separately for two subgroups—Gymnasium teachers (i.e., academically oriented teachers who had received intensive training in CK) and non-Gymnasium (i.e., general track) teachers—the two factors were more distinguishable for the non-academic track group. For academic-track teachers, the CK-to-PCK correlation was $r=0.96$; in the non-academic track, the correlation was $r=0.61$. This led the authors to speculate that the distinguishability of the two constructs might be a function of the teachers' level of expertise.

Evidence on the distinguishability between CK and PCK also comes from the TEDS-M study (Blömeke et al. 2011), which collected data from about 13,000 preservice primary and secondary school teachers from 15 countries. The analysis of these data revealed that the two multidimensional models tested (i.e., with and without cross-loadings of items

¹ Our review of the literature yielded no studies examining the dimensionality of constructs other than CK-PCK and MKT.

across knowledge factors) had better fit to the data than a unidimensional model, thus suggesting that CK and PCK represent two distinguishable factors. Interestingly, this structure was found consistently in all 15 countries. The two constructs in one of the multidimensional models tested were strongly related ($r=0.85$) with this correlation, however, ranging from $r=0.65$ (Georgia) to $r=0.97$ (Botswana).

Complementing the TEDS-M study, more recent work (Kleickmann et al. 2015) explored the dimensionality of teacher knowledge with in-service—rather than preservice—teachers in Germany and Taiwan. Using the same sample of 198 German teachers utilized in Krauss et al. (2008) and a stratified random sample of 209 Taiwanese mathematics teachers, these scholars provided evidence of a two-dimensional cross-culturally invariant structure of teacher knowledge. Like the previous study, the correlations between the two constructs varied but were generally moderate to marginally high, with $r=0.64$ and $r=0.79$ for Taiwanese and German teachers, respectively. Another study that utilized data from preservice teachers at different stages in their teacher education (Kleickmann et al. 2013) reported similar correlations between CK and PCK for first-year preservice teachers ($r=0.64$) and both third-year preservice teachers and prospective teachers during their induction period ($r=0.78$). In sum, the studies considered above suggest that although empirically discernible, CK and PCK are moderately or more strongly correlated, depending on the educational systems and the teacher populations examined.

Moving from studies that conceptualize teacher knowledge as CK and PCK, a group working at the University of Michigan (Ball et al. 2008; Hill et al. 2004) developed multiple-choice items reflecting two elements of MKT: CK and a construct originally identified as “knowledge of students and content.” Using data from almost 1500 K-5 teachers participating in professional development programs, these scholars experimented with different models to better understand the nature of the knowledge needed for teaching. Similar to the analyses above, their work revealed that teacher knowledge was multidimensional: besides a general factor identified as representing CK, two other factors also accounted for part of the variance in teachers’ answers. The first represented SCK and included items such as analyzing non-standard algorithms or procedures, using representations to illustrate mathematical ideas, and providing explanations. Answering these items required mathematics knowledge, but it was not mathematical knowledge that we expect non-teachers to know. The second factor comprised some of the knowledge of students and content items; other knowledge of students and content items loaded only on the main factor. In sum, the study results lend credence to the argument that knowledge for teaching includes elements beyond CK. However, the cross-loadings of some items led these scholars to speculate that, although there seem to be certain identifiable knowledge components, these might be imperfectly discerned from CK (see, for example, Schilling 2007).

More recently, nationally representative samples of middle school (Hill 2007) and elementary school (Hill 2010) mathematics teachers have not uncovered multiple interpretable factors of MKT. In both cases, items failed to load onto theoretically specified factors; hence, the author scored teacher responses based on a single-factor model. A secondary analysis of data collected from fourth- and fifth-grade teachers participating in the *Measures of Effective Teaching* project (Copur-Gencturk et al. 2018) also challenged the multidimensionality of MKT. The teachers’ answers to 38 items categorized as CCK, SCK, and PCK had better fit to a unidimensional model compared to a three-dimensional model. In fact, in the latter model, the SCK and PCK factors were practically indistinguishable ($r=0.96$), whereas the correlation of CCK with the other two factors was strong ($r=0.75$ in both cases)—thus further reinforcing the lack of distinguishability among the three knowledge components examined.

In all, while most theoretical models and empirical results suggest that teacher knowledge is multidimensional, the exact nature of this relationship remains an open issue for at least three reasons. First, Table 1, which provides a summary of the studies reviewed above, suggests that the existing evidence is less conclusive when considering additional factors: whereas most studies that drew on the CK and PCK conceptualization (upper panel of Table 1) provide evidence supporting the multidimensionality of teacher knowledge, studies drawing on the MKT conceptualization (lower panel of Table 1) show more inconsistent results. Second, even when multidimensionality appears, the strength of the relationship between the emerging factors (and therefore their discernibility) varies depending on the teacher population examined and the settings in which teachers' knowledge was examined. For MKT, results varied depending upon whether teachers' knowledge was assessed in professional development settings (as in Hill et al. 2004) versus more typical settings; for CK-PCK, whether the teachers assessed were preservice versus in-service, whether the teachers taught in academic or non-academic sectors, and across educational systems.

Finally, the presence or lack of multidimensionality may relate to the items that comprise these assessments. Assessments that use the CK and PCK conceptualization tend to ask straightforward, open-ended CK problems—e.g., finding the length of ribbon (Blömeke et al. 2011), or proving that $0.\bar{9}$ equals one (Krauss et al. 2008). PCK items in these sets ask teachers to solve these problems in multiple ways, to identify student misconceptions with the content, and to design tasks to help students learn. Assessment based on the MKT framework, by contrast, use the multiple-choice format and tend to include mostly CCK and SCK items (Copur-Gencturk et al. 2018; Hill 2007, 2010). Both specialized and common content items are set in classroom contexts, and thus difficult to distinguish. This subtlety of distinction may be one reason for the lack of dimensionality in later MKT papers.

Given these distinctions, we see the present work as a replication and extension study (cf. Coyne et al. 2013; White et al. 2014) on the dimensionality of MKT, one which retains the basic conditions of previous work—collecting data from MKT items and subjecting them to dimensionality analyses—while modifying other aspects to subject MKT theory to a stronger empirical test. In particular, unlike prior studies, we included CK items that are not set in classroom contexts, which is the typical approach pursued in the MKT studies reviewed above. Second, and more critically, we tapped teachers' advanced Common Content Knowledge (aCCK).² If knowledge of mathematics itself is dimensional in the way Ball and colleagues (2008) hypothesize, we argue that forming a CK scale from straightforward (i.e., not embedded in classroom contexts) yet mathematically advanced content might reveal that dimensionality³; to date, this has not been attempted by scholars. Third,

² Advanced Common Content Knowledge is distinctively different from Horizon Content Knowledge (HCK). The latter should not be equated to knowledge of the mathematics content beyond a teacher's current grade level, given that this conceptualization captures the students'—as opposed to the teachers'—horizon knowledge (see more on that in Zazkis and Mamolo 2011). This claim resonates with an elaborated definition of HCK, developed in collaboration with Ball and Bass, according to which “HCK is not about curricular development of [the] content;” rather it is an “orientation to, and familiarity with the discipline [...] that contribute to the teaching of the school subject at hand, providing teachers with a sense for how the content being taught is situated in and connected to the broader disciplinary territory” (Jakobsen et al. 2013, p. 3128).

³ Content knowledge items at teachers' grade level could be considered as prerequisites for teachers' PCK, given conceptualizations of PCK as the transformation of content knowledge into powerful forms of knowledge that are adaptive to student needs (cf. Mewborn 2003; NMAP 2008). By including content at higher grade levels, aCCK items were expected to not necessarily be prerequisites of PCK, and hence be more distinguishable from items reflecting PCK (i.e., SCK and KCT items).

Table 1 Summary of studies exploring the dimensionality of teacher knowledge

Study	Constructs ^a	Context	Teacher population		Sample	Analyses ^b	Multidimensionality	Correlations (multidimensional solution)
			Grade level	Prospective/practicing				
Blömeke et al. (2011)	CK, PCK	Botswana, Chile, Georgia, Germany, Malaysia, Norway, Philippines, Poland, Russia, Singapore, Spain, Switzerland, Taiwan, Thailand, USA	Primary and secondary	Prospective (final year of studies)	N = 13,400	CFA	Yes	$r = 0.85$ (overall) (range: $r_{\text{GEORGIA}} = 0.65$ — $r_{\text{BOTSWANA}} = 0.97$)
Kleickmann et al. (2013)	CK, PCK	Germany	Secondary	Prospective (a) 1st year (b) 3rd year (c) During induction (d) Practicing	$N_a = 117$ $N_b = 126$ $N_c = 539$ $N_d = 198$	CFA	Yes	$r_a = 0.64$ $r_b = 0.78$ $r_c = 0.78$ $r_d = 0.79$
Kleickmann et al. (2015)	CK, PCK	Germany Taiwan	Secondary	Practicing	$N_{\text{GER}} = 198$ $N_{\text{TAW}} = 209$	CFA	Yes	$r_{\text{GER}} = 0.79$ $r_{\text{TAW}} = 0.64$
Krauss et al. (2008, 2013)	CK, PCK	Germany	Secondary (Gr. 10)	Practicing	N = 198	CFA	Yes	$r = 0.79$ (entire sample) $r = 0.96$ (academic track) $r = 0.61$ (non-academic track)

Table 1 (continued)

Study	Constructs ^a	Context	Teacher population		Sample	Analyses ^b	Multidimensionality	Correlations (multidimensional solution)
			Grade level	Prospective/practicing				
Copur-Gençtürk et al. (2018)	CCK, SCK, PCK	USA	Primary (Gr. 4–5)	Practicing	N = 397	CFA	No	$r_{\text{SCK-PCK}} = 0.96$ $r_{\text{CCK-SCK}} = 0.75$ $r_{\text{CCK-PCK}} = 0.75$
Hill et al. (2004)	CK, SCK, KCS	USA	Primary (K-5)	Practicing	N = 1552	EFA, CFA	Yes (with some cross-loadings)	–
Hill (2007)	CCK, SCK	USA	Middle grades	Practicing	N = 636	EFA, CFA	No	$r_{\text{CCK-SCK}} = 0.79$
Hill (2010)	CCK, SCK, KCS, KCT	USA	Primary (Gr. 1–6)	Practicing	N = 625	EFA, CFA	No (cross-loadings)	–

^aCK, content knowledge; PCK, pedagogical content knowledge; CCK, common content knowledge; SCK, specialized content knowledge; KCS, knowledge of content and students; KCT, knowledge of content and teaching

^bCFA, Confirmatory factor analyses (including bifactor analyses); EFA, exploratory factor analyses

although the present study replicates Copur-Gencturk and colleagues' (2018) work in certain respects (e.g., sampling a similar teacher population), it departs from it in significant respects. In particular, the current study includes a more diverse and numerous set of CK items compared to those used in the former study, in which CK was measured by only five items, all tapping teachers' knowledge of the use of the equal sign. Additionally, the current study utilized a different set of items than those in the former study. Because the results drawn from any dimensionality study are unavoidably item dependent, carrying out replication studies like the present one can help explore the generalizability of the findings across a range of items.

We argue that replication and extension studies on the MKT dimensionality are warranted because, as Table 1 reveals, although studies building on the CK-PCK conceptualization converge in their results, a similar convergence is not observed in studies drawing on the MKT conceptualization. We concur with Copur-Gencturk et al. (2018) that this discrepancy might be due to the lack of clarity in the MKT literature as to how MKT is conceptualized, operationalized, and measured, especially when gauging MKT at different grade levels (see Speer et al. 2014 for an elaborated discussion). Therefore, replication studies on MKT dimensionality which utilize different operationalization and/or measurement approaches might offer insights at several fronts, pushing MKT scholars to better clarify the conceptualization of this construct as well as to rethink ways of operationalizing it.

Teacher knowledge and student learning

Scholarly focus on teacher knowledge has often been motivated by an implicit (e.g., Shulman 1986) or more explicit (e.g., Cohen et al. 2003) assumption that this construct contributes to student learning, often through differences in instructional quality. Recent years have seen concerted efforts toward understanding whether and how teacher knowledge relates to student learning. Some studies have done so using indicators that comprise teachers' scores across knowledge dimensions, or that represent only one aspect of teachers' knowledge; other studies have aimed to identify the unique contribution of different knowledge components to student learning (see Table 2).⁴ We review each type in turn.

Single-construct studies

Hill et al. (2005) examined the extent to which MKT, operationalized as a unidimensional measure containing both content knowledge and knowledge-for-teaching items, contributes to student gain scores. Their study showed that students taught by elementary school teachers who scored 1 SD above the mean on their MKT assessment experienced gains in their test scores equivalent to one-half to two-thirds of a month of additional growth compared to their counterparts taught by average-MKT teachers. Similarly, Rockoff et al. (2011) showed MKT to be among the few significant predictors of student achievement among a sample of new elementary and middle-school mathematics teachers; in particular, one standard deviation (SD) difference in MKT was associated with 0.028 SD difference in student achievement when controlling for student prior achievement. A study that did not

⁴ We limit our review to studies that obtained actual measures of teachers' knowledge, instead of using proxies for this knowledge, such as teachers' credentials, number of courses taken, or degrees obtained (e.g., Monk 1994).

Table 2 Summary of studies exploring the association between teacher knowledge and student learning

Study	Grade level	N	Teacher knowledge	Dependent variable	Student assessment type	Model	Findings
Cantrell and Kane (2013)	Gr. 4–8	Students (S): not reported Teachers (T): ≈ 3000	Content knowledge for Teaching (CKT)	Student achievement	(a) State tests (b) Assessments of higher-order thinking skills	Not specified	Nonsignificant effects (no statistics reported)
Hill et al. (2005)	Gr. 1, Gr. 3	S: 1190 Gr. 1 and 1773 Gr. 3; T: 334 Gr. 1 and 365 Gr. 3	Content Knowledge for Teaching (CKT)	Student gain scores	Terra Nova test	Three-level mixed model with student and teacher background covariates, including student prior achievement	Gr. 1: $b = 2.12$ (SE = 1.00) Gr. 3: $b = 1.96$ (SE = 0.77)
Kersting et al. (2012)	Gr. 5–7	S: 591 T: 36	MKT on fractions	Student gain scores	Short fraction quiz based on released items from State tests	Two-level model (a) with and (b) without covariates for instructional quality	Nonsignificant effects: (a) $\beta = -0.06$ (b) $\beta = -0.16$
Metzler and Woessmann (2010)	Gr. 6	S: 6461 T: 500	CK	Student achievement	National test (Peru)	Regressions jointly estimated for mathematics and reading with student, teacher, and school background covariates, including student prior achievement	$b = 0.094$ (SE = 0.019)
Ottmar et al. (2015)	Gr. 3	S: 1533 T: 88	MKT	Student achievement	State test	Two-level multi-group path analysis with covariates for student characteristics, student prior achievement, and instructional quality	Nonsignificant effects: Control group: $\beta = -0.15$ Intervention group: $\beta = 0.08$

Table 2 (continued)

Study	Grade level	N	Teacher knowledge	Dependent variable	Student assessment type	Model	Findings
Rockoff et al. (2011)	Gr. 4–8	S: not reported T: 418	MKT	Student achievement	State test	Three-level model with covariates for student and school characteristics, including prior student achievement	$b = 0.028$ (SE = 0.012)
Shechtman et al. (2010)	Gr. 7–8	S: 2669 Gr. 7 and 825 Gr. 8 T: 125 Gr. 7 and 56 Gr. 8	MKT	Student gain scores	Author developed test based on items from NAEP, TIMSS, and State tests	Two-level models (M1, no covariates; M2: student prior achievement; M3, student prior achievement *MKT)	Significant effects: M2 for the experimental group of Gr. 7: $b = 0.13$ (SE = 0.05)
Baumert et al. (2010)	Gr. 10	S: 4353 T: 181	CK PCK	Student achievement	Test covering the content stipulated in the federal states' curricula	Two-level model with student background covariates and student prior achievement (student level); track membership and instructional variables (class level)	CK: direct effect: $\beta = 0.30$; PCK: direct effect: $\beta = 0.42$

Table 2 (continued)

Study	Grade level	N	Teacher knowledge	Dependent variable	Student assessment type	Model	Findings
Campbell et al. (2014)	Gr. 4–8	S: 6413 Gr. 4–5 and 10,890 Gr. 6–8 T: 259 Gr. 4–5 and 184 Gr. 6–8	CK PCK	Student achievement	State tests	Two-level model with student background and student prior achievement covariates at the student level; teacher background and beliefs/disposition covariates at the teacher level	CK: Gr. 4–5: $b=0.071$ (SE=0.033); Gr. 6–8: $b=0.1166$ (SE=0.042); PCK: Gr. 4–5: $b=0.047$ (SE=0.038); Gr. 6–8: $b=0.1180$ (SE=0.039); CK and PCK: Gr. 4–5: $b=0.071$ (SE=0.034); Gr. 6–8: $b=0.1184$ (SE=0.042); Pearson's r (a) $r=0.06$, $p=0.54$ (b) $r=0.26$, $p=0.009$ (c) $r=0.02$, $p=0.85$
Tchoshanov (2011)	Gr. 6–8	S: ≈ 2400 T: 102	Knowledge of: (a) facts and procedures (b) concepts and connections (c) models and generalizations	Students' passing rates in state tests	State tests	Correlations	

use the original MKT measures but developed MKT-like items (Shechtman et al. 2010), however, found that in only one of the three models examined did teacher knowledge significantly contribute to student gain scores, and that was only for one of the experimental groups under consideration. In addition, several other studies found no association between teachers' MKT and student learning. For example, the *Measures of Effective Teaching* study (Cantrell and Kane 2013) and Kersting and colleagues' (2012) work reported nonsignificant associations between MKT and student gain scores. Similarly, in their randomized controlled trial, Ottmar and colleagues (2015), found no direct or indirect effects of MKT on student achievement in third grade when controlling for student achievement in second grade in either the experimental or the control group. However, the latter two studies employed significantly smaller teacher sample sizes compared to those in the studies reviewed above, which might have made it hard to detect the small effect sizes identified in Hill et al. (2005) or Rockoff et al. (2011).

Other studies focused only on teachers' CK. For example, Metzler and Woessmann (2010) estimated the causal effect of teachers' CK on students' academic achievement. Their analysis showed that a one SD increase in teacher CK increased student achievement by about 0.1 SD. This, according to the authors, implied that net of other factors, two students, one taught by a teacher at the median of the CK distribution and the other by a teacher at the 5th percentile of this distribution, would differ by 0.17SD in their achievement by the end of the school year. Similar results appear in the educational production function literature (e.g., Harbison and Hanushek 1992; Mullens et al. 1996). Given the correlations between CK and other forms of teacher knowledge found above, it is not clear in any of the above studies whether CK or knowledge for teaching is responsible for the correlations uncovered.

Multiple-construct studies

In only a handful of studies did researchers compare the contribution of different teacher knowledge dimensions to student learning. Baumert and colleagues (2010), for example, investigated the relative contribution of two knowledge dimensions, CK and PCK, to student learning. Using a representative sample of over 4300 Grade-10 students and their 181 teachers, and controlling for student achievement in Grade 9, they found CK to be less predictive of student achievement at the end of Grade 10 than PCK ($\beta=0.30$ and $\beta=0.42$, respectively). Also, the former explained a smaller proportion of the classroom-level variance compared to the latter (44% vs. 54%, respectively), leading the authors to conclude that PCK comprises an indispensable component in a teacher's toolkit, for "it makes the greatest contribution to explaining student progress" (p. 168). A similar finding arose in Tchoshanov's (2011) work, which showed secondary teachers' knowledge of concepts and connections—reflecting PCK in Shulman's terms—to have a small association with student passing rates on a standardized test ($r=0.26$), whereas teachers' knowledge of facts and procedures (CK) was not associated with these passing rates ($r=-0.06$). A more recent study (Campbell et al. 2014) utilizing data from both upper-elementary and middle-school US teachers found that for the first sample only CK had a positive effect on student achievement after controlling for student- and teacher-level characteristics: an increase of one SD in teacher CK was associated with an increase of 0.071 SDs in students' achievement. For the latter sample, both CK and PCK had positive and larger effects on student achievement; in particular, an increase of one SD in either type of knowledge was associated with an

increase of 0.22 SDs in student achievement without any controls and 0.16 and 0.18 SDs, respectively, with controls.

The preceding literature review suggests that, despite the wealth of studies on teacher knowledge and its contribution to student outcomes, uncertainty remains over this issue, especially with regards to elementary grades. Thus, in addition to replicating and extending prior studies investigating MKT dimensionality, the present study makes two other contributions. First, by concurrently attending to dimensionality and predictive validity, this study addresses a limitation of prior works (e.g., Kersting et al. 2012; Rockoff et al. 2011; Shechtman et al. 2010) in which the MKT structure was implicitly taken for granted and attention was directed only to predictive validity. Second, recently accumulated empirical evidence suggests that the test used to measure student learning plays a role in the predictive-validity conclusions drawn (see, for example, Grossman et al. 2014; Naumann et al. 2017; Papay 2011). With the exception of only one study (Cantrell and Kane 2013) that utilized two tests—a state test and a test examining higher-order thinking skills—all other studies reviewed above used only one test at any given data collection round to measure student achievement (gains). In these cases, student learning was measured using state tests (Campbell et al. 2014; Kersting et al. 2012; Ottmar et al. 2015; Rockoff et al. 2011; Tcho-shanov 2011), national tests (Metzler and Woessmann 2010), researcher-developed tests following state standards (Baumert et al. 2010; Shechtman et al. 2010) or with a commercially developed standardized test (Hill et al. 2005). Hence, there seems to be a need to concurrently draw on different tests to explore the predictive validity of teacher knowledge on student learning, especially given that state tests might be capturing more basic mathematics knowledge instead of also tapping into students' reasoning and higher-level thinking. In this study, we do so by employing two different tests, as discussed below.

Research questions

Drawing on data from upper-elementary teachers and their students, we ask:

1. Is teacher knowledge multidimensional, as advanced by different theoretical frameworks and as supported by some empirical evidence, or does it comprise a single construct?
2. To what degree does teacher knowledge (or components thereof) predict student learning, as measured by student achievement gains on two different types of tests?

Methods

To address these questions, we drew on data from the National Center for Teacher Effectiveness project. This project contained three sources of data: a teacher survey designed to explicitly assess the dimensionality of teacher knowledge, a project-administered student assessment, and district administrative data that included both student state test scores and background characteristics. In this section, we describe the instruments used to capture teachers' knowledge, the study participants, and the data collection and analysis procedures.

Instrumentation

Two teacher surveys, administered roughly one year apart (Y1 and Y2), carried the content knowledge items used in this study. We selected roughly half of these items to tap into teachers' aCCK; these items were sampled from the released items of the elementary *Massachusetts Test for Educator Licensure* (MTEL), an assessment known for its mathematical rigor and which contained content knowledge taught at both the upper-elementary grades and above (for the purposes of this study, we only used items corresponding to at least middle-grade levels). We drew a second set of items from the Michigan elementary MKT forms to represent knowledge needed for teaching, and, in particular, SCK and KCT. We illustrate the difference between these knowledge types by describing two CK items (Figs. 1, 2) and two knowledge-for-teaching items (Figs. 3, 4).

The first CK (drawn from the MTEL) item asks teachers to decide which among four different options represents a fraction that is equal to a given mathematical expression. To choose the correct answer (Option B), one simply needs substantive mathematical knowledge: specifically, that a^{-n} is equal to $1/a^n$; that a^n is equal to multiplying the base a n times; and that $a^n * b^n = (ab)^n$. The second item (also drawn from the MTEL) captures syntactic aspects of content knowledge. One possible path for correctly solving this item (i.e., selecting Option C) would be through noticing and analyzing patterns. In this particular instance, noticing that the first figure corresponds to 1 cube, the second to $1+2$ cubes, the third to $1+2+3$ cubes, and so on, can lead to the conclusion that the n^{th} figure includes $1+2+\dots+n$ cubes. More generally, these items tend to ask teachers to compute or to set up computations, to solve word problems, and to identify mathematical facts and definitions. These items pertained to place value with very large numbers; calculating sale price, tax and tip; rotation and reflection; linear functions; prime factorization of natural numbers; proportional reasoning; and exponents and scientific notation.

Items 3 and 4, in contrast, require more than CK. To answer the third item, one needs KCT to see that Option C lends itself to employing multiple unique strategies when comparing fractions (e.g., comparing fractions smaller or larger than $\frac{1}{2}$; comparing fractions with the same numerator). Similarly, in the fourth item, knowing that two different shapes cannot be used to illustrate the given multiplication unless they have the same area, which is not the case in Option C, requires SCK. Other knowledge-for-teaching items pertained to alternative algorithms for common operations, representing operations such as decimal multiplication, and providing mathematical explanations for rules and procedures (e.g., why dividing $4/2$ by $2/2$ does not change the value of the original quantity).

In total, the surveys included 72 multiple-choice items,⁵ 12 of which were presented as "testlets"; these testlets included three to five sub-items, all preceded by a common stem. As we explain below, we treated each testlet as a single item with an ordinal score equal to the number of sub-items answered correctly. A panel of four experts, all developers of MKT items (researchers in mathematics education who also worked as mathematics teachers and/or teacher educators in mathematics education) content-validated the items by assessing whether they captured any of the MKT domains; for items identified as CCK,

⁵ Although we recognize the possibility of answering an item correctly just by mere guessing or test-taking skills, a validation study (Hill et al. 2007) showed low rates of strategic test-taking and guessing, especially for the content-knowledge items (around 5% of the items taken). To the extent that such low rates were also true for the current study, the effect of guessing and test-taking skills could be thought to be minimal, especially for the aCCK items (which were fewer than the SCK/KCT items).

The expression $(5^{-8} \cdot 7^{-9})$ is equal to which of the following? (Circle ONE answer.)

(a) $\frac{1}{5(35)^8}$

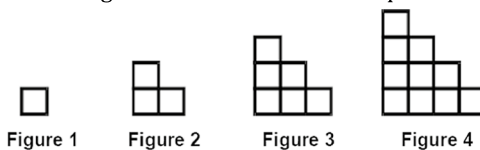
(b) $\frac{1}{7(35)^8}$

(c) $\frac{5}{(35)^8}$

(d) $\frac{7}{(35)^8}$

Fig. 1 A released MTEL item intended to capture teachers' aCCK of exponents

Use the diagram below to answer the question that follows.



If the pattern continues, how many more small squares are in figure 100 than are in figure 99? (Circle ONE answer.)

(a) 98

(b) 99

(c) 100

(d) 101

Fig. 2 A released MTEL item intended to capture teachers' aCCK knowledge of patterns

these experts were also asked to identify whether they captured at-grade level CCK or advanced CCK. After dropping three items that were either too easy or too hard (i.e., the lack of variation across teachers would impact factor analyses), we used the 36 items that at least three panel experts rated as aCCK or SCK/KCT (we collapsed the last two categories into one to reflect knowledge needed for teaching).⁶ Of these, panelists anticipated that 13 items would capture aCCK (6 from Y1 and 7 from Y2) and 23 items would reflect SCK/KCT (12 from Y1 and 11 from Y2).⁷

⁶ The present study was part of a larger study aimed at developing a strong Item Response Theory (IRT) measure of teacher knowledge. Therefore, the expert panel content validated the items for the purposes of the current study *after* they had been administered.

⁷ That only half of the items originally designed were eventually used in the study is telling of the limitations of any replication study, which is highly dependent on the items used; it is also telling of the difficulties inherent in designing items that can be unambiguously categorized in certain MKT domains.

Mr. Foster’s class is learning to compare and order fractions. While his students know how to compare fractions using common denominators, Mr. Foster also wants them to develop a variety of other intuitive methods.

Which of the following lists of fractions would be best for helping students learn to develop several different strategies for comparing fractions? (Circle ONE answer.)

(a) $\frac{1}{4}$ $\frac{1}{20}$ $\frac{1}{19}$ $\frac{1}{2}$ $\frac{1}{10}$

(b) $\frac{4}{13}$ $\frac{3}{11}$ $\frac{6}{20}$ $\frac{1}{3}$ $\frac{2}{5}$

(c) $\frac{5}{6}$ $\frac{3}{8}$ $\frac{2}{3}$ $\frac{3}{7}$ $\frac{1}{12}$

(d) Any of these would work equally well for this purpose.

Fig. 3 A released item gauging teachers’ SCK

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.)

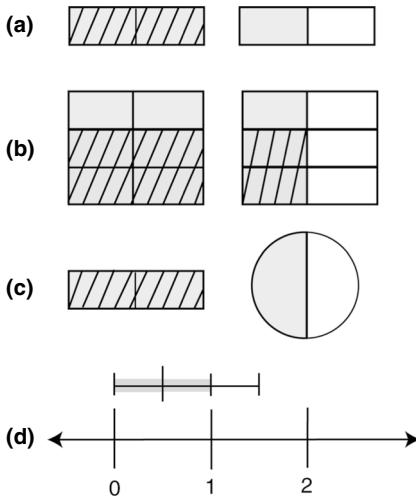


Fig. 4 A released item gauging teachers’ KCT

Participants and data collection processes

The project administered surveys to 263 fourth- and fifth-grade teachers in Y1, and 219 fourth- and fifth-grade teachers in Y2; teachers returned 247 and 214 completed surveys each year, respectively. The participants were working in four different districts in three Eastern US states. Most were White (67%) or African-American (23%) and about four out

of five were female (83%). At the time of the first survey administration, teachers had on average about 10 years of experience, with the youngest in her first year of teaching, and the most seasoned teacher having taught for 37 years. The teachers completed each survey at the beginning of each school year and received a small stipend for doing so.

Students of these teachers completed two types of tests. The first was their state-mandated assessment administered toward the end of each school year and used for school and district accountability purposes. We also administered a project-developed test twice per year (October–November and March–May); this test was written to require more complex reasoning and responses than those typically expected in traditional multiple-choice state tests (for more information on differences between the three state tests and the project test, see Lynch et al. 2017).

We did not include all students and classrooms in our analyses exploring the relationship between knowledge and student achievement gains. Of 476 classrooms (i.e., unique teacher-year observations) and 10,019 student-year observations, our final main analysis sample included 434 classrooms (teacher $n=287$) and 7890 student-year observations. We excluded classrooms that were either atypical in some way (i.e., primarily composed of special education students, classroom $n=16$, student $n=105$; classrooms with fewer than five students, classroom $n=1$, student $n=3$) or whose teacher did not have a knowledge score (classroom $n=15$; student $n=294$). The average sample classroom was composed of 18 students. We excluded students who either skipped or repeated a test grade level ($n=47$) or lacked the needed variables to estimate achievement gains (i.e., current-year test scores, $n=702$, and prior-year test scores, $n=978$). Given that a significant number of students were missing information on the dependent variable or a key control variable, we deemed it more appropriate to exclude them from our analytic sample rather than impute data for them (for a similar approach, see Chetty et al. 2014). Our final main sample of students was largely non-white (41% African–American; 24% Hispanic), and eligible for subsidized lunch (65%). Smaller proportion of students in this sample were special education students (11%) or English language learners (21%). Excluded students differed slightly from the full sample in that these students were more likely to be African-American (56%) and eligible for subsidized lunch (73%). Excluded students were also more likely to be special education students (23%) and to have lower prior achievement on the state (-0.24) and the alternative mathematics test (-0.33). (For a full set of student-level descriptives for in-sample and out-of-sample students, see “Appendix” Table 9). In discussing the study findings, we acknowledge that our analytic sample is not perfectly representative of students in these schools.

Data analyses

Data analysis unfolded in two steps, with each step corresponding to one of the two research questions. To determine the structure of teacher knowledge, we used exploratory and confirmatory factor analyses. As noted above, we decided to treat the data as ordinal rather than nominal to accommodate several testlets. We did so because items belonging to testlets are not independent—as verified by some initial explorations—and would therefore form factors reflecting the testlet to which they belonged. Given the decision to treat the testlet items as ordinal, we employed the WLSM estimator for all the confirmatory factor analysis models discussed below. Additionally, because some teachers’ responses were missing for either Y1 or Y2, we conducted analyses within year rather than across years.

Cognizant of the mixed and inconclusive evidence of prior dimensionality studies on MKT (e.g., Hill et al. 2004; Hill 2007, 2010), our investigation of the structure of teachers' knowledge began with a set of exploratory factor analyses with Geomin rotation (Hattori et al. 2017).⁸ These analyses provided initial insights into the structure of the data. We also employed exploratory factor analysis because it could help better understand any discrepancies—should these arise—in fit between our data and the theoretical models explored in confirmatory factor analyses. In the context of the latter analyses, we experimented with two types of models (see Fig. 5).⁹ The first model (M1), the most parsimonious one, assumed that within each year all items loaded onto a single factor representing a general teacher knowledge factor. The second model (M2) included two first-order factors, one encompassing all the aCCK items, and another including the SCK/KCT items. These two factors were assumed to load onto a second-order factor representing general mathematical knowledge for teaching. In running these models, we were cognizant of the fact that our sample size was relatively small, a limitation we revisit when interpreting the study findings.

We then followed the recommended structure of teachers' knowledge from the factor analyses to estimate knowledge scores for teachers. We modeled teacher survey responses to all aCCK and SCK/KCT items in either Y1 or Y2 using a generalized partial credit item response theory (IRT) model, and then predicted empirical Bayes means of the latent knowledge variables, our within-year knowledge scores, for each teacher. Using these IRT scores, we answered our second research question investigating the relationship of teacher knowledge and student learning, measured by student achievement gains on standardized tests, by estimating the following equation separately in Y1 and Y2¹⁰:

$$Y_{jkmgst} = \alpha P_{jgt-1} + \pi D_{jt} + \rho C_{kmt} + \delta S_{gst} + \sigma_g + \chi_d + \mu_{mt} + \varepsilon_{jkmgst}$$

$$\mu_{mt} = \psi TK_{mt} + \tau_{mt}$$

where the outcome of interest, Y_{jkmgst} , represents student j 's standardized score on either the (1) state mathematics exam, or, (2) the project-administered mathematics test¹¹ in either time $t=Y1$ or $t=Y2$. P_{jgt-1} represents a vector of prior achievement for student j on the state or alternative mathematics exam for the tested grade g at time $t-1$. This vector includes both a quadratic and cubic function for prior math achievement, in addition to student j 's prior achievement in an alternate subject, English Language Arts. D_{jt} represents

⁸ Geomin rotation, an oblique rotation method, is the default rotation method in MPLUS. This method was chosen because it considers the resulting factors as non-orthogonal—which resonates with our assumption about a nonzero relationship between the knowledge factors under consideration. Additionally, parameter estimates obtained with Geomin rotation have been found to be comparable with those yielded from confirmatory factor analysis and to be unbiased (cf. Hattori et al. 2017). Moreover, an appraisal of different rotation methods for dichotomous data—the data used in the current study—showed this method to compare equally well with other oblique rotation methods (cf. Finch 2011).

⁹ We also considered bifactor models (Chen et al. 2012) to examine whether more specific factors exist above and beyond a general teacher knowledge factor; however, these models failed to converge.

¹⁰ On average, teachers with data in both Y1 and Y2 ($n=160$) exhibited a slight decrease (seven percentage points) in knowledge over time, based on changes in the percentage of survey knowledge items they responded to correctly. Furthermore, teacher self-reports of mathematics-specific professional development were uncorrelated to these changes (results available upon request). Nevertheless, we opted to run this model separately across years to account for the potential of knowledge growth, and because we include different teachers in different years based on participation in the study.

¹¹ State test scores were scaled to have a mean of zero and a *SD* of one within grade and district (due to different tests across states). Project test scores were scaled to have a mean of zero and a *SD* of one within grade.

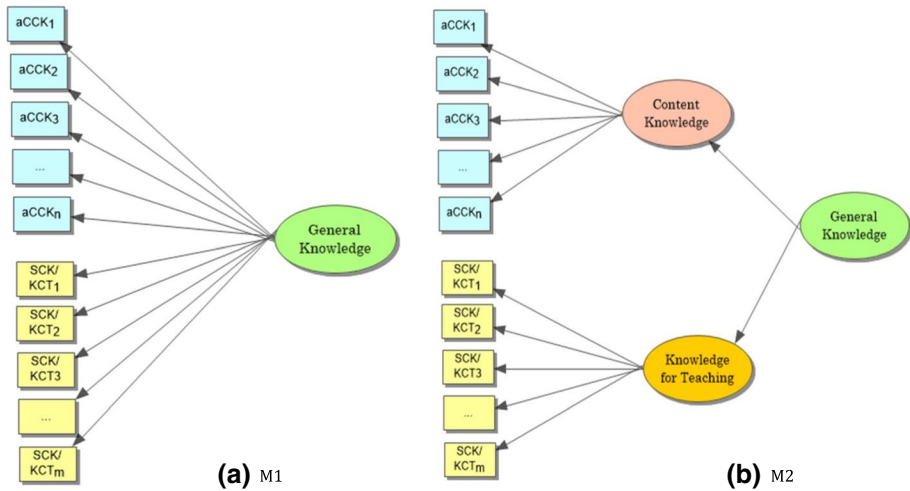


Fig. 5 Different CFA models **a** single-factor model (M1), **b** two-factor model (M2) (aCCK: advanced common content knowledge, SCK/KCT: specialized content knowledge and knowledge of content and teaching)

a vector of student demographic variables for student j at time t , including race, gender, free- or reduced-price lunch eligibility, special education designation, and level of limited English proficiency. C_{kmt} represents the aggregates of these two vectors for students in class k under teacher m at time t . S_{gst} represents the aggregates of these two vectors for students in grade g in school s at time t . Finally, two sets of fixed effects were included in the model: σ_g represented the effect of grade g (to account for potential differences in standardized tests across grades) and χ_d represented the effect of district or school.

To account for the multilevel structure of the data, where students are nested within teachers, we included one random component of variance, μ_{mt} . This component, μ_{mt} , can be decomposed into two parts: TK_{mt} , teacher m 's knowledge score(s) in time t (standardized to have a mean of zero and SD of one for ease of interpretation), and τ_{mt} , a random effect for teacher m in time t used to represent the component of teacher t 's impact on the outcome variable in either Y1 or Y2, or his or her within-year 'value-added' score (after accounting for teacher knowledge). Thus, the coefficient of interest in answering our second research question is ψ , which represents the relationship between teacher knowledge and student outcomes.

In sensitivity analyses, we included other classroom-level variables to test the extent to which the relationship between teacher knowledge and student achievement gains may be masking the contribution of other important factors for student learning. We focused on variables available in our dataset, in particular, ones constructed from responses to the same survey that contained the knowledge items. Specifically, we explored whether inclusion of teacher self-reports about the climate of their classroom (e.g., students getting along, frequency of teacher reprimands of students and student misbehavior, teacher–student rapport), their own effort preparing for class (e.g., time spent grading, gathering lesson materials, reviewing lesson content), and experience influenced the importance of teacher knowledge for student outcomes. We chose these potential confounders, in particular, because they were factors that had demonstrated significant relationships to student achievement in other work (e.g., Lavy 2009; Papay and Kraft 2015; Pianta et al. 2008) and because they were likely to influence student learning through channels other than changing teachers'

mathematical knowledge. Classroom climate comprised teacher responses to nine items and was sufficiently reliable ($\alpha = .88$); effort comprised teacher responses to five items and was also fairly reliable ($\alpha = .73$). We standardized both variables to have a mean of zero and SD of one for ease of interpretation. Finally, we included teacher experience as a dichotomous variable that flagged whether a teacher had fewer than 2 years of experience teaching mathematics.

Results

We organize the results of the study around the two research questions.

The structure of teacher knowledge

Exploratory factor analyses of the teacher survey suggested that a single-factor solution was preferable to a two-factor solution. This was implied by two pieces of evidence. First, for both years under exploration, the eigenvalue of the first extracted factor was at least twice as big as that of the second extracted factor (Y1: first factor=5.04, second factor=1.54; Y2: first factor=4.45, second factor=2.22). Ratios of this magnitude support a unidimensional rather than a multidimensional construct (cf. Kline 1994). Second, when a two-factor solution was fit to the data (Table 3) one item (SCK/KCT₇, italicized) of the 18 from Y1 had similar loadings on both factors; the highest factor loadings for another item (aCCK₅, with a loading of 0.25, underlined) was lower than 0.40, typically the lowest recommended threshold for acceptable factor loadings (cf. Field 2013; Kline 1994); and two items (aCCK₃ and aCCK₄, bold) loaded on the SCK/KCT rather than the aCCK factor, as had been expected by the expert panel. This situation was even worse for Y2. Out of 18 items, two items had similar loadings on two factors (SCK/KCT₁₉ and SCK/KCT₂₃, italicized), one had a highest loading lower than 0.40 (SCK/KCT₁₄, underlined), and five loaded onto a different factor than that theoretically assumed (in bold).¹² Together with the eigenvalue ratios, these results strongly suggest a one-factor solution.

We next turned to confirmatory factor analyses and experimented with the models shown in Fig. 5. Because for Model 2 we either had problems of convergence or the resulting matrix had non-positive variances and therefore could not be considered reliable, we mostly examined the correlation of the two first-order factors, instead of trying to form a second-order factor, as illustrated by M2 in Fig. 5. As Table 4 shows, for both Y1 and Y2, a single-factor and a two-factor solution had comparable fit to the data: they both fit the data marginally well, given that their Bentler Comparative Fit Index (CFI) was close to 0.90 (0.89 in Y1 and 0.88 in Y2 for both models) and even the upper bound of the 90% confidence interval of the root-mean-square error of approximation (RMSEA) index was close to 0.06 (0.067 for Y1 and 0.064 in Y2 for both models; see Hu and Bentler 1999; Kline

¹² Similar results were obtained when two other non-orthogonal rotation methods (Equamax and Quartimin) were applied. For example, in the Y1 two-factor solution with Quartimin rotation, one item had similar loadings on both factors; the highest factor loading for another item was lower than 0.40, and two items loaded onto a different factor than that theoretically assumed. For Y2, two items had similar loadings on two factors and seven loaded onto a different factor than that expected.

Table 3 Factor loadings of the teacher knowledge Items in a two-factor solution

Year 1 ^{a,b,c}			Year 2 ^{a,b,c}		
Items	F_1	F_2	Items	F_1	F_2
aCCK ₁	0.59	0.00	aCCK ₇	0.13	0.41
aCCK ₂	0.70	0.00	aCCK ₈	0.18	0.44
aCCK ₃	-0.10	0.55	aCCK ₉	0.49	-0.09
aCCK ₄	-0.18	0.40	aCCK ₁₀	0.64	0.03
aCCK ₅	<u>0.25</u>	<u>0.03</u>	aCCK ₁₁	0.17	0.60
aCCK ₆	0.50	0.06	aCCK ₁₂	0.37	0.50
			aCCK ₁₃	0.55	0.00
SCK/KCT ₁	0.01	0.44	SCK/KCT ₁₃	-0.54	0.41
SCK/KCT ₂	0.15	0.48	SCK/KCT ₁₄	<u>-0.20</u>	<u>0.37</u>
SCK/KCT ₃	-0.06	0.63	SCK/KCT ₁₅	-0.03	0.46
SCK/KCT ₄	0.10	0.57	SCK/KCT ₁₆	-0.10	0.40
SCK/KCT ₅	0.06	0.48	SCK/KCT ₁₇	0.12	0.54
SCK/KCT ₆	-0.20	0.58	SCK/KCT ₁₈	0.26	0.57
SCK/KCT ₇	<i>0.24</i>	<i>0.30</i>	SCK/KCT ₁₉	<i>0.30</i>	<i>0.28</i>
SCK/KCT ₈	0.10	0.56	SCK/KCT ₂₀	-0.02	0.67
SCK/KCT ₉	0.00	0.63	SCK/KCT ₂₁	-0.06	0.75
SCK/KCT ₁₀	0.07	0.49	SCK/KCT ₂₂	0.00	0.43
SCK/KCT ₁₁	0.01	0.44	SCK/KCT ₂₃	<i>0.23</i>	<i>0.29</i>
SCK/KCT ₁₂	0.15	0.48			

^aItems with factor-loading differences less than 0.10 are in italics

^bItems with their highest factor loading being lower than 0.40 are underlined

^cItems loading on a different factor than that originally theorized are presented in bold

2011).¹³ The comparison of the two models, however, implied that the more parsimonious single-factor solution ought to be preferred over the more complex two-factor solution, given that in both years, the ratio of the difference in the Chi-square values between the two models over their degrees of freedom (see lower panel of Table 4) was not statistically significant.

Tables 5 and 6, which present the factor loadings of the items utilized in our analysis for either a single- or two-factor solutions provide additional evidence in favor of the former over the latter. In particular, for both years, the factor loadings of the two-factor solution are very close to those of the single-factor solution (see Columns 4 and 7), with the exception of item aCCK₂ in Y1, which had a marginally acceptable loading of 0.40 in the two-factor solution and a loading lower than 0.40 in a single-factor solution (in all other cases, the factor loadings were consistently either below or above 0.40, regardless of the solution considered). Even more critically, the correlations of the two factors in the two-factor solutions were very strong ($r_{Y1}=0.85$ and $r_{Y2}=0.89$) suggesting that the two factors cannot be easily distinguished from one another. Strictly speaking, both solutions—regardless of the

¹³ The *Chi-square* test for the model was violated for both years (given that its *p* value was lower than 0.05), but this criterion is sensitive to the size of the analytic sample.

Table 4 Values of fit statistics for a single-factor solution and a two-factor solution

Index	Year 1		Year 2	
	$M1^a$	$M2^b$	$M1^a$	$M2^b$
χ^2_M	228.02	225.69	205.51	203.25
df_M	135	134	135	134
p	0.001	0.001	0.001	0.001
CFI	0.89	0.89	0.88	0.88
RMSEA	0.055	0.055	0.051	0.050
(90% CI)	(0.042–0.067)	(0.042–0.067)	(0.036–0.064)	(0.036–0.064)
χ^2_B	1001.13	1001.13	734.04	734.04
df_B	153	153	153	153
p	0.0001	0.001	0.001	0.001
$\Delta\chi^2_M$	2.33		2.26	
Δdf_M	1		1	
$\Delta\chi^2_M/\Delta df_M$	2.33		2.26	
p	0.12		0.13	

^a $M1$ (Model 1) as presented in Fig. 5

^b $M2$ (Model 2) pertained to correlating the two first-order factors, instead of also taking a second-order factor, as shown in Fig. 5

year of consideration—have limitations, given that in addition to marginally meeting the CFI and RMSEA index thresholds, several of their loadings were below the recommended minimum loading of 0.40.¹⁴ In *relative* terms, however, a single-factor solution seems more preferable than a two-factor solution.

Teacher knowledge and student learning

We followed the recommended structure of teachers' knowledge from the factor analysis results and estimated a generalized partial credit IRT model where teacher performance on all aCCK and SCK/KCT items in either Y1 or Y2 loaded onto a single latent knowledge factor.¹⁵ The year-specific predicted scores on this latent factor were then submitted to multilevel regression analysis predicting two measures of student learning: student achievement gains on the state standardized mathematics and project-administered tests.

Findings across the two outcomes and the 2 years were consistent: teachers' mathematical knowledge, based on their performance on both aCCK and SCK/KCT items together, were associated with mathematics score gains (see Table 7). Specifically, we observed that, in comparison with being taught by a teacher of average MKT, a student taught by a

¹⁴ We avoided dropping any items to improve the model fit or the factor loadings because any modifications made to the CFA models need to be theoretically justified (Kline 2011). Additionally, dropping items could have led to construct underrepresentation, a key limitation in exploring construct validity (cf. AERA/APA/NCME 2014).

¹⁵ Non-significant or marginally significant but negligible correlations were found between teacher knowledge and either teachers' effort or teaching experience (effort: $r_{Y1} = -0.13$, $p = 0.05$; $r_{Y2} = -0.11$, $p = 0.10$; experience: $r_{Y1} = -0.04$, $p = 0.58$; $r_{Y2} = -0.10$, $p = 0.17$).

Table 5 Factor loadings and factor correlation for a single-factor and a two-factor solution (year 1)

Parameter	M1 ^a			M2 ^b		
	Factor loadings			Factor loadings ^c		
	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
aCCK ₁	0.57	0.15	0.30	1.00 ^b	–	<i>0.34</i>
aCCK ₂	0.67	0.15	0.35	1.17	0.35	<i>0.40</i>
aCCK ₃	0.94	0.17	0.49	1.61	0.45	<i>0.55</i>
aCCK ₄	0.58	0.18	0.30	0.95	0.37	<i>0.32</i>
aCCK ₅	0.33	0.15	0.17	0.58	0.29	<i>0.20</i>
aCCK ₆	0.63	0.20	0.32	1.13	0.38	<i>0.39</i>
SCK/KCT ₁	1.00 ^d	–	0.52	1.00 ^d	–	0.52
SCK/KCT ₂	1.48	0.22	0.77	1.48	0.22	0.77
SCK/KCT ₃	0.84	0.19	0.44	0.84	0.19	0.44
SCK/KCT ₄	1.08	0.18	0.56	1.08	0.18	0.56
SCK/KCT ₅	1.15	0.18	0.59	1.15	0.18	0.60
SCK/KCT ₆	1.20	0.20	0.62	1.20	0.20	0.62
SCK/KCT ₇	0.98	0.21	0.51	0.98	0.21	0.51
SCK/KCT ₈	0.86	0.14	0.45	0.86	0.14	0.45
SCK/KCT ₉	0.82	0.21	0.43	0.82	0.21	0.43
SCK/KCT ₁₀	1.18	0.17	0.61	1.18	0.17	0.61
SCK/KCT ₁₁	1.21	0.18	0.63	1.21	0.18	0.63
SCK/KCT ₁₂	1.00	0.20	0.52	1.00	0.20	0.52
				Correlation		
aCCK–SCK/KCT				0.15	0.04	0.85

^aM1 (Model 1) as presented in Fig. 5

^bM2 (Model 2) pertained to correlating the two first-order factors, instead of also taking a second-order factor, as shown in Fig. 5

^cIn M1 all items load onto the same factor; in M2, the aCCK items load onto one factor (standardized loadings shown in italics) and the SCK/KCT load onto a different factor (standardized loadings shown in bold)

^dNot tested for statistical significance. All other unstandardized estimates are statistically significant at $p < 0.05$

teacher with a knowledge score 1 SD above average had higher gains of 0.042 SD on the state mathematics test ($SE = 0.019$, $p < 0.05$) and 0.048-SD gains on the project-administered mathematics test ($SE = 0.018$, $p < 0.01$) in Y1. Similarly, in Y2, being taught by a teacher with a knowledge score 1-SD above average was associated with higher student achievement gains of 0.046 SD on the state mathematics test ($SE = 0.019$, $p < 0.05$) and 0.036-SD gains on the project-administered mathematics test ($SE = 0.016$, $p < 0.05$) compared to a student taught by an average-MKT teacher. Omitting district fixed effects in favor of school fixed effects, which may help account for potential sorting of more/less knowledgeable teachers to specific schools but would thus be a more conservative test of our predictor, saw our knowledge measure remain marginally significant in most specifications. Notably, the point estimates for the relationship between teacher knowledge and student achievement gains across tests and years in these models were similar to those from prior work (i.e., Rockoff et al. 2011) and ranged from 0.028 to 0.051.

Table 6 Factor loadings and factor correlation for a single-factor and a two-factor solution (year 2)

Parameter	M1 ^a			M2 ^b		
	Factor loadings			Factor loadings ^c		
	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
aCCK ₇	1.87	1.39	0.36	1.00 ^b	–	<i>0.37</i>
aCCK ₈	2.30	1.61	0.45	1.24	0.43	<i>0.46</i>
aCCK ₉	2.54	1.71	0.49	1.38	0.39	<i>0.52</i>
aCCK ₁₀	1.33	1.07	0.26	0.73	0.35	<i>0.27</i>
aCCK ₁₁	3.34	2.23	0.65	1.82	0.53	<i>0.68</i>
aCCK ₁₂	3.12	2.13	0.60	1.71	0.51	<i>0.64</i>
aCCK ₁₃	0.99	0.86	0.19	0.55	0.29	<i>0.21</i>
SCK/KCT ₁₃	1.00 ^d	–	0.19	1.00 ^d	–	0.20
SCK/KCT ₁₄	1.48	1.08	0.29	1.44	1.03	0.29
SCK/KCT ₁₅	2.25	1.51	0.44	2.20	1.43	0.44
SCK/KCT ₁₆	1.78	1.20	0.35	1.74	1.14	0.35
SCK/KCT ₁₇	2.96	2.09	0.57	2.89	1.98	0.58
SCK/KCT ₁₈	3.35	2.27	0.65	3.26	2.15	0.66
SCK/KCT ₁₉	1.96	1.40	0.38	1.89	1.32	0.38
SCK/KCT ₂₀	3.27	2.19	0.63	3.25	2.11	0.65
SCK/KCT ₂₁	3.63	2.46	0.70	3.56	2.33	0.72
SCK/KCT ₂₂	2.18	1.61	0.42	2.13	1.53	0.43
SCK/KCT ₂₃	0.52	0.63	0.10	0.47	0.60	0.10
				Correlation		
aCCK–SCK/KCT				0.07	0.05	0.89

^aM1 (Model 1) as presented in Fig. 5

^bM2 (Model 2) pertained to correlating the two first-order factors, instead of also taking a second-order factor, as shown in Fig. 5

^cIn M1 all items load onto the same factor; in M2, the aCCK items load onto one factor (standardized loadings shown in italics) and the SCK/KCT load onto a different factor (standardized loadings shown in bold)

^dNot tested for statistical significance. All other unstandardized estimates are statistically significant at $p < 0.05$

In Table 7, we also provide estimates of the effect size of teachers’ knowledge using Cohen’s f^2 . We estimate Cohen’s f^2 by assessing the change in the amount of teacher-level variance in student achievement explained after including the teacher-level knowledge measure. We conclude across models that the effect size of our measure of teachers’ knowledge is, in almost all cases, small ($0.15 \geq f^2 > 0.02$).

To put the magnitude of these results in context, we compare them to the difference in performance between students who were and were not eligible for free- or reduced-price lunch on the state (0.05 SD) and project-based tests (0.04 SD). Being taught by a teacher whose knowledge was above the average teacher knowledge by 1 SD is associated with test score growth that largely offsets this difference. Finally, as shown in Table 8, when controls for the classroom climate, teacher effort, and teacher experience were added in the model, results largely remained the same for both years under consideration both with and without school fixed effects.

Table 7 Predicting student mathematics assessment achievement

	State test ^a		Project test ^a	
	District FE	School FE	District FE	School FE
<i>Panel A: Year 1</i>				
Teacher knowledge	0.042* (0.019)	0.051* (0.021)	0.048** (0.018)	0.036~ (0.019)
Cohen's f^2	0.03	0.06	0.05	0.03
<i>N</i> teachers	230	230	230	230
<i>N</i> students	4201	4201	4201	4201
<i>Panel A: Year 2</i>				
Teacher knowledge	0.046* (0.019)	0.028 (0.021)	0.036* (0.016)	0.037~ (0.020)
Cohen's f^2	0.03	0.01	0.04	0.04
<i>N</i> teachers	204	204	204	204
<i>N</i> students	3689	3689	3689	3689

^aAll models include controls for student prior achievement, demographics, peer and school aggregates, grade indicators, and random effects for teacher. Standard errors reported in parentheses

~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 8 Predicting student mathematics assessment achievement—sensitivity analyses

	State test ^a		Project test ^a	
	District FE	School FE	District FE	School FE
<i>Panel A: Year 1</i>				
Teacher knowledge	0.049** (0.018)	0.054** (0.020)	0.049** (0.018)	0.036~ (0.019)
Controls for climate, effort, and experience?	Y	Y	Y	Y
<i>N</i> teachers	226	226	226	226
<i>N</i> students	4154	4154	4154	4154
<i>Panel A: Year 2</i>				
Teacher knowledge	0.049** (0.019)	0.025 (0.021)	0.040* (0.016)	0.040* (0.020)
Controls for climate, effort, and experience?	Y	Y	Y	Y
<i>N</i> teachers	197	197	197	197
<i>N</i> students	3577	3577	3577	3577

^aAll models include controls for student prior achievement, demographics, peer and school aggregates, grade indicators, and random effects for teacher. Standard errors reported in parentheses

~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Discussion and conclusions

When considering the nature of teacher knowledge, many scholars distinguish between pure content knowledge and knowledge needed for the work of teaching. However, empirical work has yielded mixed and inconclusive findings regarding this distinction. Responding to recent recommendations (Cai et al. 2018; Makel and Plucker 2014), this work replicates earlier analyses, but also extends and complements these efforts in several respects.

First, unlike prior studies on MKT, we utilized items that captured advanced CCK (CCK above grade level) and were not framed within teaching scenarios. As such, we increased the theoretical likelihood of finding multidimensionality, effectively searching for confirmation of the knowledge structure posited by both MKT (Ball et al. 2008) and other scholars of teacher knowledge (e.g., Rowland et al. 2005). Second, we used a more diverse set of CCK items than that typically found in prior work, including Copur-Gencturk et al. (2018), which used a small number of items, and papers by Hill (2007, 2010) that used CCK and SCK items written in classroom contexts. Third, in this work, we explored both dimensionality and predictive validity, which are seldom reported concurrently in a single study. Fourth, we utilized two different types of tests to explore the predictive validity of teacher knowledge, recognizing that the test may matter for the conclusions we draw. Finally, we examined the contribution of teacher knowledge to student learning controlling for a variety of student, teacher, and classroom indicators, which is rather rare in prior pertinent studies.

A key limitation of our work pertains to the relatively small analytic sample utilized, which might have led some of our more complex models to fail to converge and which might partly account for the fact that our models were slightly below the acceptable thresholds for CFA. For example, some scholars (e.g., Tate 2002) suggest that when using categorical data, three to five subjects per correlation might be required. Despite this limitation, we think that these findings can still provide important insights into the construct of teacher knowledge, especially if one takes into consideration the small analytic sample used in prior studies as well (see Table 1). At the same time, we acknowledge that our analytic sample is not perfectly representative of students in these schools; because teachers may be more (or less) effective with specific populations, our estimates may vary from what they would be in the full population. Although we have no a priori reason to expect student characteristics to mediate MKT impacts on student outcomes, this is a topic for future research.

Challenging common assumptions, this study suggests that teacher knowledge—at least with respect to the two components investigated¹⁶ and as operationalized and measured herein—comprises a single dimension. The results of both the exploratory and the confirmatory analyses suggest this finding, in that both showed a single-factor solution for Y1 or Y2 to be more preferable than a two-factor solution. In making this argument, we recognize the limitations of the single-factor solution, including marginally acceptable fit indices and low loadings for certain items. Relatively speaking, however, this solution was better than the two-factor solution for three reasons. First, the loadings of the exploratory factor analyses for several items did not make any theoretical sense, since they departed from patterns expected based on the expert panel. Second, the confirmatory factor analyses implied that the more parsimonious model should be preferred. Third, these analyses also showed that the correlations between the two factors in the more complex solution were very strong, suggesting that the two factors were hardly distinguishable.

While consistent with prior work that also identified a unidimensional structure for MKT (Copur-Gencturk et al. 2018; Hill 2007, 2010), our results do not align with the MKT conceptualization (Ball et al. 2008), in which teacher knowledge comprises different distinct domains. Two explanations, one theoretical and one methodological, could account

¹⁶ We did not study knowledge components beyond those related to the content to be taught, such as teachers' general pedagogical knowledge. Including such components might have led to a multidimensional solution, as suggested in a recent study (Blömeke et al. 2016).

for this discrepancy, both of which require systematic future exploration. Theoretically speaking, the MKT knowledge domains might not be that distinguishable, a possibility that Ball and colleagues themselves (*ibid.*, p. 404) acknowledge as a “boundary” problem. Hence, although distinguishing these domains might serve other purposes (e.g., teacher education), unlike the apparently clear-cut distinction between CK and PCK in Shulman’s conceptualization, the MKT domains might be difficult to disentangle, at least at the elementary level. Methodologically speaking, the study results could also be an artifact of the multiple-choice format used to capture teacher knowledge. This format might better lend itself to capturing more static aspects of teachers’ knowledge, thus failing to tap into teachers’ knowledge-in-use, which is how MKT has originally been theorized. This argument is supported by both studies reporting on the difficulties in writing multiple-choice items that capture knowledge-in-use (e.g., Herbst and Kosko 2014; Hill et al. 2008) and studies documenting the potential of more dynamic approaches that engage teachers in analysis of actual (e.g., Kersting et al. 2012) or simulated (e.g., Charalambous 2008) teaching practice in capturing this type of knowledge. This latter explanation surfaces another study limitation that pertains to the fact that we measured more static aspects of teacher knowledge instead of exploring more dynamic aspects that relate to engaging teachers in certain mathematical practices.

Regardless of the interpretation given, the study findings stand in contrast to prior empirical work on CK and PCK (e.g., Blömeke et al. 2011; Kleickmann et al. 2013, 2015; Krauss et al. 2008). In fact, as Table 1 suggests, almost all studies drawing on the CK-PCK distinction provided dimensionality evidence, whereas studies utilizing the MKT conceptualization largely failed to do so. One possibility is that the CK-PCK dimensions are more distinct from one another than the CCK and SCK/KCT domains, which are closely related in the MKT conceptualization (see Ball et al. 2008). However, in CK/PCK studies, the distinguishability of the knowledge components may also be a function of several contextual parameters, including item format, level of expertise in the responding teacher population, experiences with professional development, training in CK (e.g., teachers from academic vs. non-academic sectors) and the educational system in which they serve—as our review of the literature suggested. Future work could reach stronger conclusions by systematically varying either the teacher population under consideration and/or other contextual factors (see Table 1) but holding constant the knowledge conceptualizations and operationalizations they utilize. Such studies could also employ more complex research designs compared to the correlational and cross-sectional designs; for example, research on expert and novice teachers or longitudinal studies following groups of teachers throughout different stages of their career could provide more insights into the nature of teachers’ knowledge, and most critically on the extent to which this nature changes as a function of teachers’ education, experience, and expertise.

At the same time, the study results seem to align with some conceptualizations on teacher knowledge (e.g., Bernarz and Proulx 2009; Huillet 2009) and some models of teaching proficiency (e.g., Kilpatrick et al. 2001) that view different facets of teacher knowledge as intertwined. Understanding why the MKT components—as operationalized and measured herein—appear to be intertwined, however, lies beyond the scope of the present study. Future studies that employ more qualitative approaches to explore teachers’ thought processes as they answer the study items might shed more light into this issue. Studies that compare the thinking processes of teacher and non-teacher populations could be particularly informative toward this direction, as they could help reveal the (re)sources upon which these populations draw when answering items from seemingly distinct knowledge types.

In addition to exploring the nature of teacher knowledge, in this study we also investigated the contribution of teacher knowledge to student learning. In line with prior studies documenting positive effects of teacher knowledge on student learning (e.g., Hill et al. 2005), this study showed associations ranging from 0.036 SD to 0.054 SD, depending on the model used (school vs. district fixed effects) and the type of test (state vs. project-based test) considered. Interestingly, these effects were of similar or larger magnitude than those found in similar prior studies on MKT (e.g., Rockoff et al. 2011). These effects are, of course, smaller in size than those reported in studies utilizing conceptualizations other than MKT (e.g., Baumert et al. 2010; Kersting et al. 2012).

This difference in magnitude might, however, be not only an issue of conceptualization but also an issue of modeling, since the latter studies either used almost no controls (e.g., Kersting et al. 2012) or controlled only for student-level background characteristics (e.g., Baumert et al. 2010). In contrast, in the present study we used standard econometric models, controlling for several student background and classroom-instruction characteristics but also including grade and school/district fixed effects. Indeed, we found that removing all student- and classroom-level controls from our multilevel models resulted in larger associations between teacher knowledge and student achievement (i.e., point estimates of 0.06 to 0.10, depending on the model, results not shown), though we note that even the models with uncontrolled-for associations are still fairly small given the full range of student achievement. Despite this small magnitude, our results do cast another vote in favor of the positive link between teacher knowledge and student learning, especially if one considers that teacher effects are typically small [i.e., two students taught by teachers of 1 SD apart typically exhibit learning differences of 0.10–0.15 SD, cf. Nye et al. (2004) and Rockoff (2004)] and that these effects derive from a variety of sources, including not only teacher knowledge but also the curriculum materials employed, curricular alignment with the assessment, teachers' ability to connect with and engage students, and so forth.

As such, this study carries several implications for teacher selection and education. First, it suggests selecting elementary teachers from among the more mathematically able whenever possible. In fact, recent changes in both policy and labor markets may have improved indicators of preservice teacher knowledge, including average SAT score and the academic rigor of teachers' undergraduate institution (Goldhaber and Walch 2014; Lankford et al. 2014). However, there remain additional checks during the recruitment and hiring process that could be implemented, including using mathematics certification test scores as a factor in hiring decisions.

Teacher education programs should also attend to subject-specific knowledge; in the USA, many alternative certification programs have moved to preparing teachers as generalists rather than specialists, potentially overlooking the knowledge critical to working with students in classrooms. Finally, teachers need ongoing opportunities to learn the knowledge they use in teaching—both common and otherwise—in settings that highlight the intertwined nature of that knowledge. In making this argument, we are, of course, cognizant of the fact that MKT is one among many factors that can help explain student learning gains and that other factors also need to be taken into consideration when trying to understand teacher effectiveness.

In this study, we also capitalized on two different types of tests to examine the predictive validity of teacher knowledge: the typical state tests often administered toward the end of the year, and an alternative, low-stakes but more cognitively demanding, test which required more reasoning on the part of the students (Lynch et al. 2017). We predicted that teachers' MKT might more strongly correlate with student achievement on the alternative test given empirical evidence suggesting a positive link between teachers' MKT

and the cognitive level at which mathematics is presented and enacted in the classroom (e.g., Charalambous 2010). Yet we found that the role of teacher knowledge was similar across tests even slightly favoring the state test rather than the alternative test. This result could (partly) be attributed to the fact that the state test is more high-stakes; hence, teachers might better mobilize their knowledge to support student learning in domains that are captured by this test, thus teaching to the test. In contrast, the effects of teacher knowledge on the latter test might be more difficult to detect not only because it is lower stakes, but also because developing student mathematical reasoning might be a long-term process. As such, this reasoning might be more difficult to capture in about 6 months (i.e., the time that elapsed in between the pre- and posttest administration of the alternative test, as opposed to the entire 9-month period captured by state tests). Relatedly, teachers in our sample accounted for relatively less variance in student achievement on the alternative test than the state test (i.e., 8 vs. 16%), potentially weakening relationships between the teacher-level measure of MKT and student learning.

We do not take this finding to contradict prior studies documenting that the test matters when exploring predictors of student learning (e.g., Grossman et al. 2014; Papay 2011). Rather, we interpret this finding as calling for better ways to measure teacher knowledge and student learning; more critically, the small effects found for both tests underline the need to carefully align the measures examining teacher knowledge, instructional quality, and student learning. From this respect, it appears productive to reformulate the questions we ask, from questions interrogating whether teacher knowledge matters for student learning to inquiries attending more to *how* teacher knowledge matters and for which particular types of student learning (cf. Charalambous and Pitta-Pantazi 2016). Qualitative studies which explore the mechanisms through which stronger knowledge can inform teaching and, through that, promote specific types of student learning will also be needed in the future.

The past three decades have seen significant theoretical and empirical work on understanding teacher knowledge and its different components, in essence by dissecting this construct into distinct parts. Given the knowledge that has been accumulated thus far in this domain, and provided that future studies empirically corroborate a unidimensional rather than a multidimensional character of this knowledge, scholarly attempts could be directed to bringing these different components together, seeking to understand their complex interrelationships and how they synergistically contribute to instructional quality and student learning.

Acknowledgements This research was funded by Grants IES R305C090023 and NSF 08314500. The authors would like to thank the teachers who participated in this study.

Appendix

See Table 9.

Table 9 Student sample descriptives

	Excluded	Included
Male	0.54	0.50
White	0.15	0.24
African-American	0.56	0.41
Hispanic	0.20	0.24
Asian	0.05	0.08
Other race/ethnicity	0.04	0.04
Eligible for free/reduced-price lunch	0.73	0.65
Special education	0.23	0.11
English language learner	0.20	0.21
Prior state mathematics test score	-0.24	0.09
Prior alternative mathematics test score	-0.33	-0.02
<i>n</i>	2011	7890

References

- AERA/APA/NCME. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Ball, D. L., & McDiarmid, G. W. (1990). The subject-matter preparation of teachers. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 437–449). New York: Macmillan.
- 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. <https://doi.org/10.1177/0022487108324554>.
- 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. <https://doi.org/10.3102/0002831209345157>.
- Bernarz, N., & Proulx, J. (2009). Knowing and using mathematics in teaching: Conceptual and epistemological clarifications. *For the Learning of Mathematics*, 29(3), 11–17.
- Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, 56, 35–46. <https://doi.org/10.1016/j.tate.2016.02.003>.
- Blömeke, S., Houang, R. T., & Suhl, U. (2011). *TEDS-M: Diagnosing teacher knowledge by applying multidimensional item response theory and multiple-group models*. IERI monograph series: *Issues and methodologies in large-scale assessments* (Vol. 4, pp. 109–129). Retrieved September 12, 2014 from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_04_Chapter_5.pdf.
- Cai, J., Morris, A., Hohensee, C., Hwang, S., Robinson, V., & Hiebert, J. (2018). The role of replication studies in educational research. *Journal for Research in Mathematics Education*, 49(1), 2–8.
- Campbell, P. F., Rust, A. H., Nishio, M., DePiper, J. N., Smith, T. M., Frank, T. J., et al. (2014). The relationship between teachers' mathematical content and pedagogical knowledge, teachers' perceptions, and student achievement. *Journal for Research in Mathematics Education*, 45(4), 419–459.
- Cantrell, S., & Kane, T. (2013). *Ensuring fair and reliable measures of effective teaching*. Bill & Melinda Gates Foundation. Retrieved September 15, 2014 from http://k12education.gatesfoundation.org/wp-content/uploads/2015/05/MET_Ensuring_Fair_and_Reliable_Measures_Practitioner_Brief.pdf.
- Charalambous, C. Y. (2008). *Preservice teachers' Mathematical Knowledge for Teaching and their performance in selected teaching practices: Exploring a complex relationship*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.
- Charalambous, C. Y. (2010). Mathematical knowledge for teaching and task unfolding: An exploratory study. *The Elementary School Journal*, 110(3), 247–278. <https://doi.org/10.1086/648978>.
- Charalambous, C. Y., & Pitta-Pantazi, D. (2016). Perspectives on priority mathematics Education: Unpacking and understanding a complex relationship linking teacher knowledge, teaching, and learning. In L. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (3rd ed., pp. 19–59). London: Routledge.

- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219–251. <https://doi.org/10.1111/j.1467-6494.2011.00739.x>.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632. <https://doi.org/10.1257/aer.104.9.2593>.
- Cohen, D., Raudenbush, S., & Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 1–24. <https://doi.org/10.3102/01623737025002119>.
- Committee on the Study of Teacher Preparation Programs in the United States. (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Research Council, the National Academies. Retrieved, June 5, 2013 from http://www.nap.edu/catalog.php?record_id=12882.
- Copur-Gencturk, Y., Tolar, T., Jacobson, E., & Fan, W. (2018). An empirical study of the dimensionality of the Mathematical Knowledge for Teaching construct. *Journal of Teacher Education*.. <https://doi.org/10.1177/0022487118761860>. (in press).
- Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., et al. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, *6*(1), 1–23. <https://doi.org/10.1080/19345747.2012.706694>.
- Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education*, *34*, 12–25. <https://doi.org/10.1016/j.tate.2013.03.001>.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage Publications.
- Finch, H. W. (2011). A comparison of factor rotation methods for dichotomous data. *Journal of Modern Applied Statistical Methods*, *10*(2), 549–570. <https://doi.org/10.22237/jmasm/1320120780>.
- Goldhaber, D., & Walch, J. (2014). Gains in teacher quality. *Education Next*, *14*(1). Retrieved September 7, 2017 from <http://educationnext.org/gains-in-teacher-quality/>.
- Grossman, P. L. (1990). A study in contrast: Sources of pedagogical content knowledge for secondary English. *Journal of Teacher Education*, *40*(5), 24–31.
- Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher*, *43*(6), 293–303. <https://doi.org/10.3102/0013189X14544542>.
- Harbison, R. W., & Hanushek, E. A. (1992). *Educational performance of the poor: Lessons from rural northeast Brazil*. New York: Oxford University Press.
- Hattori, M., Zhang, G., & Preacher, K. J. (2017). Multiple local solutions and Geomin rotation. *Multivariate Behavioral Research*, *52*(6), 720–731. <https://doi.org/10.1080/00273171.2017.1361312>.
- Herbst, P., & Kosko, K. W. (2014). Mathematical knowledge for teaching and its specificity to high school geometry instruction. In J. Lo, K. R. Leatham, & L. R. Van Zoest (Eds.), *Research trends in mathematics teacher education* (pp. 23–46). New York: Springer.
- 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*, 95–114. <https://doi.org/10.3102/0162373707301711>.
- Hill, H. C. (2010). The nature and predictors of elementary teachers' Mathematical Knowledge for Teaching. *Journal for Research in Mathematics Education*, *41*(5), 513–545.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, *39*(4), 372–400.
- Hill, H. C., Dean, C., & Goffney, I. M. (2007). Assessing elemental and structural validity: Data from teachers, non-teachers, and mathematicians. *Measurement: Interdisciplinary Research and Perspectives*, *5*(2–3), 81–92.
- Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, *42*, 371–406. <https://doi.org/10.3102/00028312042002371>.
- Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal*, *105*(1), 11–30. <https://doi.org/10.1086/428763>.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Huillet, D. (2009). Mathematics for teaching: An anthropological approach and its use in teacher training. *For the Learning of Mathematics*, *29*(3), 4–10.

- Jakobsen, A., Thames, M. & Ribeiro, C. M. (2013). Delineating issues related to Horizon Content Knowledge for mathematics teaching. In B. Ubuz, Ç. Haser & M. A. Mariotti (Eds.), *Proceedings of CERME 8* (pp. 3125–3134). Antalia, Turkey.
- Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, 49(3), 568–589. <https://doi.org/10.3102/0002831212437853>.
- Kilpatrick, J. J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.
- Kleickmann, T., Richter, D., Kunter, K., Elsner, J., Besser, M., Krauss, S., et al. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106. <https://doi.org/10.1177/0022487112460398>.
- Kleickmann, T., Richter, D., Kunter, K., Elsner, J., Besser, M., Krauss, S., et al. (2015). Content knowledge and pedagogical content knowledge in Taiwanese and German mathematics teachers. *Teaching and Teacher Education*, 46, 115–126. <https://doi.org/10.1016/j.tate.2014.11.004>.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Krauss, S., Blum, W., Besser, M., Brunner, M., Neubrand, M., Baumert, J., et al. (2013). Mathematics teachers' domain-specific professional knowledge: Conceptualization and test construction in COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 147–174). New York: Springer.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., et al. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716–725. <https://doi.org/10.1037/0022-0663.100.3.716>.
- Lankford, H., Loeb, S., McEachin, A., Miller, L. C., & Wyckoff, J. (2014). Who enters teaching? Encouraging evidence that the status of teaching is improving. *Educational Researcher*, 43(9), 444–453. <https://doi.org/10.3102/0013189X14563600>.
- Lannin, J. K., Webb, M., Chval, K., Arbaugh, F., Hicks, S., Taylor, C., et al. (2013). The development of beginning mathematics teacher pedagogical content knowledge. *Journal of Mathematics Teacher Education*, 16(6), 403–426. <https://doi.org/10.1007/s10857-013-9244-5>.
- Lavy, V. (2009). Performance pay and teachers' effort, productivity, and grading ethics. *American Economic Review*, 99(5), 1979–2011. <https://doi.org/10.1257/aer.99.5.1979>.
- Lynch, K., Chin, M., & Blazar, D. (2017). Relationships between observations of elementary mathematics instruction and student achievement: Exploring variability across districts. *American Journal of Education*, 123(4), 615–646. <https://doi.org/10.1086/692662>.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304–316. <https://doi.org/10.3102/0013189X14545513>.
- McCrary, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D., & Senk, S. L. (2012). Knowledge of Algebra for Teaching: A framework of knowledge and practices. *Journal for Research in Mathematics Education*, 43(5), 584–615.
- Metzler, J., & Woessmann, L. (2010). *The impact of teacher subject knowledge on student achievement: Evidence from within-teacher and within-student variation*. CESifo Working Paper Series No. 3111. Retrieved June 6, 2012, from <http://ssrn.com/abstract=1634795>.
- Mewborn, D. (2003). Teachers, teaching, and their professional development. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 45–52). Reston, VA: National Council of Teachers of Mathematics.
- Mitchell, R., Charalambous, C. Y., & Hill, H. C. (2014). Examining the task and knowledge demands needed to teach with representations. *Journal of Mathematics Teacher Education*, 17(1), 37–60. <https://doi.org/10.1007/s10857-013-9253-4>.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2), 125–145. [https://doi.org/10.1016/0272-7757\(94\)90003-5](https://doi.org/10.1016/0272-7757(94)90003-5).
- Mullens, J. E., Murnane, R. J., & Willet, J. B. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review*, 40, 139–157. <https://doi.org/10.1086/447369>.
- National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education.

- Naumann, A., Hartig, J., & Hochweber, J. (2017). Absolute and relative measures of instructional sensitivity. *Journal of Educational and Behavioral Statistics*, 42(6), 678–705. <https://doi.org/10.3102/1076998617703649>.
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3), 237–257. <https://doi.org/10.3102/01623737026003237>.
- Ottmar, E. R., Rimm-Kaufman, S. E., Larsen, R. A., & Berry, R. Q. (2015). Mathematical knowledge for teaching, Standards-based mathematics teaching practices, and student achievement in the context of the *Responsive Classroom* approach. *American Educational Research Journal*, 52(4), 787–821. <https://doi.org/10.3102/0002831215579484>.
- Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193. <https://doi.org/10.3102/0002831210362589>.
- Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130, 105–119. <https://doi.org/10.1016/j.jpubeco.2015.02.008>.
- Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal*, 45(2), 365–397. <https://doi.org/10.3102/0002831207308230>.
- Remillard, J., & Kim, O. K. (2017). Knowledge of curriculum embedded mathematics: Exploring a critical domain of teaching. *Educational Studies in Mathematics*, 96(1), 65–81. <https://doi.org/10.1007/s10649-017-9757-4>.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2), 247–252. <https://doi.org/10.1257/0002828041302244>.
- Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 6(1), 43–74. https://doi.org/10.1162/EDFP_a_00022.
- Rowland, T., Huckstep, P., & Thwaites, A. (2005). Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi. *Journal of Mathematics Teacher Education*, 8, 255–281. <https://doi.org/10.1007/s10857-005-0853-5>.
- Rowland, T., Turner, F., Thwaites, A., & Huckstep, P. (2009). *Developing primary mathematics teaching*. Thousand Oaks, CA: Sage.
- Schilling, S. G. (2007). The role of psychometric modelling in test validation: An application of multidimensional item response theory. *Measurement*, 5(2–3), 93–106.
- Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links from teacher knowledge, to classroom practice, to student learning in the instructional system of the middle-school mathematics classroom. *Cognition and Instruction*, 28(3), 317–359. <https://doi.org/10.1080/07370008.2010.487961>.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15, 4–14. <https://doi.org/10.3102/0013189X015002004>.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23. <https://doi.org/10.17763/haer.57.1.j463w79r56455411>.
- Sleep, L. (2012). The work of steering instruction toward the mathematical point a decomposition of teaching practice. *American Educational Research Journal*, 49(5), 935–970. <https://doi.org/10.3102/0002831212448095>.
- Speer, N. M., King, K., & Howell, H. (2014). Definitions of mathematical knowledge for teaching: Using these constructs in research on secondary and college mathematics teachers. *Journal of Mathematics Teacher Education*, 18(2), 105–122. <https://doi.org/10.1007/s10857-014-9277-4>.
- Steele, M. D., Hillen, A. F., & Smith, M. S. (2013). Developing mathematical knowledge for teaching in a methods course: The case of function. *Journal of Mathematics Teacher Education*, 16(6), 451–482. <https://doi.org/10.1007/s10857-013-9243-6>.
- Tate, R. (2002). Test dimensionality. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 181–213). Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Tchoshanov, M. A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics*, 76(2), 141–164. <https://doi.org/10.1007/s10649-010-9269-y>.
- White, T. G., Kim, J. S., Kingston, H. C., & Foster, L. F. (2014). Replicating the effects of a teacher-scaffolded voluntary summer reading program: The role of poverty. *Reading Research Quarterly*, 49(1), 5–30. <https://doi.org/10.1002/rrq.62>.

- Wilson, S. M., Shulman, L. S., & Richert, A. E. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp. 104–124). London: Cassell.
- Wilson, S. M., & Wineburg, S. (1988). Peering at American history through different lenses: The role of disciplinary knowledge in teaching. *Teachers College Record*, 89, 525–540.
- Zazkis, R., & Mamolo, A. (2011). Reconceptualizing knowledge at the mathematical horizon. *For the Learning of Mathematics*, 31(2), 8–13.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.