

Mathematical Knowledge for Teaching and its Specificity to High School Geometry Instruction

Patricio Herbst¹
University of Michigan

Karl Kosko
Kent State University

Abstract

This paper documents efforts to develop an instrument to measure mathematical knowledge for teaching high school geometry (MKT-G). We report on the process of developing and piloting questions that purported to measure various domains of MKT-G. Scores on a piloted set of items had no statistical relationship with total years of experience teaching, but all domain scores were found to have statistically significant correlations with years of experience teaching high school geometry. Other interesting relationships regarding teachers' MKT-G scores are also reported. We use these results to propose a way of conceptualizing how instruction specific considerations might matter in the design of MKT items. In particular, we propose that the instructional situations that are customary to a course of studies, can be seen as units that organize much of the mathematical knowledge for teaching such course.

Keywords: geometry, knowledge, teaching, instructional situations

In his description of paradigms for research on teaching, Shulman (1986a) called for a focus on teacher knowledge. With particular reference to mathematics, Ball, Lubienski, and Mewborn (2001) responded to Shulman's call by, on the one hand, reviewing research that showed that traditional measures of teachers' content knowledge (e.g., degrees obtained or mathematics courses taken) had not been shown to make a difference on students' learning (but see Monk, 1994) and, on the other hand, arguing that the kind of teacher knowledge needed to focus on was a particular kind of mathematical knowledge, mathematical knowledge for teaching (MKT). This MKT is knowledge of mathematics used in doing the work of teaching and it includes, but also goes beyond, the pedagogical content knowledge that Shulman (1986b) himself had proposed. The theoretical and empirical work on Ball's brand of MKT that followed has been vast, showing among other things that the possession of MKT can be measured, that MKT is held differently by teachers and non-teachers, that MKT is held differently by teachers of higher grade level experience than those of lower grade level experience, that it makes a

difference in students' learning, and that scores on MKT correlate with scores on an observation measure of the mathematical quality of instruction (Hill, Schilling, & Ball, 2004; Hill, Rowan, & Ball, 2005; Hill et al., 2008). The work on constructing measures of MKT has been concentrated mostly on the mathematical knowledge of elementary and middle school teachers (Hill & Ball, 2004; Hill, 2007); a more recent effort has developed MKT items in algebra (Mark Thames, personal communication, 6/15/11).

The purpose of this paper is to report on a parallel effort to develop an instrument that measures mathematical knowledge for teaching high school geometry. While we recognize that other approaches to mathematical knowledge for teaching at the secondary level have started from conceptualizing it independently (e.g., Adler & Davis, 2006; Even, 1990; McCrory, Ferrini-Mundy, Floden, Reckase, & Senk, 2012; Saderholm, Ronau, Brown, & Collins, 2010; Silverman & Thompson, 2008), we have preferred to start from the conceptualization by Ball et al. (2008). While all of the other conceptualizations had compelling features, none of them had accumulated similar critical mass of empirical research as the MKT approach spearheaded by Ball and Hill. Furthermore, only the approach by McCrory et al. (2012) had been developed in such a way that measurement instruments could be created with them, and this one did not seem to examine elements of pedagogical content knowledge that we saw as important to consider. Thus our effort has attempted to follow the theoretical conceptualization of MKT and item development procedures of Ball and Hill's group even though, as we suggest below, we anticipated that some specification of MKT might need to be developed to respond to instructional demands specific to courses of study at the high school level. This chapter describes how we developed MKT-geometry items, provides pilot data that compares high school teachers with and without experience teaching geometry in terms of their possession of mathematical knowledge for

teaching geometry as measured by selected items, and uses these results to raise some questions about the instruction specificity of the notion of mathematical knowledge for teaching.

A crucial element in our development of items to measure the mathematical knowledge for teaching high school geometry has been Ball, Thames, and Phelps' (2008) conceptualization of the different domains of mathematical knowledge for teaching. According to Ball et al. (2008), the mathematical knowledge used in teaching can be conceptualized as the aggregation of knowledge from six domains. These domains include Common Content Knowledge (CCK), which is the mathematical knowledge also used in settings other than teaching, including for example knowledge of canonical methods for solving the problems teachers assign to students. The domains also include knowledge that is specific to the work of teaching. Thus Specialized Content Knowledge (SCK) is knowledge of mathematics used particularly in doing the tasks of teaching, such as, for example, the knowledge a teacher needs to use in writing the problems they will assign to students or in figuring out "whether a nonstandard approach would work in general" (Ball, et al, 2008, p. 400). A third domain, KCT, or Knowledge of Content and Teaching is defined as a combination of knowledge of teaching and knowledge of mathematics and includes the knowledge needed to decide on the best examples and representations to use for given instructional objectives. And KCS, or Knowledge of Content and Students, includes a blend of knowledge of mathematics and of students' thinking, such as the capacity to predict what students might find confusing or what kind of errors students might make when attacking a given problem. In our effort to construct measures of mathematical knowledge for teaching high school geometry, we developed items that purport to measure each of those four domains CCK, SCK, KCT, and KCS. Ball et al. (2008) also include Horizon Content Knowledge (HCK) and Knowledge of Content and Curriculum (KCC), but our work has not targeted those domains. By

the time we started working on our MKT-G instrument, we felt that the KCC and HCK domains needed further conceptualization and that there were not sufficient examples of HCK and KCC items at other levels of schooling that could allow us to create analog items to measure teachers' knowledge of those domains in high school geometry.

Ball and Hill's Learning Mathematics for Teaching project has developed items that measure the different domains of MKT and that has included, over time, attention to different content strands, including number and operation, patterns, functions and algebra, and geometry. These instruments have also included items that purport to measure mathematical knowledge for teaching middle school mathematics as well as for teaching elementary school mathematics. The extensive item development has yielded numbers of validated items that can be put together into forms that assess MKT for particular content strands. But there has not been, as of yet, a systematic development of items to measure MKT in different content strands or deliberate theoretical consideration about how content-strand differentiation might interface with the domains of MKT (Heather Hill, personal communication, 2/8/12). As we think about conceptualizing and measuring mathematical knowledge for teaching high school mathematics it is important to ponder whether and how the specific practice of teaching particular high school mathematics courses should be considered and featured in the process of designing measures of the mathematical knowledge for teaching those courses. In this paper we present our beginning attempts to conceptualize such instruction specificity within the framework of MKT, by reporting on our development of an instrument to measure mathematical knowledge for teaching high school geometry. It's worth elaborating briefly on what we mean by instruction specificity here since it is obvious that Ball's et al. (2008) conceptualization of MKT is already rooted in a notion of instruction as the interactions among teacher, students, and content in environments

(Cohen, Raudenbush, and Ball, 2003). As explicated in Herbst & Chazan (2012), we use *instruction* along similar lines, to allude to teacher-student transactions of knowledge: While *instruction* names that phenomenon in general, we expect that the nature of the knowledge at stake (as represented in syllabi and textbooks of a year- or semester-long course of studies) makes instruction specific (that is, shapes the nature of the transactions of knowledge that take place). Thus, a first approximation to the question of what an instruction-specific MKT looks like is to ask what is the MKT that a teacher would need in order to teach a given course of studies. Thus our purpose to design an instrument to assess MKT in geometry is concerned not with geometry as a mathematical domain but with high school geometry as a course of studies; we posit that similar approaches need to be taken regarding high school algebra and other such courses of study. (Further approximations might also consider specification to particular groups of students, particular environments, and particular 'kinds' of teaching.)

Our interest in MKT originated from our attempts to contribute to a theory of mathematics teaching that accounts for what teachers do in teaching in terms of a combination of, on the one hand, individual characteristics of practitioners and, on the other hand, practitioners' recognition of the norms of the instructional situations in which they participate and of the professional obligations they must respond to (Herbst & Chazan, 2011, 2012). This effort contributes to a long-term agenda that seeks to understand the work of teaching in specific instructional systems such as high school geometry. Our earlier work had been dedicated to conceptualizing and grounding (through examining records of intact classrooms and of instructional interventions) the didactical contract in geometry (Herbst, 2002, 2003) and instructional situations in geometry (Herbst, 2006, 2010; Herbst, Chen, Weiss, & González, 2009). In the context of that work we developed a proposal for the description of teachers'

actions and decisions as responses to norms of the role teachers play in activity systems (such as didactical contracts and their instructional situations) and obligations of the position of mathematics teacher in an institution (Herbst & Chazan, 2012). In that context we had asked the question of how an individual's mathematical knowledge for teaching combines with their recognition of professional obligations and of norms of an instructional situation in accounting for the decisions they make in classroom events framed by those norms. The present paper was part of a larger effort to develop measures of the constructs that we had contributed (particularly recognition of *norms* and *obligations*) as well as measures of other constructs that would allow us to estimate the resources possessed by individual teachers.

Development of MKT-Geometry

The conceptualization and disciplined approach to measuring MKT spearheaded by Ball and Hill (Ball et al, 2008; Hill and Ball, 2004) provided us with important guidance for the development of MKT measures. This guidance included not only the conceptualization of the domains and some heuristics for how to create items, but also an awareness of the complexity of the task ahead. In the interest of proceeding toward our goal making manageable steps, we also followed the example from Ball and Hill's group in developing multiple choice and multiple response items as opposed to constructed response or open-ended items. The following sections describe the work we did to develop an MKT-G instrument.

Constructing Items for MKT-Geometry

Our item development process covered a relatively wide range of topics from the high school geometry course. We consulted curriculum guidelines in various states and on that basis sought to develop items dealing with definitions, properties, and constructions of plane figures including triangles, quadrilaterals and circles, parallelism and perpendicularity, transformations,

area and perimeter, three-dimensional figures, surface area and volume, and coordinate geometry. Those topics by themselves provided sufficient guidelines to create items of CCK. But the definitions of the MKT domains, particularly the definition of SCK, calls for items that measure knowledge of mathematics used in the tasks of teaching. To draft these items we found it useful to create a list of tasks of teaching in which a teacher might be called to do mathematical work. The list included elements such as: designing a problem or task to pose to students; evaluating students' constructed responses, particularly student-created definitions, explanations, arguments, and solutions to problems; creating an answer key or a rubric for a test; and translating students' mathematical statements into conventional vocabulary. Teachers of different courses of mathematical study might engage in those tasks of teaching. But as we sought to draft these items, we noted that those tasks of teaching could call for different kinds of mathematical work depending on specifics of the work of teaching geometry. For example, the task of designing a problem would involve a teacher in different mathematical work if the designed problem was a proof problem versus a geometric calculation. While the former might involve the teacher in figuring out what the givens should be to make sure the desired proof could be done, the latter might involve the teacher in posing and solving equations and checking that the solutions of those equations represented well the figures at hand. Thus while a list of generic tasks of teaching was useful to start the drafting of items, this list appeared to grow more sophisticated with attention to tasks that are specific to different instructional situations in geometry teaching (Herbst, 2010). As we note below, this observation led to interest in comparing responses to items that were differently related to instructional situations in geometry, and to making some conjectures about the organization of the MKT domains.

The tasks of teaching were also useful in drafting items that measured knowledge of content and teaching (KCT). To draft these items we used as a heuristic the notion that the item should identify a well-defined instructional goal and the possible answers should name mathematical items that, while correct in general, would be better or worse choices to meet the specified goal. For example, teachers often need to choose examples for the concepts they teach. While different examples might be mathematically correct, they might not all be pedagogically appropriate to meet particular instructional goals: One example may be better than others if the instructional goal is to show a first or canonical example, while another example may be better if the goal is to illustrate an extreme case.

Finally, to create items that measured knowledge of content and students (KCS), we were attentive to the definition provided by Ball et al. (2008) and sought to draft especially items that tested for knowledge of students' errors and misconceptions (Hill, Ball, & Schilling, 2008). As in the case of other domains, there were specifics of the high school geometry class that shaped the items we developed. Thus, while we did create items that probed for teachers' knowledge of students' misconceptions about geometric concepts (e.g., angle bisector), we also created items that probed for their knowledge of students' misconceptions about processes or practices that are specific to geometry—such as the notion that empirical evidence is sufficient proof (Chazan, 1993) or that definitions are exhaustive descriptions (Herbst, González, and Macke, 2005).

Our research group drafted and revised an initial set of questions including 13 CCK, 20 SCK, 26 KCT, and 16 KCS questions; this drafting and revision process relied among other things on general guidance and comments on specific items by Deborah Ball, Hyman Bass, Laurie Sleep, and Mark Thames.ⁱⁱ The questions drafted took the form of multiple-choice items,

as well as multiple-response items within a single question (e.g., a single stem with 3-4 yes/no questions following).

Teachers' Interpretations of MKT-Geometry Items

Our items were submitted to a process of cognitive pretesting (Karabenick et al., 2007), which assesses the degree to which participants interpret items in the manner intended by those who wrote them. This process involved interviewing teachers ($n = 11$) as they completed the items. Teachers were asked to read the item prompt (or stem) and then tell us what they thought the item was asking them to do. The teachers would then be asked to select their response and tell us why they selected the option they did.

Responses to these prompts were used to examine the content validity of the items, as well as improving such validity. We coded teachers' responses in regards to whether the stem was interpreted as intended, as well as whether the justification for their response was mathematical and/or pedagogical in nature. The vast majority of items were found to be interpreted as intended, with a subset of items being revised in accordance with evidence from the data. For example, one KCT item's stem was revised to make a provided theorem more explicit in the description of the task of teaching. The original and revised wordings of the stem are presented in Figure 1.

In the original stem's wording, the base angles theorem is presented explicitly, but we found that some teachers' justifications for their responses did not focus as much on the theorem. In reexamining the stem, we observed that the task of teaching was directed to choosing a proof for a theorem, but included some distracting wording (e.g., a special naming of that theorem, the description of the response options as sketches of proof). While seemingly minor, the revised wording centers the task of teaching (choosing a proof for the theorem) more clearly within the

context of the base angles theorem. To further improve the way the item targeted the intended MKT-domain of the stem, we also changed the question from a multiple-response format to a multiple choice format and noted that the proofs were all valid.

Original Wording (before Cognitive Pretesting)	Revised Wording (after Cognitive Pretesting)
<p>While preparing to teach the theorem that says that base angles of an isosceles triangle are congruent, Ms. Gomez is pondering which among the following, valid mathematical proofs to use with her 9th grade geometry students. Given the following sketches of proofs, which ones would help students understand why the base angles theorem is true?</p>	<p>Ms. Gomez is preparing to teach the theorem that says that base angles of an isosceles triangle are congruent. She is pondering which among the following, valid mathematical proofs to use with her 9th grade geometry students. Of the valid proofs below, which one is the least appropriate for Ms. Gomez to use with her students in explaining why the theorem is true?</p>

Figure 1. Revisions to KCT item stem following cognitive pretesting.

All items were revised in accordance with evidence from the cognitive pretesting data to improve interpretability and validity. While we conducted cognitive pretesting as a means of identifying potential miscues in our items, the process also provided some preliminary evidence as to the construct validity of our items. Specifically, teachers were consistently observed to describe the mathematics of an item in selecting their response (approximately 95% of the time). This was fairly consistent across all domains (CCK, SCK, KCS, KCT). Also, teachers were observed to refer to pedagogical issues much less frequently when focusing on CCK and SCK items (54% and 45% of responses, respectively) than on KCS and KCT items (91% and 72% of responses, respectively). This provided some evidence that items designed to measure KCT and KCS domains might indeed assess those aspects of pedagogical content knowledge, while items designed to measure the CCK and SCK domains might indeed assess content knowledge. Trends for particular items were used to inform revisions; for example, teachers did not discuss

pedagogical issues at all for one KCT item and so this item was revised to appeal to such reasoning. Yet, the skewing of pedagogically related justifications in KCS and KCT items suggests a relatively good fit for most items per construct, even before these items were revised.

Piloting MKT-Geometry

After the cognitive pretesting, a revised set of items was pilot tested with inservice secondary mathematics teachers from the two states (one Midwestern state and one Mid Atlantic state) between July 2011 and May 2012. The Appendix shows publicly released items. Questions from each domain were uploaded into the LessonSketch online platform (www.lessonsketch.org) and completed by participants who took them either by coming in person to a computer lab (73 participants) or by responding to the items online from their homes or workplace (10 participants). For the purposes of this chapter, all data reported is pooled from both samples ($n = 83$). Participants were predominantly female (67.5%) and Caucasian (67%). Other reported ethnicities included African American (13%), Asian American (5%), and Other (16%). Participants varied in the amount of mathematics teaching experience ($M = 14.55$, $SD = 9.75$) and mathematics content courses taken in college ($M = 11.27$, $SD = 6.08$). Regarding Geometry-specific experience, 60.2% of participants were classified as experienced geometry teachers because they had taught Geometry for 3 years or more ($M = 4.57$, $SD = 4.55$), and teachers had taken an average of 1.70 college-level Geometry content courses ($SD = 1.41$).

While we later discuss MKT-Geometry in terms of a single score that we've found to have sufficient reliability, we briefly discuss item analysis and preliminary comparisons by domain (CCK, SCK, KCS, KCT). Such a description is not generalizable in nature, but provides interesting points of reference for later discussion and investigation.

Studying MKT-G by Domain

In July and August 2011, we piloted all MKT-G items revised from cognitive pretesting to a sample of 48 secondary mathematics teachers in a Midwestern state. By design, each domain contained 10 questions. These questions included multiple-choice and multiple-response formats. In examining the statistical reliability of our questions, multiple-response questions were treated as multiple items, with the number of items per a multiple-response item dependent on the number of accepted responses. Therefore while each domain contained 10 questions, different domains contained a different number of items for the purpose of statistical item analysis (CCK = 15; SCK = 29; KCT = 10; KCS = 10). In examining the fit of items for each separate domain, we used biserial correlations (Crocker & Algina, 2006) to measure item discrimination or how well the items discriminated between higher scoring test-takers and lower scoring test-takers. Biserial correlations were examined in concert with item difficulty (percent of participants correctly answering the item) and Cronbach's alpha. This initial examination of the items allowed us to make some decisions regarding the removal of some items due to poor statistical fit. Particularly, responses for some items had near-zero or negative correlations with the overall response patterns of other items per domain. Based on this preliminary examination and pursuant to our goal to discard items that did not perform well, we removed 46.9% of all items from the MKT-G assessment and retained 34 items, with the intent to collect additional data to confirm the reliability of the retained items.

Items were removed on the basis of various reasons. Some had low discrimination power; they lacked an ability to distinguish between lower and higher scoring participants per domain. Other items that performed poorly had apparent liabilities such as being too long or having response options that were too similar to each other; these liabilities had been somewhat

anticipated in the cognitive pretesting process. For example, one KCT item prompted teachers for the most appropriate feedback to provide a student in regards to an answer they had presumably provided on a test. The correct response was answered by 34% of participants but one distracter was chosen by 40% of participants. While scores of the 34% of participants were marginally higher than the 40% who answered incorrectly on the popular distracter, this difference was not great enough to help the item discriminate between lower and higher scoring participants. In principle, revisions might be made to improve the quality of those items. However, we were ever mindful of the time commitment necessary to complete our test, and chose to put revising those items aside for the time being.

Following our initial pilot in late 2011, we collected additional data from secondary mathematics teachers in the Mid-Atlantic region during Spring 2012.¹ The sample data was pooled with that from our Midwestern sample for a combined 83 participants. We chose to have these later participants complete only those items that did not indicate poor fit in the earlier pilot. Therefore, the items per domain ranged from 7 to 11 (CCK = 9; SCK = 11; KCT = 7; KCS = 7). We reexamined item fit on the smaller subset of items using biserial correlations (with a threshold of .30) as one indicator for reliability in each domain. Item difficulty in CCK ranged from 20.5% to 92.3%. Item difficulty in SCK ranged from 21.7% to 88.0%; in KCT it ranged from 18.1% to 48.3%; and in KCS it ranged from 15.7% to 73.5%. We also calculated Cronbach's alpha to provide preliminary data regarding reliability of each domain on its own (see Table 1). While the alpha coefficients are not as high as desired (they are below the typically

¹ We note that neither the Midwestern nor the Midatlantic sample had participated in any special professional development preceding the MKT assessment that could possibly account for differences between groups.

accepted threshold of .70), these statistics provided us with valuable information regarding future steps, should we choose to focus specifically on assessing particular domains in the future. One explanation for lower reliabilities in some domains is related to both the range in difficulty of items and the number of items per domain. For example, SCK, which has acceptable reliability ($\alpha = .65$) had a large range in item difficulty, but also had the most items (13), but KCS had a much lower alpha coefficient, mainly due to having fewer items. Therefore, a more detailed study of KCS would require us to construct additional items of varying difficulty levels.

Table 1.

Descriptive Statistics by MKT Domain.

Domain	M	SD	N	α
CCK – Geometry	0.66	0.20	83	0.58
SCK – Geometry	0.59	0.21	83	0.65
KCT – Geometry	0.39	0.21	83	0.50
KCS – Geometry	0.38	0.25	83	0.49

Correlations between the domain scores are presented in Table 2, and suggest moderate to strong relationships between the different domains. These results show similar trends to those found by Hill et al. (2004) for CCK and KCS, which suggest that the different domains are, to a degree, interrelated. Thus, these correlations encourage thinking of a single MKT-G measure made of aggregating the scores per domain. The use of items from different domains to make that single score helps argue that the instrument assesses the MKT construct as conceptualized by Ball et al. (2008).

Table 2.

Correlations between MKT-G Domain Scores.

	CCK	SCK	KCT	KCS
CCK	-			
SCK	.40***	-		
KCT	.36**	.54***	-	
KCS	.69***	.48***	.43***	-

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 3.

Correlations between Experience-Type and Score.

	Years Experience		Content Coursework	
	Years Teaching Mathematics	Years Teaching Geometry	Total Math Courses	Total Geometry Courses
CCK-G	-.06	.38***	-.05	-.02
SCK-G	.09	.35**	.05	.04
KCT-G	.04	.19 ^a	-.00	-.05
KCS-G	-.02	.25*	.10	.08

^a $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

We also examined correlations, by MKT domain, with teaching experience in general and with teaching high school Geometry in particular, as well as with college coursework in mathematics. (We did not collect specific information on education classes taken; in particular we did not collect information about geometry topics covered in education classes or number of specific education courses that would be considered "methods classes", because we did not anticipate enough variability in this number that might be used to examine correlations). The correlations with experience and mathematics coursework (shown in Table 3) are preliminary as they illustrate relationships that emerge from the process of developing our assessment of MKT-Geometry. However, they provide an intriguing picture of how teachers with different experiences hold knowledge for teaching Geometry and suggest that mathematical knowledge

for teaching geometry may be learned from experience teaching geometry. These results show that neither the total number of mathematics courses taken in college nor the number of Geometry courses correlate with MKT-Geometry scores. Likewise, years of experience teaching mathematics in general also show near-zero correlations. However, for each domain in MKT-Geometry, there are indications of a relationship with experience teaching the Geometry course. Therefore, it seems that our assessment of MKT-Geometry may represent a domain of knowledge specifically related to the teaching of high school Geometry.

Preliminary Evidence for Trends in Teachers' MKT-Geometry

The previous section discussed preliminary findings for each domain (CCK, SCK, KCT, KCS) using raw scores. In order to better understand how teachers' demonstrated understandings of MKT-Geometry as a single construct were affected by varying background factors, we constructed scores using Item Response Theory (IRT). Therefore, discussion of scores and item statistics in this section examine MKT-G as a single construct (i.e., CCK, SCK, KCT, and KCS items are pooled together). While it may also be useful to construct IRT scores for specific domains of CCK, SCK, KCT, and KCS, more items would need to be included to better assess the possible variance in each specific domain. Since more items mean a longer, lengthier test, we elected not to take such an approach at this time.

As Crocker and Algina (2006, p. 339) describe,

with item response theory the test developer assumes that the responses to the items on a test can be accounted for by latent traits...[that] accounts for the responses to items on a test. At the 'heart' of the theory is a mathematical model of how examinees at different ability levels for the trait should respond to an item

This latent trait is represented by the statistic theta (θ), which is calculated both for particular items and as a test score for particular individuals. Where classical test theory approaches take the percent correct on a test as the indicator for the latent trait being measured, IRT takes into account that some items are more difficult than others, and thus a raw score is not the most accurate measure of the latent trait (see chapter 15 in Crocker & Algina, 2006, or Wilson, 2005 for introductions to IRT). We borrowed from approaches used by the LMT group led by Deborah Ball (Hill et al., 2004; Hill et al., 2008) and used a 1-parameter IRT model, which is mathematically equivalent to a basic Rasch model. According to Linacre (1994), a minimum sample of 50 examinees is sufficient for obtaining estimates for a basic Rasch model, which our sample ($n = 83$) exceeds.

Results from our IRT modeling show sufficient item reliability (.96) and person reliability (.82). Item reliability indicates how well our survey distinguishes between easier and more difficult items, and is generally considered acceptable if above .90. Person reliability indicates how well our survey distinguishes between groups of people (e.g., lower and higher scorers), and is generally considered acceptable if above .80. Possible θ scores for individuals on the MKT-G ranged between -5.50 (low demonstrated MKT-G) and 5.45 (high demonstrated MKT-G). A person scoring 0 demonstrates an average level of MKT-G. Such an individual would have a 50% chance of correctly answering an item with θ value 0 (that is an item of average difficulty), but would have a greater chance of correctly answering items of lower θ values (and smaller chance of correctly answering items of higher θ). For our sample, participants' overall scores ranged from -2.27 to 3.43 ($M = .19$, $SD = 1.03$). This suggests that our sample did not contain individuals who had extremely high (-5.50) or low (5.45) θ scores.

However, the mean is near zero and the standard deviation approximately 1.00, which are fairly good indicators of an assessment providing good data.

Building on our preliminary evidence from trends in the individual domains, and incorporating our IRT scores to ensure a more accurate depiction of demonstrated MKT-G, we conducted a multiple regression to examine the effect of being an ‘experienced’ Geometry teacher (3 or more years of experience) on MKT-G scores. To account for other potential factors, we included a dummy-coded variable indicating whether participants came from the Mid-Atlantic or Midwestern regions (course background and general experience were not considered given their near-zero correlations in the previous section). This regression equation is shown below. As modeled, the intercept, β_0 , represents the average score of non-experienced Geometry teachers (less than 3 years) who are from the Midwest.

$$MKTG = \beta_0 + \beta_1(ExperiencedGeometry) + \beta_2(dMidAtlantic) + e$$

Analysis showed that the model is statistically significant ($F_{(df=2)} = 9.10, p < .001$) with an r^2 of .19. Accounting for all independent variables, teachers who lack experience in teaching Geometry, and are from the Midwest had slightly lower than average MKT-G scores ($\beta_0 = -.10, p = .58$), but this was not found to be significantly different from an average score of 0. Being from the Mid-Atlantic, regardless of experience, was found to have a negative effect on MKT-G scores ($\beta_2 = -.39, p < .10$). Accounting for the effect of geographic location, having at least 3 years of experience teaching high school Geometry was found to have a statistically significant and positive effect on teachers’ MKT-G scores ($\beta_1 = .78, p < .001$). Further, as this effect is approximately three-quarters of a standard deviation unit, it is quite large. These results suggest that participants who have 3 or more years experience teaching high school geometry are more likely to have higher θ scores on the MKT-G assessment. The large size of the effect is double

that of the negative effect associated with our Mid-Atlantic sample; thus, experienced geometry teachers from the Mid-Atlantic sample still appeared to have significantly larger MKT-G scores than non-experienced geometry teachers (from either region).

As with the examination of the domain-specific correlations in the previous section, this analysis builds on our emerging understanding of teachers' MKT-G. First, there is evidence of potential differences in MKT-G performance depending on geographic location and as noted before none of these groups had participated in any professional development that might account for the differences found. While it may be tempting to make conjectures regarding this finding (for example, to suspect that demographic factors might account for those differences), it seems more reasonable for a larger sample to be collected in the future so that school and district level factors can be accounted for more accurately. What we can say about such findings is that where a teacher teaches may affect their demonstrated MKT-G. If, as the data suggests, experience teaching geometry matters in their MKT-G scores, it is reasonable to expect that institutional differences may account for some of the variance, just as much as demographics may. Second, while a sizeable effect attributable to geographic location was found, the effect associated with being an experienced Geometry teacher was a much stronger predictor of a teacher's MKT-G score. We believe this to be an important finding, which could be influential for teacher education practice (e.g., how to organize courses that teach mathematical knowledge for teaching to prospective high school teachers) and departmental practices in schools (e.g., how to staff different mathematics courses in a high school). Obviously, such possible practical implications hinge not only on replication of this finding but also on institutional, moral, and political considerations that only policy-makers and practitioners of particular locales can make. In what follows we explore some of the theoretical consequences of this finding.

Relationships between MKT-G Scores and Teaching Experience

Our interest in MKT contributes to a larger project that investigates the influence that individual factors (such as mathematical knowledge for teaching) and socialization to the work demands of teaching a particular high school course (in this case, high school geometry, as indicated by teachers' recognition of instructional norms and professional obligations) have in the decisions that a teacher would make (see Herbst, Aaron, Dimmel, & Erickson, 2013; Herbst, Kosko, & Dimmel, 2013, for accounts of our attempts to measure teachers' recognition of norms). A question we posed to the pilot data discussed here is what is the relationship between mathematical knowledge for teaching geometry and experience teaching high school geometry. Therefore, we correlated scores for each domain with teachers' general and Geometry-specific background, including years of experience teaching high school and college mathematics courses completed. Noticing the consistent relationship between teachers' Geometry-specific teaching experience and their MKT-G scores in each domain, we created and analyzed IRT scores to confirm this relationship.

Our findings indicated that having 3 or more years of experience teaching Geometry had a statistically significant and positive effect on teachers' MKT-G score. This effect was sizeable at approximately three-quarters of a standard deviation. To a certain degree, our findings support those by Hill (2010) who observed a near-zero correlation between elementary teachers' MKT scores and their general years of experience. Our findings also suggest a similar near zero correlation for general teaching experience. Yet, regarding general years of teaching experience, neither our findings nor those by Hill (2010) match those of Hill's (2007) examination of middle school mathematics teachers where (among other things) she found that middle school teachers with more years of teaching experience in mathematics tended to possess higher levels of

teaching-specific mathematical knowledge. What is intriguing is that in both Hill's (2007; 2010) studies, specialized forms of experience (i.e., more experience teaching mathematics at higher grade levels) were found to have highly significant correlations to MKT scores. What Hill's (2007; 2010) earlier findings have in common with ours is that when teachers have more experience at teaching the content in depth, they demonstrate higher MKT scores. Therefore, these results suggest that while teaching experience may affect MKT-Geometry scores, it is the particular experience of teaching the geometry course which matters. To the extent that mathematical knowledge for teaching is the knowledge of mathematics used in the work of teaching, the results lead us to ask how the specifics of the instructional work a teacher does in a course matter in the mathematical knowledge for teaching the teacher has: How do the specific demands of the work of geometry instruction create opportunities for teachers to learn this mathematical knowledge?

Mathematical Knowledge for Teaching and Instructional Situations

In our earlier and parallel work we have argued that the particular nature of the didactical contract (Brousseau, 1997; Herbst & Chazan, 2012) for a course of studies creates conditions of work that make the teaching of a specific course (e.g., high school geometry) different than the teaching of other mathematics courses (e.g., Algebra I). Taken together, the findings presented throughout this chapter seem to suggest that teachers of high school geometry have more mathematical knowledge for teaching geometry, while the difference does not seem to be accountable to general experience teaching secondary mathematics or to the mathematics courses teachers took in their college education. While at one level one might not find these results surprising, the fact that three of the four domains of mathematical knowledge for teaching we

tested for (SCK, KCT, and KCS) are defined as mathematical knowledge used in the work of teaching helps raise questions for future inquiry.

One interesting finding from cognitive pretesting may suggest future directions to further explore the nature of how MKT-G is developed amongst teachers. Specifically, teachers participating in cognitive pretesting were experienced geometry teachers. In providing justifications for their responses, it was anticipated that the domains of KCT and KCS would elicit a sizeable portion of pedagogically related comments, which they did. What is interesting is that approximately half of the justifications in CCK and SCK also elicited pedagogically related comments. While such evidence is limited in its breadth, when we take such evidence in context with the other data presented earlier, in particular that there was no correlation between the number of geometry courses taken in college and participants' scores on CCK and SCK. It seems that the specific experience of *teaching* geometry is important for all domains and in particular, the nature of the work of teaching geometry may be more influential than academic knowledge of the geometry domain. Thus there is reason to ask the question of how the specific instructional work a teacher does in the high school geometry course might matter in the mathematical knowledge for teaching the teacher has.

As we noted above, the current conceptualization of MKT has not addressed theoretically differentiation within MKT domains according to the content of courses of studies: How might MKT domains be organized? An easy way of thinking about that differentiation could be by deferring to the way the discipline of mathematics organizes the topical content of items: Items involving mathematical topics of the same branch of mathematics might be thought of as indicators of knowledge of a well defined region of an MKT domain. But that approach seems to apply well, that is consistently with the way MKT is defined, only to the domain of Common

Content Knowledge. To the extent that the other domains are defined in relation to the work of teaching, it is plausible that specification within each domain will also require considerations of the specifics of the teaching involved and not only of the mathematical topics referred to. Indeed, it is not immediately obvious that combining general notions in the definitions of the domains (e.g., that KCS includes knowledge of misconceptions) and mathematical domain specific topics (e.g., pentominoes, a topic that can be filed under geometry within the discipline of mathematics; Tracy & Eckart, 1990) would produce items that validly represent mathematical knowledge for teaching high school geometry: In the example, it might be possible to create an item that asked about students' errors working with pentominoes but one would have to question the extent to which such knowledge really plays any role in the actual work of teaching high school geometry. We posit instead that elements of mathematical knowledge for teaching need to be identified from the specific work of mathematics instruction in given courses of studies rather than by the pairing of generic features of instruction with lists of mathematical topics from mathematical domains.

The results from this study suggest that the teaching of high school geometry may entail specific mathematical knowledge demands. We would like to argue that those specific knowledge demands are not solely dictated by the mathematical topics that feature in the curriculum, but actually by the specifics of the actual work of teaching a course of studies (that is subject-specific instruction). In what follows we elaborate the argument using some item-to-item comparisons from our data.

Course Specificity in Specialized Content Knowledge

Specialized Content Knowledge (SCK) is defined as the knowledge of mathematics used in doing the tasks of teaching. This definition seems operational to us while an alternative one

floating in the literature (that SCK consists of that knowledge of mathematics that nobody else but teachers use) is not so operational: The definition we choose is not completely without problems but it permits creating items without having to establish the truth of an empirical negative proposition (that nobody else than teachers has the same knowledge, which seems quite hard to verify). Our chosen definition is problematic in that it generates some overlap with CCK: For example, one task of teaching is to create answer keys for problems given to students and such mathematical work involves (among other things) knowledge of canonical concepts and procedures, which is, by definition, part of CCK. We have tried our SCK items to stay away from that overlapping area or to emphasize not so much what the answer is but actually how the answer should be represented (e.g., writing an answer key that students would read, which would entail mathematical work less common outside of teaching if one thinks about the teacher's work of deciding what steps must be displayed vs. what steps can be made without displaying them). Considering our definition of SCK as the mathematical knowledge used in doing the tasks of teaching, we speculate on whether and how a course-specific version of SCK is warranted.

We grant that it is possible that not all SCK may be course-specific, in that some tasks of mathematics teaching may be generic even if they involve doing some mathematics: The task of creating a grading system, to take an extreme example, involves a teacher in making a mathematical model that feeds from grades in individual assignments; but there is no reason for this mathematical work to be different for teachers of different high school mathematics courses and quite often high school mathematics teachers use the same system across all mathematics courses they teach. There may be other such tasks of teaching that involve similar mathematical work no matter in what mathematics course they are engaged. We are not so interested in those tasks of teaching.

We are interested in other tasks of teaching where course-specificity is likely to shape the meaning of "task of teaching." Some tasks of teaching are amenable to generic statement (e.g., choosing the givens of a problem for students, checking on the correctness of what students wrote as they showed their work on a problem) but they may involve practitioners in different mathematical work depending on the specifics of the task (e.g., choosing the numbers for a word problem to be given in algebra I involves the teacher in different mathematical work than the work he or she might be involved in when constructing a geometric diagram to include among the givens in a geometry worksheet). Are those differences merely differences in mathematical domain (algebra vs. geometry) or do they also reflect differences in the activity systems to which those tasks of teaching contribute? We argue for the latter and bring in for such purpose the notion of *instructional situation* (Herbst, 2006).

Instructional situations are frames for the exchange of specific work done in a mathematical task for a claim on an item of knowledge at stake in the course of studies. For example, in Algebra I, when a student responds to $2x + 1 = x + 5$ by subtracting x and subtracting 1 to both sides of the equality, the teacher has some evidence to claim that the student knows a method for the solution of equations in one variable. Instructional situations try and capture the folk notion that every item of knowledge to be learned has one or more problem types or canonical tasks used to teach it and assess it. Instructional situations frame such exchanges by establishing norms for the elements involved: Norms that specify what those mathematical tasks look like, what doing those tasks means, who has to do each thing, etc. as well as norms that specify what the knowledge at stake is. In our earlier work (e.g., Herbst & Chazan, 2012) we have proposed that while some of those exchanges of work for knowledge may be done ad hoc (e.g., negotiated), many exchanges go without saying because they are framed by instructional

situations (that is, participants abide by the norms of a situation). These instructional situations are course-specific, which means more than mathematics specific: The roles of teacher and student in regard to exchanges of work for knowledge vary depending on whether the exchanges happen in one or another course, though the work done or the knowledge at stake may exist in both courses. For example, properties of geometric figures are part of the subject of study in elementary and middle school grades as well as in high school geometry, but these properties are studied differently in high school geometry--that is through engaging in different mathematical work. To assert that students (have the opportunity to) know those properties a high school geometry teacher needs to see students at work on mathematical tasks that are specific of the high school geometry course (e.g., constructing quadrilaterals, calculating measures of quadrilaterals, doing proofs about quadrilaterals). Reciprocally, some tasks like answering the question "what is a rhombus?" might appear not only in high school geometry but also in an earlier grade, but the knowledge at stake in that question for each course of studies is likely to be different: While a high school geometry teacher is likely to put a premium on students' knowledge of a set of necessary and sufficient conditions (e.g., the definition of rhombus), an elementary teacher might take students' capacity to list true properties as evidence of knowledge and not make much out of redundancy or conciseness.

We propose that the management of course-specific instructional situations involves teachers in singular mathematical work. The teacher's management of instructional situations includes in particular the choosing of the various mathematical tasks that students are to do, the observation of the proceeds (what students actually do), and the effecting of exchanges between such observed actions and the knowledge at stake (identifying at least for herself but possibly also publicly to the class how is it that what students have done indicates that students know the

ideas at stake). While the definition of these tasks of teaching is general, the mathematical knowledge called forth in doing them would be different across different courses, as long as the specific exchanges were different. We expect that these differences between exchanges may ensue from different kinds of knowledge at stake, different kinds of students' mathematical work being transacted for such knowledge, or even different ways of effecting those transactions. (A teacher might implement different mathematical sensibilities when appraising students' mathematical work depending on whether that work is done in second grade or in Algebra I, for example, in attending to precision.)

A case in point that helps argue that instructional situations matter comes from one SCK item in our instrument. This was a multiple-response question with two testlets; the stem spoke of a teacher needing to choose algebraic expressions for the sides of an isosceles triangle where the students would be expected to find the lengths of the sides of the triangle after solving an equation. Each item provided algebraic expressions for the three sides and asked whether or not they were an appropriate set of expressions. A quick examination of the responses to the item indicated that teachers with more or less years experience teaching geometry (≥ 3 years and < 3 years, respectively) did not respond much differently for the item where the equation could not be solved. However, the two types of teachers' responses did show differences for the item where the equation could be solved: the less experienced geometry teachers tended to answer that the set of expressions was appropriate, while the experienced geometry teachers tended to respond that the expressions were not appropriate. In fact, the numbers obtained after solving the equations of that set of expressions would not work well to represent the sides of a triangle in

that the triangle inequality² would not hold for those numbers. We conjecture that the experienced geometry teachers' familiarity with the instructional situation of "calculating a measure" (Herbst, 2010) mattered in their decision to check that the expressions would yield sides with positive lengths and that they would satisfy the triangle inequality. Our conjecture is not that the non experienced geometry teachers did not know the triangle inequality. They might or might not know it and in any case that would be an example of CCK. We argue that the non experienced geometry teachers were less likely to know that the triangle inequality had to be checked in doing the task of teaching of choosing algebraic expressions for the sides of the triangle, possibly because they only saw the problem as an exercise in algebra rather than also as an exercise on the geometry of triangles. More generally, we propose that an element of SCK specific to the instructional situation of "calculating a measure" is the knowledge that when assigning numerical (or algebraic) values to some dimensions of a geometric figure (a task that teachers have to do every time they create one such problem) one has to check that none of the geometric properties of that figure are refuted by the set of choices made. Even more generally, we conjecture that tasks of teaching that are subservient to instructional situations which are specific to a given course of studies (e.g., high school geometry) might involve teachers in mathematical work that teachers who are experienced in managing those situations would know better how to do than teachers who are not so experienced. Thus we propose that the instructional situations of a course of studies are natural containers of elements of specialized content knowledge.

Yet there is reason to suspect specialized content knowledge might include other items of knowledge than those needed to manage the instructional situations of a course. In particular, a

² The triangle inequality states that if a , b , and c are the lengths of the sides of a triangle, then $a-c < b < a+c$

given course of studies is likely to include not only recurrent instructional situations framing familiar tasks but also novel tasks, tasks that may deviate enough from what is customary that teacher and students need to negotiate what it means to work together in those (see Herbst, 2003). We argue those tasks might also call for special mathematical work on the part of the teacher (choosing the givens for a problem, interpreting what students do and say, etc.) and along the lines of the definition of SCK, that work could also imply the existence of specialized knowledge. Is that SCK expected to be general to all mathematics teachers or expected to be special to teachers of high school geometry?

Our geometry instrument also included SCK items that referred to mathematical tasks that depart from ordinary instructional situations, but that address objects of study that are part of the geometry course. In one of those items, a multiple response question including two testlets, the participants were confronted with the scenario that a teacher had asked his geometry students to provide a definition of a figure that is ordinarily studied in the course. Participants were asked to consider two definitions purportedly proposed by students and decide whether they were correct, applied only to particular cases, or could only be consequences of a definition. We contend that this task is novel in that students are rarely asked to create definitions in high school geometry. Teachers of high school geometry, however, are used to consider definitions and distinguish them from other statements: The geometry class usually confronts teachers with the task of distinguishing between definitions and their consequences, for example when they assess student proofs. Additionally teachers do ask students to recall definitions for geometric concepts in tests and quizzes and in such context they are likely to discriminate between definitions and other true statements about a concept. In contrast, in other courses the word definition is often used in much more relaxed way (i.e., defining is often used as synonymous of stating, describing,

or clarifying; see Sinclair & Moss, 2012; Zaslavsky & Shir, 2005). Thus we expected that teachers experienced in the teaching of the geometry course would do better in this item. Indeed respondents who had experience teaching geometry were much less likely to respond to this item incorrectly than respondents who did not have experience teaching geometry. In particular, one of those testlets included a purported definition that, insofar as it expressed an element of the concept image of the figure under consideration, could have been considered correct in earlier grades or by people whose notion of definition was more relaxed. Experienced geometry teachers were far more likely than non-experienced geometry teachers to single that definition out as not a correct one. Thus, while having students create their own definitions may be a novel task in most geometry classrooms, experienced geometry teachers seemed better able to use their mathematical knowledge to handle student work. In contrast, in an item that asked teachers to assess purported student work in an area of geometry not commonly emphasized in high school geometry (transformations) and in an item where they had to assess purportedly student-generated construction procedures for a figure whose construction is not often studied in the course, the differences between experienced and nonexperienced geometry teachers were not as marked (both groups were similarly likely to answer the item incorrectly).

Course Specificity in Knowledge of Content and Students and Knowledge of Content and Teaching

In regard to the other MKT domains of interest, one could ask the same theoretical question as above: How are those domains of knowledge organized? In particular, are they specific to the instructional practices of particular courses of study? We suggest that considerations of the nature of the instructional situations in the high school geometry course could lead to analogous differentiation within the domains of KCT and KCS as we argued for

SCK. As in the case of SCK we concede the possibility that some KCT and some KCS might be general to mathematics teaching, but we propose that instructional situations within a course of studies can help organize elements of KCS and KCT that are specific to the teaching of that course of studies. In particular, it is reasonable to expect that experienced geometry teachers might have more KCS if this domain is assessed with items that probe for their knowledge of students' conceptions and errors in tasks that are framed within the instructional situations of the high school geometry course, while they might have less KCS if the items concern students' conceptions or errors regarding geometric ideas that are not so central to the course. Among our KCS items, one item could be described as probing for knowledge of student errors in a task that is quite familiar to experienced geometry teachers (identifying the height corresponding to a given base of a given triangle), while another item could be described as not so familiar (determining the number of diagonals of a polygon of a large number of sides, without access to the formula, a task that we would classify as novel because it is common for geometry students to do calculate the number of diagonals of a polygon using the formula³). A quick glance at the responses to those items shows that the difference between the percentage of experienced and non experienced geometry teachers who answered the first item correctly is about three times the same difference for the second item. That is, experienced geometry teachers were a lot more aware than non experienced geometry teachers of the errors students could make in the first (familiar) task (about height of a triangle) than in the second (novel) task (about diagonals in a polygon). This in spite of the fact that both mathematical topics are equally common in the high school geometry course.

³ The number of diagonals in a polygon of n sides is $n(n-3)/2$.

In inspecting the responses to KCT items we could see similar patterns. One item asked participants to choose a figure that would help a teacher show to his or her students that the converse of a theorem is not true, an activity that we judge to be quite common when teachers are installing theorems in the geometry class (Herbst, Nachlieli, & Chazan, 2011). Another item asked participants to choose among exploratory tasks the one that would best impress upon students that two geometric concepts (angle bisectors and diagonals) are different. The difference between the percentage of experienced geometry teachers and the percentage of non-experienced geometry teachers that answered correctly the first item was more than twice that difference for the second item.

Clearly, this item-by-item commentary cannot be taken as confirmation of any general trend but it permits formulating conjectures that could be tested with more systematically constructed test forms. In particular, we conjecture that experienced geometry teachers would be more likely to answer correctly SCK items that involve a participant in doing mathematical work similar to the work they do in an instructional situation they are familiar with or related to an object of study that is a piece of the geometry course than SCK items that concern geometric objects less common in the high school geometry course or tasks that are not usually used to teach or assess an item of the knowledge at stake. We speculate that the same conjecture could be used to study how course specificity plays out in the case of MKT related to other high school courses of study, thus helping to think about how the domains of MKT could be organized.

The effort to discover an internal organization for each of the instruction-specific MKT domains is compelling theoretically. It could help us better understand the relationships between individual resources and situated activity--how different individuals may be differentially prepared for taking the teacher role in different instructional systems. The notion of instructional

situation as an organizing element for the knowledge in each domain could also be used in the construction of instruments that measure the extent to which individuals are prepared to enact instruction that departs from customary instruction--by creating items that require the practitioner to engage in the kind of mathematical thinking needed to sustain tasks that more or less deviate from the thinking they need to do in customary instructional situations. From a more practical standpoint the work of finding an internal organization to the MKT domains could inform the development of coursework in mathematics or mathematics education for future teachers, though such development also needs to attend to many other considerations of institutional nature that policy makers and practitioners of particular locales are expected to make.

Conclusion

The present chapter describes efforts to construct and validate an instrument that measures mathematical knowledge for teaching high school geometry. The construction of the items in the instrument paid attention to generic tasks of teaching and combined those with geometric content that is or could be taught at the high school level. The instrument worked relatively well and pilot data shows correlations between scores in each of the domains KCT, KCS, CCK, and SCK and the number of years of experience teaching geometry. The aggregate MKT-G scores are significantly accounted for by years of experience teaching geometry even when taking into consideration the significant effect of geographic region in which respondents taught. In contrast, neither years of experience teaching mathematics in general nor number of mathematics courses or geometry courses taken in college correlated with MKT-G scores. Our discussion of item-level differences in the sample, however, shows that experienced geometry

teachers did not perform homogeneously--being much better than non-experienced geometry teachers in some items and not so much better in others. Further item development and testing of experienced geometry teachers can help us understand individual differences and the differential difficulty individual teachers have with elements of mathematical knowledge for teaching. Our discussion of item-level differences in the sample supports a conjecture about how the three teaching-specific MKT domains organize: We conjecture that experience doing the work called for by customary instructional situations in a course of studies (such as geometry) organizes practitioners MKT more than the mere mathematical content of the items. This knowledge organization hypothesis can assist item development with which to study individual differences among experienced geometry teachers and the nature of expertise in mathematical knowledge for teaching geometry.

References

- Adler, J., & Davis, Z. (2006). Opening another black box: Researching mathematics for teaching in mathematics teacher education. *Journal for Research in Mathematics Education*, 37(4), 270-296.
- Ball, D. L., Lubienski, S. T., and Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433-456). 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.
- Brousseau, G. (1997). *Theory of didactical situations in mathematics: Didactique des mathématiques, 1970-1990*. N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield (Eds.). Dordrecht: Kluwer.
- Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational studies in mathematics*, 24(4), 359-387.
- Cohen, D. K., Raudenbush, S. W., and Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2):119-142.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Mason, OH: Wadsworth.

- Even, R. (1990). Subject matter knowledge for teaching and the case of functions. *Educational studies in mathematics*, 21(6), 521-544.
- Herbst, P. (2002). Engaging students in proving: A double bind on the teacher. *Journal for Research in Mathematics Education*, 33, 176-203.
- Herbst, P. (2003). Using novel tasks to teach mathematics: Three tensions affecting the work of the teacher. *American Educational Research Journal*, 40, 197-238.
- Herbst, P. (2006). Teaching geometry with problems: Negotiating instructional situations and mathematical tasks. *Journal for Research in Mathematics Education*, 37, 313-347.
- Herbst, P. with González, G., Hsu, H. Y., Chen, C., Weiss, M., and Hamlin, M. (2010). Instructional situations and students' opportunities to reason in the high school geometry class. Manuscript. Deep Blue at the University of Michigan, <http://hdl.handle.net/2027.42/78372>
- Herbst, P., Aaron, W., Dimmel, J., and Erickson, A. (2013, April). Expanding students' involvement in proof problems: Are geometry teachers willing to depart from the norm? Paper presented at the 2013 Annual Meeting of the American Educational Research Association. Available from Deep Blue at the University of Michigan. <http://hdl.handle.net/2027.42/97425>
- Herbst, P. Chen, C., Weiss, M., and González, G., with Nachlieli, T., Hamlin, M., and Brach, C. (2009). "Doing proofs" in geometry classrooms. In M. Blanton, D. Stylianou, and E. Knuth (Eds.), *Teaching and learning of proof across the grades: A K-16 perspective* (pp. 250-268). New York: Routledge.
- Herbst, P. and Chazan, D. (2011). Research on Practical Rationality: Studying the Justification of Actions in Mathematics Teaching. *The Mathematics Enthusiast*, 8(3), 405-462
- Herbst, P. and Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM Mathematics Education*.
- Herbst, P., González, G., and Macke, M. (2005). How Can Geometry Students Understand What It Means to Define in Mathematics? *The Mathematics Educator*, 15(2), 17-24.
- Herbst, P., Kosko, K., and Dimmel, J. (2013, November). How are geometric proof problems presented? Conceptualizing and measuring teachers' recognition of the diagrammatic register. Proceedings of the 2013 Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Available from Deep Blue at the University of Michigan <http://hdl.handle.net/2027.42/97761>
- Herbst, P., Nachlieli, T., and Chazan, D. (2011). Studying the practical rationality of mathematics teaching: What goes into "installing" a theorem in geometry? *Cognition and Instruction*, 29(2), 1-38.
- Hill, H. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis*, 29(2), 95-114.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specified knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372-400.

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.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(2), 371-406.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and mathematical quality of instruction: An exploratory study. *Cognition & Instruction*, 26(4), 430-511.

Hill, H. C., & Ball, D. L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education*, 35(5), 330-351.

McCrorry, 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.

Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, 13(2):125-145.

Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the diagnostic teacher assessment of mathematics and science (DTAMS) instrument. *School Science and Mathematics*, 110(4), 180-192.

Shulman, L. S. (1986a). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd Ed.) (pp. 3-36). New York: Macmillan.

Shulman, L. S. (1986b). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14

Silverman, J. and Thompson, P. W. (2008). Toward a framework for the development of mathematical knowledge for teaching. *Journal of Mathematics Teacher Education*, 11(6), 499-511.







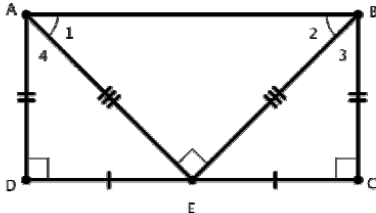
Sinclair, N. and Moss, J. (2012). The more it changes, the more it becomes the same: The development of the routine of shape identification in dynamic geometry environment. *International Journal of Educational Research*, 51&52, 28-44.

Tracy, D. M. and Eckart, J. A. (1990). Five good reasons to use pentominoes. *School Science and Mathematics*, 90(8):665-673.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum.

Zaslavsky, O. and Shir, K. (2005). Students' conceptions of a mathematical definition. *Journal for Research in Mathematics Education*, 36(4), 317-346.

Appendix: Released Items

<p>CCK released item</p>	<p>Students in Mr. Wingate’s class have been creating nets that they want to be able to fold to create a cube. For each of the student-created nets shown below, identify whether it can be successfully folded into a cube?</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 70%;"></th> <th style="width: 15%; text-align: center;">Yes</th> <th style="width: 15%; text-align: center;">No</th> </tr> </thead> <tbody> <tr> <td style="vertical-align: top;"> <p>i</p>  </td> <td></td> <td></td> </tr> <tr> <td style="vertical-align: top;"> <p>ii</p>  </td> <td></td> <td></td> </tr> </tbody> </table>		Yes	No	<p>i</p> 			<p>ii</p> 		
	Yes	No								
<p>i</p> 										
<p>ii</p> 										
<p>SCK released item</p>	<p>While proving a claim on the board about the figure below, Joe wrote “$1 + 2 = 90^\circ$.” Ms. Staples ponders how to correct that statement. Of the following, what is the best alternative?</p>  <p> A Do nothing. The statement is correct as is. B Add the degree symbol so that the statement reads $1 + 2 = 90^\circ$. C Replace what Joe wrote; write instead that “$m\angle A + m\angle B = 90^\circ$”. D Replace what Joe wrote; write instead that “$m\angle EAB + m\angle EBA = 90^\circ$”. </p>									
<p>KCS released item</p>	<p>Ms. Jamison asked students to figure out the number of diagonals in an 11-sided polygon. Of the following erroneous solutions, which one might have resulted from using the number of diagonals through one vertex?</p> <p>A 11</p> <hr/> <p>B 72</p> <hr/> <p>C 88</p> <hr/> <p>D 99</p> <hr/> <p>E 121</p>									

KCT released item	<p>After teaching her students the base angles theorem (which states that base angles of an isosceles triangle are congruent), Ms. Wellington is pondering what example she could give to illustrate how this theorem can help find new information. Which of the following would be the best example?</p> <p>A If you have a triangle that you know is isosceles and you know two of the sides are 8 and 10, then the third side will be 8 or 10.</p> <hr/> <p>B If you have a triangle that you know the sides have length 8, 10, and 12 you know it does not have two angles that are congruent.</p> <hr/> <p>C If you have an isosceles triangle and two of its angles are 30 you can calculate the third angle by subtracting 60 from 180.</p> <hr/> <p>D If you have a kite and draw its minor diagonal, the kite is divided into two isosceles triangles and those have two congruent angles each.</p>
-------------------	---

ⁱ Research reported had the support of the National Science Foundation through grant DRL-0918425 to P. Herbst. All opinions are those of the authors and don't necessarily reflect the views of the Foundation. Some of this material was presented at the 2012 Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. We acknowledge helpful comments to earlier versions of this paper by Deborah L. Ball and by members of the GRIP lab at The University of Michigan School of Education.

ⁱⁱ Daniel Chazan, a co-PI of this project was also involved in design discussions. Individuals involved in the drafting of items, in addition to the authors, included Michael Weiss, Wendy Aaron, Justin Dimmel, Ander Erickson, and Annick Rougee.