KNOWLEDGE FOR TEACHING MATHEMATICS TO LATINO ENGLISH LANGUAGE LEARNERS: AN INSTRUMENT DEVELOPMENT STUDY

DISSERTATION

Presented to the Graduate Council of
Texas State University-San Marcos
in Partial Fulfillment
of the Requirements

for the Degree

Doctor of PHILOSOPHY

by

Aaron T. Wilson, B.S., B.A.

San Marcos, Texas
May 2013

KNOWLEDGE FOR TEACHING MATHEMATICS TO LATINO ENGLISH LANGUAGE LEARNERS: AN INSTRUMENT DEVELOPMENT STUDY

Committee Members Approved:

____
M. Alejandra Sorto, Chair

____
Gilbert Cuevas

____
Joyce Fischer

____
Taewon Kim

Approved:

____
J. Michael Willoughby
Dean of the Graduate College

To the Spirit and to the bride

who both say, Come!

ACKNOWLEDGEMENTS

In the course of completing this dissertation and the work that led up to it, I have been helped and prayed for by many people. And I owe them many thanks.

I am indebted to my advisor, Dr. M. Alejandra Sorto, whose research ideas, collaboration, support, vision and pushing are deeply appreciated. I also am grateful for my other committee members, mentors that provided excellent guidance throughout this work. Dr. Gilbert Cuevas helped me to see what this study was all about. Dr. Joyce Fischer, who first introduced me to the study of the math education of ELLs, and also left no *t*'s un-crossed or *i*'s un-dotted (I think). Dr. Taewon Suh opened my mind to some powerful analytic methods that I didn't know existed and he believed that it could be done.

Many other professors and instructors from Texas State supported me with advice, motivational words, excellent and challenging instruction, or by offering their students, or a timely ear. Thank you, in no particular order, to Dr. White, Dr. Acosta, Dr. Dean, Dr. Morey, Dr. Curtin, Dr. Mireles, Dr. Thickstun, Dr. Obara, Dr. Strickland, Drs. Warshauer, Ms. Betros, and to several others that I must have overlooked. Thank you also to Dr. Edgell who suggested more than once I should look into getting a Ph.D. Also, to the kind ladies in the mathematics office, LaJuan, BeKaye, and Melinda—along with their assistants—I owe many thanks. This could also not have been completed without the Illona's timely advice and fixes.

I appreciated the mutual support of my fellow students throughout our common endeavor. Furthermore, many dear friends, brothers and sisters, have supported this work by their words, prayer and food. Thank you, saints.

My family members also supported me in ways too numerous to list here, as they know. Particular thanks go to mom and dad. Finally, this work could not have been accomplished without the continued support and love from my wife, Mandy, and from Malachi's motivational smiles, giggles and spiders. We made it, Love. We'll be moving now.

This research has benefited from the kind support of the National Science Foundation, DRL-1055067. This dissertation was submitted for final review by the committee on March 19, 2013.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

KNOWLEDGE FOR TEACHING MATHEMATICS TO LATINO ENGLISH LANGUAGE LEARNERS: AN INSTRUMENT DEVELOPMENT STUDY

by

Aaron T. Wilson, B.S., B.A.

Texas State University-San Marcos

May 2013

SUPERVISING PROFESSOR: M. ALEJANDRA SORTO

The purpose of this study was to identify the domains, as well as specific aspects of, knowledge for teaching mathematics to Latino English Language Learners (KT-MELL) and to develop a valid and reliable measure of this knowledge. Latino ELLs are rapidly becoming one of the most populous of student groups in schools—in urban cities especially, but across the nation increasingly. Furthermore, there is evidence that many mathematics teachers of these Latino ELLs may not have encountered or been given the opportunity to learn how to teach these students. A central element of preparation of teachers is the *knowledge* needed by effective teachers. While there exist studies of the knowledge needed to be effective in teaching mathematics, there is only limited, if any, study of what mathematics teachers of ELLs need to know in order to effectively teach

these students. As an initial step toward filling this gap in research, this study sought to identify domains of knowledge needed thought to be important for to teaching math to ELLs.

Based upon rigorous review of the research and practitioner literature, and upon many hours of mathematics classroom observations, this study defined a framework for the pedagogical content knowledge needed by mathematics teachers of Latino ELLs. Three domains of knowledge were proposed: KDIFF-knowledge of the difficulties that Latino ELLs may encounter in learning mathematics, KCAP-knowledge of the special capacities for learning mathematics that Latino ELLs may possess based upon their particular cultural, linguistic and learning backgrounds, and KSTRAT-knowledge of strategies for teaching mathematics to Latino ELLs. This framework is closely aligned with existing frameworks of mathematical knowledge for teaching.

Using this framework and classroom observational data, a survey instrument was developed and field-tested among both pre-service and in-service teachers from diverse regions of the state of Texas. Based upon the results of data analysis methods central to both classical test theory (CTT) and item response theory (IRT), the underlying factor structure of the instrument was identified to be bi-dimensional; KDIFF and KCAP formed a single scale and KSTRAT a separate scale. Furthermore, precise psychometric properties of the items were uncovered, which demonstrated that the measurements obtained had an acceptable degree of reliability within specific ranges of ability of the respondents. Additionally, a number of interesting response patterns were observed along with several important evidences of construct validity.

CHAPTER 1: INTRODUCTION

Statement of the Problem

Many mathematics teachers in the United States may not have adequate opportunity to learn how to effectively instruct their Latino English Language Learners (ELL). Furthermore, the knowledge needed for teaching these students has not been defined with enough precision to allow for the development of instruments designed to measure this knowledge. Although extensive work has been done in defining and measuring mathematical knowledge for teaching (MKT; Hill & Ball, 2009) among general school-aged populations of students, at least two trends call for more a precise definition of this knowledge for the special population of Latino ELLs; while the growth of the number of Spanish-speaking ELLs in public schools has far surpassed the growth of the general school population (Francis et al., 2006; Capps et al., 2005), the gap in mathematics achievement between this group of students and other students has persisted, and in some cases increased, for more than a decade (U. S. Department of Education, 2011). These trends imply that teachers, educators and educational administrators could all benefit from a more precise theory of the knowledge needed for teaching mathematics to Latino ELLs. Such theory could lead, not only to improvements in educational opportunities for mathematics teachers, but also to the development of valid and reliable knowledge measures.

Purpose of the Study

Observing this absence of a precise theory of knowledge needed for teaching mathematics to Latino ELLs as well as an absence of instruments designed to measure this knowledge as situated in specific mathematics contexts, the purpose of this study was to propose, based upon a thorough review of existing research literature, domains of knowledge for teaching mathematics to Latino ELLs (KT-MELL), and to develop an instrument capable of measuring this knowledge with validity and reliability. This study investigated the knowledge that mathematics teachers at the early secondary grades, approximately grades 4 through 8, should possess to be effective teachers of Latino English language learners. Through review of research literature and through teacher observation, this study proposed domains of knowledge for teaching mathematics that seem to be particularly pertinent for teachers of Latino ELLs. The central exploration of this study concerned developing a measure of this knowledge. Hence, the study defined possible aspects of the domains and offered questionnaire items that may serve as indicators for measurement of the knowledge. Data analysis served mainly to investigate evidence of the validity and reliability of the measurements, but also to uncover important patterns in responses.

Significance of the Study

> "Well-designed research has shown that professional development focused around such knowledge results in changed classroom performance and improved student learning." (Hill, Ball, & Schilling, 2008, p. 373).

The results of this study will contribute to the research base of theory about teaching mathematics to Latino English language learning students. By defining domains of knowledge, this study may inform both teacher education and teacher licensure.

Hence, this study may be of interest to mathematics educators involved in designing appropriate instructional experiences for pre-service teachers or in designing professional development for practicing mathematics teachers. Indeed, at the time of writing a school district in the Midwest region of the United States that was currently experiencing rapid growth of its Latino ELL population had contacted the researcher and expressed interest in administering the instrument to teachers of all levels (K – 12) with a view to using the results to inform professional development. Furthermore, because of its measurement component, the study may be of interest to policy-makers and those involved in defining criteria for evaluating the readiness of mathematics teacher candidates to enter the profession.

Another application of the results of this study is related to teacher employment and assignment decisions. Creation of valid and reliable measures of knowledge for teaching mathematics to ELLs is a step forward in the direction of assessing the quality of mathematics teachers for teaching special populations of students. Clearly, much more work would need to be completed for such measures to have predictive validity in relation to student outcomes. However, this research may constitute part of the foundational work toward that end by providing an example of a valid and reliable measure of an important knowledge construct.

A broader implication of the results of this research is the possibility of improving the quality of mathematics instructors for these students and, as a result, narrowing the long-standing mathematics achievement gap between Latino ELLs and other students. Improved teacher assessment means may imply gains in control of teacher quality for specific teaching assignments. Such a result also has implications for providing equitable access to educational opportunities and, consequently, to occupational opportunities.

Research Questions

The specific research questions that drove this inquiry were the following:

1. What are the domains of knowledge needed to teach mathematics to Latino English Language Learners?
2. Drawing from the research literature and middle school mathematics classroom observations, what are some of the aspects of these domains of knowledge when teaching Latino English Language Learners at the middle school level?
3. What evidences of reliability and validity does an instrument developed to measure these domains and aspects exhibit?

Definitions of Terms

- *Latino*(*s*) is used in reference to a person (or persons) that come from Spanish-speaking homes or communities, most of whom, in Texas and the southwestern U.S., are of Mexican descent. This usage is similar to the U.S. Census Bureau's usage, wherein the terms *Hispanic* and *Latino* synonymously refer to "those who classify themselves in one of the specific Hispanic or Latino categories… 'Mexican,' 'Puerto Rican,' or 'Cuban' - as well as those who indicate that they are 'other Spanish, Hispanic, or Latino'" (Hume, Jones, and Ramirez, 2011). For consistency, *Latino* has been used throughout this work. However, it is observed that using such generalizations may cause confusion or misunderstanding. As Téllez, Moschkovich and Civil (2011) comment, the differences among Latinos, because of places of origin, length of residence in the U.S., proficiency with English, etc., are so numerous as to make it difficult to use any label that implies generalizations.

- *English language learner*(*s*) or ELL(s) are those students whose English proficiency, though it is developing, is below average for their grade level because of primary usage at home—or at an earlier time of life—of a language other than English. Similarly, the What Works Clearinghouse of the U.S. Department of Education says that ELLs are students "with a primary language other than English who have a limited range of speaking, reading, writing, and listening skills in English" (2013, January, p. 1). The Department of Education also defines ELL students as "national-origin-minority students who are limited-English-proficient" (Francis et al., 2006, p. 3).

Delimitations

This study involved teachers of Latino English language learning students. Mainly middle school, and a very small number of high school and elementary school teachers, participated. Although some teachers from whom data were taken may have taught non-Latino ELLs, it is assumed with confidence, that most of their ELLs were Latino. Furthermore, the study also involved pre-service teachers, of whom it could not be assumed that they possessed experience teaching ELLs. Furthermore, the participating pre-service teachers may have been in preparation to teach any of elementary, middle or high school mathematics courses. Finally, both the teachers and pre-service teachers that participated in the study were exclusively from the state of Texas.

Summary

Mathematics teachers of Latino English language learners were the focus of this dissertation research. Understanding the dimensions of and creating valid and reliable measures for the knowledge needed by these teachers was the goal. The recent and projected increase in the number of Latino ELLs in schools motivates attention to the needs of these students. Furthermore, a history of underachievement on standardized tests of mathematics proficiency constitutes evidence of the present inadequacy of teachers to provide equitable access to high quality mathematics instruction and learning for these students. Hence, this study sought to understand what are the important domains and associated aspects of the knowledge essential for mathematics teachers of Latino ELLs. This inquiry does not stand alone as an attempt to improve instruction for ELLs. Many states having large percentages of students who are ELLs have structures in place for both training and assessing the proficiency of their teachers to address the needs of these students in their classrooms. Furthermore, researchers and educators in institutions of higher education have embraced the needs of ELLs by drawing up standards for teachers of ELLs and by attempting to treat ELLs' educational issues more frequently in the coursework offered to pre-service teachers. Nevertheless, many of the attempts both by states and by those in higher education have not addressed ELLs in the context of mathematics classes specifically. Current assessments and current teacher standards are exceedingly general and are, in large part, void of a precise mathematics instructional perspective. Because of this, it is questionable whether such assessments can measure teachers' readiness to design mathematics instruction for ELLs. And it is doubtful whether current teacher standards can adequately inform decision-making for teaching specific mathematics topics to ELLs. In order to help mathematics teachers of Latino ELLs there is the need that both theory and assessment be more closely tied to the mathematics that they teach.

Therefore, this study sought to define domains of knowledge that are useful for mathematics teachers of ELLs and to even identify important aspects of each domain. Possession of empirically tested knowledge constructs is essential for the effective design of curriculum materials for both teacher education and professional development. Moreover, an understanding of the proper structure of the knowledge that capacitates mathematics teachers to effectively instruct Latino ELLs is central to any attempt to assess teachers' readiness to teach this population. Therefore, this study may be significant to mathematics educators, mathematics curriculum designers and to educational policy-makers.

Beyond proposing domains of knowledge needed by effective mathematics teachers of Latino ELLs, this study also attempted to give preliminary measures of these domains. It may be inadequate to define the knowledge needed by these teachers without also proposing a method by which said knowledge may be reliably and validly measured. Hence, a major product in which this study has resulted is an empirically tested scale for measuring teachers' knowledge for teaching mathematics to English language learners.

CHAPTER 2: REVIEW OF LITERATURE

Educational Context of the Study

Mathematics teachers are encountering more and more Latino English language learners (ELLs) in their classrooms. While over the last twenty years the number of children aged 5—17 that speak only English at home has decreased by 2.6%, the number of such children that speak a language other than English at home has increased by 138.8% and the number of children that speak English with difficulty has increased by 36.7% (U.S. Department of Education, 2012). These demographic figures are reflected in schools, where ELLs are the fastest growing group of public school students. In comparison with the 12% growth of the general school population in the last two decades, ELLs have experienced a 169% increase; it is predicted that ELLs will constitute thirty percent of the school-age population by the year 2015 (Francis et al., 2006). Moreover, it is also evident that Spanish-speakers constitute the largest group of ELLs. Spanish is spoken by more than 70 percent of ELLs across all school grade levels (Francis et al., 2006; Capps et al., 2005). Thus, Spanish-speaking students, most of whom are Latino students, form the largest body of ELLs in mathematics classrooms in the United States. Statistics like these make it clear that Latino ELLs must form a significant object of the efforts of all educators and that the focus on this population will only increase.

Cognizance of another educational trend has motivated researchers to investigate ELL students and their teachers for more than three decades. Educational statistics like those published by the Center on Education Policy (2010) have shown that a persistent, and large, gap exists in the achievement levels of ELLs and of other students on state standardized tests of learning. In the report cited above the authors observe that, although trends in tests scores for all students, including ELLs, were positive in the three years preceding publication of the findings, yet the trend of the existence of "very large differences in percentages proficient", that is, meeting the standard, between ELLs and other students remains (Center on Education Policy, 2010, p. 2). Similarly, the National Center for Educational Statistics (U. S. Department of Education, 2011) reports that, over the last nearly twenty years, between white and non-ELL Latino students in grade 8 there has been a consistent difference of between 18 and 26 points in mathematics scores on the National Assessment of Educational Progress (NAEP). This trend was even more pronounced between non-ELL Latino students and Latino ELLs where the difference in scores ranged between 28 and 34 points in favor of non-ELLs.

A focus on *achievement gaps* between student groups carries social undertones that have been criticized by some (Gutiérrez, 2008). Drawing attention to disparities between student groups has been seen as offering little contribution toward the advancement of lower performing groups and, rather, as perpetuating negative stereotypes about such groups. Nevertheless, these figures lead to questions concerning the quality of mathematics instruction and of the opportunity to learn mathematics enjoyed by Latino English language-learning students (Abedi & Herman, 2010). Under what conditions are ELLs being instructed in mathematics? What qualifications do their instructors possess? Are teachers adequately prepared to teach their Latino ELL students?

Given their *limited English-proficiency* (U. S. Department of Education terminology), it seems strange that most ELLs should be given instruction in mainstream, English-only classrooms, as is current practice in most states (Wilson, 2011). Furthermore, and of great relevance to this study, since most new teachers should expect to have ELLs in their classrooms (Meskill, 2005), one would expect that structures for assessing teachers' readiness to work with this group of students would be in place.

The No Child Left Behind Act (NCLB, 2001) holds states accountable for attending to the academic progress of their ELL students and also encourages institutions of higher education to prepare beginning teachers to work with these students. How, then, are mathematics teachers being prepared to teach ELLs and what theories are driving such efforts? An investigation of the requirements for certification to teach ELLs seems warranted.

California, Texas, New York, Arizona, and Florida are among the states having the largest numbers of English language learners (Payán & Nettles, 2008). An exhaustive review of their respective departments of education websites revealed that all of these states have made some effort to serve their large number of ELLs. California, perhaps the most advanced, requires all educators that have at least one English learner (EL) in their classrooms to earn passing grades on a series of three CTEL (California Teacher of English Learners) authorization exams. These exams assess teachers' knowledge of language and its development, assessment and instruction of ELs, and ELs' culture and inclusion in classrooms. In Texas, teachers of English as a second language (ESL) are required to pass a certification exam, similar in content to the CTEL exam, but that includes separate sections on academic content knowledge as well. Florida requires its educators to pass a Professional Education exam that includes questions about research-based practices for teaching ELLs. Arizona requires a Structured English Immersion (SEI) endorsement of its teachers; the endorsement is obtained by completing 45 hours of SEI training including instruction and assessment strategies for ELLs. And New York certifies bilingual teachers to teach specific content areas in a foreign language through an exam that assesses their proficiency in the target language. Common to all of these states is assessment of teachers' knowledge of general principles and strategies across differing academic content contexts. Such assessment measures may be reflective of the content of teacher education textbooks which offer "insufficient depth to provide the beginning teacher with meaningful guidelines" (Watson et al., 2005, p. 148).

Furthermore, there is evidence that, at least at the elementary school level, these types of preparation have minimal impact on the achievement of ELL students (Williams et al., 2007). The study in question found no correlation between students' academic achievement and the number of teachers in the school holding special credentials for teaching ELLs. Thus, the question of whether *mathematics* teachers in these states may have adequate preparation for teaching ELLs remains open. More precisely, current means for assessing the preparation of mathematics teachers for teaching ELLs seem to be lacking in specificity. This apparent deficiency draws attention to theory behind teacher preparation.

Researchers concerned with the assessment of ELLs have asserted that "with the rapid growth of ELL populations, states should place a substantial focus on increasing teacher knowledge of current ELL issues…including pre-service teacher education and continuing teacher education" (Wolf, Herman, & Dietel, 2010, pp. 8—9). But, how should states do this and what are the most poignant elements of this teacher knowledge? After taking a look at teacher licensure structures in view of preparing teachers of ELLs, one could hope to find more answers in the theory behind the preparation of teachers. Knowledge needed for teaching mathematics to ELLs would seem central for the adequate preparation and professional development of mathematics teachers of this population. Furthermore, if such preparation would involve—beyond classroom practicums and observations of experienced teachers—explicit instruction, then what should be the content of such instruction and what knowledge informs the choice of said content? Three exemplary theoretical solutions are presented below.

First, researchers at TESOL (Teachers of English to Speakers of Other Languages, Inc.), jointly with NCATE (the National Council for the Accreditation of Teacher Education), have produced 11 standards (Appendix A) for recognizing ESL education programs. These standards are grouped into the five domains of language, culture, instruction, assessment and professionalism. These standards are very similar in content to those addressed by the certification exams previously discussed and they appear to be equally lacking in specificity to the task of mathematics teaching.

Second, the National Clearinghouse for English Language Acquisition (NCELA, 2010) has offered sixteen research-based "fundamentals for every successful teacher of ELLs" (Leier & Fregeau, 2010, p. 22). These sixteen content knowledge and skill areas (Appendix B) are grouped under the four categories of *language*, *culture*, *policy*, and *teaching* and are proposed to be crucial understandings for effective teachers of ELLs, regardless of the academic subject being taught. Included are basic language acquisition principles, acculturation and cultural awareness principles, legal provisions for education of ELLs, and four teaching principles that closely align with the Sheltered Instruction Observation Protocol or SIOP—making input comprehensible, including both language and content objectives, using cognitive and cultural scaffolding, and using translation services (Echevarría, Short, & Vogt, 2007). These principles, which may be important as foundational understandings for teachers of ELLs, again offer little suggestion of how mathematics teachers should think about their ELLs as mathematics learners or about their mathematics instruction for ELLs.

As a third approach to addressing the need of preparing teachers to teach ELLs, some researchers and educators have made efforts at "infusing English language learner issues throughout professional educator curricula" (Meskill, 2005, p. 739). This effort responds to the observation that "unfortunately, subject matter courses in teacher preparation programs tend to be academic in both the best and worst sense of the word, scholarly and irrelevant, either way remote from classroom teaching" (Ball, Thames & Phelps, 2008, p. 404). The Training All Teachers project, reported upon by Meskill (2005), involved university teacher educators, graduate students, and practicing school teachers in professional development centered around the topics of language acquisition, culture, regulations, and communication. The mathematics educator involved in the project, who had previously "thought that ELLs were in a 'sink or swim' position in the classroom" (p. 746), "became aware that not every ELL received adequate support" and, consequently, "considered it even more important for all teachers to design instruction based on individual learner needs" (p. 747). However, beyond the usage of hands-on materials and diagrams, there is little suggestion of how mathematics teachers should design such instruction, much less of how they should think about the task of teaching mathematics to ELLs.

Current test measures for assessing teachers' preparation for teaching mathematics to ELLs seem to be lacking in specificity to the task of mathematics teaching. More importantly, there is an apparent absence of theory that could drive the improvement of such assessments. Although some states require teachers to be conversant in general ELL strategies, they do not require that teachers know how to use them in specific mathematics instructional contexts. It is possible that many teachers are unable to do this. Furthermore, while there exists theorization about the fundamental elements of effective instruction of ELLs, it seems to be difficult to find theorization, much less empirical evidence, related to knowing how to teach mathematics to ELLs.

Theoretical Framework

To address this shortage in theory, the purpose of this study was to investigate middle school mathematics teachers' knowledge for teaching mathematics to English language learners (KT-MELL). The investigation included identification of important knowledge domains and also development of an instrument useful for measuring the knowledge. Implicit in this purpose is a framework for approaching both the theory that informs and the methodology that respond to the questions of this study. This study looked at teachers' knowledge through the lens offered by Lee S. Shulman in 1986 when he coined the term *pedagogical content knowledge* (PCK) to describe the form of knowledge that teachers invoke in their work of classroom teaching. Such knowledge is a composite of content—in this case mathematics—and of pedagogical knowledge. The theoretical framework for this study was based on the theory of teachers' knowledge offered by Shulman and, more recently, elaborated by Hill, Ball, & Schilling, (2008) in their findings concerning the domains of PCK. The methodological framework was based on measurement of teachers' knowledge for teaching mathematics (Hill, Schilling, & Ball, 2004) and will be addressed in Chapter 3. Based upon this understanding of teachers' knowledge, the first part of the literature review serves to uncover what has been learned not primarily of teachers' actions but of their knowledge applied to the mathematics teaching of Latino English language learners (ELLs). Nevertheless, in accordance with the PCK framework, this study assumed that teachers' actions are motivated by their knowledge and, thus, have direct implications for the elements of their knowledge. Since the research questions of this study concern mathematics teachers' knowledge, the ultimate section of the literature review returns to treat this topic in more depth.

Consequently, the research that informed this study came from two directions: the mathematics education of *Latino* ELLs, and mathematics teachers' knowledge. Since this study sought to determine what teachers need to know in order to help Latino English language learning students to be successful in mathematics, it necessitated consideration of the significant body of research focused specifically on the mathematical learning and teaching of ELLs. Fortunately for this study, most of the research concerning ELLs in the United States has dealt with *Latino* ELLs, since this group composes the vast majority of the ELLs in this country. Thus, such research serves as the point of departure for this attempt to shed light upon the complex nature of learning and teaching mathematics for this group.

Furthermore, since a central question of this research concerned the knowledge that teachers should possess in order to elevate the mathematical understanding of these students, research into teacher knowledge, especially into knowledge for teaching mathematics, played a central role in this study as well. Like research concerning mathematics education of ELLs, this later body of research is also quite extensive; the search for the nature of knowledge needed for teaching has resulted in many well-worn trails. Some of these trails are relatively new and appeared to offer a useful route toward answering the questions of this study. These will be discussed in the later part of the review of the literature.

Mathematics Education of Latino English Language Learners

This search to identify the knowledge needed by mathematics teachers of Latino English language learners began with careful consideration of the research concerning the mathematics teaching and learning of these students. Researchers have attended to the special needs of these students for many years. An extensive review of this literature has revealed that there is a natural division of the research that is roughly chronological. There are three sections corresponding to three foci taken by researchers concerned with the mathematics education of ELLs: *difficulties faced by Latino ELLs in mathematics classrooms*, *assets possessed by Latino ELLs for learning mathematics*, and *strategies for teaching mathematics to Latino ELLs*. Moschkovich (April 2007, 2002) has observed that earlier studies of the mathematics education of ELLs most frequently took a deficit perspective, a perspective that called attention to the difficulties and challenges that ELL students and their teachers face in the mathematics classroom. Furthermore, she noted that more recent studies have taken a *sociocultural* or affordance perspective, one that attends to the capacities that these students have for making meaning in mathematics classes. Finally, attention must be given to the large number of strategies that have been posited for helping ELLs in mathematics classes.

This categorization of the literature fits well with the research questions of this study when seen through the theoretical framework described above. As is addressed in more detail in the later part of this chapter, *pedagogical content knowledge*, as conceptualized by Shulman (1986), informs educators both of the difficulties that learners face when learning specific academic content and of things that empower them to learn said content. Hence, knowledge both of difficulties faced by ELLs in mathematics classrooms and of capacities that these students have for learning mathematics would seem central as elements of mathematics teachers' knowledge for teaching this population.

Difficulties Faced by Latino ELLs. This section discusses studies that have resulted in uncovering the difficulties that Latino ELLs can have in the mathematics classroom. Moschkovich (2002) has defined two distinct deficit perspectives taken by the research: the vocabulary acquisition perspective and the multiple-meanings perspective. From the vocabulary acquisition perspective, the knowledge—or lack thereof—of vocabulary plays a central role and difficulty in learning and teaching mathematics to ELLs. From the negotiating multiple-meanings perspective, however, a central difficulty that ELLs face in the mathematics classroom is the different meanings that words assume when used in mathematical contexts rather than in everyday contexts. This way of thinking may help to understand the difficulties of ELLs presented below. The difficulties faced by Latino ELLs in mathematics classes can broadly be categorized as: language proficiency difficulties, difficulties with word problems, difficulties with instructional format of classroom, and difficulties with assessment.

Difficulties with Language Proficiency. Latino ELLs in mathematics classrooms find themselves in the simultaneous roles of mathematics students and English language students. Since these students have been viewed as being *limited-English-proficient* and since many mathematical tasks in the classroom involve understanding problem situations set forth verbally, it is reasonable to first consider the limited proficiency as a barrier to learning math. Cuevas (1984), working with Latino ELLs in Miami, was one of the earliest researchers to inquire into the difficulties faced by these students. Two classrooms of first-grade students were selected in successive years of the study, one lower performing and one higher performing. Both classrooms contained mostly Latino students. This researcher sought to observe the difficulties that second-language learners have in learning mathematics. The two central questions of this research concerned whether the concepts were presented in a manner that matched the developmental level of the students and which language, English or Spanish, was the better language of instruction for the students. Students were evaluated in terms of their linguistic ability and their ability with number, numeration and simple mathematical operations. Cuevas (1983) concluded that some mathematical tasks were beyond the students' level of maturity and could be postponed until a later stage. Furthermore, he also found that the language of instruction caused difficulty for Spanish monolingual students. The students may have benefited more from mathematical instruction in their own language, that is, in Spanish. Finally, it is worthy to note that this research also observed that bilingual Latino students performed better in mathematics tasks than did either monolingual Spanish-speakers or Latino monolingual English-speakers. This result coincides with more recent research that suggests the advantage that bilingual math learners may possess (Téllez, K., Moschkovich, J., & Civil, M., 2011; Moschkovich, 2002).

The Cuevas (1984) study just discussed looked at the ways in which deficiencies in linguistic proficiency hindered students' access to mathematical concepts. Also taking the view that linguistic deficiency causes difficulty in learning mathematics for ELLs, Lager (2006) used an extremely narrowly focused lens to determine precisely which words caused the most difficulty for ELLs and non-ELLs on an assessment of algebraic thinking involving determining formulas from geometric patterns. A sample of 221, mostly (82%) Latino sixth and eighth grade students, 133 of whom were ELLs, took the assessment. From this sample, 26 students were interviewed concerning the errors that they had committed on the test. Lager (2006) categorized the errors that students had made as errors of confusing words or phrases, shifts of application—for example, using the term *number* in its cardinal sense versus nominal sense—, and polysemy—words having two or more different meanings. This type of error is an example of the difficulty caused by the multiple-meanings (Moschkovich, 2002) that words can take on in mathematics classes. A major finding of the work was to conclude that the words *pattern*, *previous*, and *extension* were the most troubling to students. Furthermore, the researcher noted that while *pattern* is part of the mathematical register, *previous* and *extension* are everyday—but abstract—terms being used to describe the meaning of *pattern*. The ELLs' linguistic deficiency is seen as obscuring the meaning of the words and, thus, the mathematics.

Difficulties with Word Problems. The two studies just discussed looked at the ways in which ELLs' limited English proficiency and lack of familiarity with certain vocabulary have hindered their achievement in mathematics. Other research has focused on the difficulty that word problems specifically can pose to ELLs. Martiniello (2009) investigated nonmathematical linguistic complexity as a source of difficulty for ELLs in word problems. By investigating the *differential item functioning* (DIF), that is, the ways in which the varying levels of linguistic complexity of items affects the performance of ELLs and non-ELLs of similar ability levels, the researcher found that "the greater the item nonmathematical lexical and syntactic complexity, the greater are the differences in difficulty parameter estimates favoring non-ELLs over ELLs" (p.160). This would be expected. But, what may be even more informative for this study was Martiniello's (2009) finding that inclusion of nonlinguistic representations reduced the difficulty that linguistic complexity caused for the ELLs. This could be interpreted as empirical evidence that pictures, diagrams, and other semantic mappings can be used to help ELLs make sense in mathematical learning and problem solving tasks. However, usage of visual images does not guarantee comprehension—Lager (2006) noted that many students missed a problem that was assumed to be accessible to them because of the clarity of the picture. Nevertheless, evidence that schematic representations may have the capacity to provide specific assistance to ELL students that have difficulty with verbal representations is noteworthy for this study.

Difficulties with Classroom Format. Some researchers have tried to understand the ways in which instructional practices affect language minority students. Chang (2008) used data from the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K), available from the National Center for Education Statistics (NCES), to assess the extent to which four different types of student grouping strategies—teacher-directed whole class activity, teacher-directed small-group activity, teacher-directed individual activity, and student-selected activity—were related to gains or losses in mathematics achievement among the different ethnic, socio-economic, and language proficiency groups. The comparison of grouping strategies revealed that, compared to Caucasian and African American English-only students, whose achievement improved, Latino ELL students' achievement suffered under the whole-class, teacher-centered instructional format. This study gave no indication of the reason for which Latino ELLs' scores diminished under this type of instruction. However, it would appear that teachers' usage of English as the primary language of instruction under this mode is related. Cuevas's (1983) conclusion that Spanish may be a more appropriate language of instruction for elementary school Latino ELLs would seem to confirm this reasoning.

Difficulties with Assessments. Llabre and Cuevas (1983) observed that, "Given the inevitable dependence of achievement and intelligence tests on language, it seems logical to assume that any such tests in English would have decreased validity and reliability for nonnative English speakers" (p.318). To better understand how the language in which a mathematics test is given affects student outcomes on the test, these researchers administered two equivalent versions of a mathematics achievement tests, one in Spanish and the other in English, to 408 bilingual 4th and 5th graders in Miami, Florida. Each child took both versions of the test, the language order being counterbalanced in the sample. A major finding of this work was that the students performed better on the English version of the test. This seemed to contradict the researchers' initial doubts concerning validity and reliability; one would have expected that a significant number of these bilingual students would have performed better on the Spanish version. However, the researchers were careful to note that these students were selected because of their possession of certain minimal proficiency levels in both languages.

More pertinent to the notion of mathematics teacher knowledge may be Llabre and Cuevas's (1983) finding concerning the levels of achievement on two broad types of questions. The researchers found that the items on the instruments divided themselves broadly into *conceptual* and *application* questions. While students generally performed better on conceptual problems than they did on application problems, the researchers' observed that students with higher levels of English comprehension performed more equally on application problems and comprehension problems. Although the article does not explain the difference between these two types of problems, one can guess that the difference was between computational problems that tested comprehension of mathematical concepts and word problems that tested students' ability to *apply* the computational ability in a problem situation. This interpretation is supported by Moschovich's (2002) criticism that much of the early research in mathematics education of ELLs focused on narrow definitions of classroom mathematical activity, including arithmetic computation and solving word problems. The interpretation of *application* problems in the Llabre & Cuevas (1983) study as *word* problems along with their finding that bilingual students performed more poorly on such problems echoes other research concerned with the difficulty that ELLs have with interpreting linguistically complex mathematical tasks (Lager, 2006; Martiniello, 2009). This research points to the complex issue of test-language in mathematics assessments of ELLs as well as to the difficulty that all students, and ELLs especially, have in solving word problems on assessments.

Summary. This section has looked at the difficulties that Latino ELLs in the elementary and middle school grades can have in the mathematics classroom. Broadly, these difficulties may be classified as language proficiency difficulties, difficulties with word problems, difficulties with instructional formats, and difficulties with assessments. There is evidence that ELLs' limited proficiency with English can interfere with their mathematics learning in English classrooms. When working with word problems, beyond the basic difficulty that many students have of mathematizing verbal problem situations, it appears that ELLs can struggle, more than do other non-ELL students, with certain words because of not knowing the words or because the words have been applied in ways in which they are not familiar. ELLs in the early grades appear to also have difficulty when learning mathematics in an English-speaking classroom in which the teacher directly instructs all students as a class, rather than working with individual ELLs alone. Finally, assessment poses a particular challenge for some Latino ELLs. While students that have some English fluency may perform better on assessments given in English, the language of instruction, it is also probable that they may struggle more with application problems than with more simplistic conceptual, that is computational, problems.

Mathematical Learning Capacities Possessed by Latino ELLs. The foregoing section has reviewed the difficulties in learning mathematics that research has ascribed to Latino English language learners. Although these difficulties may seem to, at least partially, explain the observed differences in mathematics achievement between Latino ELLs and other groups, studies that focus on the limited ability of these students have been criticized as offering little constructive contribution to the understanding of how these students learn and, hence, of how to elevate their achievement (Moschkovich, 2007). This section will consider studies that have drawn attention to the assets or capacities that Latino ELLs bring to the classroom, and to ways that teachers have drawn upon them, for learning mathematics. It begins with a definition of mathematical competence that is central to many of these studies, one that goes beyond mere arithmetic fluency and problem-solving skill.

Moving to More Inclusive Views of Mathematical Competence. Perspectives on what constitutes mathematical activity and mathematical competence have developed over the last twenty years. To a great extent this move has gained momentum through publications of standards for teaching and learning mathematics like those offered by the National Council of Teachers of Mathematics. The NCTM (2000) vision of mathematics education defines mathematics learning as being a communicative activity in which students should "communicate their mathematical thinking coherently and clearly to peers, teachers, and others" (p. 61). Researchers like Moschkovich (2002) have embraced this more recent, broader view of mathematics learning and have criticized earlier deficit views of ELLs' math learning as being too simplistic in their definition of what constitutes mathematics learning in the classroom. She has pointed out that mathematics learning involves more than memorizing mathematical vocabulary and solving word problems, more even than negotiating the different meanings of words in the everyday and mathematics registers. To researchers like Moschkovich (2002), mathematical communication involves more than simple arithmetic and problem solving; it is characterized by explaining solutions, describing conjectures, proving conclusions, and presenting arguments.

Linguistic Creativity and Mathematical Argumentation Ability. Building on Gee's (1999) definition of Discourse, Moschkovich's (2002) defined participation in mathematical discourse thus:

> "Participating in classroom mathematical Discourse practices can be understood in general as talking and acting in the ways that mathematically competent people talk and act." (p. 199)

This understanding has led her to observe mathematics ability in Latino ELLs that may have otherwise been overlooked if language proficiency had been the focus. Moschkovich (2002) offered two examples of Latino ELL students, between grades 6 and 9, demonstrating proficiency with mathematical objects and explanations while in neither case were all of the correct English mathematical words found. In the first case, a group of four students differentiated the properties of a quadrilateral with four right angles from other quadrilaterals and, in the process, created the word *rangle* to denote such shapes (for ignorance of the word *rectangle*). In the second case, a Latino student, engaged in a discussion concerning the slope of the graph of a linear function, used mainly Spanish to justify a statement concerning the steepness of the line. The mathematical strength of the argument was found not in the correctness of the words but in the correctness of the mathematical practice of justifying a claim by invoking assumptions about the mathematical situation and by using mathematical representations as examples to support the claim. Moschkovich (2002) used these vignettes to demonstrate the ways in which a deficit perspective of the mathematics learning of Latino ELLs can fail to capture the learning that is actually taking place. These two cases exemplify the communicative ability that Latino ELL students may possess based upon their linguistic resourcefulness—seen in the creation of new words to describe mathematical objects—and on their ability to argue mathematically in their fluent tongue.

By embracing a broader view of what constitutes mathematical discourse, some researchers (Téllez, K., Moschkovich, J., & Civil, M., 2011; Moschkovich, 2007) are attempting to move away from a deficit perspective of ELLs as being *less capable* toward an affordance perspective in which alternative forms of mathematical expression—forms that may not require the usage of correct or standard terminology—can be seen as valid demonstrations of mathematical proficiency. In their view, ELLs can show signs of advanced mathematical proficiency that are missed by focusing simply on the language used or on the precision of the arithmetic operations.

Discursive and Communicative Ability. Researchers have investigated the role that participating in mathematical discourse plays for Latino ELLs. Khisty and Morales (2004) have considered the interplay of students' academic language proficiency and mathematics teaching and learning. Although they adopt the deficit viewpoint held by language acquisition theorist Cummins (1981), believing that ELLs' limited knowledge of academic language is a major limitation to their mathematical learning, yet their case study presents examples both of how academic language deficiency hinders ELLs' math learning and of how ELLs' Spanish fluency can actually capacitate them to think mathematically and to express mathematical ideas. One high school pre-calculus student is barely able to utter a mathematical thought in response to the teacher's questioning and this is attributed both to the teacher's excessive control over the conversation and to this student's severe deficiency in academic language. However, another pre-calculus student, of the same teacher, gives rich explanation of the same exponential function question using her native Spanish tongue. It is noted that neither student had extensive command of English academic mathematical language and a conclusion of the study is that learning the mathematical words is essential, even precedent, to obtaining mathematical literacy and mathematical thought. However, that the Spanish-speaking student explained the concept with such accuracy, in Spanish, seems to warrant more attention to the ability that this student possesses than was given in the article. In fact, perhaps the most striking finding, from an affordance perspective, is that Latino ELLs can demonstrate extensive mathematical understanding in their native tongue, when allowed such expression. It would seem that more mathematical understanding could have been observed in the language deficient student, had he been given opportunity to speak in Spanish.

Linguistic and Cultural Identity. Researchers have observed that Latino students, many of whom frequently have the additional status of being from low income homes, have sometimes been perceived by their teachers as having inferior support structures for learning mathematics in their homes and as having diminished motivation for learning mathematics (Anderson & Tate, 2008). Nevertheless, some researchers, and educators, have taken a more positive view of the backgrounds of Latino students and have observed teachers that capacitate their students to learn mathematics by relating the mathematics to students' culture—making mathematics culturally relevant (Gutstein et al., 1997)—and by allowing students to work in their primary language and drawing upon their previous knowledge (Gutiérrez, 2002). Henderson & Landesman (1995) evaluated the effectiveness of making mathematics socio-culturally relevant for Mexican and Mexican-American 7th grade students by comparing levels of student achievement in two treatment groups: thematically integrated mathematics instruction and traditional mathematics instruction. The themes in the experimental treatment were jointly selected by students and teachers and were believed to be relevant and of interest for the students. While the study found no difference in computational skills gained, in attitudes about mathematics or in student self-perceptions of mathematical motivation, students participating in thematically integrated instruction showed greater achievement in mathematical concepts and applications. Both this work and that by Gutstein et al. (1997) and Gutiérrez (2002) indicate that the lived experiences and interests of Latino students can positively affect their mathematics achievement when incorporated into mathematics instruction. To the researchers and educators who view them thus, the cultural and linguistic experiences of their Latino ELL students can become assets that can both motivate and empower these students to learn mathematics.

Fischer & Perez (2008) have provided an actual example of how a mathematics teacher of high school aged Latino ELLs can effectively use students' linguistic and cultural backgrounds to strengthen their mathematical learning. The teacher in the case-study demonstrated effectiveness in improving the mathematics achievement of his Algebra 1 students. The research reported that, as a long-term trend, on average 70—90% of his students passed the state standardized mathematics test every year compared with the 15—25% passing rate at the school. The teacher's method involved a thoughtful sequencing of instructional decisions including: activating prior student knowledge through brainstorming, directly teaching and reinforcing Spanish and English academic vocabulary through definitions and cognate usage, specific instruction in translating between verbal and algebraic expressions/equations, gradually increasing the level of difficulty, and motivating students through using their names and products in the classroom activities. This teacher's method exemplified an affordance perspective of Latino ELLs' mathematics learning; the teacher drew upon students' ideas and upon aspects of students' cultural and linguistic background to capacitate them to learn mathematics, rather than seeing these as a barrier to acquiring mathematical knowledge.

The Issue of Equity. Some researchers have attended to the inequities that are evident in schools and in mathematics classrooms. Taking narrow definitions of mathematics and of what constitutes participation in classroom mathematical activity and of evidence of mathematical achievement can have a negative effects on Latino ELLs' learning of mathematics (Moschkovich, 2007; Gutiérrez, 2002, 2009; Téllez, K., Moschkovich, J., & Civil, M., 2011). Additionally, other researchers (Ball, Hill, & Bass, 2005), involved in measuring teachers' mathematical knowledge for teaching, have found that "higher-knowledge teachers tended to teach non-minority students, leaving minority students with less knowledgeable teachers who are unable to contribute as much to students' knowledge over the course of a year" (p. 44). This finding may be explained by the comparatively low level of preparation for mathematics teaching experienced by some teachers of poor students and of minority students. In their review of the research literature concerned with inequities in school mathematics, Anderson & Tate (2008) observe that students who have been traditionally marginalized are more likely to be taught by math teachers who hold no degree or certification in mathematics; this trend was most prevalent among the highest poverty schools and in classrooms with the most minority students, they say. Clearly, this situation is in direct contradiction to NCTM's (2000) first principle for school mathematics: the equity principle "requires resources and support for all classrooms and all students" (p. 14).

Summary. This section has presented an affordance perspective of Latino ELLs in the mathematics classroom. While much research has focused on the difficulties and linguistic challenges that Latino ELLs face in mathematics classes, other research has seen the unique cultural background and linguistic background of these students as powerful for capacitating them to learn mathematics. Specifically, this section has highlighted the ways in which Latino ELL students may use their linguistic resources in creative ways to communicate mathematically. It has also shown how bilingualism can capacitate these students to participate in mathematical discussions. Furthermore, this section has summarized evidence showing that Latino ELLs' linguistic and cultural identity can be positively activated in mathematics classrooms to bolster learning.

Even though they may lack the correct English words, Latino ELLs have demonstrated accurate mathematical understanding and explanation (Khisty and Morales, 2004). This observation is fundamental to the belief held by some researchers who embrace an affordance perspective of the mathematics learning of Latino ELLs, observing in Latino ELLs their communicative ability rather than inability (Moschkovish, 2002, 2005; Carrasquillo & Rodriguez, 2001). In classrooms where mathematics is seen as a communicative activity, rather than mere computation and vocabulary acquisition, the communicative ability and interests of Latino ELLs has been positively drawn upon to help these students acquire mathematical proficiency. This section also briefly reviewed research that shows ways in which deficit perspectives can lead to inequities in the opportunity that Latino ELLs have for learning mathematics. Motivated by the problem of this inequity, researchers like Mochkovich (2002) have argued that, rather than focusing on the mathematics that these students are incapable of doing, it may be more informative to observe these students using whatever communicative resources they find in their experiences and environment to learn and do mathematics in the ways that can be considered mathematically sound.

The previous two sections have presented perspectives most frequently taken in consideration of the mathematics learning and teaching of Latino ELLs: the affordance and deficit perspectives. From the theoretical framework of pedagogical content knowledge adopted for this research, teachers need "an understanding of what makes the learning of specific topics easy or difficult" (Shulman, 1986, p. 9). Hence, research results from these perspectives seem to directly inform what teachers need to know about teaching mathematics to these students, which is directly related to the central questions of this dissertation research.

Up to this point the review of literature has been concerned with the Latino ELLs' *learning* of mathematics. That is, the focus of research considered thus far has been on the qualities of these students as they affect their capacity to achieve mathematically. In response to—and actually mingled throughout—this literature is a solid body of instructional interventions that are posited to advance the mathematics achievement of ELLs. The next section will give expression to these.

Mathematics Instructional Strategies for Teaching Latino ELLs. What teachers do with their students has a powerful impact on the achievement of their students. Because of this, researchers and educators have attended not only to what teachers do, but to what they *could* or *should* do to improve the mathematics achievement of their ELLs. Therefore, this section of the review of the literature is concerned with instructional choices that are believed to be of benefit in helping Latino ELLs to learn mathematics. This section will begin by considering instructional models and strategies for usage with ELLs generally and then for teaching *mathematics* to ELLs specifically. In line with the theoretical framework adopted for this study, these strategies may be seen as forming another central component of teachers' *pedagogical content knowledge*.

Instructional Models and Strategies for ELLs Generally. In attempting to identify the instructional factors that lead to achievement for ELLs in schools, Thomas & Collier (2002) have undertaken large longitudinal comparisons of instructional models for teaching ELLs in the U.S. Their focus was on identifying the most beneficial mix of language and content instruction for the appropriate age and linguistic proficiency groups of students. After comparing state standardized test data on more than 210,000 students across the nation, these researchers determined that students, both ELL and monolingual English speakers, showed the greatest levels of achievement when enrolled in two-way bilingual immersion programs, programs in which at least 50% and up to 90% of academic content instruction is given in *L2*, that is, the second language, and the remainder in the first language. These students attained to between the 50^{th} and 83^{rd} percentile on state tests, whereas students enrolled in *pull-out* programs, in which ELLs are briefly removed from class for instruction in their native language and then returned to traditional classes, reached between the 11^{th} and 18^{th} percentiles only.

A strategy frequently associated with bilingual academic content instruction is *sheltered instruction*. Thomas & Collier (2002) define it thus: “The integration of language and content instruction, where teachers use strategies such as speaking slowly and clearly (but using natural language), using visual aids and manipulatives, and building on prior knowledge” (p. 10). They also found this instructional strategy, being a fundamental component of language immersion programs, was among the most effective for improving the achievement of ELL students. Echevarría, Short, & Vogt (2007) formalized sheltered instruction through creation of the SIOP or Sheltered Instruction Observation Protocol. Initially the SIOP was a tool to collect data on the teachers involved in their research, research aimed at refining professional development for teachers of ELLs. The SIOP contains thirty strategies for classroom practice that are believed to assist English language learners in any academic content area. The strategies are grouped into the three instructional goals of Preparation, Instruction, and Review/Evaluation, and are evaluated on a five-point Likert scale. The Instruction section contains the largest group of strategies (20). Two examples of strategies useful for *Building Background* are that teachers should explicitly link concepts with students' background experiences and link them with their prior knowledge. Examples of strategies useful for promoting *Interaction* are that teachers should use student grouping configurations that support language and content objectives and should provide *wait time* for student responses, allowing them time to think about question prompts. The SIOP is worthy of note as an important development in strategies for teaching all subjects to ELLs because of its proliferation; the framework and associated professional development programs have been widely adopted by teachers, school districts and professional developers of educators in the U.S. as fundamental to teaching the academic contents to ELLs. Notwithstanding its widespread adoption as an instructional strategy for teaching ELLs, the What Works Clearinghouse (2013, February) has issued a report calling for the need of more research that can establish the effectiveness of the SIOP framework of instruction.

Instructional Models for ELLs in Mathematics Classes. The instructional models and strategies discussed above apply to general classrooms and not to mathematics classrooms specifically. They are useful both to researchers and educators in laying a general understanding of how to help ELLs in schools. Several researchers have attempted to draw up instructional models and strategies for helping ELLs to learn mathematics specifically. One such model is the SLAMS (Second Language Approach to Mathematics Skills) (Cuevas, 1984). The SLAMS model incorporates Cummins's (1981) theory of language learning as composite both of the general, everyday language and of academic language components. The SLAMS model suggests that, parallel to planning of mathematics content interventions, teachers should consider the parts of language that will be required in learning the mathematics and should explicitly address these through language instruction. This strategy is also found in the SIOP, which assesses the extent to which both content and language objectives are "clearly defined, displayed, and reviewed with students" (Echevarría et al., 2007, p. 288).

Similar to SLAMS, Chamot & O'Malley (1986) incorporated both academic language and mathematics instruction, but added the additional component of specific instruction in learning strategies in their ELL instructional model, the Cognitive Academic Language Learning Approach (CALLA). CALLA was created as a means of assisting ELL students to transfer from bilingual education programs to mainstream classrooms in which instruction is predominantly in English. A comparison of traditional classrooms and classrooms implementing CALLA found that, when the model was implemented with a high degree of fidelity, ELL students were found to complete sequences of problem solving steps with more accuracy (Chamot et al., 1992). The authors claimed that such a finding constituted evidence of the value of providing specific problem-solving instruction for ELLs. CALLA has also been implemented in a school district in south Texas under the acronym CAPE (Content Area Program Enhancement) (Montes, 2002). In the south Texas school district, 94% of the students in the district were Latino, 25% of which were ELLs. Montes (2002) concluded that ELL students enrolled in CAPE showed greater signs of improvement on the state's standardized mathematics test than did students enrolled in non-CAPE courses. This may constitute further evidence of the effectiveness of the CALLA instructional model.

To address the challenges faced by ELLs who are simultaneously learning English and Mathematics, Diaz et al. (2011) have presented a framework for "developing mathematics literacy for bilingual learners" (title). This instructional model ties together three theoretic strands: principles of learning, effective pedagogy, and second language acquisition. The authors hypothesized—and illustrated by means of an imagined mathematics classroom scenario—that usage of these three strands results in a learner-centered environment conducive to obtaining literacy in mathematics. The authors emphasized that allowing ELLs "to discuss the mathematics requirements of the lesson in both English and Spanish is empowering and fundamentally important in supporting a learner-centered environment…. [The proposed model] is based on the premise that teachers must view bilingualism as a strength and not as an obstacle to teaching and learning" (p. 17). Thus, central to this model is the affordance perspective of ELLs that holds bilingualism as a benefit for their mathematics learning (Moschkovich, 2002; Gutiérrez, 2009).

Instructional Strategies for ELLs in Mathematics Classes. In addition to broad instructional models and frameworks for teaching mathematics to ELLs, a large number of specific instructional strategies have also been offered. Indeed, many books have been written that concern strategies for teaching ELLs *generally*. However, this section considers some of the research and summarizes many of the strategies suggested that are aimed at specifically improving the *mathematics* achievement of ELL students.

Entities that support the efforts of teachers, mathematics educators, and researchers have all generated lists of instructional strategies aimed at improving the achievement of ELLs in mathematics. Some make very general suggestions like those offered by the Center on Instruction (Francis et al., 2006). They recommend that (a) “ELLs need early explicit and intensive instruction and intervention in basic mathematics concepts and skill”, (b) ELLs struggle with academic language and (c) ELLs need support in learning academic language in order to solve word problems (pp. 35—39). Because of their emphasis on basic mathematics concepts and skills, academic language and word problems, these recommendations seem to come from an understanding of the mathematics education of ELLs that is informed by earlier research and a deficit perspective of ELLs’ learning of math (Martiniello, 2009; Lager, 2006; Cuevas, 1984; Llabre & Cuevas, 1983).

r4 Educated Solutions (2010), which provides professional development, training and information to school districts in Texas and Canada having large numbers of ELLs, offers a different focus in its suggestions for making math accessible to English language learners. Strategies are grouped according to three types of support that can be applied to learners of different English proficiency levels: affective supports, linguistic supports, and cognitive supports. The myriad specific strategies recommended for making the three types of support—such as smiling at students (affective), using *word-sorts* (linguistic) and using the *see-plan-do-reflect* model (cognitive, adapted from Polya (1957))—are largely educational strategies found in many educational methods texts.

Coggins, Kravin, Coates, & Carroll (2007) have drawn, from research concerned with educating ELLs, seven strategies for helping ELLs in mathematics classes. Based on Cummins' (2000, 1979) conceptualization of language acquisition, the first two strategies concern developing conversational language and developing academic language, respectively. Suggestions for developing conversational language include structuring a conversational classroom climate, using group work and providing ample opportunities for discourse. Suggestions for developing academic language are similar to those found in other places (r4, 2010; Kersaint et al., 2009), including, among others, using Spanish-English cognates, word walls, and modeling correct usage of academic vocabulary. Based on Vygotsky's (1978) theory of learning and on Echevarría, Short, and Vogt's (2007) SIOP model, Coggins et al. (2007) offer scaffolding instruction—by doing such things as building on students' prior knowledge using advance organizers and connecting with their prior experiences—as the third strategy for helping ELLs to learn mathematics. Fourthly, as a means of further scaffolding learning, appealing to different types of learners, and at the same time reducing dependence on verbal explanations, they suggest the usage of concrete materials in mathematics activities. The fifth strategy concerns the judicious usage of visual representations, such as diagrams, pictures, models, posters, charts, etc., to augment verbal explanations. Another strategy offered as an aid to developing mathematical and linguistic fluency is the usage of questioning strategies (Socratic, with varying levels of complexity) that elicit more than simplistic responses from students. The final strategy suggested by Coggins et al. (2007) is the provision of *comprehensible input* (Echevarría, Short, & Vogt, 2007), by which the authors signify that teachers should conscientiously monitor their own speech patterns and explanations to ensure that ELLs can make sense of what they are receiving.

Kersaint, Thompson, and Petkova (2009) have offered a lengthy list of research-based "best practices to support English language learners in the mathematics classroom" (p. 77). Categories of strategies offered include strategies for developing mathematics classroom discourse, academic language in ELLs, and problem solving ability, as well as strategies for making assessments equitable, for making mathematics instruction responsive to students' cultural background and strategies that teachers who are themselves ELLs may use to draw upon their own experience in order to strengthen their mathematics instruction for ELLs. The specific discourse strategies given by Kersaint et al. (2009) mirror some of those seen in other places (r4 Educated Solutions, 2010; Coggins et al., 2007; Francis et al., 2006) and involve making language comprehensible to ELLs through reduction of linguistic complexity by incorporating visual aids such as pictures and gestures (Shein, 2012). As central to helping ELLs to learn mathematics academic language, Kersaint et al. (2009) suggest that vocabulary instruction be based on the mathematical concepts being taught: "students can explore a mathematics concept and then attach related language to it" (p. 95). They offer a variety of well-known literacy tools such as word-walls, graphic organizers, usage of concrete objects, comparing and contrasting activities, and reading and writing strategies to aid in this endeavor. The problem-solving strategies recommended center on helping students to enter deeply into the problem situation, usually found in the context of solving word problems, by elaborating on the meanings of words, using visual models, and making connections with real-life experiences. Furthermore, they recommend specific instruction in mathematics problem-solving algorithms like Polya's (1957).

Observing that ELLs, because of their limited linguistic proficiency, may tend to perform lower than others on written assessments (Martiniello, 2009; Abedi et al., 2006;

Lager, 2006; Abedi et al., 2001), Kersaint et al. (2009) also recommend allowing ELLs to use learning tools, such as notes, organizers, and glossaries, during testing. Providing test accommodations such as reduced linguistic complexity of problems and more time, and also the usage of alternative forms of assessment, such as journals and projects, can make assessments more equitable for ELLs, they argue. The authors observe that ELLs benefit from mathematics instruction that responds to their specific cultural background (Gutiérrez, 2002; Gutstein et al., 1997) and recommend that teachers be considerate of their own culture and that of their students. Furthermore, they recommend that students be allowed to use their own language while working on mathematics (Khisty & Morales, 2004; Gutiérrez, 2002; Moschkovich, 2002) and that teachers make efforts to connect with students families. Finally, the authors offer strategies that teachers who are ELLs may use to both overcome their own linguistic limitations and to draw upon their experience as ELLs to create classrooms that are accepting of cultural and linguistic differences.

This section has presented a sampling of the many instructional strategies that have been posited to help ELLs learn mathematics. It is interesting to note that, although almost innumerable strategies have been offered, yet claims to their actual effectiveness are largely anecdotal. The next section will present empirical results concerned with assessing the effectiveness of certain strategies for teaching mathematics to ELLs.

Effectiveness of Strategies. While there exists a large number of strategies prescribed for helping ELLs to learn mathematics, there is less research that has evaluated the effectiveness of the strategies in actually obtaining the promised improvements in mathematical achievement among ELLs. Pray & Ilieva (2011) sought to find associations between strategy usage and higher achievement in mathematics among (mostly Latino) ELLs in Utah. From a sample of 11 teachers and their students, classroom observational data and interviews were collected and coded to arrive at conclusions regarding the extent of the usage of specific strategies. Then, using factorial analysis of variance, the researchers found associations between three levels of strategy usage and students' scores on the state standardized test of achievement. The most significant finding was that ELLs in the earliest stages of language development and who were enrolled in classrooms that had higher usage of visual and speech strategies, performed slightly higher on average on the exam. The visual strategies employed included using manipulatives and hands-on objects for explanation, using illustrations and visual representations, and using graphic organizers. The speech strategies included gesturing, speaking slowly with reduced usage of idioms, using simple sentence structures and clear explanations. The findings of this study may provide evidence of the specific benefit granted to ELLs by the usage of visual and speech strategies in mathematics lessons. Further evidence in support of this conclusion was provided by Martiniello (2006) in her finding that the inclusion of visual images can mitigate the difficulty of linguistically complex word problems.

As the above sources have suggested, mathematics teachers of ELLs should attend to ways in which linguistic complexity can be alleviated through the usage of other forms of nonlinguistic communication. Kersaint et al. recommend that teachers "use dramatic gestures, actions, and verbal intonations" (p. 87) as cues to assist ELLs as they grapple with unfamiliar concepts and words. Shein (2012) investigated one teacher's usage of gestures in repairing 5th grade students' mathematical errors. The study shows that teachers can use gestures for at least three pedagogical functions: grounding questions, revoicing students' strategies, and narrating the meaning of mathematical objects. This research provided an illustration of ways in which math teachers of ELLs can augment verbal explanations with gestures so as to reduce the dependence of the discourse on linguistic sources alone and, as a result, elicit the participation of ELLs in mathematical conversations.

As previously mentioned, Chang (2008) investigated the effectiveness of four student grouping models, finding that Latino ELL students performed poorest in situations of whole-class teacher directed activity, such as lecture. Conversely, the study also found that Latino ELLs showed greatest achievement in classrooms that employed teacher-directed individual activity, i.e., students working individually with one-on-one tutoring from the teacher. As a result of this finding, the authors suggest that teachers of underperforming Latino ELLs should choose this instructional method as a strategy for improving the achievement of their students.

Strategies like that offered by Chang (2008) have resulted from observations of effective practice. First, the researchers takes note of what seems to be working. Then, as is most frequently found in the discussion section of the paper, the researchers translates the observation into a suggestion. In essence, the argument is the following: *this seems to improve achievement for ELLs, therefore mathematics teachers should probably do this*. A number of mathematics strategies for ELLs have resulted in this way from research concerned with assessment, and these are presented below.

Assessment Strategies as Instructional Strategies. “As an integral part of mathematics instruction, assessment contributes significantly to all students’ learning” (NCTM, 1995, p. 13). From the perspective that assessment serves as a central component of instruction, assessment strategies may be interpreted as instructional strategies. Hence, a final source of instructional strategies for helping ELLs to learn mathematics comes from research concerned with the mathematics assessment of ELLs. Based on their analysis of the results of administering both English and Spanish versions of a mathematics test to Latino students, Llabre & Cuevas (1983) recommend that, “the primary language of instruction (English or Spanish), the level of reading proficiency in the language of instruction, and the skill being measured should be taken into account when interpreting the mathematics achievement test scores of bilingual students” (p. 323). They note that “just because a student's first language is Spanish, it does not follow that his or her performance in Spanish will be superior” (p. 322). For some students, assessing them in the language of instruction may lead to more valid inferences about their achievement than assessing them in their first language.

Researchers have also investigated the relative effectives of providing certain types of assistance to students on mathematics tests. Abedi et al. (2001), using items from the 1996 NAEP Grade 8 Bilingual Mathematics test, compared the scores of ELL and non-ELL students that were given combinations of testing accommodations: standard items, modified (simplified) English items, extra time, and access to a glossary. While all of these accommodations improved the scores of ELL students, the only one that narrowed the difference in scores between ELL and non-ELL students was the modified English items. The result strengthens the importance of an aforementioned strategy: teachers should reduce the linguistic complexity of instruction. However, in another study (Abedi et al., 2006), this strategy appeared to benefit both ELLs and non-ELLs equivalently, and not surprisingly. Thus, when teachers intentionally reduce the linguistic complexity of their instruction or of mathematics problems, a probable result will be an increase in achievement scores not only for ELLs, but for all students.

Summary. Research has investigated both the broad instructional models and specific instructional strategies for usage in teaching the academic content areas, generally, and mathematics, specifically, to ELLs. Thomas & Collier (2002) found that dual-language instruction maximizes achievement for ELLs. Other researchers (Diaz et al., 2011; Montes, 2002; Chamot and O'Malley, 1986; Cuevas, 1984) have suggested instructional frameworks for helping ELLs grasp both the language of and skills of mathematics. Furthermore, a very large collection of specific strategies for teaching mathematics to ELLs have been generated by researchers, educators and educational support organizations (r4 Educated Solutions, 2010; Francis et al., 2006; Coggins, Kravin, Coates, & Carroll, 2007; Kersaint, Thompson, and Petkova, 2009). A number of strategies for assessment have also been drafted (Abedi et al., 2006, 2001; Llabre & Cuevas, 1983). A useful synthesis of the research-based strategies for instructing ELLs in mathematics is that offered by Chval & Chávez (2011). Duplicated from page 262 of that work these are:

1. "Connect mathematics with students' life experiences and existing knowledge."
2. "Create classroom environments that are rich in language and mathematics content."
3. "Emphasize meaning and the multiple meanings of words. Students may need to communicate meaning using gestures, drawings, or their first language while they develop command of the English language and mathematics."
4. "Use visual supports such as concrete objects, videos, illustrations, and gestures in classroom conversations."

5. “Connect language with mathematical representations (e.g., pictures, tables, graphs, equations).”
6. Write essential ideas, concepts, representations, and words on the board without erasing so that students can refer to them throughout the lesson.”
7. “Discuss examples of students’ mathematical writing and provide opportunities for students to revise their writing.”

These seven strategies were chosen to conclude this section on mathematics instructional strategies for usage with ELLs because they may be seen as generalizations of the many strategies that have been posited for this purpose. Futhermore, Chval & Chávez (2011) were careful to link each of the seven with established research done with ELLs. As such, they may be seen as a suitable synopsis of the research-based strategies for instructing ELLs in mathematics.

The foregoing sections of the literature review have focused on research aimed at improving the mathematics achievement of English language learners. Framed from the three perspectives of *affordances, deficits*, and *strategies*, they have summarized the linguistic difficulties that Latino ELLs face in mathematics classes, along with the particular linguistic and cultural traits that can serve to help them to be successful in mathematics. These sections have also overviewed the many instructional strategies posited for helping these students be successful in mathematics classes. From the perspective of *pedagogical content knowledge*, knowledge of these deficits and affordances, of what makes learning mathematics difficult for ELLs as well as what empowers them to learn mathematics, seem to be fundamental elements of the teachers’ knowledge that this study seeks to determine. Furthermore, knowledge of the “strategies most likely to be fruitful in reorganizing the understanding of learners” is important for effective instruction (Shulman, 1986, p. 9). To further address the question of *what* constitutes such knowledge, the next section of the review is concerned with research that focuses on knowledge needed for teaching mathematics.

Teacher Knowledge

> "…quality of instruction depends fundamentally on what teachers *do* with students to develop their mathematical proficiency, and … what teachers *can do* depends fundamentally on their knowledge of mathematics…"
>
> (RAND, 2003, p. 15).

A central purpose of this study is to describe the type of knowledge prerequisite for the effective mathematics instruction of Latino ELLs and to evaluate the extent to which it can be measured. It is clear that what teachers *know* has an impact on their ability to instruct students. Conceptualizations of *what* teachers should know have taken a number of different forms over the years (Petrou, & Goulding, 2011; Wilson, Shulman, & Richert,1987). This section will consider one of the more significant of these and then review a recent, and related, framework for understanding knowledge for teaching mathematics.

Pedagogical Content Knowledge. Dewey (1902) asserted that:

> "Every study or subject thus has two aspects: one for the scientist as a scientist; the other for the teacher as a teacher. These two aspects are in no sense opposed or conflicting. But neither are they immediately identical (p. 29)."

Most recently, conceptions of teachers' knowledge have embraced the dichotomous nature of knowledge as proposed by Dewey. Related to this view is that held by Schon (1983) who theorized that professionals possess a special kind of knowledge, *knowledge-in-practice*; that is, knowledge of their profession that can only be obtained through the practice of the profession. Such knowledge may even be difficult for the possessors of it to disclose verbally. Calderhead (1987) applied knowledge-in-practice to teachers, the kind of knowledge that teachers gain through their classroom experience. From this perspective, the knowledge used by teachers of mathematics in the act of teaching would be qualitatively different, because of their specific professional experience, than that possessed by mathematicians and other non-teachers of mathematics.

In the same line of thinking, Lee S. Shulman (1986) proposed that, besides having knowledge of the academic subject that they teach, teachers need *pedagogical content knowledge* of the subject. By his definition, such knowledge—"subject matter knowledge for teaching"—is composite of "the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others" (p. 9). Furthermore, "pedagogical content knowledge also includes an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons" (p. 9). This conceptualization of the knowledge that teachers possess was a result of research focused on describing the ways in which teachers change through their classroom experience (Wilson, S. M., Shulman, L. S., & Richert, A., 1987). These researchers observed that teachers transform their understanding of content through teaching it. That is, teachers' pedagogy shapes their content knowledge. Shulman's concept of *pedagogical content knowledge* is currently one of the most prevalent (Ball, Thames, & Phelps, 2008) conceptions taken by researchers who study knowledge invoked by teachers in their profession.

Teachers' Mathematical Knowledge for Teaching. Building on Shulman's (1986) concept of pedagogical content knowledge, researchers (Hill, Ball, & Schilling, 2008; Hill, Schilling, & Ball, 2004) have undertaken to *unpack* the domains of pedagogical content knowledge in the context of mathematics classrooms. Based on their reviews of mathematics education literature, and on their combined research and educational experiences, these researchers have created a large number of survey measures of teachers' mathematical knowledge for teaching (MKT). The authors define mathematical knowledge for teaching as "the mathematical knowledge that teachers use in classrooms to produce instruction and student growth" (Hill, Ball & Schilling, 2008, p. 374). Figure 1 explains their understanding of the domains of knowledge for teaching mathematics.

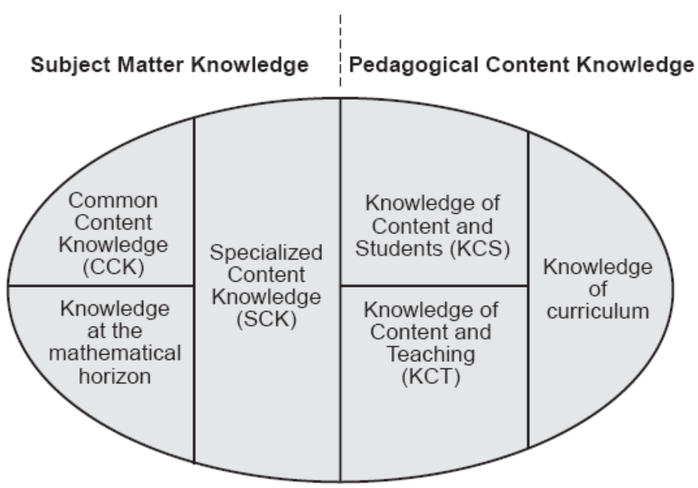


Figure 1. Domains of Mathematical Knowledge for Teaching (MKT). Adapted from "Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students," by H.C. Hill, D. L. Ball, and S. G. Schilling, 2008, *Journal for Research in Mathematics Education, 39*(4), pp. 372-400. Copyright 2008 by the National Council of Teachers of Mathematics.

This model depicts knowledge for teaching mathematics by providing three domains of knowledge for each of Shulman's (1986) subject matter knowledge and pedagogical content knowledge. Hill, Ball & Schilling (2008) proposed that within subject matter knowledge, there are three domains: common content knowledge (CCK), knowledge at the mathematical horizon, and specialized content knowledge (SCK). CCK includes "mathematical knowledge and skill used in settings other than teaching" (Ball, Thames, & Phelps, 2008, p. 399), such as might be possessed by a mathematician or "well-educated adult" (p. 398). Knowledge at the mathematical horizon, however, concerns the mathematical context of a topic, "a kind of 'peripheral vision' needed in teaching, that is, a view of the larger mathematical landscape that teaching requires" (Hill & Ball, 2009, p.70). Furthermore, SCK "is mathematical knowledge, not pedagogy. It includes knowing how to represent quantities … using diagrams, how to provide a mathematically careful explanation …, or how to appraise the mathematical validity of alternative solution methods for a problem" (Hill, Rowan & Ball, 2005, p. 377—378). Each of these domains is considered as knowledge of mathematics and as distinct from pedagogical knowledge.

Pedagogical content knowledge, as theorized for the teaching of mathematics by Hill, Ball & Schilling (2008), is also consistent of three domains: knowledge of content and students (KCS), knowledge of content and teaching (KCT), and knowledge of the curriculum. The authors define KCS as "content knowledge intertwined with knowledge of how students think about, know, or learn this particular content [mathematics]. KCS is used in tasks of teaching that involve attending to both the specific content and something particular about learners, for instance, how students typically learn…and the mistakes or misconceptions that commonly arise during the process" (Hill, Ball & Schilling, 2008, p.

375). KCT "combines knowing about teaching and knowing about mathematics" (Ball, Thames, & Phelps, 2008, p. 401). It concerns knowing how to design instruction for learning particular mathematics concepts and skills, including selection of appropriate examples and representations as well as appropriate sequencing of concepts. Finally, knowledge of curriculum, as theorized by Hill, Ball, & Schilling (2008), is equivalent with Shulman's (1986) description of it as knowledge of "the full range of programs designed for the teaching of particular subjects and topics at a given level, the variety of instructional materials available in relation to those programs, and the set of characteristics that serve as both the indications and contraindications for the use of particular curriculum or program materials in particular circumstances" (p. 10). Examples of items used to operationalize the KCS and KCT domains are given in the table below.

Table 1

Sample KCS and KCT Survey Items

Domain	Item
KCS	Mrs. Jackson is getting ready for the state assessment, and is planning mini-lessons for students focused on particular difficulties that they are having with adding columns of numbers. To target her instruction more effectively, she wants to work with groups of students who are making the same kind of error, so she looks at a recent quiz to see what they tend to do. She sees the following three student mistakes: $\begin{array}{rrr} 1 & 1 & 1 \\ 38 & 45 & 32 \\ 49 & 37 & 14 \\ \underline{+65} & \underline{+29} & \underline{+19} \\ 142 & 101 & 64 \\ \text{(I)} & \text{(II)} & \text{(III)} \end{array}$ Which have the same kind of error? (Mark ONE answer.) a.I and II b.I and III c.II and III d.I, II, and III

Table 1-Continued

Sample KCS and KCT Survey Items

	Note. Adapted from "Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students." by H. C. Hill, D. L. Ball, & S. G. Schilling, 2008, *Journal for Research in Mathematics Education, 39*(4), 372-400. Copyright 2008 by the National Council for Teachers of Mathematics, Inc.
KCT	While planning an introductory lesson on primes and composites, Mr. Rubenstein is considering what numbers to use as initial examples. He is concerned because he knows that choosing poor examples can mislead students about these important ideas. Of the choices below, which set of numbers would be best for inroducing primes as composites? (Mark one answer.) Primes — Composites a) 3, 5, 11 — 6, 30, 44 b) 2, 5, 17 — 8, 14, 32 c) 3, 7, 11 — 4, 16, 25 d) 2, 7, 13 — 9, 24, 40 *Note*: Adapted from "Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts?" by H.C. Hill, D.L. Ball, L. Sleep, & J.M. Lewis, 2007, in F. Lester (Ed.), *Second Handbook for Research on Mathematics Education*, p. 111-155. Copyright 2007 by Information Age Publishing, Inc.

Measurement of MKT. Working from this framework, which offers a more detailed explanation of pedagogical content knowledge, Hill, Schilling, & Ball (2004) have made considerable effort at providing valid and reliable, paper-and-pencil format, survey measures of their posited domains. Furthermore, the results of this work have not only given robust evidence, through survey administration and factor analysis of the data, for the existence of multiple dimensions of pedagogical content knowledge, but have also shown a correlation between teachers' level of knowledge and the quality of their mathematics instruction (Hill et al., 2008). Thus, this research seems to provide empirical evidence both of the existence and of the structure of a particular form of knowledge that capacitates teachers to effectively teach mathematics to their students.

A couple limitations of this research should be noted here. Not all hypothesized domains of MKT have performed well under statistical analysis. Many of the items designed to measure knowledge of content and students (KCS) domain showed very low reliability, $\alpha < .70$ (Hill, Ball, & Schilling, 2008). This result speaks to the difficulty of creating measures of this domain of knowledge. However, of even more interest to the present study is the fact that the MKT measures were not created with certain populations of students in view, as is the goal of this study. Extant measures of teachers' mathematical knowledge for teaching do not differentiate levels of knowledge for teaching diverse students populations. Moreover, the MKT measures make no mention of English language learners specifically; none of the items are framed in the context of instructing this special population of students.

Summary. This section of the literature review has summarized research related to one of the most influential conceptualizations of knowledge needed by teachers, pedagogical content knowledge (Shulman, 1986). PCK is different from academic knowledge because it is informed by knowledge of the instructional decisions, details of the curriculum, and qualities of their students that teachers encounter daily. Mathematical knowledge for teaching (MKT) is an influential application and explanation of PCK in the context of mathematics classrooms. The aspects of MKT include knowledge of content and students (KCS), knowledge of content and teaching (MCT), and knowledge of curriculum. Efforts to provide survey measures of these domains of knowledge have given evidence of their existence and of the potential for those teachers who possess such knowledge to improve the achievement of their students.

Summary of Literature and Research Gap. This literature review has considered two strands of research and educational theorization that inform this study. Research concerned with the mathematics education of English language Learners, many of whom are Latino, has been categorized as taking either of a *deficit*, an *affordance*, or a *strategic* perspective. Deficit perspective literature gives evidence of the difficulties that English language learners have in learning mathematics. These difficulties include limited fluency with the English language and associated deficiency with both conversational and academic language, difficulties with understanding word problems, difficulties with participating in classroom activities, and difficulties with assessments. Affordance perspective literature has drawn attention to the special qualities of ELLs that can capacitate them to learn mathematics, in spite of their perceived limitation. Their bilingualism and their cultural identity can serve them to assist their participation in mathematical conversation and in contextualization of mathematical concepts. Strategic perspective literature has provided a large number of strategies for assisting English language learners in schools generally and in mathematics specifically. Strategies for helping ELLs learn mathematics include, among many others, linguistic strategies, representational strategies, cognitive and problem solving strategies, and strategies for making instruction comprehensive to ELLs.

The second related strand of research is concerned with describing and measuring teachers' knowledge. Building upon Shulman's (1986) conception of the blending of content and pedagogical knowledge, *pedagogical content knowledge*, upon which teachers rely during instructional activities, researchers (Ball, Thames, & Phelps, 2008; Hill et al., 2008; Hill, Rowan, Ball, 2005; Hill, Schilling, & Ball, 2004) have conceived of *mathematical knowledge for teaching* (MKT) as being shaped by a combination of mathematics content knowledge and knowledge of the qualities of students, of instructional decisions, and of curriculum. This research has given rich description to the elements of pedagogical content knowledge that mathematics teachers draw upon when making instructional decisions. Furthermore, there is evidence that MKT is positively correlated with improvements in the quality of mathematics instruction.

Based upon the research concerned with teachers' knowledge, it makes sense to ask concerning the kind of knowledge needed by mathematics teachers of ELLs. In view of the rich research literature concerned with ELLs' learning of mathematics, it would seem that effective teachers of ELLs should possess, at the least, knowledge of how to teach mathematics to these students, including knowledge of instructional strategies, for example. However, what form does such knowledge take? What knowledge do ELL teachers call upon when deciding how to provide comprehensible mathematics instruction to their ELLs or when deciding which representations to use for specific mathematics concepts when choosing instructional strategies?

There appears to be a paucity of literature concerned with knowledge for teaching mathematics to ELLs. While there are studies of mathematics teachers' *perceptions* of ELLs (Hansen-Thomas, & Cavagnetto, 2010) and of efforts to improve pre-service teachers' *attitudes* about ELLs (Pappamihiel, 2007), it is difficult to find any investigation, or even theorization, about knowledge for teaching mathematics to ELLs. Indeed, at the time of writing, extensive searches of multiple research literature databases returned fewer than 40 results that contain combinations of the most important search terms related to this study, those related to ELLs and to teachers' knowledge. Several of these have already been cited in this paper. However, none specifically addresses the questions of this study which pertain to knowledge for teaching mathematics to ELLs.

Rather, the articles that concern ELLs and mathematics treat either teachers' perspectives of students, the effectiveness of certain curriculum materials, the experiences of specific mathematics teachers, equity issues, or teacher education and professional development, a closely related topic, the limitations of which have been addressed in the introduction. However, investigation or even theorization concerning teachers' knowledge for teaching mathematics to ELLs is severely lacking. It is this lack, the absence of studies of teachers' knowledge for teaching mathematics to English language learners, which motivates this study.

CHAPTER 3: METHODOLOGY

This chapter presents the research design and methods that were used to answer the questions of this study. The research questions for this study were the following:

1. What are the domains of knowledge needed to teach mathematics to Latino English Language Learners (KT-MELL)?
2. Drawing from the research literature and middle school mathematics classroom observations, what are some of the aspects of these domains of knowledge when teaching Latino English Language Learners at the middle school level?
3. What evidences of reliability and validity does an instrument developed to measure these domains and aspects exhibit?

As the title of this dissertation research indicates, this work may be classified as an instrument development study. As such, the methods employed served the three purposes of 1) identifying domains of knowledge for teaching mathematics to Latino ELLs, 2) developing an instrument to serve as the exploratory measure of the knowledge and 3) seeking validity and reliability evidence concerning the measures obtained. These purposes are in correspondence with the three questions of this study.

Selection of the methods to be used by this study involved careful consideration of methods used in related studies. Hill, Ball, Sleep, & Lewis (2007) discuss three principle methods that have been used in recent research to investigate mathematics teachers' knowledge for teaching mathematics. These methods have been classroom observations, mathematical tasks and interviews and paper-pencil, multiple-choice tests. The later method offers a number of benefits to researchers. The foremost of these is the possibility of developing valid and reliable instruments that can be used for finding relationships between teacher knowledge and other important variables, such as student achievement. A great deal of recent research has employed this methodology to the advancement of the field (Hill, Ball, Sleep, & Lewis, 2007).

The present study builds on this body of research. Moreover, the theoretical framework of teacher knowledge that guides this study has largely been advanced through research that has made effective usage of survey methodology (Hill, Schilling, & Ball, 2004). Consequently, the research design selected to answer the questions of this study was a survey design as well.

The remainder of this chapter is given to the discussion of the processes that resulted in the creation of a survey intended to serve as a measure of KT-MELL. Development of the final instrument central to this study involved two distinct instrument development and pilot-study phases. The exploratory Phase 1 involved initial theorization concerning the domains of KT-MELL as well as the writing of items intended to capture aspects of the knowledge. The first pilot-study assisted in identifying both the strengths and weaknesses of the hypothesized domains, of the items and of the instrument format. Following the first pilot-study, Phase 2 involved much more extensive review of the research literature as well as many more hours of observations of middle school classrooms. Based upon the further reading, the hypothesis of the domains of KT-MELL was refined, which resulted in formalization of the test framework for the instrument used in this study. Using this test framework as guide, a larger number of items intended to measure the hypothesized domains were developed. Many of the items developed both in this phase and in the initial phase were adapted from observed classroom situations. These contexts were seen as contributing to the instrument by providing realistic mathematics teaching situations.

As this narrative makes clear, initial theorization concerning the knowledge domains of KT-MELL proceeded simultaneous to survey item-writing. As was shown in the review of literature (Chapter 2), KT-MELL is a novel knowledge construct for which minimal prior theorization exists. Indeed, the development of items intended to capture KT-MELL served an important step in helping to identifying the domains of KT-MELL. This chapter is given to describing the two phases of instrument development including their respective pilot-studies, the implementation of the survey and the data analysis methods that were used to obtain validity and reliability information concerning the measurements obtained.

Phase 1: Exploration

Initial work toward developing theory concerning the domains of knowledge for teaching mathematics to Latino ELLs and toward developing a measure of such knowledge was begun in the fall of 2011. The details of this work, which included an exploratory first pilot-study, are presented here. Specifically, theorization about the construct in question, the development of the initial test framework and survey, the first pilot-study participants and survey administration, and the results of this first pilot-study are presented.

Initial Construct Definition and Instrument Development. Initial readings of the research literature concerned with the mathematics education of Latino ELLs, as well as informal classroom observation, resulted in the hypothesis of the existence of four distinct domains of KT-MELL. They were: knowledge of mathematical academic

Spanish (MAS), knowledge of mathematics instructional strategies for ELLs (MISE), knowledge of difficulties faced by ELLs in mathematics (DEM), and knowledge of mathematics (MATH). Guided in part by the work of Hill, Schilling, & Ball (2004), who have operationalized mathematical knowledge for teaching (MKT) in the context of elementary school classrooms, initial theorization concerning the domains of knowledge for teaching mathematics to ELLs attempted to define this knowledge as *mathematical* knowledge. For this reason, beyond including knowledge of difficulties and of strategies, knowledge of mathematical academic Spanish and basic mathematics content knowledge were initially thought to be elements of KT-MELL. As is explained later, in the section concerning the second phase of instrumentation development, further theorization concerning the domains of KT-MELL resulted in substantial transformation of the hypothesized domains.

Initial Instrument Framework. Based upon the above theorization concerning the domains of KT-MELL, an attempt was made to create an explanatory survey in the form of a paper-pencil test. Explanatory surveys are useful for, among other things, evaluating the validity of the proposed knowledge domains and for evaluating the relationships of these domains, that is, of the latent variables operationalized through the survey items, with one another.

Without exception, all of the items on the first KT-MELL survey were adapted either from research literature or from actual classroom observation. As an instance of the later, a middle school teacher at a central Texas middle school was observed during a 6th grade mathematics lesson. The final two items on the initial survey described an actual mathematical problem and student response captured during that observation. After crafting a number of survey items, the items passed through two revision processes, one concerned with content validity and the other with ascetic appearance. A scholar of the mathematics education of ELLs, who is fluent in Spanish, scrutinized the survey. This person afforded valuable insight into the salience of the items. Additionally, this reviewer provided critical comments regarding the quality of the Spanish language used in the survey. Additionally, the survey was submitted to a graphic design consultant for aesthetic review. The designer was able to comment on the clarity of the items and on the visual appeal of the document as a whole. The final initial instrument (Appendix C) had the following test-framework.

Table 2

First Pilot Study Test Framework

Knowledge Domains	Mathematical Academic Spanish (MAS)	Mathematics Instructional Strategies for ELLs (MISE)	Difficulties faced by ELLs in Mathematics (DEM)	Mathematics Content Knowledge (MATH)
Descriptions of Domains	Teacher can interpret and use mathematical academic Spanish language related to mathematical objects and operations found in the classroom.	Teacher knows and selects appropriate mathematics instructional strategies for English language learners.	Teacher understands the difficulties that ELLs face in mathematics classes.	Teacher is competent in the mathematics that is to be taught.
Items Numbers	6, 8	2, 3, 4, 5, 7		1, 8

First Pilot-Study Sample and Results. The initial survey that was designed under the above test framework contained eight items and was given to 50 individuals, 2 practicing middle school teachers in central Texas, and 48 pre-service teachers at Texas

State University - San Marcos. The group of pre-service teachers was composite of two sub-groups: 34 pre-service high school mathematics teachers who were given the survey as an in-class assignment and 14 pre-service elementary school teachers who were given the survey as an out-of-class, voluntary task. No attempt was made to collect a representative sample of any population. Rather, descriptive data, such as means and variances on responses, useful for making judgments about the qualities of the items were sought.

Answers to survey items were scored dichotomously, right or wrong, and each survey was given a score equivalent to the percent of correct responses. The mean score was 58% (median 62.5%), the maximum was 87.5% and the minimum was 25%. The test scores followed an approximately normal distribution, as seen in Figure 2 below. However, the reliability, measured by Cronbach's (1951) alpha, of the survey across all eight items was very low, $\alpha = .009$. The low reliability called for several improvements. Most importantly, there was the need to develop many more items. Additionally, the item response formats, though all multiple-choice, varied greatly, from two right-wrong options to Likert scaled items. By reducing the variation in response formats, there was an expectation of improved internal consistency of the instrument. Finally, the low reliability indicated the need of further theorization concerning the structure of the knowledge being measured by the instrument. It is also worthy of note that, while the reliability of the test among both pre-service high school and pre-service elementary school teachers was quite low, as mentioned above, the reliability of the instrument using only responses from pre-service elementary school teachers was slightly improved ($\alpha = .182$). Figure 2 below exhibits the distribution of scores fort the first pilot-study.

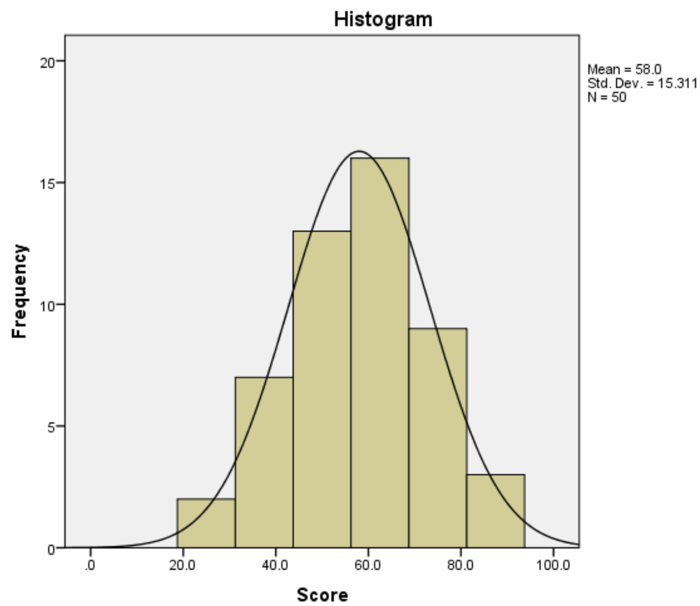


Figure 2. Distribution of First Pilot-Study Test Scores.

Although the approximately normal distribution of test scores satisfied statistical assumptions required for performing other tests with these data, as desired, yet both the low reliability of the test and the inclusion of only eight items on the first pilot-survey—a factor in the low reliability—made it difficult to answer many questions including those concerning the underlying latent variables. Furthermore, the distribution of items was unbalanced across variables; whereas MISE enjoyed five distinct items, MAS and MATH had only two, one of which (item 8) was shared between the domains, and DEM had none. Hence, the survey used for the first pilot-study did not represent all hypothesized domains equally. Nevertheless, a number of important observations were made concerning the performance of specific items.

Judging by its having the highest percentage of correct responses (90%), item 8 appeared to be the *easiest* question on the survey. This result was unexpected for the following reason. The item required respondents to identify whether a student had given the correct response when a teacher had asked concerning finding the proportion of shaded squares out of 100 squares. Although the student's response was correct and the mathematics not difficult, yet the student's answer was given in (grammatically incorrect) Spanish. Thus, respondents, only 14% of whom had self-identified as *Hispanic* and who may not have understood Spanish, had to first interpret the Spanish statement before determining its correctness. For this reason it is surprising that so many answered correctly. At least two possible explanations for this phenomenon existed. First, since respondents were required to circle their choice of "YES / NO", *primacy* may have been a factor (Dillman, Smyth, & Christian, 2009). They may have selected the correct answer because it appeared first in the list of options. The difficulty of the question may have also contributed to the primacy effect on this item (Shuman & Presser, 1996). Second, respondents may have observed the Spanish-English cognate *divisas* [you divide] in the student's response and deduced that the student had indicated the correct mathematical operation for finding proportions.

Another interesting quality of item 8 is that it was significantly correlated ($p < .05$) with items 3 and 4, having Pearson $r = .361$ and $r = -.451$, respectively. Item 3 asked concerning the effectiveness of using *key words* as an instructional strategy to remind ELLs how to solve certain types of problems. Item 4 asked respondents to determine the extent to which the image of a gymnast's parallel bars would be *culturally relevant* for ELLs as an illustration of the parallel property in geometry. Based on research literature (Kersaint et al., 2009), the theoretically correct answer was *not very relevant.* (See Gutstein et al., 1997 for a discussion of cultural relevance in the context of mathematics education.) The correlation of item 8 with item 3 could indicate that knowledge of effective instructional strategies for teaching mathematics to ELLs is positively associated with knowledge of academic Spanish, with competence in mathematics, or with both, since item 8 operationalized both of these constructs. Any of these inferences would agree with the theoretical framework for this test. However, the negative correlation with item 4 could indicate a negative correlation with either or both of these constructs. This would contradict the previous inference and disagree with theory. This apparent contradiction with theory helped to further explain the low reliability and to sound another call for more careful theorization concerning the underlying knowledge constructs.

Summary. The results of the first pilot-study gave valuable information as to how to proceed with the full study. Although the test scores were approximately normally distributed, as was hoped, the reliability of the test was quite low. Furthermore, a number of contradictory correlations indicated that theoretical problems existed with the test framework and with operationalizations of the aspects of the proposed domains. It became clear that the domains of KT-MELL were in need of further clarity and definition. This process necessitated further review of literature as well as further classroom observation. Moreover, the first pilot-study instrument was deficient both in the number of items it contained and in the number (and description) of the respondents. Without greater numbers of both of these, it proved impossible to perform the statistical analyses that could have helped to assess the validity of the survey measures. Therefore, in order to improve the quality of the instrument used to answer the research questions of this study, that is to identify the domains of KT-MELL and their aspects, and to give a measure of such knowledge, improvements in both reliability and validity were required. The following section of this chapter will explain how these issues were addressed with the goal of obtaining more useful survey data for answering the questions of this study.

Phase 2: Formalization

This section discusses the further development of the hypothesized domains of KT-MELL and of the survey instrument intended to measure such knowledge. The research theory and observational evidence that guided the development of the test framework will be presented. As illustrations of the theoretical domains of knowledge, several specific items are given. This section closes with an account of the results of a second pilot-study.

Construct Identification, Test Framework and Item Development. DeVellis (2003) explains that the first step in scale development is to "determine clearly what it is you *want* [author's emphasis] to measure" (p. 60). Accordingly, this section explains the boundaries of the knowledge construct at the heart of this study and presents the test framework that guided the development of a measure of this knowledge. The central construct of interest to this investigation is teachers' knowledge for teaching mathematics to Latino ELLs (KT-MELL). This construct is closely associated with and builds upon the theory related to a similar one, mathematical knowledge for teaching (MKT) (Hill & Ball, 2009; Hill et al., 2008; Hill, Rowan, & Ball, 2005; Hill, Schilling, & Ball, 2004). Both of these constructs represent descriptions of the kind of knowledge that is called upon by mathematics teachers *in the act of teaching*. Hence, they are domains of *applied* knowledge. Furthermore, both of these constructs intend to elucidate what Shulman (1986) meant when he referred to the *pedagogical content knowledge* that informs the instructional decisions that teachers make in the classroom.

The theorization concerning KT-MELL presented in this study builds upon, and thus, is more closely aligned with, exhaustive review of the literature that is concerned with the mathematics education of Latino ELLs and with teachers' knowledge. As has been explained in the literature review section of this paper, findings concerning the mathematics teaching and learning of Latino ELLs have been categorized according to one of three emphases: difficulties experienced by ELLs in mathematics classes, capacities that Latino ELLs bring to the classroom for learning mathematics, and strategies for teaching mathematics to ELLs. Considering that the majority, if not all, of this research has been conducted by educators, many of whom have taught mathematics in public schools and among ELLs, it was not surprising to observe that the three emphases of the research literature very closely mirrored the elements of *pedagogical content knowledge* as explained by Shulman (1986). That is, researchers have uncovered "what makes the learning of specific topics [of mathematics] easy or difficult" *for ELLs* and what are, "for the most regularly taught topics in one's subject area [i.e., *in mathematics*], the most useful forms of representation of those ideas, the most powerful analogies,…", i.e., *the most effective instructional strategies for ELLs* (Shulman, 1986, p. 9). Perhaps unconsciously, their *pedagogical content knowledge* has informed the choice of research foci taken by researchers interested in the mathematics education of Latino ELLs. Figure 3 below presents the conceptual model of KT-MELL that resulted from more careful incorporation of findings in the research literature concerned with the mathematics education of ELLs and with teachers' pedagogical content knowledge. This model depicts the situation of KT-MELL within the extant theoretical framework of mathematical knowledge for teaching.

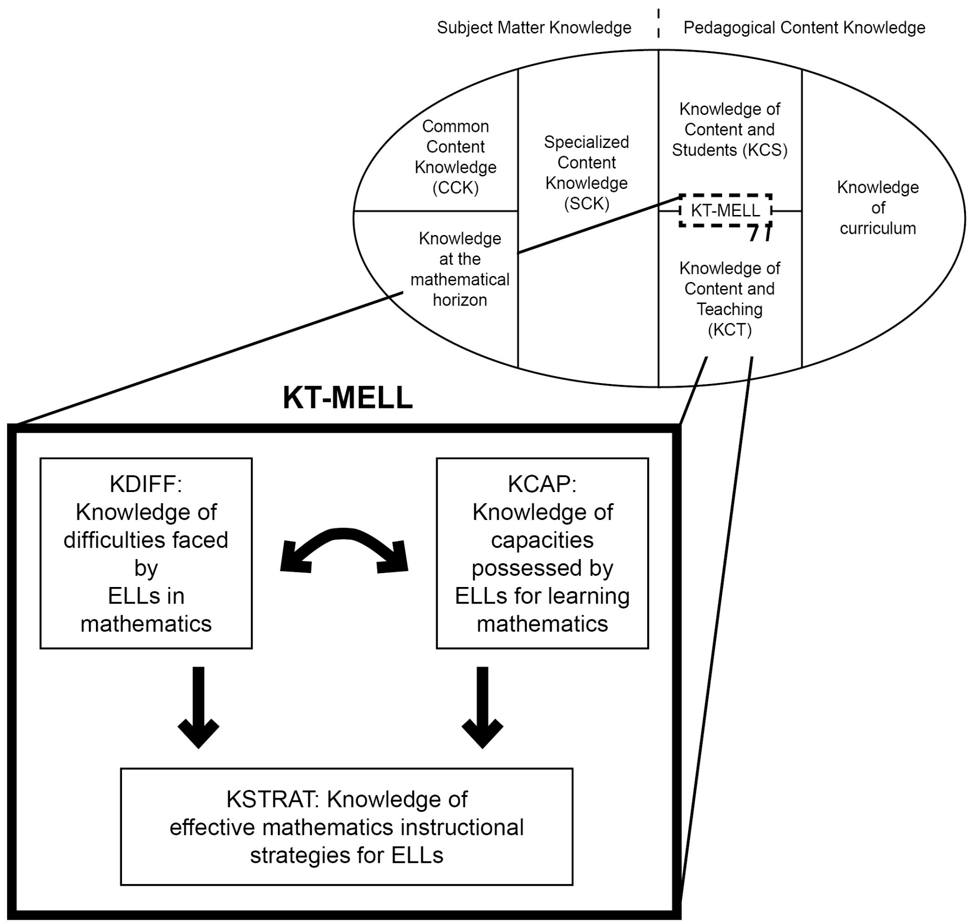


*Figure 3***.** Knowledge for Teaching Mathematics to ELLs (KT-MELL). This figure depicts KT-MELL in its relation to mathematical knowledge for teaching, MKT. The curved arrow between KDIFF and KCAP indicates the likelihood of the existence of a correlation between these knowledge domains. The straight arrows from KDIFF and KCAP to KSTRAT indicate that knowledge of difficulties and of capacities may be seen as potentially *informing* knowledge of strategies. The (ovular) MKT model was adapted from "Unpacking Pedagogical Content Knowledge: Conceptualizing and Measuring Teachers' Topic-Specific Knowledge of Students," by H.C. Hill, D. L. Ball, and S. G. Schilling, 2008, *Journal for Research in Mathematics Education, 39*(4), p. 377. Copyright 2008 by the National Council of Teachers of Mathematics.

As Figure 3 illustrates, the conceptualization of KT-MELL taken in this paper is seen, not as independent of, but as a subset of two specific knowledge domains contained within the MKT framework. Knowledge of content and teaching (KCT) has been explained by Hill et al. (2007) to be "mathematical knowledge of the design of instruction, [which] includes how to choose examples and representations, and how to guide student discussions toward accurate mathematical ideas" (p. 133). Hence, KCT may be seen as knowledge of effective mathematics instructional strategies. Moreover, examples of survey items provided by these researchers as operationalizations of KCT demonstrate strategic choices being made *in response to common student misconceptions* (see Table 1 in Chapter 2). Thus, KT-MELL is seen as being a subset of KCT. It is the application of this knowledge domain to the mathematics instruction of Latino ELLs.

KT-MELL is also seen as intersecting knowledge of content and students (KCS). Hill et al. (2007) explain that KCS is "the amalgamated knowledge that teachers possess about how students learn contents" and that KCS concerns knowledge of misconceptions frequently possessed by students (p. 133). Table 1 of Chapter 2 gave an example of a survey item used by these researchers that was intended to capture KCS. It is clear from this item that the current understanding of KCS implies that teachers who possess this knowledge should be versed in the kinds of difficulties that students encounter in learning to do mathematics. As is explained below, KT-MELL includes this kind of knowledge as applied to Latino ELL students.

As theorized in this dissertation research, knowledge for teaching mathematics to English language learners—KT-MELL—includes knowledge of the most common difficulties that Latino English language learners face in learning mathematics (KDIFF), knowledge of ways in which the linguistic, academic, and cultural background of these students can capacitate them and be an asset to the learning of mathematics (KCAP), and knowledge of the instructional strategies that can effectively respond to both of these qualities of Latino ELLs to promote their mathematical achievement (KSTRAT). This theorization is based upon exhaustive review of the research literature and upon more than 30 hours of structured observations (see Appendix D for observation protocol) of the mathematics classrooms of teachers of ELLs. Knowledge of difficulties that Latino ELLs may face in mathematics classrooms includes knowledge of linguistic difficulties that ELLs may encounter, knowledge of words that may be difficult for them, and knowledge of the ways in which verbal expressions in mathematics classes can obscure understanding for Latino ELLs. Knowledge of such difficulties also includes knowledge of the ways that the multiple meanings of English words can challenge ELLs and of the ways in which certain classroom instructional formats and decisions may limit their access and participation in mathematics in the classroom. KT-MELL also includes knowledge of the ways in which certain cultural attributes as well as fluency in the Spanish language, or being bilingual in both Spanish and English, can leverage the mathematical learning of Latino ELLs (KCAP). Furthermore, it includes knowledge of the ways in which the prior mathematical learning of Latino ELLs, including alternative algorithmic methods or notation systems, explanations and representations, may empower them to learn mathematics (KCAP) or serve as a barrier to mathematical learning (KDIFF). Finally, KT-MELL includes knowledge of the mathematics instructional strategies, the representations, models, explanations, algorithms and verbalization methods that are most effective for producing mathematics achievement in Latino ELLs (KSTRAT). These three components are the hypothesized domains of KT-MELL. For simplicity, they are given the acronyms of KDIFF, KCAP, and KSTRAT, representing knowledge of *difficulties*, *capacities*, and *strategies*, respectively.

Based upon this understanding of the KT-MELL construct, the test framework given in Table 3 was used to craft survey items intended to operationalize this knowledge construct and that, it was hoped, would result in empirical evidence concerning the dimensions of the construct, its composite domains, their aspects and relationships with one another.

Table 3

KT-MELL Test Framework

DOMAIN	ASPECTS
KDIFF Knowledge of common difficulties faced by Latino ELLs in mathematics classes	1.Limited English proficiency in speaking, reading and writing
	2.Word problems a.Specific words (vocabulary), multiplicity of words (linguistic complexity), shifts of application, polysemy
	3.Classroom format a.High speech formats (direct teaching)
	4.Assessments a.Low performance because of: i.Word problems, time limitation, high stakes, cultural-irrelevance
KCAP Knowledge of cultural and linguistic factors that can capacitate Latino ELLs in mathematics classes	1.Linguistic creativity, mathematics discursive and communicative ability a.Fluency in L_1 (i.e., first language) b.Bilingualism c.Usage of gestures, objects, and verbal *inventions* to convey meaning
	2.Linguistic and cultural identity a.Association with *Latino* culture, icons, people, values, traditions etc. b.Appreciation of Spanish language
KSTRAT Knowledge of effective mathematics instructional strategies for usage with Latino ELLs	1.Usage of students' background knowledge—academic, linguistic and cultural—to promote understanding
	2.Maintenance of classroom environment rich in linguistic and mathematics content
	3.Emphasis on meanings of words and/or provisions for students' usage of multiple modes of communication to express mathematics
	4.Usage of visual supports to—gestures, objects, illustrations—to convey the meanings of classroom conversations
	5.Connection of mathematical language with multiple forms of mathematical representation
	6.Available visual display of classroom mathematics concepts, representations and words during instruction
	7.Rich usage of students' own mathematical writings and speech with opportunity for them to make revisions

Based upon the above test framework, a large number of multiple-choice items were written that were intended to serve as indicators of particular aspects of each of the knowledge domains. During item-writing, each item underwent repeated revisions of both content and format in consultation with a mathematics educator engaged in research concerned with the quality of mathematics instruction of Latino ELLs. From among the full set of possible items, thirty-two were selected for inclusion in the final KT-MELL survey instrument. These thirty-two items underwent further content validation

Content Validation. During its development, three efforts were made to ensure the content validity of the KT-MELL survey. These were strict adherence to theory and actual practice, instrument review by a panel of experts, and investigation of the instrument by a focus group of middle school mathematics teachers of ELLs. The first of these involved careful theorization about the underlying knowledge domains of the instrument. The model of knowledge domains offered above (in Figure 3) is the result of theorization that draws heavily on established research concerning the mathematics learning of ELLs and concerning mathematics teachers' knowledge for teaching. Furthermore, all of the items that have been developed, which correspond to specific aspects of the domains of knowledge, were written in view of specific research findings. In addition to the alignment of items with theory, the items quite frequently were framed in mathematics learning contexts taken from specific instances of classroom observations. More than thirty hours of classroom observations were conducted during this research. These observations provided a rich source of material from which realistic contexts could be derived. Thus, content validation was first addressed through strict adherence to research-based theory as well as to actual mathematics classroom practice.

A second means of ensuring content validity was through review of the KT-MELL survey by a panel of experts in closely related fields. In his recommendations of the steps involved in scale development, DeVellis (2003) recommended that after generating a pool of items, researchers should submit the items to a panel of experts knowledgeable about the construct of interest. Although KT-MELL is a novel construct, as has been explained, its conceptualization was closely tied to research concerned with the mathematics education of ELLs and with teachers' knowledge. For this reason, the panelists that reviewed the instrument were chosen for their expertise in closely related fields. From the seven letters (Appendix E) that were sent to elicit participation as expert panelists, four researchers, not all from the same institution, participated. Two of the experts were foremost researchers in the mathematics education of ELLs. One of the panelists was expert in the education of bilingual students. Another expert had experience in investigating both mathematics teachers' knowledge and the mathematics instruction of ELLs.

The content validation process requested of the experts consisted of two steps: categorization of items according to one of the three proposed domains of knowledge, and provision of optional comments about the items. Thus, information gained from the expert review process provided evidence of the extent to which each of the items seemed to serve as an indicator of the domain of knowledge for which it was intended. Results from the expert review indicated overwhelmingly that the items represented aspects of the domains of knowledge that they were intended to represent. For only three of the thirty-two items that were submitted for review was the consensus of the experts at odds with the theoretical orientation. Specific categorizations of the items, provided by each of the experts on all items, are given in Appendix F.

The final means of content validation involved presentation of the KT-MELL survey to a panel of eight practicing middle school mathematics teachers of ELLs from a mid-sized school district in central Texas. This collection of teachers was purposefully selected; the teachers were participating in professional development sessions focused on understanding the quality of mathematics instrument of ELLs. The eight teachers were asked to both take the survey and to comment on the extent to which the items in the survey captured the most important aspects of KT-MELL. In the context of whole group discussion, the teachers verbally agreed that the survey was comprehensive in the sense of its capturing the intended knowledge domain; they suggested no substantive changes.

Exemplary Survey Items. An exemplary selection of items is given below. This sample is not representative of all of the aspects of the knowledge domains measured by the entire instrument. Rather, it is offered to give a sense of the format, style and contents of items used as indicators of the three domains of KT-MELL: KDIFF, KCAP, and KSTRAT.

Mrs. Tash's students, many of whom are Latino English Language Learners, have been working in groups to discover how surface area changes when the dimensions of an object are changed by a scale factor. On the previous class day, students used rulers and a rectangular prism to complete the table below.

Picture of Solid	Name of Solid	Dimensions	Scale Factor	New Dimensions	New Surface Area, *SA*	Conclusion
			$k = 2$			

Which word in the table may cause the most trouble for the English Language Learners in the classroom?

- Solid
- Dimensions
- Conclusion
- Surface

Figure 4. Example KDIFF Survey Item.

The item, which was adapted from the observation of an actual mathematics classroom containing a large number of Latino ELLs, begins with a common stem from which a series of survey items draw. In the item above, respondents are asked to select the word that could cause difficulty for ELLs. The correct answer is the fourth option, the only option that is not a Spanish-English cognate. This item represents an aspect of the KDIFF domain—namely, knowledge of words that can be unknown or misunderstood by ELLs—and should differentiate between respondents that are familiar with the linguistic challenges faced by ELLs in mathematics classes and those that are not. The following item is an example of an item that operationalizes an aspect of the KCAP domain.

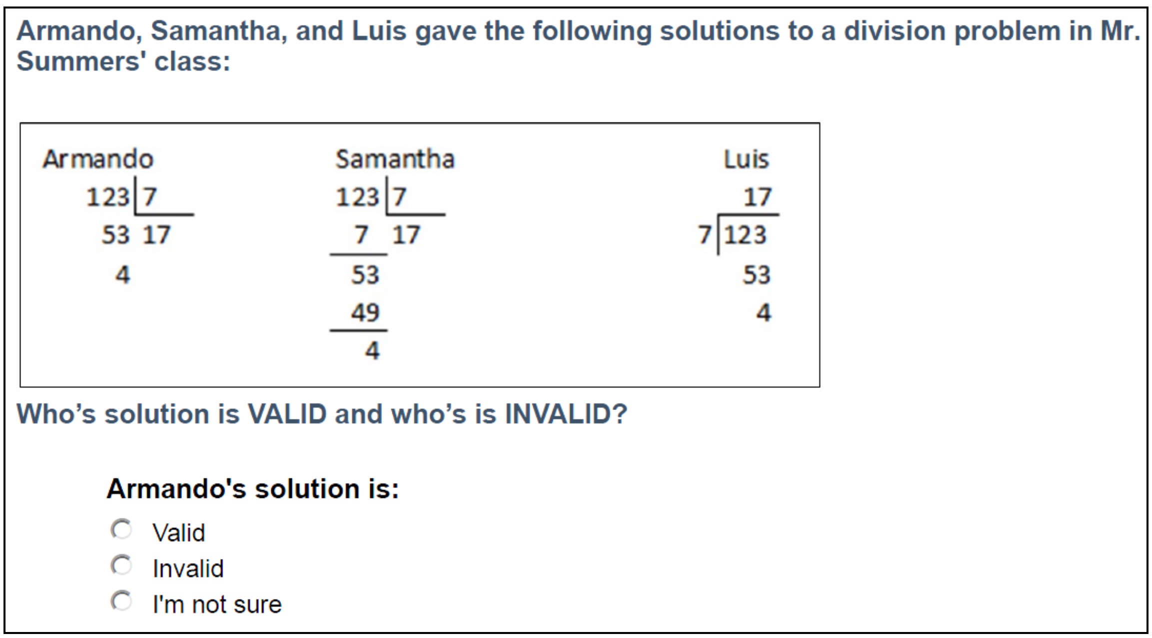
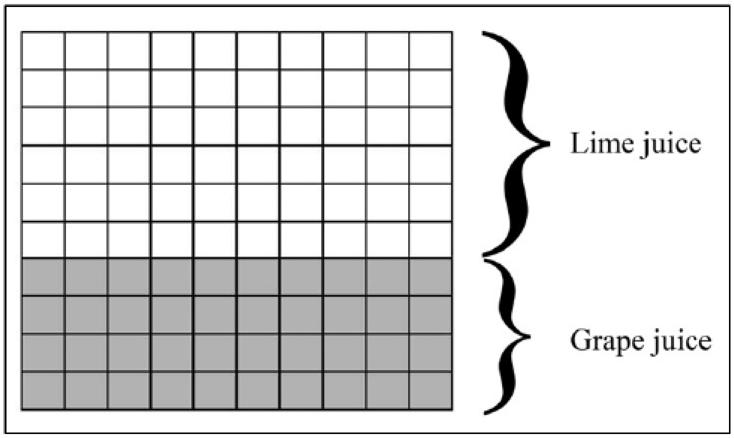


Figure 5. Example KCAP Survey Item.

The item above is intended to elicit knowledge of alternative, yet no less mathematically valid, mathematical notation that ELL students from Latin America may use to perform long division. In this item, Luis has used the long division notation that is typically taught in the United States while Armando and Samantha have used notations commonly found in Central and South America. All methods are valid. Although they differ in appearance, Armando's and Luis's work are quite similar; both students have shown the same number of steps in their computations. This KCAP item is intended to differentiate between respondents that can recognize when a Latino ELL has given a valid, though non-traditional, mathematical solution and those that cannot.

A third item, one that operationalizes knowledge in the KSTRAT domain, is the following:

Mr. Garza asks his class a question about the contents of a grape-limeade drink depicted in the image below: "How do you find out what proportion the grape juice is of the drink?"



Ciarra, an English Language Learner, responds: "Cuenta los grises y lo divide entre cien." Which is the BEST way for Mr. Garza respond?

- He should ask Ciarra to rephrase her answer using English
- He should acknowledge that Ciarra tried to answer his question, but then ask another student to respond in English
- He should acknowledge that Ciarra tried to answer his question, but then explain the correct answer in English so that all of the students understand
- He should ask Ciarra to give her response again, but to use the pictured image and gestures to show what she means

Figure 6. Example KSTRAT Survey Item.

The item above is intended to measure teachers' knowledge of strategies that may improve ELLs' achievement in mathematics. In this item, an ELL has used the Spanish language to give a mathematical response to the teacher's English mathematical question. Respondents must then select the best way for the teacher to respond to the student. Based upon research that has exhibited the value of allowing ELLs to use their native language to express mathematics (Moschkovich, 2002) and that has recommended, as an instructional strategy, that teachers make usage of students' mathematical productions (Chval & Chavez, 2011), the correct answer to this item should be the fourth option. This item serves as an indicator of knowledge in the KSTRAT domain; it should differentiate between respondents that can select the most appropriate mathematics instructional strategy for teaching Latino ELLs and those that cannot. These few exemplary items represent a sample of the many items that were drafted and of the thirty-two items use in the final KT-MELL survey.

Second Pilot-Study Sample and Results. Following the review of the KT-MELL survey by a panel of experts, a pilot-study sample of more than 150 university students who were pre-service elementary, middle or high school teachers was selected to take the survey during the summer and early fall semesters of 2012. Students were selected based upon their enrollment in mathematics courses offered to future teachers. While the instrument was written to assess the knowledge of middle school teachers and not necessarily of pre-service middle school teachers, the attempt to select only future middle school teachers at the university level proved impractical for several reasons. The first of these reasons was that several mathematics courses served multiple levels of pre-service teachers, either both elementary and middle school, or middle school and high school. Another reason was that many pre-service teachers could not be certain of their future teaching assignments at the time of taking the content courses in which they were enrolled. Hence, the pilot-study sample represented a large variety of mathematical ability levels and of (intended) mathematics teaching levels. The instrument was administered in a paper and pencil format during students' regularly scheduled mathematics classes. After removal of surveys containing missing data (i.e., large numbers of unanswered items), 146 survey responses were retained for analysis of the pilot-study data.

While the second pilot-study sample was not representative of the intended sample of this study (i.e., practicing mathematics teachers), nevertheless this sample was purposefully selected for a number of reasons. One of the reasons for selecting pre-service teachers to participate in the pilot-study was that it was relatively simple to obtain a moderately large number ($N = 146$) of responses to the survey, from which valuable information concerning the clarity of the items and of response formats could be obtained. Responses from pre-service teachers also gave a sense of the response and scoring patterns that could possibly result from administration to the intended sample of teachers, and of the psychometric properties of the items. However, a most important goal in selecting pre-service teachers as the pilot-study sample was the possibility of using the pilot-study group of pre-service teachers for comparison with the intended sample of practicing teachers. It was hoped that results of this comparison would provide evidence of the validity of the measurements obtained from practicing teachers.

Results from this second pilot-study indicated that items and response patterns posed no difficulty in terms of linguistic clarity or of understanding the cognitive task being required of respondents. This inference was made based on verbal and written feedback received from participants upon the completion of the survey. Additionally, the second pilot-study indicated that respondents required approximately twenty minutes to complete the survey. This finding was important as it later informed the communications used to elicit survey responses from practicing mathematics teachers.

One of the principal purposes of conducting the second pilot-study was to obtain preliminary estimates of the distribution of the test scores, of the reliability (internal consistency) of the instrument, and of the properties of item responses. As with the first pilot-study so with the second pilot-study, test scores were approximately normally distributed as seen in Figure 7 below.

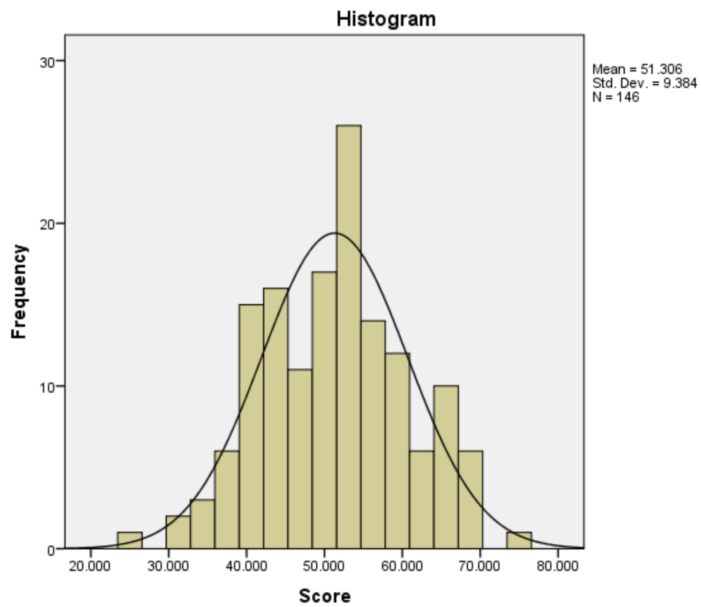


Figure 7. Distribution of Scores on Second-Pilot Study

The measure of internal consistency found by Cronbach's (1951) alpha for the second pilot-study data was α = .345. Although this statistic implied that the survey had low reliability, it was significantly improved over the reliability (α = .009) of the survey used in the first pilot-study. Furthermore, since alpha reliability assumes unidimensionality of the latent variable, and the test framework upon which this survey was built implied the existence of three variables, this finding was taken as potentially confirmative of the underlying factor structure of the survey. To investigate this possibility, it was also observed that a smaller, conceptually related subset of items obtained reliability α = .538. Futhermore, a two-parameter item response theory model (IRT is discussed in the section of this chapter concerning data analysis) that included this reduced set of items demonstrated acceptability of the psychometric properties of these items. Based upon these anecdotal indications that the underlying latent variables were multiple, conforming with the theory of the survey, and that many items demonstrated good psychometric properties, and also based upon the goal of comparing pre-service teachers to in-service teachers, the decision was made to go forward with collection of survey data from practicing teachers using the same thirty-two items that were used in the second pilot-study.

Summary. This section has recounted the means employed to develop a measure of KT-MELL. The first two questions of this dissertation study concern defining domains of KT-MELL and identifying aspects of these domains, respectively. To a great extent the answers to these questions provided by this study are offered in this chapter concerning the methods used in this study. More specifically, the domains and aspects of KT-MELL identified in this study are embodied in the instrument that was developed and find expression in the theoretical model given in Figure 3 (above) and in the test framework given in Table 3 (above). Hence, answering the first two of the research questions mainly involved review of the research literature, classroom observation and careful theorization. As described above, the instrument that resulted from these processes also passed through a stage of content validation, as required in part by the third of the research questions of this study, and through a pilot-study phase.

The third of the research questions concerned assessing the reliability and validity of the measures obtained using the instrument developed in this study. This question was addressed by collecting responses from actual, practicing mathematics teachers. To this end, the remaining sections of this chapter describe the survey administration methods, the population and sample, as well as the data analysis methods that were used to answer the questions of this study.

Survey Administration, Population and Sample

With a view toward increasing the quality (independence) of responses, minimizing cost and facilitating data collection, internet survey administration was selected over the paper and pencil method. (Telephone surveys were impractical since many items required that respondents view mathematical symbols and representations.) An additional benefit of the internet survey method was the possibility of avoiding the difficulty of handling missing data by requiring that all fields receive responses for completion. Data were collected in the fall of 2012.

The means of obtaining contact information for practicing middle school teachers involved sending two documents to administrators of the school districts: a formal letter explaining the research request, and also eliciting teacher contact information, was accompanied with a flyer specifically designed to attract the participation of mathematics teachers. It was hoped that interested districts would distribute this flyer among schools and teachers. The letter that was sent to school districts is attached as Appendix G and the flyer is attached as Appendix H. Once teacher contact information was obtained, eliciting responses from teachers was done by sending a series of email communications addressed to teachers by name. These communications described the kind of participation being requested and also offered the possibility of a small token financial incentive. (Such incentives have been shown to improve response rates (Dillman et al., 2009)). The survey software used to collect survey data also enabled teachers to receive personally addressed emails inviting their participation and providing a link to the internet survey. Teachers received a total of three requests of participation—one initial contact and two reminders—provided they had not completed the survey after the first or second contact. These participation letters are found as Appendices I, J, and K respectively.

Population. Having a very large number of Latino ELLs, Texas is an appropriate setting for this study. Practicing mathematics teachers of middle school students in Texas form the population under investigation for this study. This study sought to define domains of knowledge for teaching mathematics to Latino ELLs and to attempt to develop a measure of such knowledge. This knowledge is of importance to practicing teachers; knowledge possessed by such teachers may be informed by both formal and experiential learning and could potentially lead to greater capacity to teach mathematics to Latino ELLs.

Teachers of *middle school* were chosen because, for many ELLs, the classrooms of these teachers may be the students' first experience in being instructed in English only. In Texas, with very few exceptions, only elementary school students have access to bilingual education programs. Therefore, middle school (and high school) teachers are expected to teach mathematics using English, even though their students may lack English proficiency. As a result, proficiency with teaching ELLs is sought in these teachers, making them an appropriate population for investigation. Intuitively, it would seem that more experienced teachers would perform better on the measure than would inexperienced teachers. This hypothesis was one justification for the comparison of the results of practicing teachers with those of the pre-service teachers obtained in the second pilot-study. Such findings assisted in answering the third research question that is concerned with the validity of the measure.

Intended Sampling and Sample. This study sought to draw a stratified random sample of 300 middle school mathematics teachers in Texas. The purpose for selecting this sampling method was so that survey responses could be used to make generalizations about teachers in Texas. Stratification would enable the separation of teachers into subgroups defined by percentages of Latino ELLs in their school districts. Assumedly, this classification of strata would be useful as it would facilitate the comparison of survey responses from groups of teachers exposed to differing numbers of Latino ELLs; it was expected that patterns would exist in survey responses across such strata. Random selection of participants within strata would ensure, at much as possible, representation of the population.

As it would result in a larger pool from which the random samples within strata could be drawn, the method selected for obtaining lists of potential survey respondents was to select Texas school districts according to their student population, their population of Latino students, and finally their population of ELLs. These data were publicly available on the Texas Education Agency website. Larger school districts would be contacted first because of their access to more teachers. Once the lists of teachers had been obtained, equivalently sized simple random samples of teachers would be drawn from within the separate strata.

The decision of sample size for this study was driven by the requirements of the type of statistical analyses to be performed, which included, among others, confirmatory factor analysis (CFA) and item response theory (IRT). Responding to the third research question of this investigation, a primary goal of the study was to determine the extent to which the measures of KT-MELL obtained in this study had validity and reliability. CFA is fundamental to verifying the underlying structure of the knowledge constructs in question and to developing scales for such knowledge. Based upon recommendations given by DeVellis (2003) for CFA, a “good” sample size would be 300 participants, and even 200 may be adequate (p. 137). The required sample sizes for IRT are more ambiguous. de Ayala (2009) emphasized the importance of not imposing “hard-and-fast rules” to determine appropriate sample sizes for IRT (p. 42). Furthermore, while both de Ayala (2009) and Orlando (2004) agree that, for calibration of a measure, larger sample sizes (more than 500) are better, the later author conceded that “one does not need large sample sizes for a clear picture of response behavior” (p. 8). Based upon the needs of CFA and upon the desire to use IRT for the realistic purposes of obtaining a sense of item response patterns rather than full calibration, it was decided that a sample of 300 practicing middle school mathematics teachers would suffice. Based upon evidence that internet surveys (a discussion of the selection of this method is given in a later section) can suffer very low response rates (Dillman et al., 2009), a conservative estimated response rate of 5% was assumed. Given this response rate, obtaining the desired 300 respondents would require collection of contact information for 6,000 teachers. Furthermore, data publicly available from the website of the Texas Education Agency indicated that, in the 2010-2011 school year, there were more than 99,000 mathematics teachers employed in the state of Texas, making this figure seem obtainable.

Obtained Sample. Complete survey responses were obtained from forty-two mathematics teachers from diverse parts of Texas. Early in the survey administration process, it became clear that severe barriers, given the temporal and financial limitations of this study, stood in the way of contacting teachers and would obfuscate the collection of the desired stratified random sample. In following the plan outlined in the previous subsection of this chapter, many school districts of varying sizes, having large numbers of Latino ELLs were contacted by phone and by email to request participation in the study. Contacts were made at different levels of administration, beginning at the superintendent level and down to the individual school principal level. While some of this effort was fruitful in terms of obtaining teacher contact information, because of the overwhelming infrequency of returned phone calls and emails, it became evident that this method of seeking distribution lists would not suffice. (Fewer than 50 contacts were obtained in this way.) As a result, efforts were redirected toward seeking support from state-level administrators. Separate communications with two representatives at high levels in the state department of educational administration indicated that this state entity, although concerned with the mathematical education of ELLs, would not provide formal support for the study. However, one of the representatives referred the researcher by name to the president of a state professional organization of mathematics teachers, suggesting that this could be a potentially profitable alternative route. Following this guidance, several dozen educational leaders in executive or administrative roles of diverse statewide professional organizations concerned with mathematics and bilingual education were contacted to request support for this study. These requests of support involved asking by email for names and contact information of practicing middle school mathematics teachers. Ultimately, this method of eliciting teacher contact information was the most effective; a few key leaders provided distribution lists of mathematics teachers.

Given the difficulty of obtaining contact information for teachers, the 199 names and email addresses of teachers obtained during the four months of survey administration were gratefully received. Because of the small number of contacts, the intention of drawing a stratified random sample was abandoned and letters of request for participation (Appendices I, J, and K) were sent to all of these teachers. From these 199 contacts, 42 complete survey responses were received, constituting a 21.1% response rate. Although generalization of the results should, at best, be made with caution and, at worst, not be made at all, it can be observed that the characteristics of the sample, in terms of gender, years of teaching experience, and ethnicity, hold some resemblance to the population of Texas mathematics teachers. The properties of the sample are given in Table 4 below. For comparison purposes, statistics concerning the sample of pre-service teachers used in the second pilot-study, as well as statistics publicly available on the Texas Education Agency website concerning mathematics teachers in the state of Texas are given in this table.

Table 4

Characteristics of Sampled Teachers, Pilot-Study Sample of Pre-Service Teachers, and State of Texas Mathematics Teachers

	Practicing Teachers, $N = 42$ Frequency *(Percent)*	Pre-service Teachers, $N = 146$ Frequency *(Percent)*	State of Texas mathematics teachers *(Percent)*
Gender			
Female	31 *(73.8)*	115 *(78.8)*	*(63.3)*
Male	11 *(26.2)*	31 *(21.2)*	*(36.7)*
Years of teaching experience			
0 – 9 years	25 *(59.5)*	-	*(57.3)*
10 – 19 years	14 *(33.3)*	-	*(25.6)*
more than 20 years	3 *(7.1)*	-	*(17.1)*
Grade level taught			
Elementary grades	3*(7.1)*	-	-
Middle school grades	34*(81.0)*	-	-
High school grades	5*(11.9)*	-	-
Courses taught			
K-5 math course	2 *(4.8)*	-	-
6th grade math	14 *(33.3)*	-	-
7th grade math	22 *(52.4)*	-	-
8th grade math	23 *(54.8)*	-	-
Algebra I	13 *(31.0)*	-	-
Geometry	10 *(23.8)*	-	-
Algebra II	6 *(14.3)*	-	-
Mathematical Models	5 *(11.9)*	-	-
Precalculus	4 *(9.5)*	-	-
Calculus	2 *(4.8)*	-	-
Statistics	1 *(2.4*	-	-

Table 4-Continued

Characteristics of Sampled Teachers, Pilot-Study Sample of Pre-Service Teachers, and State of Texas Mathematics Teachers

Extent of mathematical study in college			
Basic math, like College Algebra	14 *(33.3)*	-	-
Several higher level math courses	15 *(35.7)*	-	-
Bachelor or higher degree in math	13 *(31.0)*	-	-
Largest percent of ELLs taught			
0 – 20%	16 *(38.1)*	-	-
20 – 40%	11 *(26.2)*	-	-
40 – 60%	6 *(14.3)*	-	-
60 – 80%	4 *(9.5)*	-	-
80 – 100%	5 *(11.9)*	-	-
Types of professional development experiences for teaching ELLs			
None	13 *(31.0)*	-	-
College course-work	13 *(31.0)*	-	-
Sheltered Instruction	13 *(31.0)*	-	-
ELPS Academy	12 *(28.6)*	-	-
LPAC	8 *(19.0)*	-	-
Other	5 *(11.9)*	-	-
Knowledge of the Spanish language			
None at all	3 *(7.1)*	64 *(43.8)*	-
A few words	13 *(31.0)*	-	-
The basics	11 *(26.2)*	63 *(43.2)*	-
Conversation	9 *(21.4)*	9 *(6.2)*	-
Very fluent	6 *(14.3)*	10 *(6.8)*	-
Knowledge of languages other than Spanish or English			
German	1 *(2.4)*	-	-
French	1 *(2.4)*	-	-
Chinese	1 *(2.4)*	-	-
Russian	1 *(2.4)*	-	-
Japanese	2 *(4.8)*	-	-
Sign language	1 *(2.4)*	-	-
Filipino	1 *(2.4)*	-	-
Malay	1 *(2.4)*	-	-
Portuguese	1 *(2.4)*	-	-

Table 4 indicates that the gender of sampled mathematics teachers was very similar in proportion to the sample of pre-service teachers used in the second pilot-study and to the population of Texas mathematics teachers (within ten percentage points). The years of experience held by the sampled teachers was also very similar to the population of Texas mathematics teachers. While the instrument designed in this study was developed in view of measuring the knowledge of middle school mathematics teachers, which composed the majority of the sample, a small number of mathematics teachers of high school and elementary school students were also included as well. Reflective of teacher level was the finding that the specific mathematics courses most taught by this group of teachers were the middle school level courses of 6^{th} grade, 7^{th} grade, 8^{th} grade math, and Algebra 1. (Some teachers taught various other mathematics courses as well.) Furthermore, the mathematical training enjoyed by the teachers ranged nearly equivalently between having taken "basic math like College Algebra" to "several higher level mathematics courses" to having a "bachelor or higher degree in math". The statistics just presented concern the teachers' gender, mathematical training and mathematical teaching experience.

Data also gave indications concerning other variables potentially related to performance on the KT-MELL instrument, such as teachers' experience of teaching ELLs and of participating in professional development focused on teaching ELLs. For instance, slightly more than half of the teachers (27) had obtained experience teaching classrooms composed of between 0% and 20% ELLs, or between 20% and 40% ELLs. The remainder had taught in classrooms with higher percentages of ELLs. Furthermore, while approximately one third of respondents claimed to have received no formal training related to working with ELLs, the same numbers of respondents claimed to have been exposed to ELL issues during their college course work. Additionally, one third of respondents claimed to have received professional development concerned with the well-known Sheltered Instruction principles (see Chapter 2) for teaching ELLs. Smaller numbers of respondents also admitted to having participated in professional development programs specific to Texas such as the ELPS Academy (English Language Proficiency Standards Academy) or related to LPAC (the Language Proficiency Assessment Committee), or in “Other” professional development experiences.

The survey also asked respondents concerning their proficiency in the Spanish language. Among the mathematics teachers sampled, only three of them (7.1%) claimed total ignorance of the Spanish language; interestingly, 43.8% of pre-service teachers from the second pilot-study made such a claim. The majority of practicing teachers (57.2%) claimed to know “a few words” or “the basics”, and a few admitted to being conversant in Spanish or to being “very fluent”.

Finally, although the ethnicity of teachers was not assessed by means of this survey instrument, based upon using a combination of the names of the teachers and their Spanish language fluency as anecdotal evidence, a cautious estimate of the count of sampled teachers who were of Hispanic ethnicity may be given. Because of the possibility of name-changes, this figure, 13 (31%) should probably be considered an upper bound on the number of Hispanic teachers in the sample. This figure is slightly higher than the percentage of Hispanic mathematics teachers (19.2%) reported for the state of Texas. Furthermore, more than 16 school districts and 23 schools were represented, which covered a wide geographic sector of the state. A map depicting the geographic distribution of responses is given in Figure 8 below.

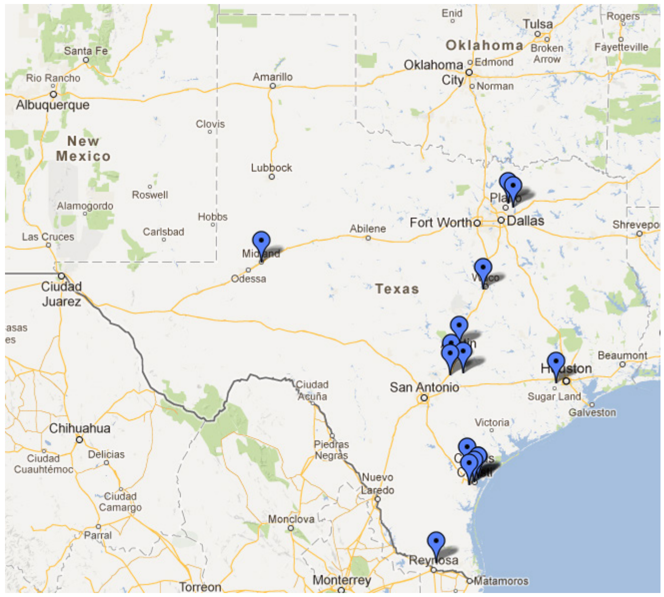


Figure 8. Map of Participating School Districts.

Data and Data Analysis

This study involved the collection of survey data from both pre- and in-service mathematics teachers. Thus, data collected were almost exclusively quantitative. The KT-MELL survey instrument administered in both the second pilot-study and in the full study of practicing teachers contained thirty-two multiple-choice knowledge items in addition to several multiple-choice items intended to capture descriptive data concerning respondents. It also contained one open-ended response question.

The knowledge items were coded in two ways. First, each survey item response was coded with a positive integer that indicated which option had been chosen. (For example, "1" indicated that the first option had been chosen, "2" the second, etc.) Following this coding of all responses, the thirty-two KT-MELL items were re-coded with a "0" or "1" indicating whether the respondent had given the correct or incorrect response, respectively. Thus, the final data set was composite of two distinct tables, one containing categorical data and the other containing dichotomously scored data. The open-ended question was coded according to whether or not a response had been given, and responses to this item were recorded separately as text, with no other coding schema applied.

As described earlier in this chapter, the first two of the research questions, which concern identification of the knowledge domains and aspects of KT-MELL, were addressed through careful reading of the research literature, observation of mathematics classrooms, and thoughtful theorization. Hence, the data analysis methods described in this section apply principally to the third research question, which concerns the properties of the instrument that was developed in this study and of the measurements obtained using this instrument. The data analysis methods employed in this study were chosen for their applicability to the goal of developing a measurement instrument of KT-MELL. Methods were needed that could identify the psychometric properties of the survey items, but that could also lead to evidence concerning the reliability and validity of the measurements obtained by the instrument. Furthermore, as this work was exploratory in the sense that it involved the development of an instrument to measure a novel knowledge construct, several methods were chosen for their utility in identifying patterns in responses from different populations of respondents and in identifying the factor structure of the KT-MELL instrument. The data analysis methods employed in this study were founded in two well-known measurement theories related to test development: classical test theory and item response theory.

Classical Test Theory and Item Response Theory. Both classical test theory (CTT) and item response theory (IRT) provided frameworks for considering the results of and evaluating the properties of the measurements obtained from the KT-MELL instrument. Whereas CTT is the older of the two and was “the mainstay of psychological test development for most of the 20th century", IRT has become a central theoretical framework for the development of psychological measures (Embretson & Reise, 2000, p. 13). These frameworks differ in a number of important ways. The fundamental difference between CTT and IRT is this: whereas in CTT the unit of consideration is the total test score, that is, the sum of correct responses to all items in an instrument, in IRT the individual responses to items (as the name suggests) are the units of consideration (de Ayala, 2009; Embretson & Reise, 2000). Thus, researchers working from the standpoint of CTT investigate the properties of and relationships between respondents' total scores on an instrument. Respondents' ability levels on the underlying trait of investigation, then, are described by the percentages of correct responses to all items. Researchers working from the standpoint of IRT, however, wish to model respondents' ability on the underlying trait by understanding how the individual items are related to the underlying trait. Estimations of the properties of items, such as their levels of difficulty on the continuum of ability and their capacity to discriminate between respondents of varying levels of ability, are used to obtain estimates of the abilities of the respondents. The possibility of estimating both item and person locations on an ability continuum of the latent trait is an example of one of the practical and substantive advantages that IRT offers over CTT (Embretson & Reise, 2000).

This dissertation study employed analysis methods pertaining to both CTT and IRT as they found utility in understanding the properties of the measurements obtained from the KT-MELL instrument. From the perspective of CTT, it was meaningful to consider the percentage of correct responses to items for a number of reasons. Analyses of test scores were useful for understanding both the distribution of scores and the possible sources of variation in scores. These and other usages of CTT are detailed in subsequent sections of this chapter.

IRT is more complicated than CTT. As Embretson & Reise (2000) point out, "few readers would want to understand IRT solely for its psychometric elegance" (p. 39). Nevertheless, IRT was useful for understanding how individual items on the KT-MELL instrument behaved and how they contributed to the quality of the measurements obtained. This study used a two-parameter IRT model, described below.

The Two-parameter IRT Model. The IRT model selected for usage with the data in this study is the two-parameter logistic IRT model (2PL). The two parameters from which this model obtains its name are the item *difficulty* and *discrimination* parameters. (In comparison, one-parameter logistic IRT models (1PL) include only estimates of the *difficulty* of the items.) Item *difficulties* describe the locations of the items on an ability continuum of the latent trait: "In general, the items that are considered to be 'easy' are the ones that persons with low proficiencies have a tendency to answer correctly. Conversely, the 'harder' items are the items that persons with high proficiencies tend to get correct" (de Ayala, 2009, p. 15). Clearly, item difficulties constitute important psychometric properties.

However, the *discrimination* of the items is at least (and perhaps more) important. Discrimination refers to the capacity of the items to differentiate between respondents of varying ability levels of the latent trait, in this case, knowledge. As seen below, the discrimination parameter, α, is a multiplier in the model that results in different predictions for persons of differing abilities. Hence, the discrimination parameter allows researchers to distinguish between respondents of different ability levels (de Ayala, 2009). Since this research is an instrument development study, and it cannot be assumed that all items in the instrument discriminate equally well, selection of the 2PL model allowed for varying estimates of discrimination. Indeed, the exploratory nature of this work implies the possibility of the existence of profoundly different levels of both difficulty and discrimination in the items. Hence, the intended result of using a 2PL IRT model was a fuller picture of the psychometric properties possessed by the survey items.

IRT models estimate the probability of correct responses to the items, given certain parameters. The 2PL IRT models used to analyze the data collected in this study took the following form:

$$p\left(x_j = 1 \mid \theta, a_j, \delta_j\right) = \frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}}, \text{ (de Ayala, 2009).}$$

This model predicts the probability of a correct response on the j^{th} item by a person of θ ability, given that item j has discrimination and difficulty parameters a_j and δ_j, respectively. The above function p is called the item probability function and its graph, which has an ogival (*S*-shaped) form, the item characteristic curve (ICC). Estimates for the difficulty and discrimination parameters of all items were accomplished by means of the method of marginal maximum likelihood estimation using Gauss-Hermite quadrature (Rizopoulos, 2006). This method has a number of advantages, among which are its applicability to the 2PL model and also its efficiency for tests of differing length (Embretson & Reise, 2000). Usage of IRT models also requires satisfaction of a series of assumptions.

Assumptions Required of the Two-parameter IRT Model. Using the 2PL IRT model above requires satisfaction of three *strong* assumptions: the unidimensionality assumption, the conditional independence assumption and the functional form assumption (de Ayala, 2009). Under the unidimensionality assumption, the latent trait of interest, in this case knowledge for teaching mathematics to Latino ELLs, is taken as containing only one dimension or factor. That is, all survey items should serve as manifest indicators of a single latent variable, a single domain of knowledge. Since, the theoretical test framework of the KT-MELL instrument implied multiplicity of latent variables, it seemed likely that, for the full set of items, this assumption would fail and that multiple 2PL models would be applied to separate unidimensional subsets of items. Indeed, IRT models proved invaluable in helping to validate the structure of the knowledge domains underlying the KT-MELL instrument of this study.

The second assumption of IRT models, conditional independence, requires that responses to items are entirely determined by the respondents' locations on the continuum of ability and not by their responses to other items of the instrument (de Ayala, 2009). Difficulties with conditional independence occur when responses to certain items inform responses given to other items. When this happens, responses may not only be determined by the respondents' locations but by responses that they have given to other items on the instrument. In the context of attitude surveys, Schuman & Presser (1996) investigated such *order effects*. They found that "merely placing two questions with similar content next to each other does not necessarily create an order effect" (p. 35). They argued that responses to similar questions will only influence one another if respondents have a *need* to seem consistent. Since respondents to the KT-MELL survey were assured of their anonymity and that the results would be used only for research, such a need for consistency would seem to be absent. Additionally, since the KT-MELL survey items represent a variety of the aspects of knowledge found in three hypothetically separate domains, and are contextualized in many different teaching situations, it would seem that most items would be substantially different from others so as to satisfy conditional independence by invoking responses based upon knowledge and not upon other survey responses. While a very few of the items do draw upon common classroom situational strands, the cognitive tasks required of most such items vary substantially. Upon these bases and for the purposes of applying the 2PL IRT model above, this study assumed the conditional independence of responses to the KT-MELL survey items.

The third assumption, the functional form assumption, required of IRT models is that the data follow the *S*-shaped curve given by logistic models (de Ayala, 2009). This shape implies that, at the extremes of the ability continuum, small changes in ability result in very small changes in the probability of responding correctly to items. In the center of the continuum, that is, at the item's location on the continuum, small changes in ability can result in substantial changes in the probability of correct responses. In applying the 2PL models, this study assumed that the KT-MELL items have this functional form, while admitting along with de Ayala (2009) that both this assumption and the unidimensionality assumption "are rarely ever exactly met in practice" (p. 21). The sections that follow describe the specific analytic methods that were used for observing and assessing the psychometric properties of the items, the reliability of the measurements, patterns in responses, and also the validity of the measurements.

Assessing the Factor Structure of the KT-MELL Instrument. One of the analyses that involved application of methods pertaining to both CTT and IRT was the assessment of the underlying factor structure of the KT-MELL instrument. Since both the calculation of the reliability coefficient α and the application of the 2PL IRT model required the assumption of unidimensionality, it was necessary to determine whether this assumption was met. Indeed, the hypothetical factor structure of the KT-MELL instrument implied the existence of three distinct knowledge domains and survey items were constructed with the intent of measuring these domains. To the end of assessing whether this goal had been achieved the step of confirmatory factor analysis (CFA) preceded other analyses.

"CFA is almost always used during the process of scale development to examine the latent structure of a test instrument" (Brown, 2006). Assessing the latent structure of the KT-MELL instrument was essential for understanding the kind of knowledge that was being measured by the instrument. This step was helpful for informing both the kinds of analyses to be done as well as their interpretations. In addition to evidence concerning the multiplicity of the underlying factor structure that could be obtained from computations of α and from inspections of multiple IRT models, this study employed the CFA method of principal component analysis with Promax rotation. This factor analytic method aligns with methods used to understand the underlying factor structure of the closely related construct mathematical knowledge for teaching (Hill, Schilling, & Ball, 2004).

Uncovering the Psychometric Properties of the KT-MELL Items. The psychometric properties of the KT-MELL survey items found and reported in this study were the following: the percentage of correct responses received by each of the items, point-biserial correlations of items responses with total test score, inter-item correlations, the item difficulty and discrimination parameters, and the item and total test information estimates. Calculating the percentages of correct responses involved simple computation of the ratio of correct to incorrect responses per item. This item property is given merely to give a sense of the characteristics of the entire instrument; it cannot be construed as a measure of the difficulty of the items. Point-biserial correlation is a special case of Pearson product-moment correlation for usage when considering the correlation of a dichotomous variable (such as a graded response) with a continuous variable (such as a total test score) (Nunnally & Bernstein, 1994). The interpretation of this correlation is that it measures the strength of the association of the variation of individual item responses with the variation of the total test score.

The difficulty, discrimination and information estimates were found using the 2PL IRT model described above. As mentioned, estimation of difficulty and discrimination parameters was accomplished by means of the method of marginal maximum likelihood estimation using Gauss-Hermite quadrature (Rizopoulos, 2006). The interpretation of the difficulty parameter is that it locates the item on the continuum of ability, indicating the level of ability above which respondents tend to give correct responses and below which they tend to answer incorrectly. The interpretation of the discrimination parameter is that it indicates the extent to which an item serves to differentiate between respondents of higher and lower abilities.

The IRT information estimate is found as a function of an item's discrimination parameter as follows:

$$I_j(\theta) = \alpha_j^2 p_j (1 - p_j),$$

where p_j is the probability of a correct response on item j, and α_j is the item's discrimination (DeAyala, 2009). The interpretation of the information estimate is that it serves as an indicator of the precision to which an individual's location on the ability continuum—in this case the continuum is knowledge—may be estimated. More precisely, greater information provided by items and instruments implies reduction in the uncertainty about the position of the individuals on the continuum of interest. Since IRT information estimates concern the quality of the measurements obtained using the instrument for which data are modeled, IRT *information* is discussed further in the following section that concerns the means employed by this study for assessing reliability.

Assessing Reliability. As defined by Carmines and Zeller (1979), reliability "concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials" (p. 11). Researchers depend on having reliable measurement instruments since, as the above definition implies, an unreliable instrument would potentially produce differing measurements even though the object being measured has experienced no qualitative change. In the case of psychometric measures such as that being developed by this study, this kind of reliability implies that reliable instruments will give similar results for different respondents who have similar levels of knowledge. But as the above definition of reliability also implies, reliability can be improved by increasing the number of very similar items: "A scale's reliability is increased by reduncancy" (DeVellis, 2003, p. 138). This understanding of reliability pertains to classical test theory. However, from the IRT perspective, "Reliability is enhanced not by redundancy but by identifying better items" (DeVellis, 2003, p. 138). Because of its long-standing acceptance in psychometric research, this study first addresses reliability from the perspective of internal consistency that is central to CTT. However, because of its more recent acceptance and because it is especially suited to developing scales for tests of ability (DeVellis, 2003), emphasis in the discussion of reliability is given to IRT methods. These frameworks treat reliability in different ways as described below.

Internal Consistency Analysis. The standard CTT measure of reliability is Cronbach's (1951) alpha: the "Alpha formula is constructed to apply to data where the total score … will be taken as the person's observed score" (Cronbach, 2004, p. 13).

Precisely speaking, Cronbach's alpha is a generalization of Kuder-Richardson's formula 20, KR-20, which is a measure of internal consistency useful for sets of dichotomously scored items, such as those found in the instrument used in this study (Cronbach, 1951). That is, in the case of dichotomously scored items, KR-20 is equivalent to Cronbach's alpha. Hence, as a measure of internal consistency, this study used Cronbach's alpha, the computation of which is given by the following formula:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_{y_i}^2}\right).$$

In this equation, k gives number of items on the instrument, $\sum \sigma_i^2$ gives the sum of individual (non-communal) variances across all items, and $\sigma_{y_i}^2$ gives the total variance, both individual and covariances (communal). This alpha is, thus, the proportion of total variation across all items that is accounted for by covariation between items. That is, it is a measure of the extent to which variation in item scores is explained by covariation among the items. It is considered to be a *lower-bound* on the internal consistency of the instrument. The multiplier $k/(k-1)$ serves to ensure that values of alpha range between 0 and 1, with a value of 1 indicating that all items are perfectly correlated and 0, that there is absolutely no correlation.

A scale that is internally consistent has the property that all of the items in the scale measure the same thing; they are indicators of the same phenomenon. As a result, the items are highly inter-correlated. An obvious assumption of α as a measure of reliability is the unidimensionality of the latent construct. In the case of the KT-MELL survey instrument, this measure of internal consistency may have limitations. As addressed in an earlier section of this chapter, knowledge for teaching mathematics to

Latino ELLs is hypothesized to be multidimensional and the KT-MELL instrument contains items intended to capture the different dimensions. Indeed, Lee J. Cronbach (2004) has recently stated, "I doubt whether coefficient alpha is the best way of judging the reliability of the instrument to which it is applied" (p. 3). To address this limitation, in addition to investigating its internal consistency, assessing the reliability of the KT-MELL instrument involved a detailed investigation of the qualities of the items and of the total information obtained from the instrument as seen from the perspective of IRT.

IRT Information. In addition to the difficulty and discrimination parameters offered by IRT models as psychometric properties of items, these models also offer an additional psychometric property useful for making judgments about the quality of the measurements obtained by items individually and by sets of items as scales: "To obtain an idea of how well an item and the entire instrument can estimate person locations [on the continuum of knowledge], we examine the item and total information" (de Ayala, 2009, p. 31). Item and total information estimates are indicative of the extent to which the uncertainty about person locations has been reduced. That is, information is inversely related to the standard error of estimation given by IRT models. Mathematically, the item information function is just the first derivative of the item probability function (Revelle, 2013). As mentioned above, an item's information is given as: $I_j\left(\theta\right)=\alpha_j^2 p_j(1-p_j)$, where p_j is the probability of a correct response on item j, and α_j is the item's discrimination (DeAyala, 2009). Furthermore, the total information function for a set of items is just the sum of the item information functions, that is, $I\left(\theta\right)=\sum_{j=1}^{L} I_J(\theta)$. Both the item information and the total information are descriptive of the capacity with which the item and total test, respectively, are useful for estimating person's locations on the ability continuum.

Furthermore, as indicated by the notation $I(\theta)$, information is a function of the location, θ, on the ability continuum. Hence, the item and total information estimates not only indicate the extent to which persons' locations may be accurately determined, but they indicate the specific ranges on the ability continuum within which these estimates of person locations are most accurate. It is possible that items may serve differently for measuring ability levels at different locations along the continuum. For example, an "easier" item that discriminates well may give good information at a lower level of ability than would a more "difficult", but equally discriminating item.

Whereas reliability, as measured by Cronbach's (1951) α, gives an indication of the internal consistency of the instrument to which it is being applied, IRT information gives an indication of the extent to which the instrument (and specific items) serve well in measuring persons' ability levels, *at specific locations along the continuum*. An advantage of IRT information over conventional reliability is that, "unlike the concept of reliability that depends on both instrument and sample characteristics, an instrument's total information is a property of the instrument itself" (De Ayala, 2009, p. 29). Hence, in addressing the issue of reliability—a central component of the third research question of this instrument development study—the focus is on understanding what information properties the KT-MELL instrument possesses.

Construct Validation. Beyond investigating the reliability of the measurements obtained using the KT-MELL instrument, their validity was the second major focal point of the third research question. This study employed a number of means intended to avoid common threats to validity. As has been explicated in the section of this chapter concerned with the KT-MELL test framework, efforts to ensure the *content validity* of the KT-MELL instrument involved three directions: strict adherence to research findings, usage of actual classroom observation for item-contextualization, and review of survey items by a panel of experts in fields related to the mathematics education of ELLs. The effort to ensure the *external validity* of the results involved the attempt to collect a large enough stratified random sample so that generalization of results to the population of Texas middle school mathematics teachers might be possible. These validity issues were addressed through the *design of the study*. An additional source of validity that was addressed using data analytic means was *construct validity*.

Construct validity concerns the extent to which an intended measure of a construct actually measures the construct that it is intended to measure. More precisely, it involves "Determining what psychological constructs account for test performance" and, by implication, what constructs *do not* account for test performance (Cronbach & Meehl, 1955, p. 176). This interpretation of construct validity has implications for the behavior of the measures obtained and assessing the extent to which the measures behave accordingly is central to assessing construct validity. The two means for assessing construct validity employed by this study involved assessing A) both the convergent and discriminant validity of the measurements (Campbell & Fiske, 1959) and B) the extent to which the KT-MELL measures are described by a nomological network (Cronbach & Meehl, 1955).

"…for the establishment of construct validity, *discriminant* as well as convergent validation is required" (Campbell & Fiske, 1959, p. 81). Convergent validity refers to the extent to which different measures of the same construct are correlated while discriminant validity refers to the extent to which measures of different, theoretically unrelated, constructs are uncorrelated. Since KT-MELL is a novel construct and the instrument developed in this study stands alone as a measure of KT-MELL, at the time of this research, no other measures of KT-MELL were extant. Hence, for this study, the assessment of convergent and discriminant validity served the purpose of assessing the extent to which separate sets of theoretically related items served as measures of distinct domains (scales) within KT-MELL. Whereas the correlation among items in the same scale should be relatively high (convergent validity), "Tests can be invalidated by too high correlations with other tests from which they were intended to differ" (Campbell & Fiske, 1959, p. 81). That is, construct (discriminant) validation of the separate scales involved assessing the extent to which items and scores were uncorrelated across scales. The method chosen for assessing both convergent and discriminant validity was an adaptation of the multitrait-multimethod (MTMM) matrix (Campbell & Fiske, 1959). While the different scales could be considered as different *traits,* the *method* for all measurements was the same. That is, measurements were not obtained using different methods, as required of MTMM. As a result, this method gave limited evidence of convergent and discriminant validity.

The second method of assessing construct validity involved the *nomological* validity of the measures. "Construct validation takes place when an investigator believes that his instrument reflects a particular construct, to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses, which are a means of confirming or disconfirming the claim" (Cronbach & Meehl, 1955, p. 187). These authors proposed that validation involves describing a *nomological* network, an "interlocking system of laws which constitute a theory" (p. 187). This understanding of validity is closely related to that offered by DeVellis (2003): "It is the extent to which a measure 'behaves' the way that the construct it purports to measure should behave with regard to established measures of other constructs" (p. 53). As a knowledge construct situated within the construct *mathematical knowledge for teaching* (MKT), a network of variables that could potentially be related to KT-MELL at once emerges. As researchers concerned with measuring MKT have done, this study has also investigated the relationship of KT-MELL with such teacher factors as years of mathematics teaching experience, certification status, and mathematical training (Hill, Rowan, & Ball, 2005). Additionally, many other variables that could potentially be related to KT-MELL (experience teaching ELLs, Spanish language proficiency, ELL professional development experiences, etc.) were also investigated.

Ross et al. (2003) recommend that, "In determining construct validity, the instrument developer describes a network of relationships that includes the survey score. If the observed relationships correspond to the hypothesized relationships, the validity of the survey is confirmed" (p. 352). For simplicity, and because of its conventional usage for assessing relationships between teacher knowledge and other factors (Hill, Ball, Schilling, 2008), this investigation was done using linear-regression modeling. The different variables of interest were regressed on teachers' KT-MELL scale scores. The application of linear-regression models requires the assumption that data be approximately normally distributed; distribution of scale scores is treated in the results section. The results of the linear-regression models were sought as evidences concerning the construct (nomological) validity of the measures obtained as measures of the knowledge construct KT-MELL.

Observing Response Patterns. A number of additional statistical methods were useful for uncovering and attempting to explain patterns of responses that existed within the KT-MELL survey data. Pearson's chi-squared test of independence (Pearson, 1900) was used to understand the ways in which different categories of respondents gave correct and incorrect responses to individual survey items. This statistical method requires that a number of assumptions be made that were, generally, easily satisfied by the data. The foremost of these concern the independence of the observations (individual responses to items) and the required minimum cell count required for the test (5 or more were sought for usage in this study). Another method used to understand response patterns was the calculation of Pearson's r as a measure of the correlation between pairs of items. A final statistical method that was used to understand response patterns was analysis of variance, ANOVA. ANOVA was used to test the null hypothesis of the equivalence of mean scale scores between differing categorical groups of teachers. In the case of two groups, this method is equivalent to a linear regression model involving only two variables. Hence, the results of ANOVA were also useful for investigating possible sources of construct validity.

Software Used for Data Analysis. For this study, data collection, data handling and data analysis were accomplished by using five software programs. SNAP 10 Professional was used to both email teachers with requests for survey participation and to collect survey data. Microsoft Excel was used to assemble and "clean" spreadsheets of data for analysis. IBM SPSS Statistics (Version 20) and *R* (version 2.13.1) were used for computations related to a large number of statistical tests. ltm (Rizopoulos, 2006) is a software package available for *R* that was used to compute IRT models.

Summary. This study involved the design and field-testing of an instrument for measuring mathematics teachers' knowledge for teaching Latino English Language Learners (KT-MELL). Presented in this chapter were the processes that resulted in a three-dimensional framework of knowledge, as the testing framework from which operationalizations in the form of survey items were developed. This conceptual model and testing framework (Figure 3 and Table 3, respectively) are the results of careful theorization based upon results found in the research literature concerned with the mathematics education of ELLs and with mathematics teachers' knowledge, as well as of more than thirty hours of the observation of the classrooms of middle school mathematics teachers in central Texas. During their development, survey items underwent two successive pilot-study phases that involved samples of pre-service mathematics teachers, as well as content validation by a panel of experts in related fields, before being administered to actual practicing middle school mathematics teachers.

The quantitative survey data collected from teachers in this study were analyzed using methods pertaining to two theoretical perspectives of test development: classical test theory and item response theory. Methods were chosen for their utility in giving evidence of the reliability and validity of the measurements obtained using the thirty-item KT-MELL instrument. Specifically, data analysis methods were selected that could answer questions regarding the underlying factor structure of the instrument, the psychometric properties of the items, the reliability and validity of the measurements, and the particular patterns of responses that emerged. Several well-known statistical software packages available at the time of the study were used, which were capable of performing computations needed for the required analyses.

CHAPTER 4: RESULTS

Responding to the methodological framework presented in the previous chapter, this chapter presents the results of the data collection and data analysis. To more clearly focus the treatment of these, the research questions of this study are reconsidered here.

1. What are the domains of knowledge needed to teach mathematics to Latino English Language Learners (KT-MELL)?
2. Drawing from the research literature and middle school mathematics classroom observations, what are some of the aspects of these domains of knowledge when teaching Latino English Language Learners at the middle school level?
3. What evidences of reliability and validity does an instrument developed to measure these domains and aspects exhibit?

The first two questions have been addressed by the chapter concerned with reviewing the research literature that treats the mathematics teaching and learning of ELLs, and even more directly, by the methodology chapter that gave a theoretical framework of domains of knowledge potentially needed for teaching mathematics to ELLs as well as specific aspects of these domains. (See Figure 3 and Table 3 of that chapter for domains and aspects, respectively.) The focus of this chapter concerns the third research question. Specific topics addressed in this chapter include the internal consistency and empirical factor structure of the instrument given both by classical test theory and item response theory analyses, properties of the items and scales, analysis of item response patterns, and validity of the measurements.

Evidence of Internal Consistency and Factor Structure.

The third of the research questions that guided this study concerned the reliability and validity of the measurements obtained using the exploratory KT-MELL survey instrument. This section addresses reliability while the penultimate section of this chapter addresses validity. Interweaved in this section is the assessment of the underlying factor structure of the KT-MELL instrument. The methods for evaluating the worth of this instrument's measurements in terms of their capacity for accurately estimating the knowledge possessed by respondents drew from classical test theory (CTT) and item response theory (IRT). Both of these frameworks assume the unidimensionality of the underlying latent factor structure of the instrument being investigated. Hence, determining whether the KT-MELL instrument was unidimensional or multidimsional (as hypothesized) was crucial before further analyses could be accomplished. This determination was accomplished based upon the theoretical orientation of the items and by using evidence gained from the measures of internal consistency. That is, the assessment of the reliability of the measures lead to results concerning the underlying latent factor structure.

Internal consistency refers to "the homogeneity of the items within a scale" (DeVellis, 2003, p. 27). That is, a scale that is internally consistent has the property that all of the items in the scale measure the same thing; they are indicators of the same phenomenon. As a result, the items are highly inter-correlated. Indeed, measures of reliability such as Cronbach's (1951) coefficient alpha are, equivalently, measures of internal consistency. That is, "a measure's [alpha] reliability equals the proportion of total variance among its items that is due to the latent variable and thus is communal" (DeVellis, 2003, p. 35). This measure of internal consistency is central to CTT.

Computation of Cronbach's coefficient alpha assumes unidimensionality of the latent construct. As explained in the chapter on the methodology of this study, knowledge for teaching mathematics to Latino English Language Learners was, hypothetically, composite of three distinct domains of knowledge. These domains were: knowledge of difficulties encountered by these students in mathematics classes (KDIFF), knowledge of their capacities for learning mathematics (KCAP), and knowledge of strategies for teaching them mathematics (KSTRAT). Hence, it was conceivable and even expected that the instrument being used in this study would exhibit the psychometric properties of a multidimensional scale rather than a unidimensional one. Usage of traditional measures of reliability that depend on unidimensionality of the instrument then became questionable. Indeed, Lee J. Cronbach (2004) has recently stated, "I doubt whether coefficient alpha is the best way of judging the reliability of the instrument to which it is applied" (p. 3). To address this limitation, this assessment of the reliability of the instrument was accomplished from two perspectives: classical test theory and the more recent, item response theory. Indeed, assessing the reliability of the instrument proved invaluable as a means of understanding it underlying factor structure.

Classical Test Theoretic Evidence of Reliability. Classical test theory (CTT) takes, as the unit of analysis, the individual respondent's observed score; that is, the sum of the item scores on the entire instrument (de Ayala, 2009). Indeed, as a measure of the construct in question, knowledge for teaching mathematics to Latino ELLs, it would seem useful to obtain a single descriptive score. (The later section concerning the properties of scales addresses with the properties of this score.) It is important to note here that Cronbach's "Alpha formula is constructed to apply to data where the total score … will be taken as the person's observed score" (Cronbach, 2004, p. 13). Thus, from a CTT viewpoint, it is meaningful and appropriate to consider what evidence of reliability can be seen from calculations of alpha on the full set of items and on subsets of items. These findings had direct implications for the internal consistency of the instrument, and indirect implications for it underlying factor structure. This section does just that. Precisely speaking, Cronbach's alpha is a generalization of Kuder-Richardson's formula 20, KR-20, which is a measure of internal consistency useful for sets of dichotomously scored items, such as those found in the instrument used in this study (Cronbach, 1951.) That is, in the case of dichotomously scored items, KR-20 is equivalent to Cronbach's alpha. The formula for computing this alpha is given below:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_i^2}{\sigma_{y_i}^2}\right)$$

In this equation, k gives number of items on the instrument, $\sum \sigma_i^2$ gives the sum of individual (non-communal) variances across all items, and $\sigma_{y_i}^2$ gives the total variance, both individual and covariances (communal). This alpha is, thus, the proportion of total variation across all items that is accounted for by covariation between items. That is, it is a measure of the extent to which variation in item scores is explained by covariation among the items. It is considered to be a lower-bound on the internal consistency of the instrument. The multiplier $k/(k-1)$ serves to ensure that values of alpha range between 0 and 1, with a value of 1 indicating that all items are perfectly correlated and 0, that there is absolutely no correlation.

To evaluate the internal consistency of the instrument, and also to evaluate the validity of the unidimensionality assumption, Cronbach's alpha was computed for the full set of thirty-two items used in this study. Based upon responses from the 42 participating teachers, it was found that $\alpha = 0.450$. This amounted to saying that less than 50% of the variation of scores on the items is attributable to covariation among the items. There is disagreement about acceptable cutoff points for α. Nunnally and Bernstein recommend that "In the early stages of predictive or construct validation research, time and energy can be saved using instruments that have only modest reliability, e.g., .70" (p. 264). Indeed, based upon the argument given by these authors, .70 is frequently used as a rule of thumb in social science research. However, others consider that, for dichotomously scored items, KR-20 greater than .5 is acceptable (McGahee, T. W., & Ball, J., 2009).

One interpretation of the alpha found for the items on the instrument in this study is that they are imperfectly representative of the construct in question. This is an issue of content validity and has been addressed in the chapter on methodology. Based upon the positive results of review of the instrument by a panel of experts, an alternative interpretation seemed in order. One such alternative interpretation of this alpha was that the instrument contained too few items. Indeed, the Spearman-Brown prophecy formula showed that an additional fifty-nine similar items (total of 91 items) would bring reliability to .70. It should also be noted here that reliability for the full set of items was higher for practicing teachers, $\alpha = .450$, than it was for pre-service teachers in the second pilot-study, $\alpha = .345$. This difference seemed to indicate that another interpretation was appropriate.

An interpretation that seemed to conform more closely to the test-framework from which the instrument was created was that variation of item scores could probably not be explained by variations of the true score *on a single underlying construct*. This interpretation implied that knowledge for teaching mathematics to Latino ELLs would be composite of multiple factors. To investigate this possibility, a set of items that was maximally reliable in terms of coefficient alpha was found.

Through a process of item-deletion, a reduced set of 13 items was found that were conceptually related and that had a measure of internal consistency $\alpha = .642$. Thus, 64.2% of the variation of item scores among this reduced set of items was attributable to covariation among the items. This amounted to a nearly 20% increase in internal consistency. Furthermore, the 13 items that composed this subset were almost exclusively taken from the KDIFF and KCAP domains; they were concerned with knowledge of the difficulties that Latino ELLs face and with knowledge of the capacities for learning mathematics that such students may have because of their background knowledge and cultural-linguistic attributes.

This finding was important in at least two ways. First, it seemed to confirm the multiplicity of the underlying latent constructs, negating the assumption of unidimensionality. But this finding also seemed to confirm an earlier suspicion (see Chapter 3) that knowledge of difficulties and capacities may not be distinct domains. Rather, based upon the great improvement in internal consistency obtained using items in these domains, it appeared that they pertained to a single domain. That is, this result gave evidence that the two hypothetical domains of KDIFF and KCAP could, in fact, be indistinct.

While, this set of items was interesting as a scale in and of itself. However, a more encompassing theory of internal consistency, and one that could make usage of a larger number of the items in the instrument, was sought. Item response theory (IRT) was useful as an aid in the search for the appropriate factor structure and understanding of reliability. The next section presents the results of an analysis of internal consistency that made usage of (IRT).

Item Response Theoretic Evidence of Reliability. In comparison with CTT, which takes the observed score on the entire instrument as the unit of analysis, IRT takes the item as the unit of analysis (De Ayala, 2009). Thus, IRT models aim to predict observed responses based upon characterizations of the respondents' and of the items' positions on a continuum of ability level of the latent trait, that is, the underlying construct of interest. (In this case, KT-MELL.) In comparison with CTT models that compute correlations of items as a measure of reliability, IRT models are useful for computing item and test *information* estimates. Such estimates of information serve as indicators of the precision to which an individual's location on a continuum of ability—in this case the continuum is knowledge—may be estimated. That is, greater information provided by items and instruments implies a reduction in the uncertainty about the position of the individuals on the continuum of interest. An advantage of IRT information over conventional reliability is that, "unlike the concept of reliability that depends on both instrument and sample characteristics, an instrument's total information is a property of the instrument itself" (De Ayala, 2009, p. 29). Hence, in this instrument development study, IRT information estimates are extremely useful as indicators of the *worth* of the instrument in question.

As was explained in the chapter concerning methodology, the IRT model selected for usage with the data in this study is the two-parameter logistic IRT model (2PL). While a 1PL model includes only estimates of the *difficulty* of the items—and this is an important statistic—the *discrimination* of the items is at least (and perhaps more) important. Discrimination refers to the capacity of the items to differentiate between ability levels of the latent trait, in this case, knowledge. Since this research is an instrument development study, it cannot be assumed that all items in the instrument discriminate equally well. The exploratory nature of this work even implies the possibility of profoundly different levels of both difficulty and discrimination. For this reason, 2PL IRT models are considered in this section, as they add to the difficulty parameter the additional parameter of discrimination. The intended result is a fuller picture of the psychometric properties possessed by the survey items.

The IRT models considered in this chapter take the following form:

$$p\left(x_j = 1 \mid \theta, a_j, \delta_j\right) = \frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}} \ .$$

This model predicts the probability of a correct response on the j^{th} item by a person of θ ability given that item j has discrimination and difficulty and values a_j, δ_j, respectively.

Satisfying the Assumptions Required of the Two-parameter IRT Model. Usage of the 2PL IRT model above required satisfaction of three *strong* assumptions: the unidimensionality assumption, the conditional independence assumption and the functional form assumption (de Ayala, 2009). Under the unidimensionality assumption, the latent trait of interest, in this case knowledge for teaching mathematics to Latino ELLs, is taken as containing only one dimension or factor. That is, all survey items should serve as manifest indicators of a single latent variable, a single domain of knowledge. The relationship of the IRT models that are considered in this chapter with the unidimensionality assumption was critical. Indeed, the models themselves were partly instrumental in evaluating the assumption and arriving at conclusions regarding the underlying factor structure of the KT-MELL instrument. Since, the theoretical test framework of the KT-MELL instrument implied multiplicity of latent variables, it seemed likely that, for the full set of items, this assumption would fail and that multiple 2PL models would be applied to separate unidimensional subsets of items.

The second assumption of IRT models, conditional independence, required that responses to items be entirely determined by the respondents' locations on the continuum of ability and not by their responses to other items of the instrument (de Ayala, 2009). Difficulties with conditional independence occur when responses to certain items inform responses given to other items. When this happens, responses may not only be determined by the respondents' locations but by responses that they have given to other items on the instrument. In the context of attitude surveys, Schuman & Presser (1996) investigated such *order effects*. They found that "merely placing two questions with similar content next to each other does not necessarily create an order effect" (p. 35). They argued that responses to similar questions will only influence one another if respondents have a *need* to seem consistent. Since respondents to the KT-MELL survey were assured of their anonymity and that the results would be used only for research, such a need for consistency was assumed absent. Additionally, since the KT-MELL survey items represent a variety of the aspects of knowledge found in three hypothetically separate domains, and are contextualized in many different teaching situations, it seemed that most items were substantially different from others so as to satisfy conditional independence by invoking responses based upon knowledge and not upon other survey responses. While a very few of the items draw upon common classroom situational strands, the cognitive tasks required of most such items vary substantially. Upon these bases and for the purposes of applying the 2PL IRT model above, conditional independence of responses to the KT-MELL survey items was assumed.

The third assumption, the functional form assumption, required of IRT models is that the data follow the *S*-shaped curve given by logistic models (de Ayala, 2009). This shape implies that, at the extremes of the ability continuum, small changes in ability result in very small changes in the probability of responding correctly to items. In the center of the continuum, that is, at the item's location on the continuum, small changes in ability can result in substantial changes in the probability of correct responses. In applying the 2PL models, this study assumed that the KT-MELL items have this functional form, while admitting along with de Ayala (2009) that both this assumption and the unidimensionality assumption "are rarely ever exactly met in practice" (p. 21).

A Unidimensional 2PL IRT Model. As a help in understanding how the survey items behaved together under the unidimensionality assumption, a two-parameter IRT model that took all thirty-two of the survey items as a single conceptual factor, estimated on the set of forty-two practicing teachers, was computed. This model yielded the item difficulties and discriminations given in Table 5 below.

Table 5

2PL IRT Item Difficulty and Discrimination Estimates on Full 32 Items Instrument

Item	**Difficulty**	**Discrimination**	**Item**	**Difficulty**	**Discrimination**
1	0.079	2.695	**17**	0.178	-0.684
2	0.789	-0.903	**18**	2.249	-1.090
3	8.777	0.033	**19**	4.611	-0.515
4	1.764	-0.565	**20**	0.552	0.845
5	6.733	-0.304	**21**	6.058	0.032
6	-0.292	0.666	**22**	.4.038	0.410
7	9.839	-0.275	**23**	-2.159	0.274
8	4.327	-0.275	**24**	10.192	-0.090
9	1.749	1.169	**25**	-8.099	0.113
10	-1.032	-0.485	**26**	12.273	-0.132
11	-2.174	-0.497	**27**	-54.865	0.026
12	2.247	-0.636	**28**	-1.802	-0.533
13	3.158	0.384	**29**	-0.796	-0.235
14	-5.183	0.315	**30**	5.732	0.162
15	-6.190	0.211	**31**	2.797	0.557
16	1.059	25.942	**32**	-0.149	0.595

Inspection of the item difficulties yielded evidence that the items varied widely in the level of difficulty; negative values could be considered "easier" while positive values were "more difficult". Of more interest in Table 5, though, were the signs found on the discrimination figures. Positive discriminations indicated that respondents of higher ability had a better probability of responding correctly to these items, as would be expected. However, negative discriminations indicated that respondents of higher ability had reduced probability of responding to these items correctly. Another way to consider these data was by looking at the item characteristic curves (ICCs) for the set of items under this IRT model. The ICCs shown in Figure 9 below model the relationship between ability level (*x*-axis) and the probability of a correct answer (*y*-axis). The item discriminations are seen in the slopes of the curves. Items with positive discrimination have ICCs with positive slopes while items with negative discriminations have ICCs with negative slopes. Furthermore, steepers slopes indicate higher discriminations.

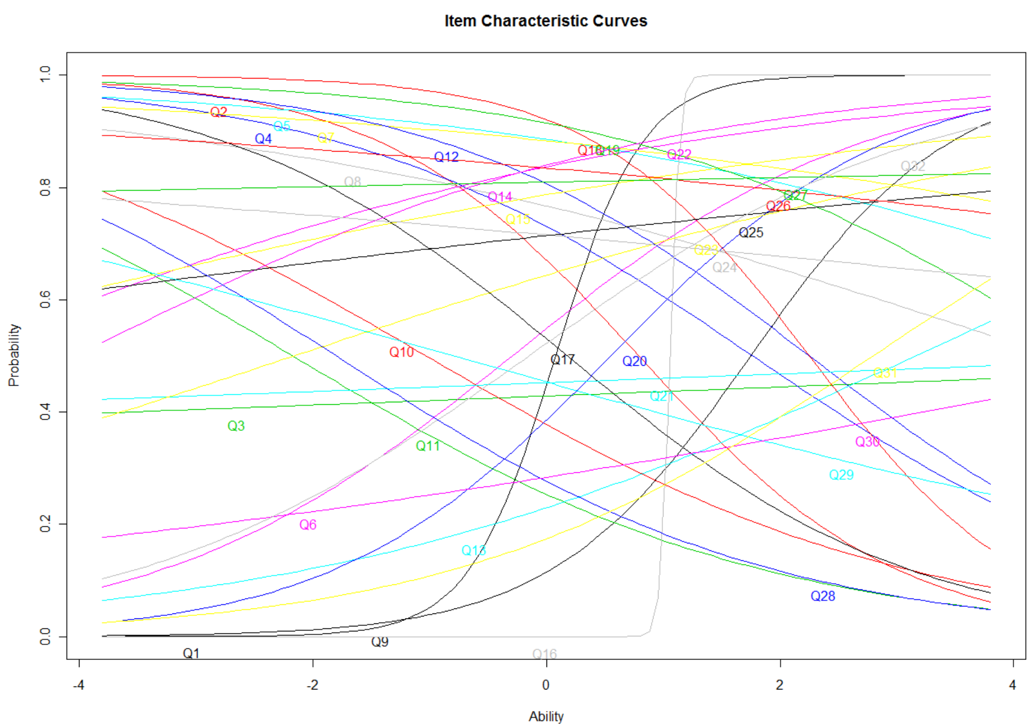


Figure 9. 2PL IRT Item Characteristic Curves (ICCs) of the Full 32 Item Test

As mentioned above, the large number—about half—of items that have negative discriminations indicated that many of the items performed counterintuitively; higher ability teachers tended to have less likelihood of responding correctly to these items. As discussed in the section concerning alpha reliability, one interpretation of this result was that items are poorly representative of the construct of interest. However, just as with alpha reliability, so with IRT models, an assumption of unidimensionality was being made. Therefore, as evidenced by the results of alpha reliability on the reduced test discussed in the previous subsection, by the difference in signs of IRT discriminations on these items, and as supported by the theoretical framework of the instrument under study, a serious look at multiple scales was in order. The remainder of this discussion of internal consistency of the instrument presents the IRT results of separating the items into two distinct domains of knowledge.

Multidimensionality: Creating Separate IRT Models. Based upon the results of the IRT model just described, the complete set of items was partitioned into two distinct scales according to the sign of the discrimination coefficient: scale 1 contained items with positive discriminations and scale 2 contained items with negative discriminations. Intuitively, these sets of items appeared to be measuring distinct kinds of knowledge since the probability of giving a correct response to the items held similar relationships with the contiuum of ability under the unidimensionality assumption. The alpha reliability of these two factors was $\alpha = .5325$ and $\alpha = .5168$ respectively. It was at once apparent that separating the items into two distinct scales resulted in improved reliability, as predicted.

However, a number of the items in each of these scales were arguably divergent in conceptual orientation from other items in the scales. Several items that did not appear to be conceptually related and that were not highly correlated as a result, or that appeared to constradict theory by having negative item discriminations, were removed. A total of eleven items were omitted from inclusion in either domain. The two resulting scales were composite of items that pertained mainly to the KDIFF and KCAP domain or mainly to the KSRAT domain. Table 6 below identifies the scales, lists the items composing the scale and gives α as a measure of internal consistency.

Table 6

Knowledge Domains, Items and Internal Consistency of the Two Separate Scales

Scale	Items											α
KDIFF/ KCAP	Q1	Q6	Q9	Q13	Q15	Q16	Q20	Q22	Q25	Q31	Q32	0.6214
KSTRAT	Q5	Q7	Q8	Q10	Q11	Q18	Q19	Q24	Q26	Q29		0.6056

This table shows that a total of twenty-one items were included in these scales, eleven items in the KDIFF/KCAP domain and ten items in the KSTRAT domain. This table also shows substantial improvement in the reliability of the scales over the reliability of the full set of thirty-two items. Although this figure falls below that recommended by Nunnally and Bernstein (1994), the Spearman-Brown prophecy formula (op. cit.) indicates that adding as few as five similar items to each of the scales would result in both of them achieving α = .70.

As a means of verifying the validity of this partition of items, a confirmatory factor analytic (CFA) step computed factor loadings on two distinct factors. Factor loadings are given in Table 7 below. Based upon CFA, the items were then categorized as pertaining to the factor upon which the items loaded more heavily. For example, Q1 loaded more heavily on the first factor (.584) than on second (-.313). As is seen in Table 7, the results of this factor analysis, which used the method of principal component analysis, very closely coincided with the partitioning of items according to their conceptual orientation and the sign of their IRT discrimination parameters. (Component 1 corresponds with the KDIFF/KCAP domain and Component 2 corresponds with the KSTRAT domain.)

Table 7

CFA Item Loadings on Two Factors

	Structure Matrix	
	Factor	
Item	2	1
Q1	-.313	.584
Q5	.595	.087
Q6	.073	.341
Q7	.556	.044
Q8	.467	-.066
Q9	-.019	.592
Q10	.492	-.092
Q11	.293	-.202
Q13	.027	.276
Q15	.367	.497
Q16	-.273	.738
Q18	.444	-.354
Q19	.564	-.040
Q20	-.073	.448
Q22	.130	.317
Q24	.488	.177
Q25	.340	.259
Q26	.342	.020
Q29	.309	-.209
Q31	-.077	.273
Q32	.100	.500

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

It is important to note here that CFA typically requires sample sizes that are significantly greater than the sample obtained for this study ($N = 42$). Hence, this factor loadings are very likely to be innacurate. Notwithstanding this limitation, it was notable that factor loadings were in overwhelming agreement with the the assignment of items to scales obtained from their conceptual alignment and using their IRT discrimination coefficients. That is, for each of the items, the value of the factor loading was greater on the factor that corresponded with the scale to which the item had been assigned using its conceptual orientation and the sign of its disrimination coefficient, excepting only item Q25 (discussed below). The agreement of these three evidences—the conceptual alignment of the items, the parity of the signs of IRT disrimination coefficient, and factor loadings—were taken as strong evidence of the existence of the two scales.

Furthermore, it came as no surprise that the resulting assignment of items agreed fairly well with the hypothesized test framework of the instrument. Items with positive discriminations, and that were more heavily loaded on Factor 2, were mostly concerned with knowledge of the ways in which language can cause difficulties *or* capacitate learning in mathematics classes. Items with negative discriminations, more heavily loaded on Factor 1, were mostly concerned with knowledge of appropriate strategic decisions involving ELLs in specific mathematics tasks. The only item that differed in assignment location based upon its discrimination coefficient and factor loading was item Q25. This item concerns knowledge of linguistic complexity in mathematics as a source of difficulty for understanding problems. Thus, it aligns more closely with the items in the KDIFF/KCAP domain and was so assigned.

As discussed extensively in Chapter 3, the IRT analog to CTT's *reliability* is *information*, that is, the reduction in the uncertainty with which person's ability levels can be estimated. The continuation of this discussion of reliability is found in the appropriately subsections of the following section, which fully explicate the properties of the two scales identified during the process of reliability assessment. The following section of this chapter is given to presentation of the two scales of knowledge that were found to compose the KT-MELL instrument.

Two Scales: Difficulties/Capacities and Strategies

The section that precedes this section has explained how the process of reliability assessment resulted in the identification of two scales forming the underlying structure of the KT-MELL instrument. The three goals of this study were identification of the knowledge domains required for teaching mathematics to Latino ELLs, documentation of specific aspects of these domains, and development of an instrument that could measure the aspects of these domains in a valid and reliable way. The initial conceptual framework (Figure 3) expressed the hypothesis of the existence of three, distinct domains of knowledge: knowledge of difficulties that Latino ELLs may encounter in mathematics classes (KDIFF), knowledge of ways in which their bilingual, cultural, and prior learning can capacitate Latino ELLs to learn mathematics (KCAP), and knowledge of strategies for teaching mathematics to Latino ELLs (KSTRAT). However, evidences obtained from methods pertaining to CTT and to IRT indicated that the KDIFF and KCAP domains seemed to form a single underlying latent construct; items belonging to these domains seemed to be drawing on one body of knowledge. For this reason, the KDIFF and KCAP domains were combined, KDIFF/KCAP. The remaining subsections of this chapter treat the KDIFF/KCAP and KSTRAT domains separately, first explaining the psychometric properties of the items and scales as well as the descriptive scale statistics. After this, item analyses begins with descriptions of the items and are followed by observations of inter-item correlations and teacher factors that were associated with item responses.

The KDIFF/KCAP Scale. Items forming the KDIFF/KCAP scale are concerned with how language can be both a source of difficulty and an avenue for communication for Latino ELLs in mathematics classrooms. They call upon knowledge of the ways in which linguistic complexity in mathematics problems can be a source of difficulty for Latino ELLs. They also call upon knowledge of the ways in which students can use their first language to express mathematics effectively and that teachers can use students' written and spoken productions to positively affect learning. Item specifications for the KDIFF/KCAP scale are given in Appendix L.

Psychometric Properties of the KDIFF/KCAP Scale. The items forming the KDIFF/KCAP scale showed improvements over the full set of 32 items, not only in alpha reliability, but also in the IRT coefficients of difficulty and discrimination. For this set of items, all discriminations were positive. Hence, there was evidence that, for all of the items in this scale, respondents of greater ability on the underlying construct were more likely to give correct responses. This formed part of the confirmation that these items formed a unidimensional scale. A summary of the psychometric properties of items in this scale is given in Table 8 below.

Table 8

Psychometric Properties of KDIFF/KCAP Items

Item	Percent Correct (%)	Point-biserial correlation	Difficulty	Discrimination	Information	Percent (%) of total information
Q1	47.6	.378	0.099	1.365	1.35	4.92
Q6	54.8	.223	-0.284	0.716	0.72	2.62
Q9	16.7	.371	1.738	1.169	1.17	4.26
Q13	23.8	.270	5.447	0.217	0.15	.55
Q15	78.6	.293	-1.195	1.485	1.49	5.43
Q16	16.7	.473	1.034	19.160	19.16	69.83
Q20	40.5	.317	0.886	0.467	0.46	1.68
Q22	83.3	.166	-2.148	0.843	0.84	3.06
Q25	71.4	.178	-2.769	0.337	0.31	1.13
Q31	19.0	.177	2.854	0.542	0.53	1.93
Q32	52.4	.272	-0.094	1.260	1.26	4.59
				Total	27.44	100.00

In addition to the percentage of correct responses received by each item, the table above gives the IRT difficulty estimates as well. Although percentage of correct responses and difficulty are related, they are not equivalent. Percentage of correct responses may give an intuitive conception of the difficulty of the items. However, the IRT difficulty parameters are more descriptive because they estimate locations of the items on the continuum of ability. Table 8 illustrates that item difficulties ranged from very easy (-2.769) to very difficult (5.447). An item's difficulty parameter is the point on the continuum at which, as ability increases, the probability of a correct response to the item becomes fifty percent. Thus, although items Q9 and Q16 received the least percentage of correct responses of items in this scale, item Q13 was more *difficult* that both of these items because its location on the ability continuum is greater; the probability of answering item Q13 correctly was associated with a greater level of ability than was the same probability for items Q9 and Q16. Futhermore, excepting item Q16, their discriminations were approximately quite similar, varying between 0.217 and 1.485. (More is said about Q16 in the analysis of items forming this scale.) An additional statistic given in Table 8 is the point-biserial correlation of items with the total scale score (sum of correct responses). That all point-biserials were positive indicates that correctly responding to each of these items was correlated with obtaining higher total scale scores. This is another indicator of the unidimensionality of the scale.

Another way to consider the difficulty and discrimination properties is by looking at the graphs of the item probability functions, that is, at the item characteristic curves (ICCs). Figure 10 gives the ICCs for the items in this scale, which are also exhibitive of the improved coefficients described above. The slopes of all curves, item discriminations, are positive. Furthermore, as evidenced by the steepness of the slopes, many of the items have similar levels of discrimination. Items with significantly different discriminations are apparent by their differences in slope, either more or less steep.

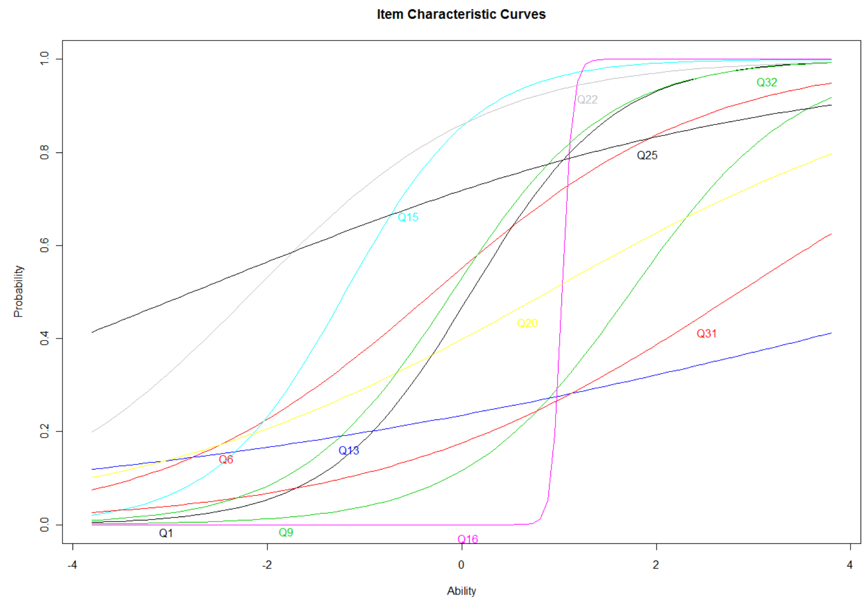


Figure 10. ICCs for the KDIFF/KCAP Items.

Of interest to this discussion of the reliability of this scale, and closely related to the item discriminations seen in Figure 10, is the *information* given by the items in the scale and by the *total information*. The item information and total information functions give estimates of how well the items and instrument can estimate the locations of respondents on the knowledge continuum. Thus, it is important to assess the level of *information* provided by this set of items. On the item level, "the maximum amount of information provided by an item varies as a direct function of the magnitude of α_j", that is, of the item's discrimination (De Ayala, 2009, 101). The item information function is just the first derivative of the item probability function, represented by the ICC (Revelle, n.d.). Hence, an item gives its maximum information when the probability of a correct reponse to the item is 50%. An item's information is given as: $I_j(\theta) = \alpha_j^2 p_j (1 - p_j)$, where p_j is the probability of a correct response on item j, and α_j is the item's discrimination (DeAyala, 2009).

Among the psychometric properties given in Table 8, an important figure to notice is the percentage of total information given by the scale that is provided by the individual items. For the eleven items in the KDIFF/KCAP scale, the amount of total information obtained by the scale is 27.44, which indicates that this collection of items functions well as a scale in reducing the uncertainty about individual person locations on the ability continuum. The percentage of total information given by items in this scale ranges from 0.55% (Q13) to 69.83% (Q16). Item Q16 clearly gives the most information in this scale. Items Q15, Q1, Q32, and Q9 all have information estimates greater than 4. The remaining six items, in the order of decreasing information, are Q22, Q6, Q31, Q20, Q25, and Q13. There is a direct relationship between discrimination and information. Items that discriminate better reduce the uncertainty about the ability of the respondents, and thus provide more information concerning the level of knowledge of respondents.

Of chief importance to the question of reliability is the total information given by this scale. The total information gives an indication of how well the scale functions in estimating person locations. Furthermore, it also describes precisely for which kind of respondents—that is, respondents possessing which level of knowledge—the scale is most reliable. The total information function is just the sum of the item information functions. Hence, $I(\theta) = \sum_{j=1}^{L} I_J(\theta)$ (de Ayala, 2009). The total information function for the KDIFF/KCAP scale is given in Figure 11 below.

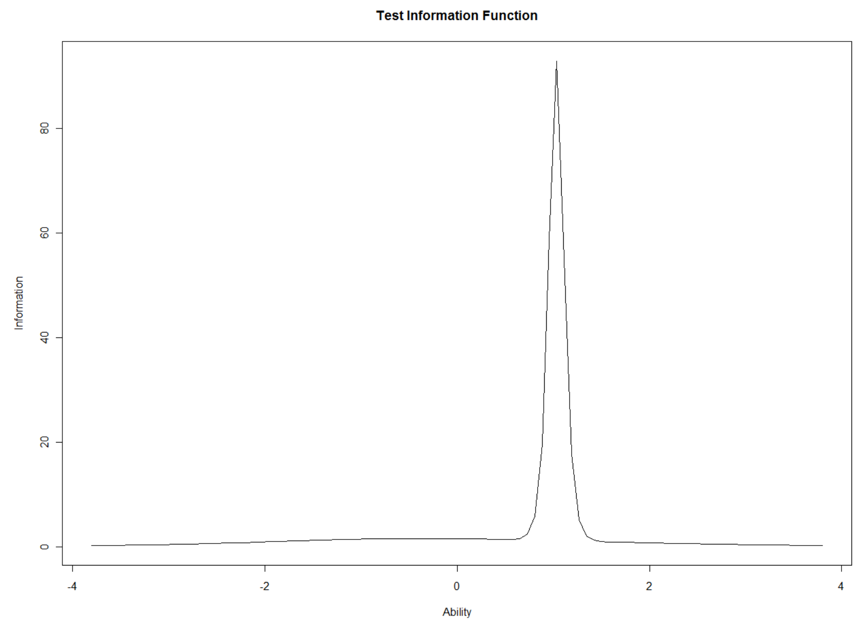


Figure 11. KDIFF/KCAP Total Information Function.

Figure 11 shows that the KDIFF/KCAP scale provided maximum information for respondents who had ability slightly greater than zero, that is, slightly greater than average ability. In other words, the uncertainty of person locations for respondents who were of slightly greater than average ability was minimized. Important to notice is that the high peak evident in this figure implies that the items in this domain functioned very well in reducing the uncertainty with which estimates of person abilities could be made, *at these given ability levels*. For persons of less than average ability, or of very great ability on the KDIFF/KCAP scale, very little could be assessed of their level of knowledge. This estimation capacity is a property of this scale of items and responds directly to the question of the reliability of the measurements.

Analysis of KDIFF Items. This subsection of the results takes a much closer look at responses to the KDIFF scale items in the attempt to better understand what factors influenced teachers to choose correct and incorrect survey responses. The purpose of making such an analysis is to attempt to answer questions such as the following: how were items conceptually related in the minds of teachers?, what characteristics of the sampled teachers were related to correct and incorrect item responses within the respective knowledge domains and across domains?, what patterns seemed to exist in the ways that the sampled teachers thought about the aspects of knowledge presented in the survey items? As a result of this analysis, it is hoped that a better understanding of the teacher characteristics associated with higher levels of knowledge will be gained. Based upon empirical identification of the two scales that composed the survey instrument, the analysis is separated into two sections: knowledge of linguistic issues that either cause difficulty for or capacitate mathematics learning for Latino ELLs (KDIFF/KCAP), and knowledge of strategies for helping Latino ELLs to learn mathematics (KSTRAT). Each section begins with a description of the items composing the scale. Inter-item correlations are then discussed. Finally, significant associations between teacher factors and the items are addressed.

KDIFF/KCAP: Knowledge of Linguistic Issues that Either Cause Difficulty for or Capacitate Mathematics Learning for Latino ELLs. This section presents an item analysis of the eleven items composing the KDIFF/KCAP scale: knowledge of linguistic issues that either cause difficulty for or capacitate mathematics learning for Latino ELLs. Survey items in this domain require that teachers choose specific linguistic elements of mathematics tasks that can be troublesome for ELLs, or that ELLs linguistic background or prior learning can hinder mathematics learning. They also require that teachers identify ways in which ELLs' linguistic capacity, such as knowledge of the Spanish language, can help them to express mathematics ideas.

KDIFF/KCAP Item Descriptions. The item analysis begins with brief descriptions of the items composing the KDIFF/KCAP scale. These item descriptions are intended to provide a context for the descriptions of psychometric properties given above and of statistical associations with teacher factors that follow. Each description begins by framing the task posed to the respondent and closes with a statement of what a correct response to the item should indicate about teachers' knowledge.

Item Q1 presents a situation in which a mathematics teacher has posed a question to students regarding the solution to a common proportions problem. A student has responded with the correct procedural answer to the teacher's question, but has done so in Spanish. The item, which belongs to the set of items concerned with strategies for ELLs, requires teachers to determine the best response to be given to the student. This item requires that teachers be knowledgeable of the ways in which ELL students' own mathematical expressions, verbal or written, (Chval and Khisty, 2009) can be used pedagogically to improve mathematical learning for the students. Correct responses to this item may indicate that teachers understand that requiring ELL students to explain their own mathematical reasoning, by using additional means of expression such as gesturing (Shein, 2012), are beneficial for their mathematical learning.

Item Q6 is similar to Q1 in that it presents a students' solution to a mathematics problem and then requires the respondent to choose the best way to respond to the student. The student has solved a long division problem, yet using an algorithm that is different than the one traditionally taught to students in the United States; this algorithm is more commonly used in Central America. The respondent must choose between options that either lead the student to learn a different method or require the student to give an explanation of the method that has been used. As with Q1, correct responses to this item may indicate that teachers understand that requiring ELL students to explain their own mathematical reasoning are beneficial for their mathematical learning.

Item Q9 is an item which elicits knowledge of ways in which linguistic complexity in a mathematics problem can cause difficulty for ELLs. In this item, the teacher has given to the students a mathematics problem that requires students to read and interpret several instructional phrases before approaching the problem with their peer groups. The item asks teachers to decide which part of the task could pose particular difficulty for Latino ELLs. Correct responses to this item may indicate that teachers understand that interpreting written mathematics instructions can be a significant source of difficulty for ELLs.

Q13 is an additional item that elicits knowledge of linguistic complexity as a source of difficulty for Latino ELLs. In this item student groups have been given a geometric task involving measuring the surface area of a geometric solid and then determining how the area changes by the application of a scale factor. The respondent must choose which specific mathematics word used in the instructions may be difficult for Latino ELLs to understand. Correct responses to this item may indicate that teachers are familiar with the relative difficulty for Latino ELLs of certain English words (Spanish-English cognates and non-cognates) used in mathematics classes.

Item Q15 elicits knowledge of the ways in which students' bilingualism can capacitate, rather than impede, their mathematical reasoning (Moschkovich, 2002). The item presents an example of a student's mathematical reasoning (algebraic manipulation) about the relationship between the surface area of a sold and the surface area that results from application of a scale factor to the dimensions of the sold. In her work the student has used both mathematical symbols and Spanish mathematical words. The respondent must decide whether or not the student's usage of Spanish is reflective of poor understanding. Correct responses to this item may indicate that teachers are aware that

ELLs may prefer to use their first language to express mathematics and that teachers are observant of when such usage has assisted the student to express mathematics concepts.

Item Q16, which draws upon the same problem situation as that posed in Q15, asks teachers to decide whether the ELL student's mathematical reasoning about the relationship between the surface area of a three-dimensional solid and a scale factor is mathematically precise or not. In this item, the student has used a mixture of both Spanish language and mathematical symbols to carry out the steps of the reasoning. Teachers must determine whether the reasoning is mathematically precise. The item is intended to capture teachers' knowledge of the ways in which bilingualism can facilitate mathematical expression for some Latino ELLs. Correct responses to this item may indicate that teachers are able to correctly understand and evaluate the precision of ELL students' mathematical logic.

Item Q20 elicits knowledge of the ways in which ELL students' bilingualism can capacitate them to understand mathematics (Moschkovich, 2002). In this item a mathematics teachers has required students to read and speak a series of different English words that explain the concept of slope. The respondent to this item is asked to select, from a list of words, the word that ELL students may grasp more readily than other students. Correct responses to this item may indicate that teachers are familiar with specific English mathematics words that have Spanish cognates that assist Latino ELLs to readily grasp the meaning of the English words.

Item Q22, like Q20, elicits knowledge of the ways in which ELL students' bilingualism can capacitate them to understand mathematics (Moschkovich, 2002). In this problem situation, the teacher has asked students to describe the properties of rectangles. Three students have responded in succession to the question, one of whom has used the

Spanish word *paralela* and gestures to describe the relationship of the sides. The respondent must make a judgment, based upon students' responses, concerning the extent to which the students understand the properties of rectangles. Correct responses to this item may indicate that teachers are aware that ELL students may use multiple modes of expression, including their native language and gestures, to correctly express mathematical concepts.

Item Q31 elicits knowledge of linguistic complexity in mathematics as a source of difficulty for some ELLs. This item requires the respondent to select the correct words, phrases, or a combination of both words and phrases, in a particular mathematics word problem, that could be confusing to ELLs. Correct responses to this item may indicate that teachers are aware of ways in which linguistically complex mathematics problems, especially involving polysemy (Martiniello, 2009), can cause difficulty for ELLs.

Item Q32 calls upon knowledge of the ways in which connecting mathematical symbols with English language can be difficult for some ELLs. More specifically, it calls upon knowledge of the difficulty that ELLs may have in translating between symbolic mathematical expressions and verbal mathematical expressions. To answer this item, the teacher is required to decide which incorrect phrase would most likely have been provided by an ELL as the verbal translation of a given algebraic expression. Correct responses to this item may indicate that teachers are familiar with a specific way in which translating mathematical symbols to English statements can be difficult for ELLs.

KDIFF/KCAP *Inter-item Correlations.* As a further means of shedding light on patterns of responses that could be observed in the data, all pair-wise inter-item correlations were considered. This part of the analysis was aimed at understanding how responses to different items were related. Table 9 below gives the correlation matrix (Pearson's *r*) for the eleven KDIFF/KSTRAT items.

Table 9

Inter-Item Correlations for KDIFF/KCAP Items.

	Q1	Q6	Q9	Q13	Q15	Q16	Q20	Q22	Q25	Q31	Q32
Q1	1	.196	.341*	.139	.033	.469**	.282	.171	.075	.023	.050
Q6	.196	1	.150	.059	.108	.278	-.030	-.021	.166	.075	.091
Q9	.341*	.150	1	.050	.234	.314*	.282	.029	.283	-.054	.043
Q13	.139	.059	.050	1	.019	-.100	.564**	.100	.106	.156	.085
Q15	.033	.108	.234	.019	1	.234	.076	.078	.183	.106	.315*
Q16	.469**	.278	.314*	-.100	.234	1	.152	.200	.000	.108	.426**
Q20	.282	-.030	.282	.564**	.076	.152	1	.108	.092	.094	-.088
Q22	.171	-.021	.029	.100	.078	.200	.108	1	-.141	.054	.213
Q25	.075	.166	.283	.106	.183	.000	.092	-.141	1	.038	.030
Q31	.023	.075	-.054	.156	.106	.108	.094	.054	.038	1	.220
Q32	.050	.091	.043	.085	.315*	.426**	-.088	.213	.030	.220	1

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

As can be seen in the table, among the KDIFF/KCAP items, the only correlations that were statistically significant were positive ones; no negative correlation was statistically significant. The two items that were the most highly correlated were Q13 and Q20. This correlation was expected since both of these items involve knowing that Spanish-English cognates can mediate the difficulty of, or in other words, capacitate the learning of mathematics concepts for ELLs. Items Q16 and Q1 were also positively correlated. Q16 involved assessing the precision of a students' mathematical reasoning, a mixture of Spanish language and mathematical symbols, and Q1 involved responding to a students' mathematical explanation given in Spanish. As an explanation for this correlation, it is conceivable that the two thought processes could be related. An item that was positively correlated with both Q1 and Q16 was Q9. This item involves knowledge that interpreting written instructions in English can be a source of great difficulty for ELLs. Q16 was also positively correlated with Q32. Q32 involved knowledge of a specific way in which translating mathematical symbols to English statements can be difficult for ELLs. An additional item that was positively correlated with Q32 was Q15. The latter of these involves knowledge of bilingualism as a capacity for the expression of mathematical thought. The remaining items, Q6, Q22, Q25, and Q31 held no specific inter-item correlations.

Teacher Factors Associated with the KDIFF/KCAP Items. The final section of the analysis of the KDIFF/KCAP scale was aimed at identifying the teacher factors that were associated with correct responses given to the items in this scale. Pearson's chi-squared test of independence (Pearson, 1900) was used to determine whether differences in numbers of correct responses seemed to be associated with teacher factors. A summary of the statistically significant results, is presented in Table 10 below.

Table 10

Teacher Factors Associated with Difference in Percentages of Correct Responses to KDIFF/KCAP Items.

Item	taught math for more than 5 years?		Taught a class with more than 40% ELLs?		Possess other certifications		Response given to open question	
	NO	*YES*	*NO*	*YES*	*NO*	*YES*	*NO*	*YES*
Q1								
Q6	14.3%	75.0%	40.7%	80.0%				
	$\chi^2 = 13.888$***		$\chi^2 = 5.999$**					
Q9			7.4%	33.3%			9.1%	44.4%
			$\chi^2 = 4.667$**				$\chi^2 = 6.364$ **	
Q13							18.2%	44.4%
							$\chi^2 = 2.689$*	
Q15								
Q16								
Q20			29.6%%	60.0%				
			$\chi^2 = 3.692$*					
Q22								
Q25					60.0%	88.2%	63.6	100.0
					$\chi^2 = 3.953$**		$\chi^2 = 4.582$**	
Q31							24.2%	0.0%
							$\chi^2 = 2.695$*	
Q32								

*. χ^2 statistic is significant at the 0.10 level.
**. χ^2 statistic is significant at the 0.05 level.
***. χ^2 statistic is significant at the 0.01 level.

These data should be interpreted with caution. Strictly speaking, a significant Chi-square statistic only gives evidence that the percentage of correct (or incorrect) responses received by a particular category of teacher appears to be different than what may have been expected if responses had been evenly, and randomly, distributed across categories of teachers. Although it cannot be said that any teacher factor has *caused* a correct response, it is informative to observe that a number of teacher factors have statistically significant differences in correctness of responses on the items of the survey. Furthermore, excepting gender, all teacher factors that appeared to indicate greater experience (years of teaching, percentage of ELLs taught, multiplicity of educational certifications), or greater interest in teaching ELLs (response given to open question) were statistically associated with a greater percentage of correct responses to the items. Among these teacher factors, no statistically significant result was associated with a lower percentage of correct responses.

Summary of KDIFF/KCAP Item Analysis. Up to this point the analysis of the KDIFF/KCAP scale has presented the psychometric properties of the items, including the item information, the inter-correlations, and several teacher factors that were significantly associated with item responses. The KDIFF/KCAP scale concerns knowledge of issues that either cause difficulty for or capacitate learning in mathematics for Latino ELLs. Brief descriptions of the items preceded these results of item analysis to aid in their interpretation. As a way of summarizing the results of the KDIFF/KCAP item analysis, Figure 12 presents a graphic representation of the items composing the scale. In this weighted-node and weighted-edge graph, the relative magnitude of the information provided by each items is indicated by the size of its node and the strength of the inter-item correlations are given adjacent to the edges connecting nodes ("*" denotes significance at the .05 level and "**" at the .01 level). Furthermore, teacher factors that were associated with differences in percentages of correct responses are represented by an appropriate letter placed adjacent to the node, as explained in the footer.

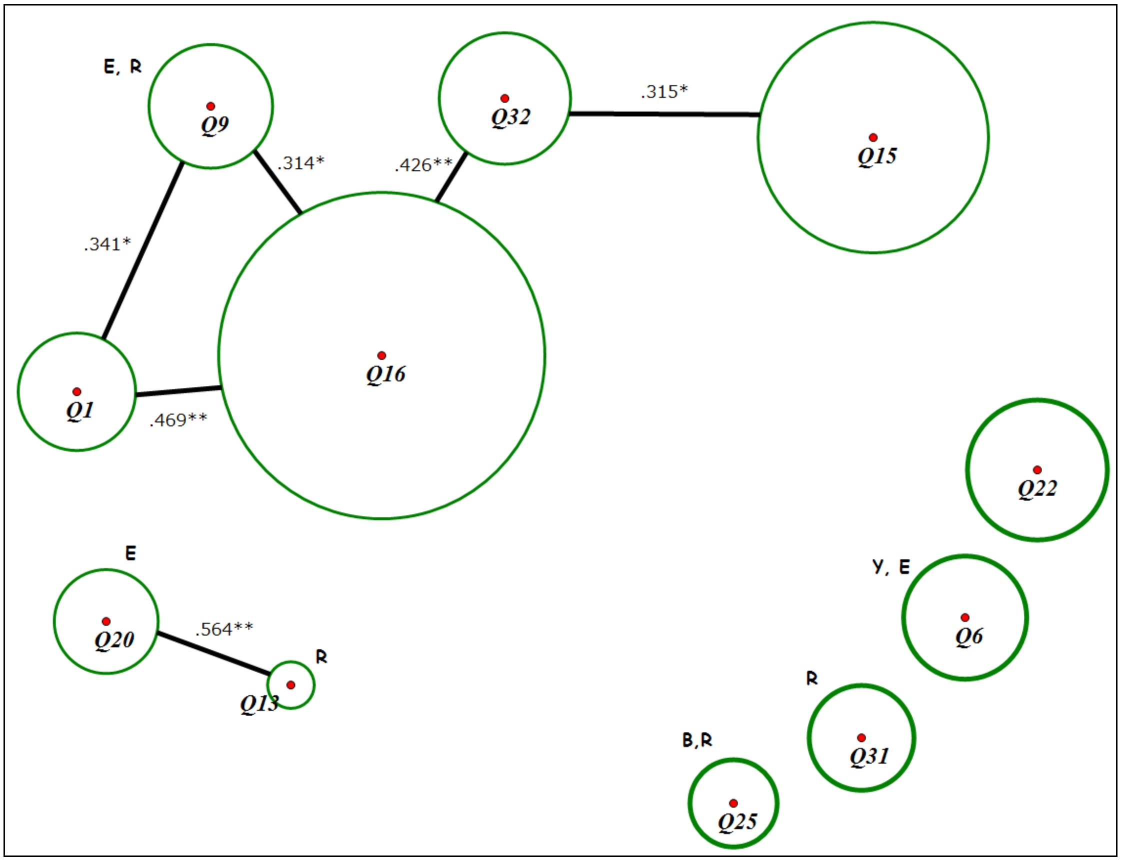


Figure 12. Semantic Map of Information, Inter-item Correlations, and Significant Teacher Factors for KDIFF/KCAP Items. The size of the nodes corresponds to the relative amount of information provided by the items. Furthermore the size of the connecting edges corresponds to the strength of the correlation; large edges indicate stronger correlation. Teacher factors associated with differences in numbers of correct responses to the items are indicated by the following abbreviations: *Y*-more than 5 years of mathematics teaching experience, *E*-experience teaching classrooms containing 40% or more ELLs, *B*-breadth of educational licensure (multiplicity of educational certifications), *R*-response given to the open question concerning teaching ELLs.

Figure 12 may offer a semantic map of the KDIFF/KCAP scale. Three clusterings of items can be identified. The most informative cluster, containing items Q1, Q9, Q15, Q16, and Q32, exhibits of number of inter-item associations. Items Q6, Q22, Q25, and

Q31 provide nearly equivalent amounts of information, and shared no statistical correlation between themselves or with other items in the scale. Finally, two items that were strongly inter-correlated, and that contributed smaller amounts of total information to the scale, were Q13 and Q20. These relationships were investigated in the attempt to better understand the patterns of correct responses to the items in this scale. The semantic map given in Figure 12, though incapable of explaining causation, is merely offered as a synthesis of the psychometric properties of the KDIFF/KCAP items and of item relationships and as possibly indicative of directions for future cognitive research related to the analyses which follow.

This subsection of the results has offered an explanation of the properties of the KDIFF/KSTRAT items and scale. It began with a summary of the psychometric properties of the items, offering comparisons of the item difficulty and discriminations parameters, and focusing on the information given by the items and the total scale. The item and total information obtained by this scale indicated that the scale performed acceptably in reducing the uncertainty with which persons' locations on the knowledge continuum could be estimated, for persons of greater than average ability, and very poorly elsewhere. The interpretation of this IRT statistic is analogous to reliability (internal consistency) central to CTT. The purpose of this investigation was to respond to third research question of this study as it asks concerning the reliability of the KT-MELL measures.

This subsection also gave a thorough item analysis. It began with detailed descriptions of the items, followed by observations of inter-item correlations and of teacher factors that were associated with differences in numbers of correct responses. The findings concerning item information, inter-item correlations and significant teacher factors were presented in synthesized form in Figure 12, which may serve as a semantic map of the KDIFF/KCAP scale. The following subsection is given to a detailed analysis of the items forming the KSTRAT scale.

The KSTRAT Scale. Items forming the KSTRAT scale are concerned principally with knowledge of teaching strategies that result in the creation of environments that are rich in language and mathematics (Chval and Chavez, 2011). These items required that teachers justify strategic decisions regarding the teaching of specific mathematics concepts to ELLs. They also required knowledge of alternative mathematics notation and of algorithms that Latino ELLs may bring with them as prior knowledge and that can, as a result, capacitate or hinder their learning of mathematics. Item specifications for the KSTRAT scale are given in Appendix M.

Psychometric Properties of the KSTRAT. This section describes the psychometric properties of the ten items that form the scale for Domain 2. The items forming the KSTRAT scale showed improvements over the full set of 32 items, not only in alpha reliability, but also in the IRT coefficients of difficulty and discrimination. For this set of items, all discriminations were positive. Hence, there was evidence that, for all of the items in this scale, respondents of greater ability on the underlying construct were more likely to give correct responses. This formed part of the confirmation that these items formed a unidimensional scale. A summary of the psychometric properties of items in this scale is given in Table 11 below.

Table 11

Psychometric Properties of KSTRAT Items

Item	Percent Correct (%)	Point-biserial correlation	Difficulty	Discrimination	Information	Percent (%) of total information
Q5	88.1	.354	-1.446	2.366	2.37	20.36
Q7	88.1	.309	-1.816	1.471	1.47	12.63
Q8	76.2	.257	-1.447	0.939	0.94	8.08
Q10	38.1	.320	0.421	1.787	1.79	15.38
Q11	26.2	.266	1.255	0.984	0.98	8.42
Q18	88.1	.265	-2.301	1.027	1.03	8.85
Q19	90.5	.332	-2.329	1.184	1.18	10.14
Q24	71.4	.274	-1.625	0.608	0.6	5.15
Q26	83.3	.250	-2.481	0.712	0.71	6.10
Q29	45.2	.240	0.360	0.576	0.57	4.90
				Total	11.64	100.00

In comparison with the items in KDIFF/KCAP scale, the items forming the KSTRAT scale were evidently *easier.* Seven of the ten items received correct responses from more than 70% of respondents. Furthermore, the difficulty parameters of these seven items, which are more precise estimates of difficulty, are nearly all negative. This means that, on the ability continuum, these items are located below the average ability level. Hence, respondents with less than average ability on this scale have a greater than 50% likelihood of giving correct responses to many of these items. An additional statistic given in Table 11 is the point-biserial correlation of items with the total scale score (sum of correct responses). That all point-biserials were positive indicates that correctly responding to each of these items was correlated with obtaining higher total scale scores. This was taken as another indicator of the unidimensionality of the scale.

As another way to consider the difficulty and discrimination properties given in Table 11, Figure 13 depicts the item characteristic curves (ICCs). These curves are also exhibitive of the improved coefficients described above. The slopes of all curves, item discriminations, are positive. Furthermore, as evidenced by the steepness of the slopes, many of the items have similar levels of discrimination. Items with significantly different discriminations are apparent by their differences in slope, either more or less steep.

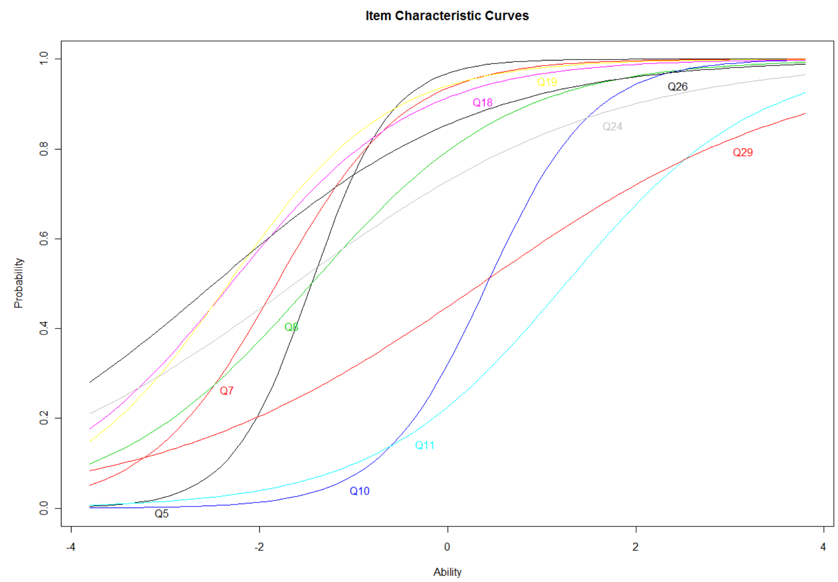


Figure 13. ICCs for the KSTRAT Items.

Another indication that this scale of items is "easier" than the KDIFF/KCAP scale is the locations of the ICCs. An item's difficulty is the point on its ICC as which the probability of a correct response to the item becomes greater than 50%, hence, the point of inflection. As seen in Figure 13, nearly all ICC have inflection points to the left of zero, indicating that nearly all items could be correctly answered by persons of less than average ability. Another interesting property of these items was that their discriminations were more uniform. This property can be seen in the approximately parallel curves.

Of interest to the discussion of the reliability of this scale, and closely related to the item discriminations seen in Figure 13, is the *information* given by the items in the scale and by the *total information*. There is a direct relationship between discrimination and information. Items that discriminate better reduce the uncertainty about the ability of the respondents, and thus provide more information concerning the level of knowledge of respondents. The item information and total information functions give estimates of how well the items and instrument can estimate the locations of respondents on the knowledge continuum. Thus, it is important to assess the level of *information* provided by this set of items.

The information estimates for the KSTRAT items, like the discrimination parameters, were more evenly distributed than those of the KDIFF/KCAP scale. They ranged in value from 0.57 to 2.37. Alternatively, no item provided less than 4.9% (Q29) of the total information and none provided more than 20.36% (Q5). Of chief importance to the question of reliability is the total information given by this scale. The total information gives an indication of how well the scale functions in estimating person locations. Furthermore, it also describes precisely for which kind of respondents—that is, respondents possessing which level of knowledge—the scale is most reliable. The total information provided by this scale is 11.64, which indicates that the scale functioned well in reducing the uncertainty about person locations on the ability continuum. The total information function for the KDIFF/KCAP scale is given in Figure 14 below.

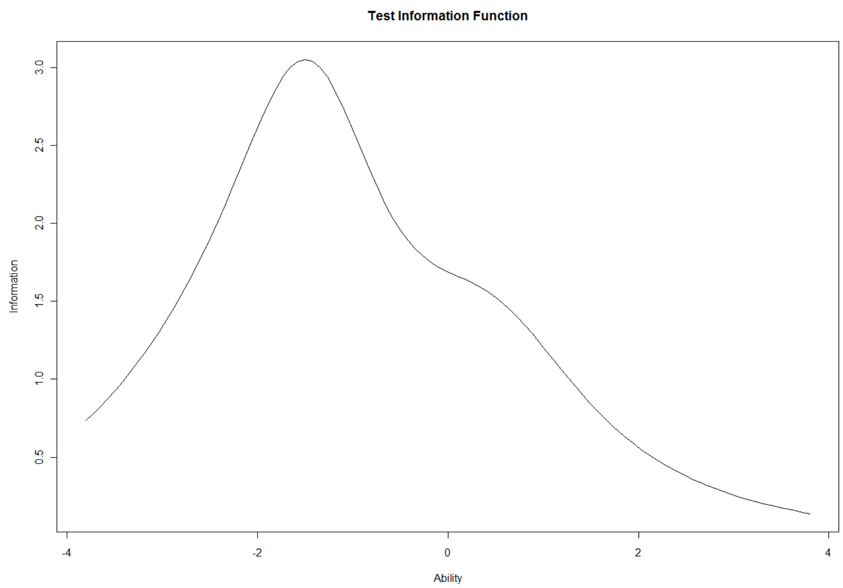


Figure 14. KSTRAT Total Information Function.

Figure 14 shows that the KSTRAT scale provided information for respondents who had a wide range of ability levels. (The range of measurment capacity for this scale is greater than that of the KDIFF/KCAP scale.) However, information was maximized for persons who were of slightly less than average ability. In other words, the uncertainty of person locations for respondents who were of slightly less than average ability was minimized. Important to notice is that the peak evident in this figure implies that the items in this domain functioned well in reducing the uncertainty with which estimates of person abilities could be made, *at these given ability levels*. As indicated by the curve, information slightly leveled off for persons of around average ability before. This small rise should indicate that a few items of greater difficulty give their maximum information for persons of average or greater ability. For persons of extremely low or extremely high ability on the KSTRAT scale, very little could be assessed of their level of knowledge. This estimation capacity is a property of this scale of items and responds directly to the question of the reliability of the measurements.

Analysis of KSTRAT Items. This subsection of the results takes a much closer look at responses to the KSTRAT scale items in the attempt to better understand what factors influenced teachers to choose correct and incorrect survey responses. As a result of this analysis, it is hoped that a better understanding of the teacher characteristics associated with higher levels of knowledge will be gained. This section begins with a description of the items composing the scale. Inter-item correlations are then discussed. Finally, significant associations between teacher factors and the items are addressed.

KSTRAT: Knowledge of strategies for instructing Latino ELLs in Mathematics. This section presents a thorough item analysis of the ten items composing the KSTRAT: knowledge of strategies for instructing Latino ELLs in mathematics. Items in this scale are concerned principally with knowledge of strategic teaching decisions that may improve the opportunity for students to communicate about mathematics. They are also concerned with teachers' knowledge of mathematical notations that are commonly taught in the United States or in Central America that can, because of the students' familiarity (or lack of familiarity) with these, facilitate mathematics communication or become a barrier to such for Latino ELLs.

KSTRAT Item Descriptions. The analysis of KSTRAT items begins with brief descriptions of the items composing the scale for this domain. These item descriptions are intended to provide a context for the descriptions of psychometric properties given above and of statistical associations with teacher factors that follow. Each description begins by framing the task posed to the respondent and closes with a statement of what a correct response to the item should indicate about teachers' knowledge.

Item Q5 requires respondents to evaluate the validity of a student's solution to a division problem. In this item the student's work has been presented in such a way as to allow comparison with the work of two other students. The two other students have used alternative notation for the division algorithm. The student in item Q5 has used the long division algorithm and notation commonly taught in the United States. Correct responses to this item may indicate that teachers have knowledge of division algorithms used by ELL students and can evaluate the validity of a student's work.

Item Q7 involves knowledge of teaching strategies that may encourage the creation of a classroom environment that is conducive to the rich usage of language and mathematics (Khisty & Chval 2002). In this item, the teacher has given students a linguistically complex problem involving arithmetic sequences. Respondents must choose, from a list of options, the best reason for the teacher to use cooperative grouping as an instructional mode for this mathematics task. Correct responses to this item may indicate knowledge of the benefit to ELLs that encouraging them to speak about mathematics brings.

Item Q8 is conceptually attached to Q7. In this item respondents must decide whether having students work in groups on a mathematical task was a good idea on the part of the teacher or a bad idea. Two options give reasons for which group work may be a good idea and two options give reasons for which it may be a bad idea. Correct responses to this item may indicate knowledge of strategies that can promote mathematical communication among ELLs.

Item Q10 is similar to Q7 and Q8 in that it assesses knowledge of strategies for promoting a linguistically and mathematically rich classroom environment for ELLs. In this item the respondent must decide from a list of instructional decisions, all of which have been implemented in the particular mathematics task, which of the decisions is most supportive of ELLs' understanding of the nature of the task. Correct responses to this item may indicate knowledge of strategies that support ELLs' learning of mathematics by involving them in active discourse about mathematics. Items Q7, Q8, and Q10 are designed to serve as multiple indicators of this particular strategic knowledge.

Item Q11 presents a situation in which an ELL from Central America has responded incorrectly to a subtraction problem involving the English numeric word *billion*. The respondent must identify which incorrect solution the ELL would most likely have provided. Correct responses to this item may indicate knowledge of ways in which differences in commonly used mathematical language can be a potential barrier to mathematics achievement for Latino ELLs.

Item Q18 assesses teachers' knowledge of an alternative notation, commonly used in much of Latin America, for the division algorithm. In this item student work is given for a division problem. Based solely upon the student's work, the respondent must correctly determine which division operation the student intended to perform. Correct responses to this item may indicate knowledge of mathematical capacities that Latino ELLs may possess for solving mathematics problems using different, yet no less valid, notation.

Item Q19 is a further item intended to measure knowledge of strategies that encourage ELLs to practice speaking mathematically. In this item the teacher has presented to the students a series of mathematics learning objectives, in the form of English sentences, to the class and has required students to read them aloud. The respondent must then select the appropriate reason for which this instructional decision would be beneficial for ELLs. Correct responses to this item may indicate knowledge of visual displays that promote the connection of mathematics concepts with mathematics language and of strategies that promote a language-rich environment for ELLs (Chval and Chavez, 2011).

Item Q24 is intended to measure knowledge of ways in which mathematical language can cause difficulty in carrying out mathematical tasks for Latino ELLs. This item presents a mathematical task in which students must select the appropriate probability of the occurrence of an event, given certain conditions. The problem and conditions are linguistically complex and involve polysemy; number words have been used in their demonstrative sense rather than numerical sense. Correct responses to this item may indicate that teachers have knowledge of ways in which words with multiple meanings can cause difficulty for ELLs in mathematics classes.

Item Q26 is a further item intended to measure knowledge of ways in which ELLs may express themselves using valid, yet non-traditional, notation. Similar to items Q5 and Q18, this item presents examples of two students' work of mathematical division. Respondents must decide which student has given a correct solution to the division problem. Correct responses to this item may indicate knowledge of mathematical capacities that Latino ELLs may possess for solving mathematics problems using different, yet no less valid, notation.

Item 29 presents a linguistically complex multi-step mathematics problem. The problem involves four different sentences, multiple subordinate clauses, and differing monetary notations. Respondents are told that the problem causes difficulty for ELLs. They must then select, from a series of options, the best way to reword the mathematics problem so as to make it more understandable to the students. Correct responses to this item may indicate knowledge of strategies for reducing linguistic complexity while retaining mathematical integrity (Truball & Solano-Flores, 2010).

KSTRAT Inter-item Correlations. The next section of the analysis of KSTRAT items seeks to identify inter-item correlations. Inter-item correlations may be helpful in explaining response patterns. Table 12 below gives the correlation matrix (Pearson's *r*) for the eleven KSTRAT items.

Table 12

Inter-Item Correlations for KSTRAT Items

	Q5	Q7	Q8	Q10	Q11	Q18	Q19	Q24	Q26	Q29
Q5	1	.319*	.312*	.288	.052	.319*	.131	.093	.033	.039
Q7	.319*	1	.312*	.137	.052	.092	.382*	-.070	.033	.186
Q8	.312*	.312*	1	.093	.079	.140	.200	.265	-.100	-.053
Q10	.288	.137	.093	1	.313*	.137	.087	.062	.219	.075
Q11	.052	.052	.079	.313*	1	.052	.009	.017	.121	.329*
Q18	.319*	.092	.140	.137	.052	1	.131	.093	.230	.039
Q19	.131	.382*	.200	.087	.009	.131	1	.333*	.073	.132
Q24	.093	-.070	.265	.062	.017	.093	.333*	1	.283	.151
Q26	.033	.033	-.100	.219	.121	.230	.073	.283	1	.150
Q29	.039	.186	-.053	.075	.329*	.039	.132	.151	.150	1

*. Correlation is significant at the 0.05 level (2-tailed).

This correlation matrix demonstrates that many items were significantly correlated with others at the $p \leq .05$ level, as would be expected among conceptually related items. Furthermore, no negative correlations were statistically significant. Item Q7 was correlated with Q8 and Q19. All of these items concern knowledge of strategies for creating language rich mathematics classroom; they are conceptually related. Q5, which concerns knowledge of division algorithms used by students, was also correlated, surprisingly, with Q7 and Q8. The correlation of Q5 with Q18 was expected, since both of these items are related to knowledge of division algorithms used by students. Items Q11 and Q10 were correlated. These items concern, respectively, the difficulty that false Spanish-English cognates (words that appear similar in Spanish and English but that carry different meanings) in a mathematics problem can cause for Latino ELLs and strategies for creating language-rich mathematics classrooms. Furthermore, item Q11 was also correlated with item Q29, which concerns knowledge of strategies for reducing the linguistic complexity of mathematics problems while retaining mathematical integrity. Q19 and Q24 were correlated. The former concerns a strategy for promoting mathematics discourse among ELLs and the latter concerns the potential difficulty for Latino ELLs that polysemy can cause. Item Q26 alone was not significantly correlated with any other item. This was surprising, since the item is related to knowledge of division algorithms used by students, as were items Q5 and Q18.

Teacher Factors Associated with the KSTRAT Items. The final section of the analysis of the KSTRAT scale was aimed at identifying the teacher factors that were associated with correct responses given to the items in this scale. As was mentioned in the similar section of the KDIFF/KCAP scale analysis, a battery of Chi-squared tests of independence (Pearson, 1900) were used to determine whether differences in numbers of correct responses seemed to be associated with teacher factors. A summary of the statistically significant results is presented in Table 13 below.

Table 13

Teacher Factors Associated with Difference in Percentages of Correct Responses to KSTRAT Items

Item	Taught math for more than 5 years?		Possess other certifications?		Taught classes with more than 40% ELLs		Response given to open question	
	NO	*YES*	*NO*	*YES*	*NO*	*YES*	*NO*	*YES*
Q5					96.3%	73.3%		
					$\chi^2 = 4.848$**			
Q7	71.4%	96.4%	96.0%	76.5%				
	$\chi^2 = 5.562$**		$\chi^2 = 3.680$*					
Q8	57.1%	85.7%						
	$\chi^2 = 4.200$**							
Q10								
Q11			16.0%	41.2%				
			$\chi^2 = 3.318$*					
Q18					96.3%	73.3%		
					$\chi^2 = 4.848$**			
Q19								
Q24			60.0%	88.2%				
			$\chi^2 = 3.953$**					
Q26								
Q29							36.4%	77.8%
							$\chi^2 = 4.896$**	

*. χ^2 statistic is significant at the 0.10 level.
**. χ^2 statistic is significant at the 0.05 level.

Table 13 illustrates the statistical association (or lack thereof, if no data given) of the percentages of correct responses on the KSTRAT items with teacher factors: having taught mathematics for more than 5 years, possessing multiple educational certifications, having taught mathematics classes containing more than 40% ELLs, and offering a response to the open question (*In your experience, is there anything else that effective teachers of Latino ELLs need to know?*). Percentages of correct responses to items Q7 and Q8, which concern knowledge of strategies for creating linguistically rich mathematics classrooms, were greater ($p \leq .05$) for teachers who had more than five years of teaching experience. Possession of multiple educational certifications was statistically associated with higher percentages of correct responses to items Q24 ($p \leq .05$) and Q11 ($p \leq .10$). These two items concern knowledge of difficulty that Latino ELLs may find in solving mathematics problems that involve words that can have confusing meanings to ELLs, either words with multiple meanings (polyemy) or false Spanish-English cognates. For Q7, possession of multiple educational certificates was also associated ($p \leq .10$) with a slightly lower percentage of correct. It is interesting to note that on both items Q5 and Q18, teachers that had experience teaching in mathematics classrooms containing more than 40% ELLs tended to answer incorrectly to these items. Both of these items concern valid division algorithms that may be used by Latino ELLs in mathematics classes. Finally, teachers who had given a response to the open question that asked about knowledge that teachers need for teaching mathematics to ELLs tended to answer item Q29 correctly. This item required that respondents select the most appropriate way to re-word a linguistically complex mathematics task so that ELLs could understand it better.

Summary of KSTRAT Item Analysis. Up to this point the analysis of the KSTRAT scale has presented the psychometric properties of the items, including the item information, the inter-correlations, and several teacher factors that were significantly associated with item responses. The KSTRAT scale concerns knowledge of strategic teaching decisions that may improve the opportunity for students to communicate about mathematics and of mathematical notations that are commonly taught in the United States or in Central America that can, because of the students' familiarity (or lack of familiarity) with these, facilitate mathematics communication or become a barrier to such for Latino ELLs. Brief descriptions of the items preceded these results of item analysis to aid in their interpretation. As a way of summarizing the results of the KSTRAT item analysis, Figure 15 presents a graphic representation of the items composing the scale. In this weighted-node and weighted-edge graph, the relative magnitude of the information provided by each items is indicated by the size of its node and the strength of the inter-item correlations are given adjacent to the edges connecting nodes ("*" denotes significance at the .05 level). Furthermore, teacher factors that were associated with differences in percentages of correct responses are represented by an appropriate letter placed adjacent to the node, as explained in the footer.

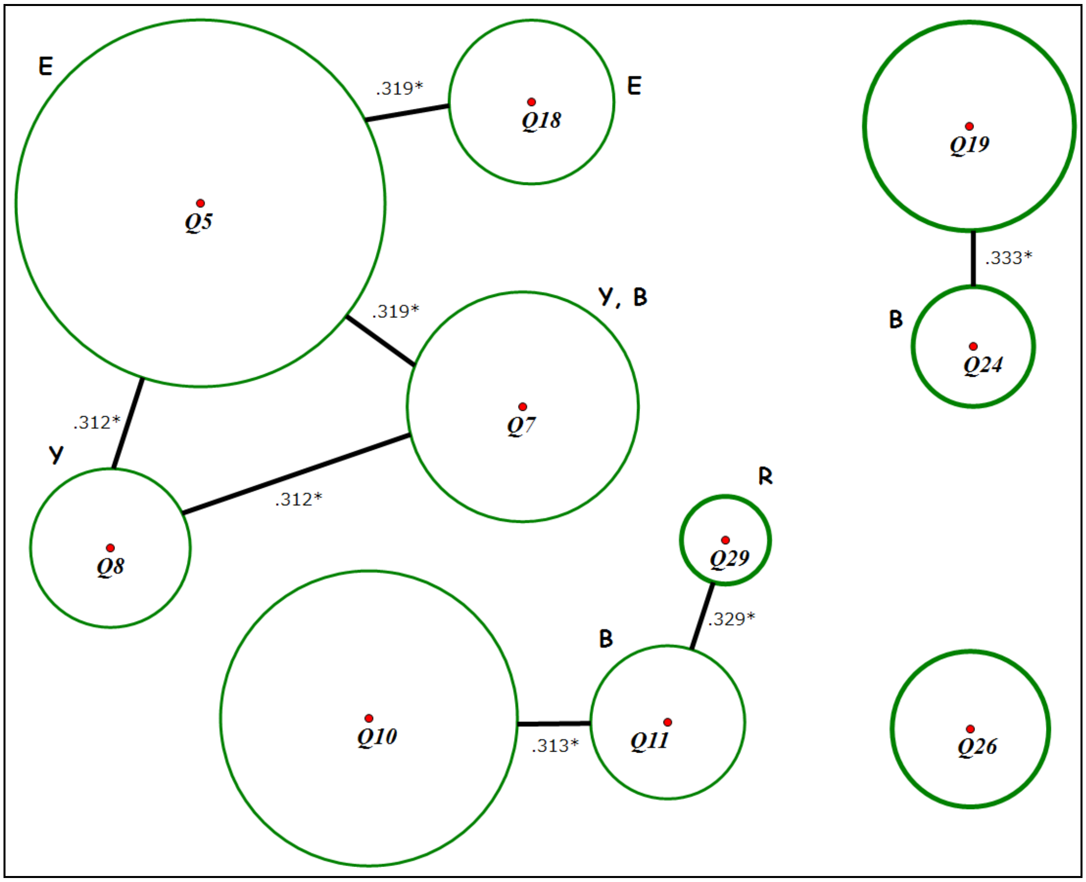


Figure 15. Semantic Map of Information, Inter-item Correlations, and Significant Teacher Factors for KSTRAT Items. The size of the nodes corresponds to the relative amount of information provided by the items. Furthermore the size of the connecting edges corresponds to the strength of the correlation; large edges indicate stronger correlation. Teacher factors associated with differences in numbers of correct responses to the items are indicated by the following abbreviations: *Y*-more than 5 years of mathematics teaching experience, *E*-experience teaching classrooms containing 40% or more ELLs, *B*-breadth of educational licensure (multiplicity of educational certifications), *R*-response given to the open question concerning teaching ELLs.

Figure 15 may offer a semantic map of the KSTRAT scale. In this graph, four clusterings of items can be identified. The most informative cluster, containing the items Q5, Q7, Q8, and Q18 are a highly correlated set, with the most informative item in the scale (Q5) occupying a place of prominence because its correlation to the other three items in the cluster. Items Q10, Q11, and Q29 appear to follow this cluster in terms of amount of information contributed to the total information and in terms of the strength of inter-item correlations. Items Q19 and Q24 were correlated and provided smaller amounts of combined information. Finally, item Q26, though not minimal in terms of information contributed to the scale, was uncorrelated to any other item in the scale. As with the similar graph given for the KDIFF/KCAP scale, these relationships were investigated in the attempt to better understand the patterns of correct responses to the items in this scale. The semantic map given in Figure 15, though incapable of explaining causation, is merely offered as an observation of item relationships and as possibly indicative of directions for future cognitive research related to the analyses which follow.

This subsection of the results has offered an explanation of the properties of the KSTRAT items and scale. It began with a summary of the psychometric properties of the items, offering comparisons of the item difficulty and discriminations parameters, and focusing on the information given by the items and the total scale. The item and total information obtained by this scale indicated that the scale performed acceptably in reducing the uncertainty with which persons' locations on the knowledge continuum could be estimated, for persons of a wide range of ability levels. However, the scale, which was comparably *easier* than the KDIFF/KSTRAT scale, functioned especially well for persons of slightly less than average ability on the continuum. It served as a poor measure of knowledge for persons of extremely high or low ability. The interpretation of this IRT statistic is analogous to reliability (internal consistency) central to CTT. The purpose of this part of the investigation was to respond to third research question of this study as it asks concerning the reliability of the KT-MELL measures.

This subsection also gave a thorough item analysis. It began with detailed descriptions of the items, followed by the observations of several inter-item correlations and of teacher factors that were associated with differences in numbers of correct responses. The findings concerning item information, inter-item correlations and significant teacher factors were presented in synthesized form in Figure 15, which may serve as a semantic map of the KSTRAT scale.

This concludes the analysis of the items that compose the KDIFF/KCAP scales. The final section of this chapter addresses evidences of the validity of the KT-MELL measures that were found.

Evidence of Construct Validity.

The final section of this chapter presents evidences that were found of the construct validity of the measurements of the KT-MELL instrument. The instrument central to this study was designed to measure teachers' knowledge for teaching mathematics to Latino English Language Learners (KT-MELL). *Content validation* of the instrument was addressed during the processes of instrument development, as reported in the methodology chapter (Chapter 3) of this study. To summarize the details of those processes here, the content validity of the KT-MELL instrument was established through three steps. First, in its conceptual design and item development, the conceptual framework (Figure 3), test-framework (Table 3) and items were developed in close alignment with theory that drew from related bodies of research and from more than 30 hours of classroom observations. Secondly, the entire instrument underwent review by a panel of four, widely-respected scholars of the mathematics education of ELLs and of the education of bilingual students. Finally, the entire instrument was submitted to a panel of practicing middle-school mathematics teachers who commented positively on the extent to which the items in the instrument captured the essential elements of KT-MELL. These three steps of validation provided the basis upon which the instrument was assumed to have an acceptable degree of content validity.

Three aspects of the *construct* validity of the KT-MELL measures are addressed this section: convergent and discriminant validity, and the nomological validity of each of the two separate scales. The first part of the section presents evidences of convergent and discriminant validity.

Convergent and Discriminant Validity. As was discussed in Chapter 3 concerning the methods used in this study, limited evidence of convergent and discriminant validity was available at the time of completion of this study, since KT-MELL was a novel construct for which no other measures existed. Hence, results presented below serve mainly to assess the construct (convergent and discriminant) validity of the separate scales that composed the KT-MELL instrument. Using a modified multitrait-multimethod (MTMM) matrix (Cambell & Fiske, 1959), wherein the KDIFF/KCAP and KSTRAT scales occupy the places of the different traits, the measurements of which were both obtained using the same (singular) method, the Pearson's *r* correlations of items within and across scales was considered (see Table 14 below).

Table 14 indicates that items within scales were more highly correlated (convergent) than were items across scales (divergent). Within the KDIFF/KCAP scale, six pairs of items were positively correlated at (at most) the .05 level. Furthermore, only six of the (insignificant) correlations were negative. Within the KSTRAT scale, eight pairs of items were positively correlated at (at most) the .05 level and only three of the (insignificant) correlations were negative.

Table 14

Modified MTMM Table Including Separate Scales as Traits

		KDIFF/KCAP Scale											KSTRAT Scale									
	Items	Q1	Q6	Q9	Q13	Q15	Q16	Q20	Q22	Q25	Q31	Q32	Q5	Q7	Q8	Q10	Q11	Q18	Q19	Q24	Q26	Q29
KDIFF/KSTRAT Scale	Q1	1																				
	Q6	.196	1																			
	Q9	.341*	.150	1																		
	Q13	.139	.059	.050	1																	
	Q15	.033	.108	.234	.019	1																
	Q16	.469**	.278	.314*	-.100	.234	1															
	Q20	.282	-.030	.282	.564**	.076	.152	1														
	Q22	.171	-.021	.029	.100	.078	.200	.108	1													
	Q25	.075	.166	.283	.106	.183	.000	.092	-.141	1												
	Q31	.023	.075	-.054	.156	.106	.108	.094	.054	.038	1											
	Q32	.050	.091	.043	.085	.315*	.426**	-.088	.213	.030	.220	1										
KSTRAT Scale	Q5	.056	-.039	-.033	.033	.346*	-.033	-.146	.230	.093	-.009	.091	1									
	Q7	-.238	.257	-.033	.033	.166	-.033	.004	.033	.093	.178	.091	.319*	1								
	Q8	-.027	-.059	-.050	-.213	.117	-.050	-.222	.050	.018	-.014	.027	.312*	.312*	1							
	Q10	-.159	-.075	.044	.137	.171	-.219	.052	.351*	-.047	-.131	-.136	.288	.137	.093	1						
	Q11	-.026	.106	-.121	.048	-.085	-.266	.060	.121	.017	-.013	-.083	.052	.052	.079	.313*	1					
	Q18	-.238	-.039	-.230	.033	-.013	-.230	-.146	-.164	.093	-.196	-.056	.319*	.092	.140	.137	.052	1				
	Q19	-.178	.194	-.073	-.200	.226	-.073	-.063	-.145	.154	-.049	.015	.131	.382*	.200	.087	.009	.131	1			
	Q24	-.136	-.045	.141	-.018	.183	.000	.199	.141	.300	-.096	.136	.093	-.070	.265	.062	.017	.093	.333*	1		
	Q26	-.085	-.150	.200	.250	-.078	-.143	.108	-.029	.141	-.108	.213	.033	.033	-.100	.219	.121	.230	.073	.283	1	
	Q29	-.005	.057	-.021	.053	-.225	-.278	.030	.021	.257	-.075	-.187	.039	.186	-.053	.075	.329*	.039	.132	.151	.150	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

In comparison to the larger number of positive inter-item correlations within the KDIFF/KCAP and KSTRAT scales (shaded), only two pairs of items were positively correlated across scales. Furthermore, more than half of the insignificant inter-item correlations across scales were negative.

The relative frequency of significant positive inter-item correlations within the separate scales may be taken as limited evidence of the convergent validity of these scales, whereas the relative infrequency of significant positive inter-item correlations across scales may be taken as limited evidence of the discriminant validity of the two scales. A further evidence of discriminant validity is found in the Pearson product-moment correlation of the total scale scores between scales, $r = -.005$. That the scale scores between the KDIFF/KCAP and KSTRAT scales were so uncorrelated gives some indication that they were not measuring the same constructs.

Nomological Validity. As was discussed in Chapter 3 concerning the methodology, a second aspect of construct validity that was addressed in this study concerned the extent to which the KT-MELL measures were related to a nomological network of theoretically related constructs (Cronbach & Meehl, 1955). KT-MELL is conceptually related to *pedagogical content knowledge*, PCK, (Shulman, 1986) and to *mathematical knowledge for teaching,* MKT (Hill, Schilling, & Ball, 2004). (See Figure 3 of Chapter 3 for a conceptual model.) The theoretical underpinnings of both of these constructs imply that, while teachers gain knowledge through formal educational experience, the experience of teaching students plays an integral part in the formation of their knowledge.

Hence, for this study, it was hypothesized that the formation of KT-MELL in the minds of teachers should take a similar path as for other forms of PCK. That is, formal training and education should play a formative role and actual mathematics teaching experience should play a perhaps more important role. Although this research cannot assess the extent to which these types of experiences play the roles of forming, growing or of transforming KT-MELL in the minds of teachers, statistical associations between these types of experiences and the measurement of KT-MELL may be informative as construct validity (nomological) evidences related to the instrument under study.

This analysis of such associations employed two methods. As an investigative step, one-way analyses of variance (ANOVA) were computed to test the null hypothesis of equivalence of scale means between all groups of teachers that could be defined according to teacher factors that were assessed. (Appendix N gives a complete table containing the F-statistic and associated *p*-values for the comparisons of means on both scales against all possible teacher factors.) To obtain a more descriptive sense of how teacher factors were related to scale scores, linear regression (teacher factors regressed on teachers' scale scores) was used. To determine which factors seemed to be most related to the scale score a series of linear regression models were estimated. To begin, all of the possible factors were included in the model. Following this, factors that did not add to the total variance explained by the model or that were statistically insignificant, were eliminated until an optimal model was obtained.

The results of these analyses are presented in two sections according to the two scales of KT-MELL that were identified: KDIFF/KCAP and KSTRAT. Each section begins with descriptive statistics concerning the scale scores and is followed by presentation of the full and reduced regression models.

Nomological Validity of the KDIFF/KCAP Scale. As a context for the analysis of KDIFF/KCAP scale scores and associated teacher factors, the distribution of scale scores is first presented.

Distribution of KDIFF/KCAP Scale Scores. Descriptive statistics concerning the KDIFF/KCAP scale scores are given in Table 15 below. The mean (as well as minimum and maximum) scores represent the percentage of items that received correct responses out of the eleven items that composed this scale.

Table 15

Descriptive Statistics for KDIFF/KCAP Scale Scores

N	Min	Max	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
							Std. Error		Std. Error
42	.09	.91	.4589	.2027	.041	.381	.365	-.683	.717

As Table 15 demonstrates, on average teachers correctly responded to fewer than 50% of the items in this scale. This figure is related to the difficulty of the items as seen in the ICCs for these items (Figure 10). Several of the items in this scale had difficulties indicating that they would only receive correct responses from respondents of higher ability. Based upon the variance, as well as the measures of skewness and kurtosis, the scale scores may be said to have been approximately normally distributed. This observation is confirmed by Figure 16 below.

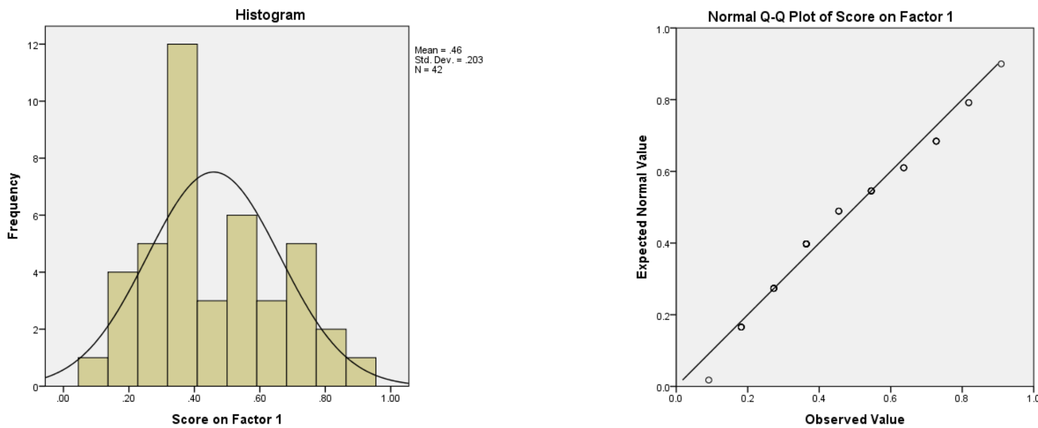


Figure 16. Distribution and Normal Q-Q Plot of KDIFF/KCAPP Scale Scores.

Teacher Factors Associated with KDIFF/KCAP Scale Score. Based upon the satisfaction of the assumption of the normality of the distribution of the response variables (scale scores), as seen in Figure 16 and with a view to identifying the teachers factors associated with differences in scores, teacher factors were regressed linearly on the KDIFF/KCAP scale scores. The tables that follow exhibit both the full and the reduced linear regression models for the KDIFF/KCAP scale.

The linear regression model given in Table 16 indicates that several variables thought to be associated with KT-MELL were in fact statistically associated with greater KDIFF/KCAP scale scores. With their associated percent increases, these were: number of years teaching mathematics (11.6%), having taught mathematics classes containing 40% or more ELLs (10.7%), and breadth of educational licensure--indicated by having multiple educational certifications—(9.6%). Furthermore, teachers who had offered a response to the open survey question (*In your experience, is there anything else that effective teachers of Latino ELLs need to know?*), tended to score approximately 14% higher on this scale than others. It is reasonable to conjecture that teachers who gave responses to this question did so because of strong convictions founded in experiences of teaching mathematics to ELLs. If this conjecture holds, then the association of this variable with higher KDIFF/KCAP scores, in so far it implies a greater measure of experience teaching ELLs, may also constitute construct (nomological) validity. Another finding, seemingly unrelated to construct validity, was that gender was statistically associated with differences in scale scores; controlling for other factors in the model, men tended to perform 24.3% better than women.

Several variables that seemed to potentially be related to KT-MELL were found to be either unrelated or statistically associated with lower scale scores. For example, the experience of having participated in professional development related to teaching ELLs did not predict differences in scale scores. Other insignificant variables included the teachers' proficiency in other languages (Spanish or other) and the duration of time (more than one day) taken to complete the survey. Furthermore, two variables potentially related to greater knowledge in this domain, were in fact statistically associated with lower scale scores. These were whether or not the teacher held a degree in mathematics (9.1% reduction) and whether or not the teacher was certified to teach English as a second language (5.9% reduction). This finding, taken along with findings of the above variables that were statistically associated greater scale scores, seemed to imply that knowledge in the KDIFF/KCAP domain could be more related to actual teaching experience rather than to experiences of formal education. The full model is presented below.

Table 16

Full Linear Regression Model for the KDIFF/KCAP Scale

Coefficients							
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	.025	.104		.240	.812	-.188	.238
have you taught math for more than 5 years?	.116	.066	.272	1.740	**.092**	-.020	.252
taught more than 3 different courses?	.028	.090	.057	.312	.758	-.156	.211
taught a class with more than 40% ELLs?	.107	.060	.257	1.789	**.084**	-.015	.230
have a degree in math?	-.120	.069	-.278	-1.750	**.091**	-.261	.020
certified to teach ESL?	-.174	.088	-.281	-1.966	**.059**	-.354	.007
possess other certifications?	.096	.054	.234	1.758	**.089**	-.016	.207
had ELL professional development?	-.007	.057	-.018	-.124	.902	-.124	.110
how well do you know Spanish? (low or high)	-.026	.054	-.062	-.484	.632	-.135	.084
do you speak a language other than Eng/Span?	-.001	.076	-.002	-.016	.987	-.157	.155
anything else effective teachers of Latino ELLs need to know?	.139	.076	.284	1.829	**.078**	-.016	.294
more than one day to complete survey?	.038	.075	.081	.514	.611	-.114	.191
are you female or male	.243	.068	.533	3.566	**.001**	.104	.382

This model had acceptable measures of goodness of fit; the *R* and *R*-squared statistics were 0.762 and 0.581, respectively. The *R*-squared statistic indicated that more than 58% of the variation in scale scores was explained by the variation in the variables included in the model. Since a number of the variables in this model were insignificant or added little to the variance explained, the reduced model in Table 17 was found to synthesize the most important variables without much loss of variance explained.

Table 17

Reduced Linear Regression Model for the KDIFF/KCAP Scale

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	.025	.087		.292	.772	-.152	.203
have you taught math for more than 5 years?	.106	.051	.249	2.072	**.046**	.002	.210
taught a class with more than 40% ELLs?	.094	.050	.225	1.873	**.070**	-.008	.196
have a degree in math?	-.117	.053	-.269	-2.195	**.035**	-.225	-.009
certified to teach ESL?	-.164	.075	-.265	-2.197	**.035**	-.316	-.012
possess other certifications?	.093	.048	.227	1.942	**.060**	-.004	.189
anything else effective teachers of Latino ELLs need to know?	.153	.058	.314	2.635	**.013**	.035	.272
are you female or male	.249	.057	.547	4.336	**.000**	.132	.366

This model, which had R and R-squared statistics 0.756 and 0.571, respectively, indicates an improvement in the statistical significance of all variables. Variables that persisted in their association with higher scale scores were mathematics teaching experience (10.6% increase), teaching larger groups of ELLs (9.4% increase), breadth of educational licensure (9.3% increase), offering a response to the open-ended item (15.3%

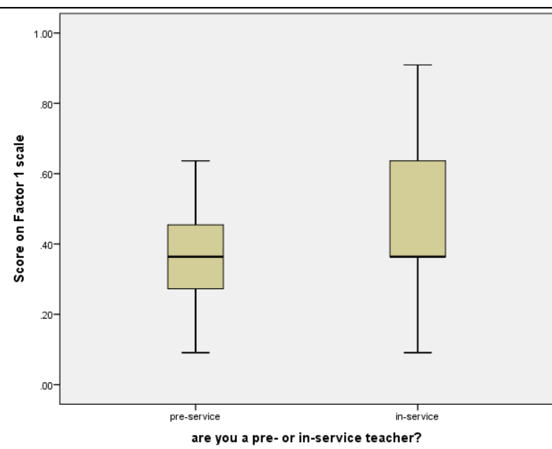
increase), and gender (24.9% increase). Possessing a degree in mathematics and being certified to teach ESL were still negatively correlated with scale scores (11.7% and 16.4% reduction in scale score, respectively).

Based upon the hypothesis that knowledge in the KDIFF/KCAP domain, as a subset of PCK, should be, at least in part, the result of actual teaching experience, a comparison of scale scores of the 42 practicing teachers with those of the 146 pre-service teachers that composed the second pilot-study was performed using ANOVA. The boxplot and ANOVA table, given in Table 18 below, indicate that practicing middle school teachers scored an average of 4.5 percentage points higher (eta-squared) than did pre-service mathematics teachers. This difference was significant at the .01 level.

Table 18

Comparison of Practicing Teachers' and Pre-Service Teachers' KDIFF/KCAP Scale Scores

KDIFF/KCAP Scale Score



	SS	df	MS	F	Sig.
Between Groups	.204	1	.204	8.94	.003
Within Groups	4.24	186	.023		
Total	4.45	187			

Effect size

$\eta^2 = 0.045843$
$\omega^2 = 0.040465$

Summary of KDIFF/KCAP Nomological Validity Evidence. The analysis given in this section compared a number of teacher factors potentially related to knowledge in the KDIFF/KCAP domain with scale scores in the attempt to ascertain whether evidences of construct (nomological) validity existed. This analysis uncovered convincing evidence that knowledge in the KDIFF/KCAP domain seems to be related to actual mathematics classroom teaching experience. Specific teacher factors that were associated with increases in scale scores were years of mathematics teaching experience, experience teaching greater percentages (40% or more) of ELLs, and breadth of teaching experience. Additionally, teachers that responded to the open-ended item that requested their input concerning knowledge for teaching mathematics to ELLs also scored higher, on average. Two factors that seemed to potentially be related to KT-MELL—possession of a mathematics degree and certification to teach English as a second language—were in fact unassociated. Since both of these credentials—a mathematics degree and the certification to teach ESL—are generally the results of formal study rather than of teaching experience, this result was taken to imply that knowledge in the KDIFF/KCAP domain may be a product of actual teaching experience more than of education or training. In so far as experience is thought to be related to KT-MELL, this interpretation would constitute another evidence of construct (nomological) validity. A final compelling evidence of construct (nomological) validity that was found concerned the comparative performance of actual middle school mathematics teachers and pre-service teachers. Actual teachers tended to score higher than did pre-service teachers, which agreed with other findings concerning the association of classroom teaching experience with higher levels of KDIFF/KCAP knowledge.

Nomological Validity of the KSTRAT Scale. The analysis now turns to consideration of evidences of nomological validity related to the KSTRAT scale. The distribution of scale scores is first presented.

Distribution of KSTRAT Scale Scores. Descriptive statistics concerning the KSTRAT scale scores are given in Table 19 below. The mean (as well as minimum and maximum) scores represent the percentage of items that received correct responses out of the ten items that composed this scale.

Table 19

Descriptive Statistics for KSTRAT Scale Scores

N	Min	Max	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
							Std. Error		Std. Error
42	.00	1.00	.6952	.1899	.036	-1.184	.365	2.890	.717

As Table 19 demonstrates, on average teachers correctly responded to approximately 70% of the items in this scale. This figure may be explained by the difficulty (i.e., easiness) of the items as seen in the ICCs (Figure 13). Several of the items in this scale had difficulties indicating that they could receive correct responses from respondents of less than average ability. Although the variance of KSTRAT scale scores was small, the measures of skewness and kurtosis indicate that data were skewed slightly left. However, the normal Q-Q plot gives evidence that this effect was not extreme and that data were still approximately normally distributed. Figure 17 below gives both the distribution of scale scores and the normal Q-Q plot.

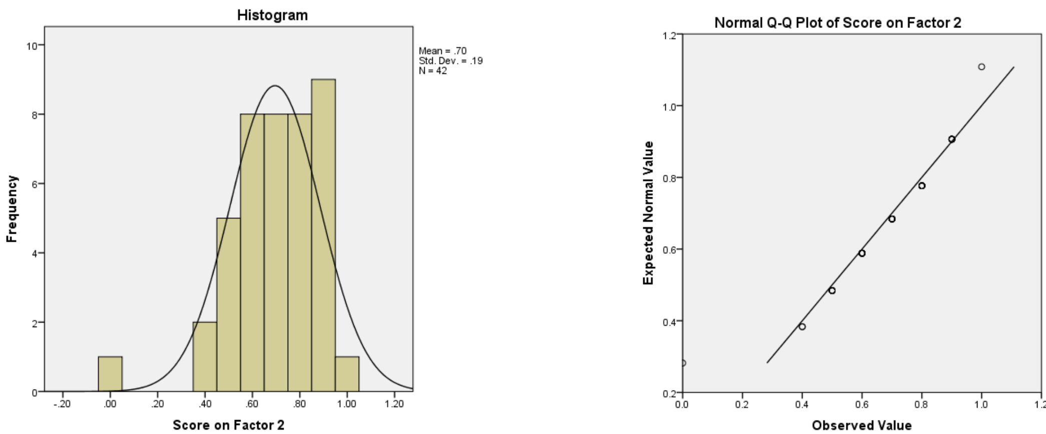


Figure 17. Distribution and Normal Q-Q Plot of KSTRAT Scale Scores.

Teacher Factors Associated with the KSTRAT Scale Scores. Based upon the results found in Figure 17 above, approximate normality of the distribution of KSTRAT scale scores was assumed with a view toward regressing teacher factors on the scale scores. The tables that follow exhibit both the full and the reduced linear regression models for the KSTRAT scale.

The linear regression model given in Table 20 below indicates that two variables thought to be potentially associated with KT-MELL were in fact statistically associated with greater KSTRAT scale scores. With their associated percent increases, these were: having 5 or more years of mathematics teaching experience (13.6%) and breadth of educational licensure (12.2%). A variable that could have been related to greater KSTRAT scores but that actually behaved counter intuitively was whether or not teachers had experienced professional development related to ELL issues. Such experience was actually associated with a 12.9% reduction in KSTRAT scale score. One interpretation of this finding, and that seems to agree with a similar result concerning knowledge in the KDIFF/KCAP scale, is that knowledge in the KSTRAT domain is more a product of classroom teaching experience than of formal training. All other factors were not significantly associated with differences in scale scores.

Table 20

Full Linear Regression Model for the KSTRAT Scale

Coefficients							
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	.666	.124		5.384	**.000**	.413	.919
have you taught math for more than 5 years?	.136	.079	.341	1.717	**.097**	-.026	.297
taught more than 3 different courses?	.038	.107	.083	.356	.725	-.180	.256
taught a class with more than 40% ELLs?	-.012	.071	-.031	-.169	.867	-.158	.134
have a degree in math?	-.024	.082	-.059	-.295	.770	-.191	.143
certified to teach ESL?	-.022	.105	-.037	-.205	.839	-.236	.193
possess other certifications?	.122	.065	.318	1.881	**.070**	-.011	.254
had ELL professional development?	-.129	.068	-.340	-1.897	**.068**	-.268	.010
how well do you know Spanish? (low or high)	-.008	.064	-.021	-.131	.897	-.138	.122
do you speak a language other than Eng/Span?	-.098	.091	-.206	-1.087	.286	-.284	.087
anything else effective teachers of Latino ELLs need to know?	.073	.090	.161	.815	.422	-.111	.258
more than one day to complete survey?	.029	.089	.065	.322	.750	-.153	.210
are you female or male	-.025	.081	-.058	-.303	.764	-.190	.141

This model was limited in its measure of goodness of fit; the *R* and *R*-squared statistics were 0.570 and 0.325, respectively. The *R*-squared statistic indicated that slightly more than 32% of the variation in scale scores was explained by the variation in the variables included in the model. A model that includes only the variable that have statistical significance in relation to the scale score is given in Table 21 below.

Table 21

Reduced Linear Regression Model for the KSTRAT Scale

Coefficients							
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	.628	.056		11.275	.000	.516	.741
have you taught math for more than 5 years?	.142	.059	.358	2.410	.021	.023	.262
possess other certifications?	.130	.054	.339	2.398	.022	.020	.239
had ELL professional development?	-.141	.056	-.372	-2.528	.016	-.254	-.028

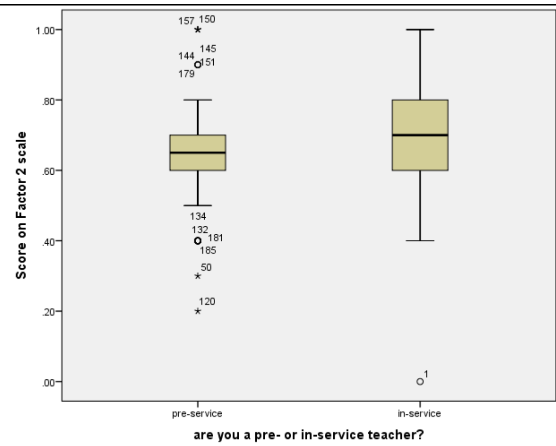
This model, which had *R* and *R*-squared statistics 0.509 and 0.259, respectively, indicates that years of teaching experience and breadth of educational licensure retained their statistically significant association with greater scale scores (14.2% and 13.0%, respectively) and that having had ELL professional development persisted in its association with (14.1%) reduced scale scores.

As with knowledge in the KDIFF/KCAP domain, it was hypothesized that knowledge in the KSTRAT domain should be, at least in part, the result of actual teaching experience. To evaluate the association of these kinds of knowledge and experience, a comparison of KSTRAT scale scores of the 42 practicing teachers with those of the 146 pre-service teachers that composed the second pilot-study was performed using ANOVA. The boxplot and ANOVA table, given in Table 22 below, indicate that practicing middle school teachers scored an average of 1.5 percentage points higher (eta-squared) than did pre-service mathematics teachers. This difference was significant at the .1 level.

Table 22

Comparison of Practicing Teachers' and Pre-Service Teachers' KSTRAT Scale Scores

KSTRAT Scale Score



	SS	df	MS	F	Sig.
Between Groups	.067	1	.067	2.87	.092
Within Groups	4.32	186	.023		
Total	4.39	187			

Effect size

$\eta^2 = 0.015262$
$\omega^2 = 0.009971$

Summary of KSTRAT Nomological Validity Evidence. This section of the results has collected evidence of the construct (nomological) validity of the KSTRAT scale. It has done so by observing that certain variables thought to be related to knowledge in the KSTRAT domain were in fact statistically associated with higher scale scores. These were years of mathematics teaching experience and breadth of educational licensure. This scale was not associated with measures of teaching experience specifically related to teaching mathematics to ELLs. Indeed, it is possible that the strategic knowledge assessed by this measure is general knowledge that experienced mathematics teachers would possess, and that it does not apply strictly to ELLs. Another finding of this analysis was that professional development focused on ELLs did not seem to improve teachers' knowledge in the KSTRAT domain. On the contrary, having such PD was negatively correlated with scale scores. This seemed to imply the superiority of actual teaching experience over formal training or perhaps the inferiority of current professional development models related to ELLs. The former interpretation could be taken as a further evidence of construct (nomological) validity of the KSTRAT scale. Another finding that lent slight support to the suggestion that KSTRAT is knowledge that depends on teaching experience was the finding that the sample of practicing middle school mathematics teacher performed approximately 1.5% percentage points better than did the pre-service teachers.

Summary. This chapter has presented results that respond to the elements of the third research question of this study. That question concerns the reliability and validity of the measurements obtained using the KT-MELL instrument as well as other observable response patterns. Data analyzed in this chapter were collected from the administration of the KT-MELL survey to a sample of 42 practicing middle school mathematics teachers from wide geographic sectors of Texas. Although this sample of teachers was not obtained using the intended method of stratified random sampling, nor was the size as large as had been hoped, yet the sample bore resemblance to the population of Texas middle school mathematics teachers in several ways.

The beginning of this chapter addressed the matter of reliability from the two theoretical perspectives of CTT and IRT. Investigation of the internal consistency of the instrument based upon Cronbach's (1951) alpha and upon a two-parameter IRT model led to the conclusion that the full set of items failed to meet the assumption of unidimensionality, as the theoretical factor structure of the instrument would indeed imply. Based upon the theoretical orientation of items, and upon evidence obtained from the measure of internal consistency and, in a limited way, from CFA, two scales were found that exhibited greater internal consistency and conceptual unity. These scales KDIFF/KCAP and KSTRAT, respectively, are concerned with knowledge of both difficulties that Latino ELLs can face in learning mathematics and also of the capacities for learning mathematics that the students bring with them, and of strategies for teaching ELLs mathematics. Based on the IRT information exhibited by each of these scales, it was observed that the scales served to an acceptable degree in estimating person locations on the continuums of knowledge represented by the scales. This IRT finding, along with the improved CTT measure of internal consistency, was taken as evidence that the scales had a reasonable degree of reliability, given the novelty of the knowledge constructs and the exploratory nature of the study.

A number of important associations between item responses and between teacher factors and item responses were observed. Semantic mappings of response patterns were offered for each of the scales.

The final section disclosed evidences of construct (convergent, discriminant, and nomological) validity that were found in connection to the KT-MELL measures. Positive correlations within scales were more numerous than across scales, which was taken as limited evidence both of the convergent and discriminant validity of the separate KDIFF/KCAP and KSTRAT scales. This section also discussed a number of important associations between teacher factors and scale scores that were observed. Several factors, such as years of experience, experience teaching classrooms with 40% or more ELLs, and breadth of educational licensure were seen to be associated with gains in scale scores on one or both scales. Some other factors, typically related to formal learning about teaching, such as having a mathematics degree or participating in professional development, were seen to be negatively correlated with scale scores. All of these findings, together with the observation that practicing teachers performed better than did pre-service teachers (second pilot-study sample) on both scales, seemed to indicate that the knowledge measured by the KT-MELL scales was experiential knowledge. This finding was taken as construct (nomological) validity of the measures obtained by the KT-MELL instrument.

CHAPTER 5: CONCLUSIONS AND IMPLICATIONS

Motivated by the rapidly growing numbers of Latino English Language Learners in mathematics classrooms in the United States and by evidence that mathematics teachers may not fully understand how to serve this body of mathematics learners, this study has attempted to define the domains of knowledge needed for teaching these students and to develop an instrument capable of yielding valid and reliable measurements of this knowledge. Both the work of identifying the domains and of developing items to serve as indicators of the knowledge were grounded in educational research concerned with the mathematics learning of ELLs and built upon research concerned with well-defined domains of knowledge for teaching mathematics. Furthermore, this work was also informed throughout the conceptualization and instrument development phases by observations of the mathematics classrooms of teachers who taught this population of students. To conclude this study and to elucidate the place of this research in the broader context of mathematics education research, this chapter will briefly review what the intended goals of the study were and what outcomes were obtained. Following this, a discussion of the limitations of the study precedes and helps to understand the implications of the research for mathematics teacher educators and educational policy-makers. The final section offers suggestions for further research.

Goals and Results of this Study

The central objectives of this study, motivated by the three research questions that guided the inquiry, were two-fold: identification of the domains of knowledge for teaching mathematics to ELLs (KT-MELL) and development and field-testing of survey items capable of obtaining valid and reliable measurements of KT-MELL. Based upon careful analysis of the research and practitioner literature, this study identified three domains of knowledge potentially needed by mathematics teachers of Latino ELLs. These are: knowledge of common difficulties experienced by Latino ELLs in mathematics classes (KDIFF), knowledge of ways in which Latino ELLs' background knowledge and experiences (including bilingualism) can capacitate them to learn mathematics (KCAP), and knowledge of strategies for instructing Latino ELLs in mathematics (KSTRAT). These three domains were seen as theoretically situated within Hill, Ball, and Schilling's (2008) model of mathematical knowledge for teaching, MKT (see Figure 3). More precisely within that model, the domains of KT-MELL were seen as intersecting both knowledge of content and students and knowledge of content and teaching.

Using the three proposed domains of KT-MELL as a test framework, along with aspects identified in the literature as well as observed in classrooms, survey items were developed. In addition to their strict alignment with theory and actual classroom practice, evidence of content validity of the items was obtained by submission of the complete set of items to a panel of experts in the fields of mathematics education and bilingual education, all of whom had studied educational issues related to Latino ELLs extensively. Their responses indicated strong agreement that the items would serve as indicators of the intended domains of knowledge, as proposed by the test framework; the consensus of the experts was in agreement with the theoretical assignment of items as indicators of their intended constructs on more than 90% of the items. The instrument was also submitted to a focus group of eight practicing teachers whose comments indicated agreement that the items served as a measure of KT-MELL.

After revision, expert review and pilot-testing of thirty-two items among 146 pre-service teachers, the same thirty-two items were retained for the full study, including ten items intended to capture KDIFF, twelve items intended to capture KCAP, and another ten items intended to capture KSTRAT. The instrument was then administered via internet survey to a sample of 42 practicing, predominantly middle school, mathematics teachers from wide geographic sectors of Texas. This sample may be considered semi-representative, as it bore resemblance to the general population of middle school mathematics teachers in Texas.

Once domains and aspects of the knowledge construct in question had been identified—responding to the first two questions of this study, respectively—and the instrument developed and administered, it was possible to answer the third research question of this study. This question concerned the extent to which the instrument provided a reliable and valid measure of KT-MELL. Analysis of the reliability of the measure employed methods pertaining to both classical test theory (CTT) and item response theory (IRT).

The test framework used to guide creation of the instrument implied that the construct in question, KT-MELL, was probably a multidimensional knowledge construct. Therefore, it was not surprising that the measure of internal consistency, Cronbach's (1951) alpha, of the full 32-item instrument indicated that less than 50% of the variation of item responses was accounted for by covariation among items. Since this alpha assumes unidimensionality of the latent variable and multidimensionality was implied by the theoretical test framework, the low measure of internal consistency was taken as potentially indicative of multiple dimensions. This conclusion was confirmed by the result of IRT analysis of the 32 items. Under the assumption of unidimensionality of knowledge construct, a two-parameter IRT model that included all 32 items resulted in an unstable solution in which a large number (about half) of items had negative item discriminations. Such discrimination coefficients seemed to imply that higher levels of knowledge were negatively correlated with selecting the correct survey response. On more careful investigation, it seemed that different types of knowledge were at play in this model, which stood in direct contradiction of the unidimensionality assumption of this IRT model.

Based upon these evidences of low internal consistency under the unidimensionality assumption and upon the multidimensional nature of the theoretical test framework, the items were separated into two distinct scales using their conceptual orientations and their IRT discrimination coefficients. The two scales included, respectively, items intended to measure knowledge of both difficulties and capacities for learning mathematics experienced by Latino ELLs (KDIFF/KCAP) and knowledge of strategies for teaching these students mathematics (KSTRAT). Items which were poorly correlated with either scale, a total of eleven items, were omitted entirely from the analysis. The resulting scales agreed very well with confirmatory factor analysis (CFA) of the items: although CFA typically requires sample sizes of at least 200 (DeVellis, 2003), that the assignment of items to factors given by the factor loadings using the sample of 42 teachers only disagreed by a single item with the conceptual and IRT assignment of items to factors was taken as evidence of the suitability of the composition of the scales.

An important finding of this analysis pertains to the theoretical structure of KT-MELL. It was observed that KDIFF and KCAP seemed to pertain to a single latent variable, rather than to two, as hypothesized; many items in these separate domains were in fact strongly correlated with one another. Hence, whereas theory seemed to indicate the existence of three separate domains of KT-MELL, the data analysis gave evidence that this theoretical model could be replaced with a bi-dimensional model, composite of two distinct scales.

Two-parameter IRT models were then computed for each of the separate scales, the scale concerned with both ELLs' difficulties and capacities for learning mathematics and with strategies for teaching mathematics to ELLs. These models gave estimations of the difficulty and discrimination parameters of each of the items in the scales. For each scale, all of the items possessed positive discrimination coefficients, indicating that the probability of a correct response on all items was positively correlated with level of ability, as theory would suggest. (This finding was taken as further confirmation that the "right" underlying factor structure had been identified.) These models also gave evidence, in the form of IRT information, of the capacity of the scales to estimate respondents' levels of knowledge. The information for the items varied widely, with some items providing considerably more information than others. The total information estimates for each of the scales (27.44 and 11.64 respectively for the KDIFF/KCAP and KSTRAT domains) indicated that they both served well in reducing the uncertainty of the estimation of respondents' levels of knowledge, *within specific ranges of ability of the respondents*. These information estimates along with Cronbach's alpha for each scale (α

= .6214 and α = .6056, respectively) indicated that the scales had an acceptable level of reliability, given the exploratory nature of this study.

Once reliability of the scales had been established, evidence of construct validity was sought. Convergent, discriminant and nomological validity were all addressed. Limited results were obtained to address convergent and discriminant validity, because of the exploratory nature of developing an instrument for a novel construct. Nevertheless, correlations within the scales were more numerous than across the scales, which gave some evidence that the KDIFF/KCAP and KSTRAT constructs were being measured separately by the separate scales.

Concerning nomological validity, separate models, with teacher factors regressed linearly on the scale score, for each of the two scales gave evidence that scores were associated with a number of important factors. For the scale comprising knowledge of difficulties experienced by Latino ELLs and of capacities for learning mathematics possessed by these students (KDIFF/KCAP), several measures of teachers' experience were significantly associated with higher scores. These factors were years of teaching experience, experience teaching classes containing 40% or more ELLs, breadth of educational experience (indicated by multiplicity of educational licenses), and responsiveness to ELL issues (indicated by whether or not teachers gave a response to the open question concerning knowledge needed for teaching ELLs). The association of years of experience and experience teaching ELLs with higher scores for this scale seemed to indicate that this domain of KT-MELL is related to the extent of experience that mathematics teachers have in teaching ELLs. Assuming that teachers' knowledge for teaching mathematics is, in large part, a product of their teaching experience, this finding may be interpreted as an indication of the (nomological) validity of the scale for measuring the KDIFF/KCAP domain of KT-MELL. If multiplicity of educational licenses is an indication of breadth of educational experience, then the association of this variable with higher scale scores also seems to give evidence of the (nomological) validity of the measure. Furthermore, although merely having an opinion about knowledge needed for teaching mathematics to ELLs would not seem to be an obvious indicator of higher levels of knowledge, it is conceivable, and can be conjectured, that teachers who gave responses to the open question did so because of having extensive experience teaching ELLs which may have resulted in the possession of strong opinions about the issue. In this case, the association of responses to the open question with higher scale scores may also be taken as construct (nomological) validity. Finally, that a number of variables indicative of types of formal training experienced by teachers—having a degree in mathematics, having had ELL-related professional development, or being certified to teach English as a second language—were either unassociated with scores or even negatively associated with scores seemed to give further evidence that this domain of KT-MELL may be primarily related to teachers' experience of teaching rather than to formal training.

Factors associated with higher scores on the scale comprising knowledge of strategies for teaching mathematics to Latino ELLs (KSTRAT) included years of teaching experience and breadth of educational licensure. As with the association of these factors with higher scores on the KDIFF/KCAP scale, this association may also be taken as evidence of the (nomological) validity of the scale as a measurement of the KSTRAT domain. That is, an association of greater experience with higher scale scores would be expected on this domain. Another factor that was negatively associated with the scale score on this domain was having had ELL-related professional development. There was evidence that, controlling for years of experience and breadth of educational licensure, teachers who had had such professional development tended to score lower than those who had not. It appears that actual experience of teaching mathematics is associated with higher scores on this scale while formal training is associated with lower scores. This may be evidence that the KSTRAT scale serves as a valid measure of a kind of pedagogical content knowledge gained through actual teaching experience rather than formal training.

To further understand teachers' responses, three-step analyses of the items comprising the two scales were conducted that included: comparison of the relative levels of information given by items, analysis of inter-item correlations, and also chi-square tests of independence for responses to the items from the different groups based upon the teacher qualities found to be significantly associated with scale scores. The results of these analyses indicated which items served best in reducing the uncertainty of teachers' levels of knowledge. Furthermore, some indication of *why* teachers gave correct responses to items may have been gained.

On the KDIFF/KCAP scale, correlations observed between KDIFF and KCAP items seem to indicate that teachers who are aware of the difficulties that ELLs may have with symbolic and linguistic mathematical expressions are more likely to have the capacity to discern when ELLs have used alternative, and valid, ways of expressing mathematics. That is, it seems that familiarity with the difficulties faced by Latino ELLs in mathematics may be related to being able to see the capacities that these students may have for expressing mathematics effectively using their own language, notation and gestures. Also on this scale, experience teaching percentages of ELLs greater than 40% was associated with higher percentages of correct responses on three types of items, KDIFF, KCAP, and KSTRAT. This finding seemed to indicate that perceiving Latino ELLs' difficulties and their capacities in mathematics, as well as making appropriate instructional decisions for their benefit, may be a product of the experience of working with ELLs. Furthermore, several KDIFF items on this scale received more correct responses from teachers who had responded to the open question regarding what mathematics teachers need to know in order to be effective in teaching Latino ELLs. This finding may indicate that teachers who hold strong opinions about this topic tend to be more conscious of difficulties faced by these students in mathematics classes than of capacities that the students bring to class for learning mathematics. It is possible that teachers are more likely to be cognizant of deficiencies in mathematics among ELLs than of their capacities for learning mathematics. Researchers have observed that this same propensity is seen in studies concerned with ELLs' learning of mathematics (Gutiérrez, 2008; Moschkovich, April 2007;).

On the scale concerned with strategic mathematical knowledge, a number of important correlations between knowledge of difficulties and capacities with strategic knowledge were observed. It seemed that teachers who had knowledge of valid mathematical algorithms used by Latino ELLs were more likely to be aware of the benefit that these students can receive from working in classroom environments that lend themselves to communication about mathematics with the teacher and with peers. One reason for this result could be that teachers who employ collaborative group work and, as result, are aware of its benefit for ELLs, have had more opportunity to observe the different algorithms used by their Latino ELLs, since a portion of their classroom time may be given to circulating among working student groups rather than to teaching the whole class. Alternatively, it is possible that teachers who are cognizant of the valid alternative mathematical notation and algorithms used by their Latino ELLs are more likely to also know that allowing these students to communicate their mathematics with their peers benefits both the ELLs and their peers. Another finding of the analysis of items in this scale was that knowledge of the ways in which variations in mathematical notation between Latin America and the United States can cause difficulty for Latino ELLs appeared to be associated with knowledge of two important strategies that may alleviate this difficulty: using collaborative grouping as a means of promoting mathematical communication and reducing the linguistic complexity of mathematics problems.

Aside from assessing the reliability of the two scales, finding associations between teacher factors and scale scores, and analyzing item responses, further (nomological) validity evidence was sought by comparison of the pilot-study sample of pre-service teachers with the sample of practicing teachers. These analyses provided compelling evidence that KT-MELL, as measured by the research instrument, seems to be a product of actual teaching experience. On both scales separately, as well as on the full set of 32 items under investigation, practicing teachers responded correctly to a higher percentage of items than did pre-service teachers. Combined with the results of linear regression analyses of scale scores for the sample of practicing teachers—in which years of experience teaching mathematics, experience teaching higher percentages of ELLs, breadth of educational licensure, and offering advice for teaching ELLs were all associated with higher scores—, these results seem to indicate that actual classroom experience of teaching mathematics to ELLs is significantly associated with higher levels of KT-MELL. Theoretically, KT-MELL, as a sub-domain of PCK, should be, at least in part, knowledge that is gained through teaching practice. Hence, the finding that, within the sample of practicing teachers, the level of experience predicts the percentage of correct responses, along with the finding that, across samples of pre-service and in-service teachers, in-service status predicts the percentage of correct responses seems to indicate that the instrument used in this study may offer a valid measure of KT-MELL.

Limitations of this Study

A principal outcome of this study is an instrument that measures teachers' knowledge for teaching mathematics to the special population of Latino ELL students with reasonable reliability and validity. Nevertheless a number of limitations of both the methods and the instrument of this study need to be addressed so that the full applicability of the research may be properly understood. These limitations relate, broadly, to the generalizability of results, the validity and reliability of the measure, and the purposes of measurement.

The sample of practicing teachers that participated in this study was both relatively small ($N = 42$) and non-random. Although the teachers came from wide geographic areas of the state (see Figure 8) and the characteristics of these teachers bore similar qualities as those of the Texas middle grade mathematics teachers generally and (see Table 4), the non-randomness and smallness of the sample imply that results of the study apply to the sampled teachers alone and any generalization should be done with extreme caution. Furthermore, that the sample included almost exclusively middle grade mathematics teachers and only those from the state of Texas, generalization of results to other levels of mathematics teachers and to teachers of other regions may not be possible. The small sample size also implies the possibility of measurement errors. All of the statistical tests used in this study benefit from—and some require—larger sample sizes for their results to be accepted as trustworthy and generalizable.

The content validity of the central instrument of this study was sought through strict adherence of the items to published research and to contexts observed in actual classroom practice. The instrument also passed through an expert panel review phase in which experts commented on items and assigned them to their respective knowledge domains, as well as through panel review by a group of practicing mathematics teachers. Notwithstanding these steps to ensure content validity, the classroom observations that informed the development of many items were all conducted in a narrow region of central Texas. Therefore, it is conceivable that some aspects of the three knowledge domains composing the test framework of the instrument may be absent. The items are neither exhaustive in the mathematical curricular contexts in which they are situated nor in the many types of classroom interactions among ELL mathematics learners. The items in the instrument are thus representative of *some* of the aspects of the knowledge domains in question, as required by the second of the research questions of this study. They are neither representative of *all* of the mathematics teaching situations encountered involving ELLs nor of *all* of the mathematics topics that ELLs encounter.

The survey instrument in this study is also limited by its reliability. Whereas both of the subscales identified exhibited suitable psychometric properties, the measures of internal consistency obtained—$\alpha = 0.6214$ and $\alpha = 0.6056$, respectively, for the KDIFF/KCAP and KSTRAT domains—may serve mainly to demonstrate the plausibility of developing measures of KT-MELL. Indeed this was one of the most central intended results of the study. However, researchers wishing to make usage of the instrument as a measurement tool will probably prefer to wait until greater internal consistency has been achieved through more extensive item-writing and testing. This instrument may serve as a valuable starting point for that work.

An additional limitation of this study concerns the understanding of the structure and the growth of the knowledge that the instrument is intended to measure. This study provided three hypothesized domains of knowledge needed for teaching mathematics to Latino ELLs. It also gave evidence that two of these domains may be indistinct. However, the results of this study cannot identify the specific types of interactions, such as causal relationships, that may exist between the constructs. That understanding would require, at least, alternative analysis methods (such as structural equation modeling), that necessitate larger sample sizes. Furthermore, this study does not identify precisely how KT-MELL develops and grows in teachers or how this knowledge changes. Results of the study imply that KT-MELL may be related to both duration and breadth of experience teaching mathematics to Latino ELLs. Yet, the contribution of specific types of teaching experiences to the growth of the knowledge is not yet understood. Furthermore, the connection of KT-MELL with student outcomes, an important question, was not addressed in this study.

Implications for Mathematics Teacher Educators and Policy-Makers

A key finding of this study was that knowledge for teaching mathematics to Latino ELLs seems to be more related to teaching experience than to formal training. Whereas years of teaching experience, experience teaching larger percentages of ELLs, and breadth of educational licensure were all associated with higher scale scores, measures of formal training, such as having a mathematics degree or having had ELL-related professional development, were either not associated with scores or were associated with reduced scores. This finding seems to imply the superiority of actual teaching experience over established means of mathematics teacher education concerning

ELL issues. Such a result seems to have implications for mathematics teacher educators and for educational administrators and policy-makers.

These results probably should not be taken as implicative of the futility of formal teacher training or of professional development methods. However, reconsideration of the contents of trainings offered to mathematics teachers of ELLs may be in order. This study has offered a framework for teacher knowledge domains *hypothesized* to be crucial to effective mathematics instruction of Latino ELLs. If further research were to establish an association of this knowledge with positive student outcomes, then it would be conceivable that these domains could be informative of professional development models aimed at improving teachers' knowledge for teaching mathematics to ELLs. Furthermore, the instrument used in this study may have value as a measure of knowledge growth resulting from professional development experiences. Indeed, at the time of writing a school district outside of Texas that was currently experiencing rapid growth of its Latino ELL population had contacted the researcher and expressed interest in administering the instrument to teachers of all levels with a view to using the results to inform professional development.

The result that KT-MELL seems to be more closely related to duration and breadth of teaching experience may be important for educational administrators and policy-makers as well. If this knowledge is indeed fundamental to effective mathematics teaching of Latino ELLs and if the growth of this knowledge is primarily a product of teaching experience as implied by this study, then administrators and policy-makers may wish to make decisions that promote both teacher-retention and broad teaching experiences for teachers. In this study, years of experience, multiplicity of educational licenses, and experience teaching larger numbers of ELLs seemed to be associated with greater knowledge. Educational decision-makers may wish to impose structures that can improve not only teacher-retention, but also the opportunity that all teachers have for teaching in classrooms containing significant populations of ELLs. For example, whereas most states require that pre-service mathematics teachers participate in a semester-length teaching internship program before they obtain teacher licensure, states may wish to bolster the duration of teaching experience before granting licensure. A year-length (possibly paid) full-time internship, in which teachers-in-training are required to teach in at least two distinct mathematics courses involving ELLs, could potentially improve their knowledge for teaching these students. Alternatively, to improve the knowledge of existing mathematics teachers, administrator may wish to require that teaching assignments vary annually. Requiring teachers to have experience teaching many different courses may result in the breadth of experience that elevates teachers' KT-MELL as implied by this study.

Directions for Further Research

This study provides a number of immediate avenues for mathematics education researchers. Firstly, it would be informative to investigate whether the measure of KT-MELL provided by this study holds any associations with measures of student outcomes, such as with the scores of teachers' students on standardized tests of mathematics achievement. This may indeed be the most important direction for research that builds upon this study to take. Although this study has shown that the knowledge measured by the instrument under investigation seems to be a product of mathematics teachers' experience in teaching ELLs, it has not shown whether said knowledge is indeed fundamental to the effective teaching of Latino ELLs. One way to answer this question would be to investigate the association between teachers' KT-MELL scale scores and their students' standardized mathematics test scores.

Secondly, to accomplish this investigation would require a measure of KT-MELL that has greater reliability and validity than does the current measure. While the limitations of the current measure in terms of reliability and validity have been addressed, it is even more important to note that this research has provided rich evidence concerning the kinds of items that can be used to improve both. This research involved a body of KT-MELL survey items, the psychometric properties of which have been fully evaluated. These properties can be used to improve the instrument. Indeed, "This knowledge of how an instrument will behave in estimating person locations permits the design of an instrument with specific estimation properties" (de Ayala, 2009, p. 32). An immediate direction for future research is to use these research findings to increase both the number of valuable items and the breadth of the ability continuum for which they provide good information.

Thirdly, and closely related to developing a more valid and reliable measure of KT-MELL, this study has contributed to a theory of the structure of this knowledge. An essential step to advance the understanding of this structure would be confirmatory factor analysis. Accomplishing this calls for, at least, a replication study that would involve a larger number of participants.

Fourthly, this study has demonstrated the possibility of measuring knowledge for teaching mathematics to a special population of students, Latino English Language Learners. Another implication of this work is that it may be possible to develop measures of knowledge for teaching mathematics to any of a large number of special populations of students—of language groups or of ability groups, for example. Such measures may have value for informing the contents of professional development protocols for mathematics teachers of such student groups. Furthermore, if the measures are found to have predictive capacity for student outcomes, then the measures may be informative for teaching assignment decisions as well. Investigation of the development of instruments to measure knowledge for teaching mathematics to various student populations is a direction for future research in mathematics education.

Fifthly, another important direction for researchers interested in mathematics teachers' knowledge is to investigate the association that the measure of the knowledge given by this study may have with at least two other important measures. First, it would be informative to investigate whether there exist any associations of this measure of KT-MELL with other more established measurements of mathematical knowledge for teaching, such as those offered by the learning mathematics for teaching project (Hill & Ball, 2004). These measures have received a great deal of attention in the last decade and have also been shown to have predictive capacity for measures of student outcomes (Hill, Rowan, & Ball, 2005). A better understanding of the relationship of the proposed knowledge domains of this study—knowledge of the difficulties faced by Latino ELLs, of the capacities possessed by Latino ELLs, and of strategies for teaching Latino ELLs in mathematics classes—with other previously studied domains of pedagogical content knowledge would advance current understandings of the dimensions of knowledge needed for teaching mathematics.

Finally, since this study has shown that the actual experience teaching mathematics to classrooms of ELLs seems to be related to this knowledge, an important direction for this research to take is in better understanding how KT-MELL develops and how it can be changed. If indeed this knowledge is primarily a product of teaching experience, then what kinds of teaching experiences are most beneficial for mathematics teachers of Latino ELLs? How should these experiences be sequenced to optimize the acquisition of this knowledge? What are the most beneficial teaching contexts in which teachers should be placed for their professional development? Are any professional development experiences valuable for the development of KT-MELL? If so, which ones? Answers to these questions could have significant impact for the improvement of opportunities for teachers to be adequately prepared to serve Latino ELLs.

APPENDIX A

TESOL P-12 TEACHER EDUCATION PROGRAM STANDARDS

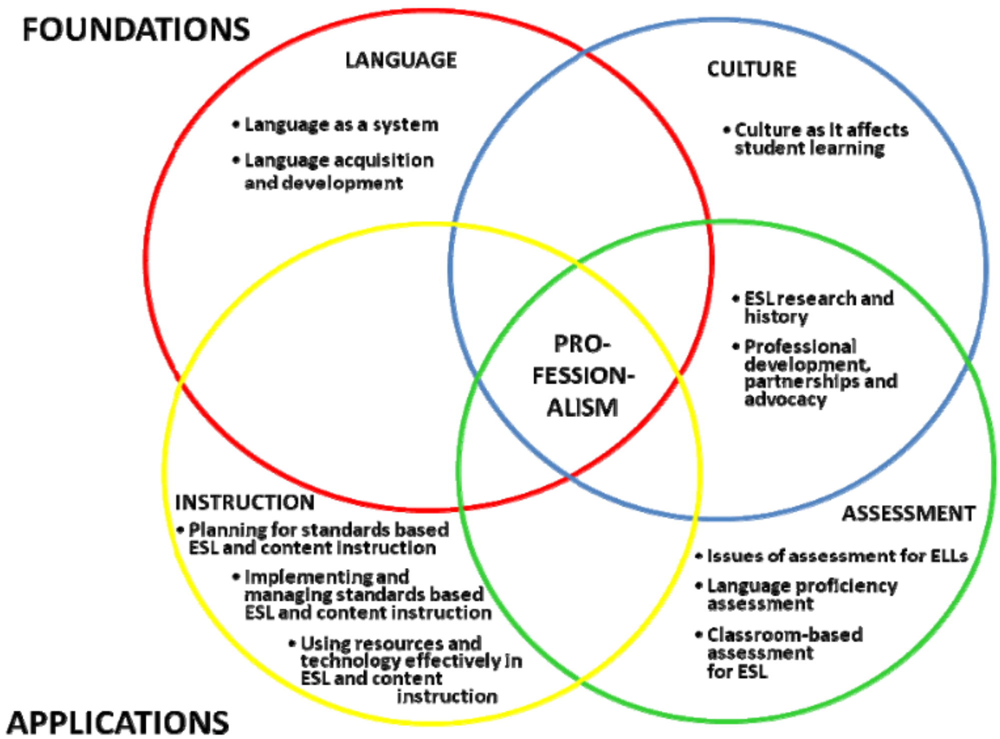


Figure A1. The Five Domains of the TESOL-NCATE P-12 Teacher Education Program Standards. The standards are used to evaluate and recognized programs that educate of teachers of English for speakers of other languages (TESOL). Adapted from "TESOL/NCATE Standards for the Recognition of Initial TESOL Programs in P-12 ESL Teacher Education," by Teachers of English to Speakers of Other Langauges, Inc. – A Global Education Assocation. Copyright 2010 byTeachers of English to Speakers of Other Languages, Inc

APPENDIX B

NATIONAL CLEARINGHOUSE FOR ENGLISH LANGUAGE ACQUISITION (NCELA), FUNDAMENTALS FOR EVERY SUCCESSFUL TEACHER OF ELLS

Language

1. *First and second language acquisition*
   Teachers understand the similarities and differences between first and second language acquisition.
   For more see http://homepage.ntlworld.com/vivian.c/SLA/L1%20and%20L2.htm.

2. *Language acquisition stages*
   Teachers understand the natural progression of language acquisition stages.
   For more see http://www.everythingesl.net/inservices/language_stages.php.

3. *BICS and CALP*
   Teachers understand and apply the theories of Basic Interpersonal Communication Skills (BICS) development for social language applications, and Cognitive Academic Language Proficiency (CALP) development for academic language applications.
   For more see http://www.everythingesl.net/inservices/bics_calp.php.

4. *Applied linguistics applications*
   Teachers use the knowledge of applied linguistics.
   For more see http://talktime.wordpress.com/2009/08/07/teaching-pronunciation-to-adult-english-language-learners/.

Culture

5. *Cultural adaptation and culture shock*
   Teachers understand and address cross-cultural issues.
   For more see http://www.asu.edu/clas/shesc/projects/bajaethnography/shock.htm .

6. *Learning styles and culture*
   Teachers frequently use culturally specific learning styles.
   For more see http://www.colorincolorado.org/educators/content/cooperative.

7. *ELL parental involvement*
   Teachers understand cultural communication patterns and cultural differences when communicating with parents.
   For more see http://www.palmbeach.k12.fl.us/MULTICULTURAL/MulticulturalNew/ProceduresManual/19%20PARENT%20INVOL.pdf.

8. *Inclusion of ELLs in classroom and school cultures*
   Teachers realize that ELLs feel alienated, especially when they are isolated from peers of their linguistic and cultural background.
   For more see http://www.colorincolorado.org/educators/reachingout/welcoming.

Figure A2. National Clearinghouse for English Language Acquisition (NCELA), Fundamentals for Every Successful Teacher of ELLs. Adapted from "Sixteen Fundamentals for Successful Teachers of ELLs," by R. D. Leier and L. A. Fregeau, winter 2010, *AccELLerate! The quarterly newsletter of the National Clearinghouse for English Language Acquistion*, *2*(2), pp. 22—23. Copyright 2010 by the National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs (NCELA).

Policy

9. Laws and policies governing the education of ELLs
Teachers are aware of the obligation to provide instructional accommodations to ELLs and assess their yearly progress in language and content areas.
For more see http://www.ncela.gwu.edu/faqs/view/6.

10. ELLs' juxtaposition to Special Education
Teachers do not view ELLs as "disabled" (needing Special Education accommodations) as this can hold back their language and academic progress.
For more see http://www.geneseeisd.org/edlearn/docs/ESL/ELLs%20-%20Special%20Education.pdf.

11. Consequences of language discrimination on ELL learning and retention
Teachers are aware of how discrimination based on language can negatively affect the social and academic achievement of ELLs.
For more see http://faculty.weber.edu/rwong/edu3200/articles/ELLClassMgt.pdf.

12. Attitudes towards English language learners
Teachers view working with ELLs as a rewarding and unique opportunity to learn about another language and culture. They do not view ELLs as "problems" to be avoided or fixed.
For more see http://www.ed.psu.edu/educ/pds/intern-resources/ESL_handbook.pdf.

Teaching

13. Make input more comprehensible
Effective teachers use visuals such as photos, pictures, illustrations, graphs, charts, graphic organizers, and even gestures to augment comprehension.
For more see http://esl.fis.edu/teachers/support/sum.htm

14. Include both content and language objectives when planning lessons
Mainstream and sheltered content instruction teachers include both language objectives and content objectives in lesson planning for ELLs.
For more see http://www.newhorizons.org/spneeds/ell/wallace.htm.

15. Appropriate language translation services
Teachers are aware of free online translation services and pre-translated letters and forms available to them through TransAct Library (online language translation of school forms). It should be noted that online translation services translate text literally and may not transfer meaning accurately and, therefore, need to be reviewed by a native speaker before dissemination!
For more see http://babelfish.yahoo.com/translate_txt and http://www.transact.com/.

16. The use of ELLs' cultural knowledge in scaffolding new content
Teachers realize that ELLs' participation in class is not dependent only on language proficiency but also on the inclusion of the ELLs' culture and background knowledge through scaffolding.

Figure A2-Continued. National Clearinghouse for English Language Acquisition (NCELA), Fundamentals for Every Successful Teacher of ELLs.

APPENDIX C

FIRST PILOT STUDY QUESTIONNAIRE WITH INVITATION LETTER

Mathematics and ELL Survey

Dear Texas State Student:

You are among a small group of students that have been selected to help us better understand how to close the mathematics achievement gap between English Language Learning (ELL) students and other students. Your thoughtful responses to this survey will help mathematics educators who are seeking to improve the opportunity of all students to excel in mathematics. We very much hope you will participate!

We ask that you return your completed survey to the instructor of the class in which you received the survey at the next class period. Or you may also return it to:

Department of Mathematics, MCS 470
ATTN: Aaron Wilson

Thank you for your time and consideration of this important educational issue!

1. At the beginning of the school year, a sixth grade math teacher divided her students into groups and gave them this problem. Read the problem and answer the highlighted questions below:

> *Antonio and María have two different ways to arrange square tables.*
> *Only one person can sit on each side of one of the tables.*
> *Antonio's Way: Keep the tables separate*
> *Maria's Way: Push all of the tables together in a long, narrow row*
>
> *Your group's job:*
> 1. *Investigate Antonio's table arrangement plan and then Maria's plan.*
> 2. *With your group, make a two-column chart for each plan*
> 3. *Show the number of tables and the number of seats for people if Antonio and Maria use 1, 2, 3, 4, 5, or 10 tables.*

(*adapted from Coggins, Kravin, Coates, & Carroll, 2007)

CIRCLE the ONE correct statement:

A)Both Antonio and Maria's methods imply geometric sequences
B)Both Antonio and Maria's methods imply arithmetic sequences
C)Antonio's method is arithmetic while Maria's is geometric
D)Antonio's method is geometric while María's is arithmetic

2. This activity contains two strategies that are intended to appeal to English language learners (ELLs) in the classroom: using the names Antonio and María makes the activity culturally relevant for the students, and using collaborative student groups increases opportunities for speech development.

STATE which strategy is *MORE* and which is *LESS* effective, for this activity, as a support for English language learners?

Cultural relevance _______Collaborative work _______

3. A math teacher of ELLs teaches his students that there are certain "special" English words that are used in mathematics that have "magical" properties: when such words are used in a literal mathematical sentence, the order of the words in the symbolic translation of the sentence is "magically" switched.

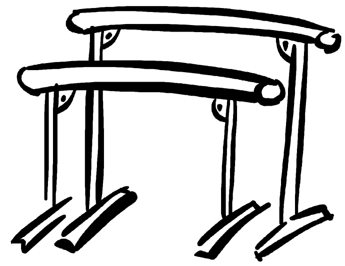
One of such words is the word "than". For example, "five less *than* an unknown number" means "$x - 5$". The number and the unknown have "switched" places.

Would this teaching strategy be effective in causing students to correctly obtain "$2x + 10$" from "Ten more than twice x"?

Effective	Not Effective	I'm not sure
1	2	3

(***adapted from Fischer & Perez, 2008)

4. In a geometry lesson, the image below was used as an example of the property of being parallel.



RATE the image in terms of cultural relevance for English language learners:

5 Very relevant4 321 Not very relevant

(* Adapted from Kersaint et al., 2009)

5. Two teachers illustrate the process of evaluating an expression for different values of the variable using slightly different methods.
CIRCLE the method that seems to be more supportive of English language learners?

A.Mr. Worther's method:

Evaluate $4n + 5$

n = 2 $4n + 5 = 4(2) + 5 = 13$

$n = 7$ $4n + 5 = 4(7) + 5 = 33$

$n = 10$?

n	$4n + 5$
2	13
7	33

B.Mrs. Vasquez's method:

Evaluate $4n + 5$, for $n = 2$, 7, and 10:

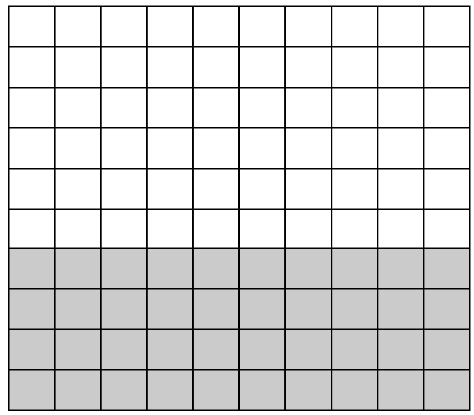
n	$4n + 5$	Simplify
2	$4(2) + 5$	$8 + 5 = 13$
7	$4(7) + 5$	$28 + 5 = 33$
10	?	?

C.Either method; they are equally effective.
(* Adapted from Kersaint et al., 2009)

6. In a class discussion on proportional reasoning, Eduardo gave *la razón inversa*.

Which did he give: the inverse ratio ?or the opposite reason? (Circle ONE)
(*adapted from Fischer & Perez, 2008)

7. Mr. Garza asks the whole class concerning the proportion of grape found in the "Grime" drink. ("Grime" is a grape-limeade.) "How do you begin to determine what proportion the grape is of the Grime?" he asks, while the students are viewing the image below:



Ciarra responds: "Cuentas las grisas y divisas por los cien." Mr. Garza responds, "Did somebody say something? I didn't understand you." But Mr. Garza speaks Spanish fluently.

Is Mr. Garza's requirement that Ciarra respond in English beneficial for Ciarra, whose first language is Spanish?

Beneficial	Not Beneficial	I'm not sure
1	2	3

8.Did Ciarra accurately answer his question? (CIRCLE ONE)YES / NO

Your background

•Are you a certified teacher?YES /NO

If YES, please check all that apply:

___All subjects___Mathematics___Science ___Other:_______________

•Are you:

___ Hispanic, regardless of race___ Black, not of Hispanic origin
___ White, not of Hispanic origin___ Asian or Pacific Islander
___ American Indian or Alaskan Native___ Biracial/multiracial
___ Other

•Are you:___ Female___ Male

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE!

If you have any comments about the questions, please write them in the space below.

APPENDIX D

CLASSROOM OBSERVATION PROTOCOL

The form below was used to collect classroom observation data that informed the development of survey items for this research.

Mathematics Instruction for English Language Learners
Teaching Observation Protocol
CAREER Project
Texas State University

I.BACKGROUND INFORMATION

Name of teacher ______________________________
Announced Observation?________________________
(yes, no, or explain)

Location of class
____
(district, school, classroom, computer lab)

Grade: ____________________
Course name:____
(Algebra I, Algebra II, Geometry)

Number of students observed _____________
Type of Class_____________________________________
(Regular or Pre-AP)

Observer _____________________________________
Date of observation ___________________________

Start time ____________________________________
End time ____

Observation number _________________________________

Second Observer: ______________________________

II.DESCRIPTION OF TEACHING CONTEXT

In the space below please give a brief description of the lesson observed, the classroom setting in which the lesson took place (regular classroom, computer lab, setting arrangements, etc.) Capture, if you can, the defining characteristics of this situation that you believe provide the most important context for understanding what you will describe in greater detail in later sections. Use diagrams if they seem appropriate.

III.DESCRIPTION OF EVENTS

Record here events which may help in documenting the ratings. For the immediate rating, code for each event that you describe.

Time	Description of Events	Work connected To mathematics?	Some students talking in two languages?

Mathematics Language

1. Mathematical Language in English This code captures how fluently the teacher (and students) use mathematical language and whether the teacher supports students' use of mathematical language. Examples: Fluent use of technical language Explicitness about mathematical terminology Encouraging students to use mathematical terms			
N/A ☐	**Low** ☐	**Mid** ☐	**High** ☐
Code here if English is *not used* for the duration of the lesson.	Teacher does not demonstrate fluency in mathematical language. Teacher uses non-mathematical terms to describe mathematical ideas and procedures AND/OR teacher talk is characterized by sloppy/incorrect use of mathematical terms. If there is little mathematical language used, score as low.	Teacher uses mathematical language as a vehicle for conveying content, but has few or none of the special features listed under high. This is the default score when teacher is using mathematical language neither sloppily nor outstandingly. Also score as mid when lesson has both features of high but includes some linguistic sloppiness.	Teacher uses mathematical language correctly and *fluently*. May include explicitness about terminology, reminding students of meaning, pressing students for accurate use of terms, encouraging student use of mathematical language. Density of mathematical language is high during periods of teacher talk. Dense, fluent, and accurate student talk can also count here.

2. Mathematical Language in Spanish This code captures how fluently the teacher uses mathematical language in Spanish to support the learning and understanding of mathematics in the classroom. Examples: Instructional conversation in Spanish that is fluent and promotes complex mathematical language. Explicitness about mathematical terminology in Spanish. Encouraging students to use mathematical terms in Spanish.			
N/A ☐	**Low** ☐	**Mid** ☐	**High** ☐
Code here if Spanish is *not used* for the duration of the lesson.	Teacher does not engage in instructional conversation in Spanish regarding mathematics. OR Teacher does engage in instructional conversation in Spanish, but the use of mathematical terms is incorrect or conceptually misleading.	Teacher engages in instructional conversation in Spanish, but has few or none of the special features listed under high. This is the default score when teacher is using mathematical language in Spanish sparingly, or when a combination of correct and incorrect terms in Spanish is used during instruction.	Teacher engages in instructional conversation in Spanish that is fluent and promotes complex mathematical language. The use of mathematical terms in Spanish is correct. Teacher encourages the use of mathematical terms in Spanish.

3. Imprecisión en el uso del español		
• Errores en el uso del lenguaje matemático. • Errores en el uso de lenguaje común. ***Definiciones*** ***Lenguaje matemático.*** Se refiere a términos usados en matemáticas, tales como *ángulo*, *radio*, *correlación*, etc. También se refiere al nombre de símbolos matemáticos, como “>” (*mayor que).* Si el maestro hace uso incorrecto de términos matemáticos o si comete errores ortográficos o gramaticales al definir un término matemático, registre el acontecimiento. Errores de traducción también deben ser registrados, ya sean del español al inglés o viceversa. ***Lenguaje común.*** Se refiere al lenguaje usado por el maestro para explicar conceptos matemáticos sin recurrir a términos técnicos. Incluye el uso de analogías, metáforas y narraciones. El maestro debe ser particularmente cauteloso en la diferenciación entre el significado de un término en el lenguaje común y su significado en lenguaje matemático. Registre cualquier acontecimiento en el que el maestro no haya sido capaz de expresarse en forma eficaz en español al explicar conceptos matemáticos.		
Low ☐	**Mid** ☐	**High** ☐
La enseñanza estuvo libre de errores en el uso del lenguaje matemático y el lenguaje común. Si hubo errores, el maestro fue capaz de detectarlos y corregirlos.	El maestro cometió errores ocasionales en el uso del lenguaje matemático y el lenguaje común.	La enseñanza se caracterizó por el uso incorrecto del lenguaje matemático y el lenguaje común, aún cuando su uso fue esporádico.

Proficiency in Instructing ELLs in Mathematics

4. Connections of mathematics with students’ life experiences		
Low ☐	**Mid** ☐	**High** ☐
In no part of the lesson are any of the elements of High or Mid found; i.e., no connections of mathematical words to student’s life experiences.	Either teacher or students make brief mention of connections of mathematical objects, concepts and words to students’ life experiences. However, these strands of conversation are not multiple, or are not exhaustively used to enrich instruction.	Instruction and/or classroom discussion includes multiple connections of mathematical objects, concepts and words to other domains of students’ life experiences.

5. Connections of mathematics with students’ existing knowledge		
Low ☐	**Mid** ☐	**High** ☐
In no part of the lesson are any of the elements of High or Mid found; i.e., no connections of mathematical words to student’s existing knowledge are found.	Either teacher or students make brief mention of connections of mathematical objects, concepts and words to students’ existing knowledge. However, these strands of conversation are not multiple, or are not exhaustively used to enrich instruction.	Instruction and/or classroom discussion includes multiple connections of mathematical objects, concepts and words to other domains of students’ existing knowledge.

6. Connection of mathematical concepts with multiple representations Connect language with mathematical representations. For example, pictures, tables, graphs, equations.		
Low ☐	**Mid** ☐	**High** ☐
No, or almost no, connections of mathematical utterances to other forms of expression are found.	Teacher or students make connections of mathematical words to pictorial, graphic, tabular, symbolic, or alternative verbal representations. However, such connections do not characterize the lesson, or there is obvious room for additional usage of such connections.	Both teachers and students make explicit connections of mathematical concepts, symbols and objects to linguistic representations, and these connections are reinforced by repetition. That is, mathematical utterances find both verbal and other representations, such as graphs, tables, equations/expressions, illustrations.

7. Meaning and multiple meanings of words. Students may need to communicate meaning by using gestures, drawings or their first language while they develop command of the English language and mathematics.		
Low ☐	**Mid** ☐	**High** ☐
No opportunities for students to explore the meaning of the mathematical words and objects through speech and other forms of expression are found.	Some, limited opportunities for students to explore the meaning of the mathematical words and objects through speech and other forms of expression are found. Multiple meanings of mathematical words are not expounded with depth.	Opportunities for students to explore the meaning of the mathematical words and objects through speech and other forms of expression are abundant. Conversation about the multiple meanings of mathematical words (mathematical meanings and/or colloquial meanings) is found.

8. Use of visual supports For example concrete objects, videos, illustrations, and gestures in classroom conversations		
Low ☐	**Mid** ☐	**High** ☐
The teacher makes no attempt to convey mathematical meaning through visual supports, or an attempt is made, yet it obscures, rather than enhances, mathematical understanding.	An attempt to teach or reinforce teaching through usage of visual supports is made. The visual object(s) selected, or the instructional usage of the object(s), is(are) moderately helpful for solidifying students' comprehension.	Teacher supplements instruction with powerful visual media that enhance (not detract from) comprehension of mathematical concepts. Visual media may include illustrations, videos, gestures, manipulative, or other visual objects. The emphasis is on the power of the visual object in conveying or reinforcing the mathematical meaning and students' understanding.

9. Record of written essential ideas, concepts, representations, and words on the board ELLs need to see written record of the lesson without erasing so that they can refer to them throughout the lesson.		
Low ☐	**Mid** ☐	**High** ☐
The teacher (and/or students) fail to record the important information of the lesson in a place that is visible to all students, or removes the information from view at a time when it is most needed.	The teacher (and/or students) displays some, but not all of, the important information of the lesson, or removes some portion of the information from view before students have had opportunity to record it in notes or apply it in practice.	The teacher (and/or students) makes careful and conscientious usage of visual display media (chalk/dry-erase board, computer projection, etc.) and students have access to pertinent information throughout instruction and practice portions of the lesson.

APPENDIX E

LETTER REQUESTING PARTICIPATION OF EXPERT PANEL REVIEWERS

Dear Expert,

Because of your particular teaching and research experience, you are being asked to participate as a content expert in the validation of a survey instrument aimed at capturing a particular domain of teacher knowledge. This survey, which has been pilot-tested with pre-service mathematics teachers and which has shown promising psychometric properties, is designed to be administered to practicing mathematics teachers. The items in the survey have been written from the background of extensive review of research literature and a large number of them are adaptations of actual classroom instances. They are intended to capture selected elements of the knowledge that is needed by effective teachers of Latino English Language Learners. Your feedback will be used to improve the quality of this instrument.

The knowledge under consideration has been categorized into the following three domains:
Difficulties: Knowledge of common difficulties faced by Latino ELLs in mathematics classes
Capacities: Knowledge of particular cultural and linguistic factors that can capacitate Latino ELLs in mathematics classes
Strategies: Knowledge of effective mathematics instructional strategies for usage with Latino ELLs

You are being asked to ***read through each item of the instrument***. Following each item is a prompt requesting your feedback ***Please choose*** **(by marking the appropriate box)** ***the knowledge domain to which the item, in your judgment, most directly relates***. Also, (optional) opportunities to provide comments regarding each item are found in the areas labeled "Comments".

SELECT ONE: Difficulties Capacities Strategies
Comments:

Please save the completed form to your computer and then email it as an attachment to [EMAIL]@txstate.edu by [DATE].

Please contact me if you have questions or comments about this request. Also, please contact me in the future if I may return this favor by assisting you in your research endeavors. Your thoughtful attention to this task is *VERY APPRECIATED!!!*

Sincerely,
Aaron Wilson

APPENDIX F

EXPERT REVIEWERS' CATEGORIZATION OF ITEMS ACCORDING TO KNOWLEDGE DOMAINS

Table A1

Expert Reviewers' Categorization of Items According to Knowledge Domains

Item	Responses					
	Reviewer A	Reviewer B	Reviewer C	Reviewer D	Reviewer Consensus	Theoretical Orientation
1	KSTRAT	KCAP	KSTRAT	KSTRAT	KSTRAT	KSTRAT
2	KCAP	KSTRAT	KCAP	KCAP	KCAP	KSTRAT
3	KDIFF	KDIFF	KCAP	KDIFF	KDIFF	KDIFF
4	KCAP	KSTRAT	KSTRAT	KSTRAT	KSTRAT	KSTRAT
5	KCAP	KCAP	KCAP	KCAP	KCAP	KCAP
6	KCAP	KCAP	KCAP	KCAP	KCAP	KCAP
7	KCAP	KCAP	KCAP	KCAP	KCAP	KCAP
8	KCAP	KDIFF	KSTRAT	KSTRAT	KSTRAT	KSTRAT
9	KDIFF	KCAP	KSTRAT	KDIFF	KDIFF	KDIFF
10	KSTRAT	KSTRAT	KSTRAT	KSTRAT	KSTRAT	KSTRAT
11	KCAP	KSTRAT	KSTRAT	KSTRAT	KSTRAT	KSTRAT
12	KDIFF	KDIFF	KCAP	KDIFF	KDIFF	KDIFF
13	KDIFF	KDIFF	KSTRAT	KDIFF	KDIFF	KDIFF
14	KSTRAT	KSTRAT	KSTRAT	KSTRAT	KSTRAT	KSTRAT
15	KCAP	KCAP	KCAP	KCAP	KCAP	KCAP
16	KCAP	KDIFF	KSTRAT	KCAP	KCAP	KCAP
17	KCAP	KCAP	KSTRAT	KCAP	KCAP	KCAP
18	KCAP	KCAP	KCAP	KCAP	KCAP	KCAP
19	KSTRAT	KSTRAT	KCAP	KSTRAT	KSTRAT	KSTRAT
20	KCAP	KCAP	KCAP	KCAP	KCAP	KCAP
21	KDIFF	KDIFF	KSTRAT	KSTRAT	KDIFF	KDIFF
22	KCAP	KCAP	KSTRAT	KCAP	KCAP	KCAP
23	KSTRAT	KSTRAT	KCAP	KSTRAT	KSTRAT	KSTRAT
24	KDIFF	KDIFF	KSTRAT	KDIFF	KDIFF	KDIFF
25	KDIFF	KDIFF	KCAP	KDIFF	KDIFF	KDIFF
26	KCAP	KCAP	KCAP	KCAP	KCAP	KCAP
27	KDIFF	KCAP	KCAP	KCAP	KCAP	KCAP
28	KCAP	KCAP	KCAP	KDIFF	KCAP	KDIFF
29	KSTRAT	KDIFF	KCAP	KDIFF	KDIFF	KSTRAT
30	KCAP	KCAP	KSTRAT	KCAP	KCAP	KCAP
31	KDIFF	KDIFF	KSTRAT	KDIFF	KDIFF	KDIFF
32	KDIFF	KDIFF	KCAP	KDIFF	KDIFF	KDIFF

NOTE: Highlighted sections indicate divergence of expert consensus from theory.

APPENDIX G

LETTER REQUESTING SURVEY PARTICIPATION BY SCHOOL DISTRICTS

TEXAS STATE UNIVERSITY
SAN MARCOS
The rising STAR *of Texas*

November 28, 2012

Subject: Request for participation in KT-MELL survey

Dear School District Administrators and Principals,

I am writing to request your help in contacting middle and high school mathematics teachers from your school district to take a survey. This survey is part of my dissertation research and is aimed at understanding Knowledge for Teaching Mathematics to English Language Learners (KT-MELL). It contains items that ask teachers to choose between different instructional decisions, to identify difficulties that students can have with specific topics, to understand students' mathematical work, or to identify student traits that can help them to learn math. The survey will be administered online and is designed to take 20 minutes to complete. A sample of the online survey, including a subset of the items, is available upon request.

In this letter, I am requesting the names and email addresses of middle and high school mathematics teachers, to whom the offer to take the survey will be extended. The results of this survey research will be used to help researchers and educators understand how to better serve this underserved population of students in mathematics classes and how to support their teachers. This research has satisfied the requirements of Texas State's Institutional Review Board for protection of human subjects (IRB# EXP2012K4335). Teachers' participation in the survey is completely voluntary and they can withdraw from participating at any time. Furthermore, all responses will be kept confidential and no personally identifiable information will be associated with responses in any reports of the data. No information concerning specific students is requested. Also, no information that can be used to identify the specific school district will be given in any reports of the data. As a token of appreciation, two teachers will be randomly selected from all participating districts to receive $50 each.

For your convenience, I am attaching a flyer that may be used to advertise this opportunity to principals and their teachers. Should the school district be interested in my research findings, I will make the results available to any participating district. Please feel free to request any other information that may be of help to you in your decision to support this research effort. Alternatively, you may also contact my supervising professor, Dr. Maria Alejandra Sorto (sorto@txstate.edu, 512-245-4722), if you would like more information about my study.

I look forward to hearing from you and hope that I may expect the support of your school district in helping to advance our knowledge of ways to improve the educational opportunities for English Language Learners in Texas.

Please accept my sincere appreciation for your assistance in my research efforts.

Aaron Wilson

Aaron Wilson
Doctoral Research Assistant
Department of Mathematics
Texas State University--San Marcos
Phone: 512-245-4753
Email: atwilson@txstate.edu

DEPARTMENT OF MATHEMATICS
601 University Drive | MCS 470 | San Marcos, Texas 78666
phone 512.245.2551 | *fax* 512.245.3425 | WWW.TXSTATE.EDU

This letter is an electronic communication from Texas State University-San Marcos, a member of The Texas State University System.

APPENDIX H

FLYER FOR ELICITING TEACHER PARTICIPATION IN SURVEY

have you got kt-mell?

Knowledge for teaching mathematics
to English Language Learners Survey

Your knowledge and experience in teaching ELLs makes all the difference.

What is this survey about?

How do Latino ELLs learn math? What things help or hinder these students in learning math? What do effective teachers of Latino ELLs do with their students? And what do YOU KNOW ABOUT IT? That's what this survey is about.

Why should I take this survey?

Your participation by answering the questions of this survey can guide future efforts to support Latino ELLs in Texas and their teachers. **Furthermore, as a small token of appreciation two teachers will be randomly selected to receive $50!**

If you have further questions about participation, please contact:

Aaron Wilson
Texas State University-San Marcos
Mathematics Department,
Doctoral Research Assistant,
atwilson@txstate.edu
512-245-4753

How long will this survey take?

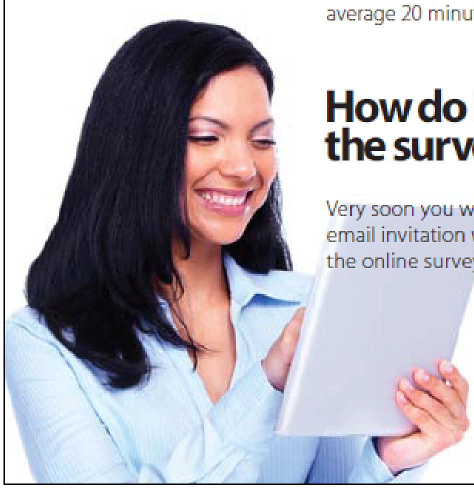
This survey is designed to take on average 20 minutes to complete.

Is this survey confidential?

Yes. Individual responses to this survey are completely confidential and no personally identifiable information will be associated with responses in any reports of the results of the survey. (Texas State IRB# EXP2012K4335)

How do I take the survey?

Very soon you will receive an email invitation with a link to the online survey.



TEXAS STATE UNIVERSITY
SAN MARCOS
The rising STAR of Texas

APPENDIX I

INITIAL SURVEY INVITATION EMAIL SENT TO TEACHERS

Dear [Teacher],

I am writing to ask for your participation in a survey that I am conducting for my dissertation as part of a doctoral program in Mathematics Education at Texas State. I am asking middle school math teachers like you to answer questions about teaching math to Latino English Language Learning students. Your responses to this survey are very important and will help in understanding what teachers know about teaching this significant population of students in Texas.

This is a short survey and should take you about 20 minutes to complete. I have already received the approval of your administrators and they encourage your participation. Please click on the link below to go the survey website (or copy and paste the survey link into your Internet browser).

<survey link>

Your participation is voluntary (you can withdraw at any time without penalty) and your responses are confidential. No personally identifiable information will be associated with your responses in any reports of this data. Should you have any further questions or comments, please feel free to contact me at atwilson@txstate.edu or [phone number].

As a small token of appreciation, two teachers will be randomly selected from all participating schools and districts to receive $50 each. Your time and consideration in completing the survey are valuable. I hope that you enjoy taking it! Through the help of teachers like you I think that we can make better decisions about teacher training and the math curriculum taught in schools.

Thank you,
Aaron Wilson
Doctoral Student
TEXAS STATE UNIVERSITY
SAN MARCOS
The rising STAR of Texas

APPENDIX J

SECOND SURVEY INVITATION EMAIL SENT TO TEACHERS

Dear [Teacher],

I am math teacher and graduate student at Texas State University. I am asking math teachers like you to answer questions about teaching math to Latino English Language Learning students.

- This is a short, 20 minute survey and it doesn't have to be completed all at once; just click on the link in this email to resume at any time.
- As a small token of appreciation, two teachers will be randomly selected receive $50.
- Your responses will contribute to helping us make better decisions about teacher training and the math curriculum taught in schools.
- Your administrators approve of this survey.

To begin, click on this link or paste it into your browser: <survey link>

Your time and consideration in completing the survey are greatly appreciated. I hope that you enjoy taking it!

Please contact me if you have questions or if you would like to be removed from further mailings. (Please DO NOT REPLY to this email. You can contact me at the email address below.)

Thank you,

Aaron Wilson
Doctoral Research Assistant
Department of Mathematics
Texas State University - San Marcos
atwilson@txstate.edu
512-245-4753

Confidentiality Notice: Your participation is voluntary (you can withdraw at any time without penalty) and your responses are confidential. No personally identifiable information will be associated with your responses in any reports of this data.

APPENDIX K

Third and Final Survey Invitation Email Sent to Teachers

Dear [Teacher],

As my final email request to you, I ask you to consider all of the ways in which you may be helping by taking this survey:

- You will be participating in research funded by the National Science Foundation that is attempting to improve opportunities for ELLs in mathematics.
- You will be helping a fellow math teacher to pursue a graduate degree.
- You may help yourself to become $50 richer (or less poor, as we teachers often think of it). Two participating teachers will be randomly selected to receive a $50 reward.

To take the KT-MELL survey, simply click on this link or paste it into your browser: <survey link>.

The survey takes about 20 minutes and can be completed in parts. Your time and consideration in completing the survey are greatly appreciated. I hope that you enjoy taking it! This is my final request of participation; I will send you no further emails. Please contact me if you have questions about this study.

Thank you and happy holidays!

Sincerely,

Aaron Wilson
Doctoral Research Assistant
Department of Mathematics
Texas State University - San Marcos
atwilson@txstate.edu
512-245-4753

Confidentiality Notice: Your participation is voluntary (you can withdraw at any time without penalty) and your responses are confidential. No personally identifiable information will be associated with your responses in any reports of this data.

APPENDIX L

KDIFF/KCAP ITEM SPECIFICATIONS

Table A2

Item Specifications for the KDIFF/KCAP Knowledge Domain

Item	**Measurement Specification**
Q1	KSTRAT: Measures knowledge of the usage of students' own mathematical writings and speech as a strategy for teaching math to ELLs.
Q6	KSTRAT: Measures knowledge of the usage of students' own mathematical writings and speech as a strategy for teaching math to ELLs.
Q9	KDIFF: Measures knowledge of linguistic complexity in mathematics as a source of difficulty for understanding problems.
Q13	KDIFF: Measures knowledge of linguistic complexity in mathematics, and of specific English words (non-Spanish cognates), that can cause difficulty for ELLs.
Q15	KCAP: Measures knowledge of students' first language as a resource (and not obstacle) to mathematical reasoning.
Q16	KCAP: Measures knowledge of usage of Spanish language in deductive mathematical reasoning concerning geometric area and a scale factor.
Q20	KCAP: Measures knowledge of students' first language (especially of Spanish to English cognates) as a resource (and not obstacle) to mathematical reasoning.
Q22	KCAP: Measures knowledge of students' first language (especially of Spanish to English cognates) as a resource (and not obstacle) to mathematical reasoning.
Q25	KDIFF: Measures knowledge of linguistic complexity in mathematics as a source of difficulty for understanding problems.
Q31	KDIFF: Measures knowledge of difficulty in mathematics for ELLs found in linguistic complexity and in polysemy.
Q32	KDIFF: Measures knowledge of specific ways in which translation of mathematical symbols into English language can cause difficulty for ELLs.

APPENDIX M

KSTRAT ITEM SPECIFICATIONS

Table A3

Item Specifications for the KSTRAT Knowledge Domain

Item	Measurement Specification
Q5	KCAP: Measure knowledge of the traditional U.S. long division algorithm as a valid algorithm.
Q7	KSTRAT: Measures knowledge of the benefit for ELLs of group work as an instructional format that promotes classroom environments that are rich in language and mathematics content.
Q8	KSTRAT: Measures knowledge of the benefit for ELLs of group work as an instructional format that "promotes classroom environments that are rich in language and mathematics content.
Q10	KSTRAT: Measures knowledge of the benefit for ELLs of group work as an instructional format that "promotes classroom environments that are rich in language and mathematics.
Q11	KDIFF: Measures knowledge of an alternative mathematical notation (commonly used in Central America) that can be a barrier to comprehending mathematics for ELLs.
Q18	KCAP: Measures knowledge of a valid alternative representation of long division algorithm, typically used in central America.
Q19	KSTRAT: Measures knowledge both of visual displays that support ELLs' in mathematics and of strategies for promoting an environment that is rich in language and mathematics content.
Q24	KDIFF: Measures knowledge of a specific example of polysemy (numerical versus demonstrative) as a difficulty for ELLs in mathematics.
Q26	KCAP: Measures knowledge both of the traditional U.S. long division algorithm and of an alternative representation of the long division algorithm (typically used in central America) as valid.
Q29	KSTRAT: Measures knowledge of effective means of reducing linguistic complexity while retaining mathematical integrity in mathematics.

APPENDIX N

F-STATISTICS FOR THE ANOVA TEST OF DIFFERENCES OF MEANS: SCALE SCORES BY TEACHER FACTOR

Table A4

F-statistic and Associated p-Values for the ANOVA Test of Equivalence of Means on Both Scales Against all Possible Teacher Factors

Teacher Factor	*Domain 1: Knowledge of Linguistic Difficulties and Capacities*		*Domain 2: Strategic mathematical knowledge*	
	F(df)	*p*	*F(df)*	*p*
Years of teaching experience	*F(5,36) = 1.138*	*p = .358*	*F(5,36) = .913*	*p = .484*
1 – 5 years				
6 – 10 years				
11 – 15 years				
16 – 20 years				
More than 20 years				
Years of teaching experience	*F(1, 40) = 3.683*	***p = .062***	*F(1, 40) = 1.621*	*p = .210*
5 or fewer years				
More than 5 years				
Teacher grade level	*F(2, 39) = .387*	*p = .682*	*F(2, 39) =1.262*	*p = .295*
Elementary				
Middle				
High school				
Breadth of teaching experience	*F(1, 40) = 3.361*	***p = .074***	*F(1, 40) = 1.160*	*p = .288*
Taught 3 or fewer different courses				
Taught more than 3 different course				
Level of courses taught	*F(1, 40) = .472*	*p = .496*	*F(1, 40) = .079*	*p = .780*
6^{th} and 7^{th}				
8^{th} grade or above				
Experience teaching ELLs	*F(1,40) = 1.982*	*p = .167*	*F(1, 40) = .137*	*p = .713*
Taught 20% or fewer ELLs				
Taught classes with more than 20% ELLs				

Table A4-Continued

F-statistic and Associated p-Values for the ANOVA Test of Equivalence of Means on Both Scales Against all Possible Teacher Factors

Experience teaching classes with ELLs *0 – 20% ELLs* *20 – 40% ELLs* *More than 40% ELLs*	*F(2, 39) = 2.639*	***p = .084***	*F(2, 39) = .515*	*p = .601*
Experience teaching classes with ELLs *40% or fewer ELLs* *More than 40% ELLs*	*F(1, 40) = 5.394*	***p = .025***	*F(1, 40) = .305*	*p = .584*
Extent of formal mathematics training *"basic math like College Algebra"* *"several higher level math courses"* *"bachelor degree or higher in math"*	*F(2, 39) = 1.411*	*p = .256*	*F(2, 39) = 3.066*	***p = .058***
Having a degree in mathematics *Yes* *No*	*F(1, 40) = .008*	*p = .928*	*F(1, 40) = .058*	*p = .812*
Having certification to teach English as a second language *Yes* *No*	*F(1, 40) = 1.812*	*p = .186*	*F(1, 40) = .192*	*p = .664*
Multiple educational licenses *Yes (additional license possessed–other content area, principal, etc.")* *No*	*F(1, 40) = .757*	*p = .389*	*F(1, 40) = 3.387*	***p = .073***
Having professional development (beyond college) for teaching ELLs *Yes* *No*	*F(1, 40) = 1.943*	*p = .171*	*F(1, 40) = 2.729*	***p = .106***
Knowledge of the Spanish language *"not at all" or "a few words"* *"know the basics"* *"make conversation" or "very fluent"*	*F(2, 39) = .287*	*p = .752*	*F(2, 39) = .138*	*p = .871*
Knowledge of other languages *Yes* *No*	*F(1, 40) = 1.839*	*p = .183*	*F(1, 40) = .289*	*p = .549*
Response given to the question, "Is there anything else teachers need to know to teach ELLs?" *Yes* *No*	*F(1, 40) = 4.095*	***p = .050***	*F(1, 40) = 2.228*	***p = .143***

Table A4-Continued

F-statistic and Associated p-Values for the ANOVA Test of Equivalence of Means on Both Scales Against all Possible Teacher Factors

Time to complete survey	$F(1, 40) = .006$	$p = .939$	$F(1, 40) = .446$	$p = .508$
Same day				
More than 1 day				

REFERENCES

Abedi, J., Courtney, M., Leon, S., Kao, J., and Azzam, T. (2006, November). *English Language Learners and Math Achievement: A Study of Opportunity to Learn and Language Accommodation*. (Technical Report 702). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Herman, J. (2010). Assessing English language learners' opportunity to learn mathematics: Issues and limitations. *Teachers College Record, 112*(3), 723-746.

Anderson, C. R. & Tate, W. F. (2008). Still separate, still unequal: Democratic access to mathematics in U.S. Schools. In L. D. English (Ed.), *Handbook of international research in mathematics education. 2nd Ed*. (pp. 299—318). New York: Routledge.

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.

Calderhead, J. (1987). Introduction: Teaching as a ‘professional’, thinking activity. In J. Calderhead (Ed.), *Exploring teachers’ thinking* (pp.1—19). Sussex: Holt, Rinehart & Winston.

Cambell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81—105.

Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: The Urban Institute.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Thousand Oaks, CA: Sage.

Carrasquillo, A., & Rodriguez, V. (2001). *Bilingual education and bilingualism, 33 : Language minority student in the mainstream classroom*. Clevedon, GBR: Multilingual Matters Limited.

Center on Education Policy. (2010). *State test score trends through 2007-08, part 6: Has progress been made in raising achievement for English language learners?* Washington, DC: Chudowsky, N. & Chudowsky, V.

Chamot, A., Dale, M., O'Malley, M., & Spanos, G. (1992). Learning and problem solving strategies of ESL students. *Bilingual Research Journal. 16*(3&4)*, 1—34*.

Chamot, A. U., O'Malley, J. M., & National Clearinghouse for Bilingual Education. (1986). *A cognitive academic language learning approach: An ESL content-based curriculum.* Washington, DC: National Clearinghouse for Bilingual Education. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED338108&site=ehost-live

Chang, M. (2008). Teacher instructional practices and language minority students: A longitudinal model. *Journal of Educational Research, 102*(2), 83-97.

Chval, K. B., & Chávez, Ó. (2011). Designing math lessons for English language learners. *Mathematics Teaching in the Middle School*. *17*(5), pp. 261-265.

Coggins, D., Kravin, D., Coates, G. D., & Carroll, M. D. (2007). *English language learners in the mathematics classroom*. Thousand Oaks, CA: Corwin Press.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4)*,*281—302. doi: 10.1037/h0040957

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391-418.

Cuevas, G. J. (1984). Mathematics learning in English as a second language. *Journal for Research in Mathematics Education, 15*(2, Minorities and Mathematics), 134-144.

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, No. 19, 121-129.

Cummins, J. (2000). Putting language proficiency in its place: Responding to critiques of the conversational/academic language distinction, in J. Cenoz and U. Jessner (eds.) *English in Europe: The acquisition of a third language.* Clevedon: Multilingual Matters Ltd.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In California Office of Bilingual Bicultural Education. (Ed.), *Schooling and language minority students: a theoretical framework* (pp.3—49). Los Angeles, CA: Evaluation, Dissemination and Assessment Center, California State University Los Angeles.

De Ayala, R.J. (2009). *Theory and practice of item response theory*. New York, NY: The Guilford Press.

DeVillis, R. F. (2003). *Scale development: Theory and applications. Applied social research methods series, vol. 26.* Thousand Oaks, CA: Sage.

Dewey, J. (1902). *The child and the curriculum*. Chicago, IL: University of Chicago Press.

Diaz, Z., Esquierdo, J. J., Ramirez, O., & Almaguer, I. (2011). Developing mathematics literacy for bilingual learners: A framework for effective learning. *Teaching for excellence and equity in mathematics, 1*(3), 10—19.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. Hoboken, NJ : John Wiley & Sons, Inc.

Echevarría, L., Short, D., & Vogt, M. (2007). *Making content comprehensible for English learners: The SIOP model* (3rd ed.). Boston: Pearson Educational Press.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Fischer, J., & Perez, R. (2008). Understanding English through mathematics: A research-based ELL approach to teaching all students. In R. S. Kitchen & E. Silver (Eds.), *Research monograph of TODOS* (pp. 43-58). Tempe, AZ: TODOS; Washington, DC: National Education Association.

Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). Practical guidelines for the education of English language learners: research-based recommendations for instruction and academic interventions. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Gee, J. (1999). *An introduction to Discourse analysis: Theory and method*. New York: Routledge.

Gutiérrez, R. (2002). Beyond essentialism: the complexity of language in teaching mathematics to latina/o students. *American Educational Research Journal, 39*(4), 1047-1088.

Gutiérrez, R. (2008). A "gap-gazing" fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education, 39*(4), pp. 357-364.

Gutstein, E., Lipman, P., Hernandez, P., & Reyes, R. d. l. (1997). Culturally relevant mathematics teaching in a Mexican-American context. *Journal for Research in Mathematics Education, 28*(6), 709-737.

Hansen-Thomas, H., & Cavagnetto, A. (2010). What do mainstream middle school teachers think about their English language learners? A tri-state case study. *Bilingual Research Journal, 33*(2), 249-266.

Henderson, R. W., & Landesman, E. M. (1995). Effects of thematically integrated mathematics instruction on students of Mexican descent. *The Journal of Educational Research, 88*(5), 290-300.

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.

Hill, H. C., & Ball, D. L. (2009). The curious--and crucial--case of mathematical knowledge for teaching. *Phi Delta Kappan, 91*(2), 68-71.

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., Ball, D.L., Sleep, L. & Lewis, J.M. (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts? In F. Lester (Ed.), *Handbook for Research on Mathematics Education (2nd ed)*, p. 111-155. Charlotte, NC: Information Age Publishing.

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 the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511.

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

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105*(1), 11.

Hume, K. R., Jones, N. A., & Ramirez, R. R. (2011). *Overview of race and Hispanic origin: 2010* (Publication No. C2010BR-02). Washington, DC: U. S. Census Bureau.

Kersaint, G., Thompson, D. R., and Petkova, M., (2009). *Teaching Mathematics to English Language Learners*. New York, NY: Routledge.

Khisty, L. L., & Morales Jr., H. (2004). Discourse matters: Equity, access and Latinos learning mathematics.

Lager, C. A. (2006). Types of mathematics-language reading interactions that unnecessarily hinder algebra learning and assessment. *Reading Psychology, 27*(2-3), 165-204.

Leier, R. D., & Fregeau, L. A. (2010). Sixteen fundamentals for successful teachers of ELLs. *AccELLerate! The quarterly newsletter of the National Clearinghouse for English Language Acquisition, 2*(2), 22—23.

Llabre, M. M., & Cuevas, G. (1983). The effects of test language and mathematical skills assessed on the scores of bilingual hispanic students. *Journal for Research in Mathematics Education, 14*(5), 318-324.

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment, 14*(3-4), 160-179.

McGahee, T. W., & Ball, J. (2009). How to read and really use an item analysis. *Nurse Educator, 34*(4), 166-171.

Meskill, C. (2005). Infusing English language learner issues throughout professional educator curricula: The training all teachers project. *Teachers College Record, 107*(4), 739-756.

Montes, F. (2002). Enhancing content areas through a cognitive academic language learning based collaborative in South Texas. *Bilingual Research Journal, 26*(3), 697–716.

Moschkovich, J. N. (2002). A situated and sociocultural perspective on bilingual mathematics learners. *Mathematical Thinking and Learning*, 4(2 & 3), 189—212. doi: 10.1207/S15327833MTL04023_5

Moschkovich, J. N. (2005). Using two languages when learning mathematics. *Educational studies in mathematics*. 62, 121—144.

Moschkovich, J. N. (2007). Bilingual mathematics learners: how views of language, bilingual learners, and mathematical communication affect instruction. In Nasir, N.S. and Cobb, P. (Eds.) *Improving Access to Mathematics: Diversity and Equity in the Classroom* (pp.89—104). New York: Teachers College Press.

Moschkovich, J.N. (2007, April). "Latinos, mathematics learning, and language: A review of the empirical research literature." Paper to be presented as part of the symposium *Mathematics achievement of Latino English learners: The impact of language, culture, curriculum, and school context.* at the American Educational Research Association (AERA) Annual Conference, Chicago, IL.

Moschkovich, J. N. (2010). *Research in mathematics education: Language and mathematics education : Multiple perspectives and directions for research*. Charlotte, NC, USA: Information Age Publishing.

National Council of Teachers of Mathematics (NCTM). (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115 Stat. 1425 (2002). Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York, NY: McGraw-Hill.

Orlando, M. (2004, June). *Critical issues to address when applying item response theory (IRT) models.* Paper presented at the Drug Information Association meeting, Bethesda, MD: Hyatt.

Pappamihiel, E. (2007). Helping preservice content-area teachers relate to english language learners: An investigation of attitudes and beliefs. *TESL Canada Journal, 24*(2), 42-60. r4 Educated Solutions. (2010). *Making math accessible to English language learners: Practical tips and suggestions, grades 9-12*. Bloomington, IN: Solution Tree Press.

Payán, R. M., & Nettles, M. T. (2008, January). *Current state of English-language learners in the U.S. K–12 student population* (English-Language Learners Symposium Fact Sheet). Princeton, NJ: Educational Testing Service. Retrieved on April 4, 2007 from http://www.ets.org/Media/Conferences_and_Events/pdf/ELLsympsium/ELL_factsheet.pdf.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, *50*(5), 157–175.

Petrou, M. & Goulding, M. (2011). Chapter 2: Conceptualizing teachers' mathematical knowledge in teaching. In T. Rowland & K. Ruthven (Eds.), *Mathematical knowledge in teaching* (9—25). New York: Springer.

Polya, G. (1957). *How to solve it.* 2nd edn. Princeton, NJ: Princeton University Press.

Pray, L., & Ilieva, V. (2011). Strategies for success: Links to increased mathematics achievement scores of english-language learners. *Teacher Education and Practice, 24*(1), 30-45.

Revelle, W. (n.d.). *An introduction to psychometric theory with applications in R*. Retrieved from http://www.personality-project.org/r/book/.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*. *17*(5), pp. 1—25.

Ross, J. A., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education*, *34*(4), 344—363.

Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Shein, P. P. (2012). Seeing with two eyes: A teacher's use of gestures in questioning and revoicing to engage English language learners in the repair of mathematical errors. *Journal for Research in Mathematics Education. 43*(5), pp. 182—222.

Schuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, working, and context.* London: Sage.

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

Téllez, K., Moschkovich, J., & Civil, M. (2011). Latina/os and mathematics education: Why this book now? In K. Téllez, J. Moschkovich, & M. Civil (Eds.), *Latinos/as and Mathematics Education* (pp.1—19). Charlotte, NC: Information Age Publishing, Inc.

Thomas, W.P., & V.P. Collier. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Santa Cruz, CA: Center for Research on Education, Diversity, and Excellence.

Trumbull, E., & Solano-Flores, G. (2011). Addressing the language demands of mathematics assessments: Using a language framework and field-based research findings. In Basterra, M. R., Trumbull, E., & Solano-Flores, G. (Eds.), *Cultural validity in assessment: A guide for educators* (pp. 218-253). New York: Routledge.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, (2011, June). *Achievement gaps: How Hispanic and white students in public schools perform in mathematics and reading on the national assessment of educational pogress.* (NCES Publication No. 2011-459). Retrieved from http://nces.ed.gov/nationsreportcard/pdf/studies/2011459.pdf

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, (2012). *The condition of education: Participation in education – elementary/secondary enrollment.* (Table-A-6-1. Number and percentage of children ages 5-17 who spoke only English at home, who spoke a language other than English at home and who spoke English with difficulty, and percent enrolled in school: Selected years, 1980-2009). Retrieved from http://nces.ed.gov/programs/coe/tables/table-lsm-1.asp.

U.S. Department of Education, What Works Clearinghouse, Institute of Education Sciences. (2013, January). *Evidence review protocol for interventions for English language learners* (Version 2.2). Retrieved from http://ies.ed.gov/ncee/wwc/document.aspx?sid=22#definitions.

U.S. Department of Education, What Works Clearinghouse, Institute of Education Sciences. (2013, January). *WWC intervention report: Sheltered Instruction Observation Protocol ® (SIOP®)*. Retrieved from http://ies.ed.gov/ncee/wWc/pdf/intervention_reports/wwc_siop_022013.pdf.

Vygotsky, L. S. (1978) *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Watson, S., Miller, T. L., Driver, J., Rutledge, V., & McAllister, D. (2005). English language learner representation in teacher education textbooks: A null curriculum? *Education, 126*(1), 148.

Williams, T., Hakuta, K., Haertel, E., et al. (2007). *Similar English learner students, different results: Why do some schools do better? A follow-up analysis based upon a large-scale survey of California elementary schools serving high proportions of low-income and EL students.* Mountain View, CA: EdSource.

Wilson, D. M., (2011, March/April). Dual language programs on the rise. *Harvard Education Letter, 27*(2). Retrieved from http://www.hepg.org/hel/article/496.

Wilson, S. M., Shulman, L. S., & Richert, A. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring teachers' thinking* (pp.104-124). Sussex: Holt, Rinehart & Winston.

Wolf, M. K., Herman, J. L., Dietel, R. (2010). *Improving the validity of English language learner assessment systems. Policy brief 10, spring 2010.* National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

VITA

Aaron T. Wilson was born in Atlanta, Georgia on March 19, 1976, the son of Loren and Dawn Wilson. He graduated from Albemarle High School in Charlottesville, Virginia in 1994. After brief travels within and without the United States he entered Texas State University-San Marcos from which he received the two degrees of Bachelor of Science in Mathematics and Bachelor of Arts in Spanish in 2005. From 2006 to 2009 he taught different mathematics courses at Hays High School in the Hays Consolidated Independent School District. In August of 2009, he entered the Graduate College of Texas State.

Permanent E-mail Address: aarontwilson@hotmail.com

This dissertation was typed by Aaron T. Wilson.