OKEYO, SHERRAD R., Ph.D. Development and Validation of an Instrument to Assess Mathematics Educators' Visions of Mathematics Instruction. (2019)
Directed by Dr. Holt Wilson. 138 pp.

This study reports on the development of a survey instrument designed to assess mathematics educators' instructional visions along what is identified in the research literature as significant dimensions of mathematics classroom practice. This survey instrument is intended to be a tool for informing implementation research and policymaking. This paper describes the extensive instrument development process that was undertaken to ensure that the final product could be applied with sufficient reliability to assess individuals' visions of math instruction. Data analysis procedures included using respondents' feedback, cognitive interviewing, Cronbach's alpha statistics and factors analysis for each dimension, and follow-up interviews. The final instrument consists of 28-items, in three dimensions, all with acceptable levels of reliability. Future directions include the addition of a pool of items to the existing instrument.

DEVELOPMENT AND VALIDATION OF AN INSTRUMENT TO ASSESS MATHEMATICS EDUCATORS' VISIONS OF MATHEMATICS INSTRUCTION

by

Sherrad R. Okeyo

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2019

Approved by

____
Committee Chair

I dedicate this dissertation to my mother, Martha Okeyo, who is the source of my strength, inspiration, and an exemplar of hard work and perseverance. To my beloved wife, Kissah Okeyo (Agaltoto), and children, Barack, Asajile, and Amani, for patience, unfaltering love, prayers, and emotional support during the pursuit of my graduate education. To the rest of the family, my mentor Dr. Holt Wilson, and friends like Dr. Jason Mose, for sharing words of advice and encouragement to finish this research.

APPROVAL PAGE

This dissertation, written by Sherrad R. Okeyo, has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair ______________________
Holt Wilson

Committee Members ______________________
Ye He

____
Kerri Richardson

____
Vicki Jacobs

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

For decades, the mathematics education community has endeavored to reform mathematics education. In essence, the reform agenda as evidenced in the publication of several “standards” documents has garnered considerable attention (Curriculum and Evaluation Standards for School Mathematics [NCTM], 1989; Principles and Standards for School Mathematics [NCTM], 2000); National Governors Association Center for Best Practices and Council of Chief State School Officers [NGA & CCSSO], 2010). These “standards” describe the kinds of mathematics students doing, and since they are ambitious, they require new forms of instruction. Research has documented the types of instruction that assist all students in meeting these standards, collectively representing Visions of High-Quality Mathematics Instruction (VHQMI) that describe rigorous mathematics knowledge and authentic ways to engage students as active participants in their learning (Munter, 2014). Specific practices that are often associated with HQMI include collaborative learning groups, inquiry-based activities, and the use of cognitively demanding tasks. All of these practices are intended to encourage active rather than passive learning, to promote critical thinking skills, justifying ideas, generalizing, and connecting mathematical relationships. In attempting to give students the opportunities to engage in the behaviors outlined above, policymakers have implemented multiple standards-based initiatives for teachers to develop practices that are assumed to facilitate students' learning (Smith, 2000; Spillane & Zeulli, 1999). Their efforts to support teacher learning of HQMI have resulted in successes (e.g., Bell, Wilson, Higgins, & McCoach, 2010; Borko, 2004; Carpenter, Fennema, & Franke, 1996), but also reveal some implementation challenges (e.g., Cohen, 1990; Weimers, 1990).

Mathematics reform efforts have generally faced enormous local implementation challenges, often failing to produce even modest changes in instruction (Coburn, Hill, & Spillane, 2016). A large body of research has explained these challenges by point to the plethora of mixed messages, increased expectations that pull educators in different directions, and conflicting interpretations (Gamoran, 2003; Spillane, 2000; Spillane, Reiser, & Gomer, 2006). For example, in a study analyzing conceptions that local district leaders constructed from a mathematics reform initiative, Spillane (2000) highlighted how the leaders' conflicting images of instruction led them to focus on superficial features, missing more profound structural ideas that were proposed by the reform. In this study, some leaders viewed hands-on activities as serving the purpose of invoking increased student interest and motivation to learn, rather than as tools for reasoning.

More generally, to address such challenges, researchers have argued for alignment of supports (Gamoran, 2003; O'Day & Smith, 1993), and opportunities to "craft coherence" (Honig & Hatch, 2004, p. 16), especially when implementing large-scale instructional improvement efforts. Specifically, policy researchers argue for situations where all members of the school system (teachers, school and district leaders, state leaders & research communities) work in partnership to craft and coordinate initiatives and develop new material resources geared towards strengthening students' opportunities to learn. For example, Cobb and Jackson (2011) found that district leaders' framing of the ways to support students' learning influences what they hold school leaders, coaches, and teachers accountable for, and thus, the prospects for district-wide instructional improvements.

Research has indicated that shared values and a common vision for teaching and learning are vital components for systemic coherence and improvements at scale (Bryk, 2010; Cobb & Jackson, 2011; McDonald, Kazemi, & Kavanagh, 2013; Munter, 2014). Instructional improvements can be significantly sustained when teachers, school, and district leaders not only hold a shared vision that is based on current reform recommendations, but that improvement efforts are tied to such a shared vision. However, professional development, collaboration, and collegiality are limited when teachers, school, and district leaders' visions of instruction are not shared (Cobb & Jackson, 2011; Cobb & Smith, 2008). Furthermore, disparate visions of instruction can result in inconsistent expectations and practices across classrooms (Munter, 2014) as well as the degree to which teachers are supported to improve their instructional practices (Cobb & Jackson, 2011). Large-scale instructional improvement efforts, therefore, should present opportunities for individuals to articulate their vision looks from their perspective. Understanding the various visions held could assist leaders, and professional development designers tailor supports for improvements in instruction.

To support the development of a shared vision of HQMI, it is reasonable to assume that assessing individuals' visions of instruction across all relevant dimensions of classroom instruction may present opportunities to gauge the extent to which their vision is compatible with those of their colleagues. Surfacing these visions may provide a measure of how far one's vision is from the shared vision of HQMI that the mathematics education community promotes (Hammerness, 2001; Munter, 2014). Also, knowledge of various visions and their quality can assist in professional development initiatives and present foundations for learning new forms of instruction (Munter, 2014; Wilhelm, 2014).

As the role that vision plays in the reform of instructional practices gains an increasingly high profile, studying vision poses a unique challenge. This challenge is particularly compounded when venturing into large-scale improvement efforts. In the book *Seeing through Teachers' Eyes*, Hammerness (2006) described this challenge in the following way:

> In our everyday experience, visions held by political or religious leaders are highly public and accessible-shared in public speeches, on television, in newspapers, or even sometimes accessible in advertisements. But teachers' [and others'] visions are . . . far from being readily apparent in practice, . . . and [are] publicly invisible. (p. 89)

In other words, the context in which mathematics educators work provides limited opportunities for them to share or sharpen their visions. Such realities present the research community with a methodological problem: What methods are used to uncover these visions? Specifically, are the methods that are currently being used to study visions efficient in large-scale reform efforts? Investigating these questions was the underlying motive of the current study.

Despite its relevance, research on educators' visions shows that efficient instruments that can be utilized on a large scale to publicize these visions have not been readily available. To date, most researchers have relied extensively on case studies and interviews to study vision. However, as Munter (2014) noted, "choosing to employ interviews to assess change in visions of high-quality mathematics instruction undoubtedly requires more time and expertise . . . than using pencil-and-paper assessments or electronic surveys" (p. 616). To further expand the efforts that have been carried out to study vision, efficient instruments aimed at assessing visions, especially on a large scale, are necessary. I contend that self-reporting surveys can provide fast, cost-effective, efficient, and reasonably accurate methods of studying vision.

Additionally, surveys are easy to administer, could be delivered remotely, and can reach a much broader audience or a more substantial number of respondents. Although there have been several critiques about the accuracy of using surveys to collect data (e.g., Hamilton et al., 2003; Spillane & Zeulli, 1999), others have argued persuasively that a thoroughly validated instrument can equally produce accurate measures as other methods of data collection (Groves et al., 2009; Hinkin, 1998). Described in the study is the development of a survey instrument that could be used to characterize teachers' and others' (principals, math coaches, and district leaders) visions of mathematics instruction.

This study uses a mixed methods approach to optimize the development of an instrument to measure visions of math instruction. Mixed methods research is an approach that allows for the collecting, analyzing, and integrating of both qualitative and quantitative measures in a single study (Creswell & Clark, 2011; Creswell & Creswell, 2017; Teddlie & Tashakkori, 2009). Mixing qualitative and quantitative techniques within the same study is advantageous since neither quantitative nor qualitative methods by themselves are necessarily sufficient to gain a full understanding of a given situation (Ivankova, Creswell, & Stick, 2006). This study begins with a broad survey to collect initial data and then focuses, in subsequent phases, on a detailed analysis of quantitative and qualitative data to validate the survey instrument. The rationale for mixing approaches was to ensure "instrument fidelity" (i.e., to outline the "steps taken by the researcher to develop an instrument and to maximize its appropriateness and/or utility" (Onwuegbuzie, Bustamante, & Nelson, 2010, p. 59). Generally, the instrument development and validation approaches in this study are similar to studies and journal briefs that were focused upon instrument development processes (Burić, Slišković, & Macuka, 2018; Hinkin, 1998; Onwuegbuzie et al., 2010; Ross, McDougall, Hogaboam-Gray, & LeSage, 2003).

Contextualizing the Study

The current investigation is a part of an ongoing research and development efforts of a research-practice partnership that includes six campuses of a State university system, the State educational agency, and leaders from State school districts. Through this partnership, teachers, school district and state leaders, mathematicians, and researchers are working together to advance math teaching and learning across the State by examining the resources and structures that promote systemic coherence when implementing state academic standards. A broad goal of the partnership is to improve the process of implementing state academic standards. As part of this more general goal, developing a shared vision of high-quality mathematics instruction (VHQMI) was identified as one of the paramount strategies to address implementation challenges. Hence, the project provided an opportunity to begin the work of assessing the instructional visions of the mathematics education community in the State. As part of this effort, this study reports progress in the development of a reliable and valid instrument for assessing VHQMI.

The partnership holds that surfacing visions of mathematics instruction would be beneficial for the following reasons. First, knowing these visions can inform the resources developed to support the implementation of new math standards. Second, an understanding of the visions held by mathematics educators may help align multiple interpretations and messages across the system and bring more coherence. Third, math teachers and leaders identified a mismatch between their visions of HQMI and those of other education supervisors and leaders as a major challenge in standards implementation specifically, and instructional improvements more generally.

The work described here is an essential first step of seeking to understand whether the development of this new instrument is possible and whether it has promise as a new tool for examining math educators' visions of instruction. The development of a valid and reliable measurement instrument for vision will be useful as one component of a successful large-scale implementation agenda.

This study uses "instructional vision" in a similar manner to Hammerness's (2001) notion of teachers' vision, which she defined as "a set of images of ideal classroom practice for which teachers [and others'] strive" (p. 143). The study reported here takes the position that instructional vision is one aspect of existing or prior knowledge that every individual possesses, and that influences teachers' instruction, and other educators' expectations for teaching and learning. These visions can help explain how individuals make meaning of new stimuli, and the decisions that educational leaders make concerning the types of supports they avail to encourage teacher learning. Because instructional vision is considered as an aspect of prior knowledge, it can be surfaced and productively influenced. Thus, an instrument will be developed to attempt to captures these visions of mathematics instruction.

Statement of the Problem

Despite research that shows the promise of harnessing individuals' perceptions or articulations of their math instruction (Jackson, Cobb, Wilson, Webster, & Appelgate, 2015; Munter & Correnti, 2017; Walkowiak, Lee, & Whitehead, 2015), existing methods of collecting these data are inefficient for large-scale improvement efforts. In looking back at the work that the mathematics research community has done in recent years, the field still needs research in developing instruments that could be used to capture individuals' perceptions or images of practice. Almost 2 decades ago, Ross, Hogaboam-Gray, and McDougall (2002) were already advocating for research on developing such instruments. In their review of the literature on mathematics education reform between the years 1993-2000, they noted that research was needed to develop surveys that could be used to evaluate and track the progress of large groups of educators as they implemented math standards. This study addresses this need by focusing on the development of an instrument that can be deployed at a large scale to capture individuals' visions of instruction.

Purpose of Study and Research Question

The purpose of this study was to develop an instrument that captures practitioners' instructional visions in three specific areas: the role of the teacher, the role of tasks, and the role of classroom discourse. By focusing on mathematics educators' instructional visions, this study seeks to illuminate the types of understandings they have constructed about high-quality math instruction. Consequently, the primary aim of this study is to develop validation evidence for an instrument that can be used to assess mathematics educators' visions of mathematics instruction. Expressly, it assumes that a valid and reliable measurement instrument of educators' visions is possible and will be useful in supporting efforts at improving the implementation of standards at scale. Understanding the development of such an instrument is essential in assessing whether the tool is valid (Groves et al., 2009). Therefore, this study describes in detail the steps that were carried out in developing the survey and seeks to answer the following question: To what extent does the collection of survey items based on an established interview rubric create a reliable and valid measure of VHQMI?

Significance of the Study

The contribution of this study is the development of a reliable instrument that measures mathematics educators' visions of instruction. The developed instrument will be relevant to the academic, practitioner, and policy communities who hold an interest in improving the implementation of reform-based initiatives. This study seeks to make a theoretical contribution to the body of knowledge related to implementation research and math education reform.

This study aims to also contribute to methodological advances in implementation research. Insights gained by this investigation may provide opportunities for those interested in developing valid and reliable surveys to get a road map for what the validation process might look like. Moreover, the development of a survey to assess individuals' visions may play a significant role in lowering research costs, while also providing a quick and reliable progress-monitoring tool for ongoing or new implementation initiatives.

Summary

As mathematics reform continues to be a national and local focus, and innovations for math teaching and learning emerge, an instrument that can be utilized on a large scale to assess math educators' visions of instruction efficiently must take center stage. The motivation for the work reported in this dissertation arose from a need to reliably document the instructional visions among the mathematics education community in the larger project. The goal of surfacing these visions was to inform and support the ongoing efforts of implementing new state mathematics standards.

This research proposal is organized into five chapters, references, and appendices. The first chapter provided introductory information for this dissertation, including the purpose significance of the study, and concludes with a list of definitions used throughout the dissertation. The second chapter presents a review of the literature and presents the framework for this study. Chapter 3 presents a description of the research design, participants, the methods for data collection, and how the data was analyzed. Chapter 4 presents the results and findings, and Chapter 5 presents a summary, recommendations, implications, and conclusions of the study.

Definitions of Terms

The following is a list of terms and definitions that will be used throughout this study.

Standards—Policies that define ambitious and challenging goals for students' learning.

Standard-based instruction—Math instruction based on the "standards" and that incorporates challenging learning experiences, high-level tasks, active co-participation, critical thinking, meaningful discourse, and knowledge construction.

Instructional visions—A set of images of ideal classroom practice.

Implementation—A specified set of activities designed to put policy into practice.

Mathematics reform—Radical shifts in instructional practices and transformations of classroom roles and expectations to promote learning mathematics with understanding for all students.

Practitioner—When this term is used, it refers to teachers, principals, math coaches, and district curriculum leaders.

CHAPTER II

LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

The goal of this study is to develop a survey to examine the instructional visions of math educators. The interest in studying math educators' visions comes from a broader curiosity about how teachers may be supported to improve the quality of math instruction for all students. Specifically, the more general objective of this study is to promote an initiative for fostering conditions for crafting coherence in the educational system to facilitate implementing state academic standards. In this section, I first define instructional vision and review its origins and evolution as a research construct. Second, I present a view of standards-based reform that forms the basis for instructional improvements. Next, I examine how professional and instructional vision impacts the implementation of reform-based ideas and various methods of assessing vision to justify why the use of a survey is warranted. Next, a survey development viewpoint is presented. Lastly, I articulate the conceptual framework that informed the development of the survey in this study.

Defining Instructional Vision

The construct of vision is not new. Goodwin (1994) introduced the term professional vision to describe how members of a professional group perceive and interpret phenomena central to their work. In Goodwin's words, professional vision involves "ways of seeing and understanding events that are answerable to the distinctive interests of a particular social group" (p. 606). He uses the examples that a lawyer's professional vision allows him to make sense of an eyewitness account in court, and an archeologist's professional vision provides a lens through which human history may be studied to illuminate the concept. For educators, a central aspect of the profession is teaching and learning in classrooms. Thus, educators' professional vision involves particular ways of seeing and interpreting that are distinct in the field of education (Van Es & Sherin, 2002).

Since the seminal work of Goodwin (1994) was published, various conceptualizations of vision that are characteristic of teaching and learning have been proposed. For example, Drake and Sherin (2009) examined curricular vision to explain the understanding that teachers develop about the goals of curriculum materials. Others have developed instructional visions as a way to understand what individuals value or view as significant in their work of instruction (Hammerness, 2001), and ways that these visions can help us better understand the development of their work (Munter, 2014).

Hammerness (2001) explored how the concept of "teachers' vision" or "teachers' images of ideal classroom practice" (p. 143) may be used to understand better the perceptions that teachers have of high-quality instruction. Drawing upon analysis of interviews with 16 high school teachers, she characterized vision into three dimensions by examining how the differences in focus (the primary areas of concentration of ideas), range (the span or extent of the focus), and distance (the gap between one's vision relative to one's current practice) of those visions give meaning to teachers' work, as well as a means of measuring how far teachers may be from their ideals. Hammerness suggested that teachers' visions could help account for the assumptions they make, what they learn, and how they may choose to change their practice. In the process, she concluded that vision might function as a powerful tool for helping teachers surface and interrogate their existing knowledge as well as imagine an elaborate pathway to new knowledge.

Munter (2014) investigated teachers', principals', mathematics coaches', and district leaders' evolving perceptions of high-quality mathematics instruction. As part of a longitudinal study of four large districts working to promote instructional practices aligned with mathematics education research, Munter integrated Hammerness's (2001) work, previous findings from research literature, and analysis of more than 900 interviews to develop an interview-based instrument that can be used for assessing and tracking the changes in individuals' visions of instruction. His work led to the Visions of High-Quality Mathematics Instruction (VHQMI) rubrics that delineate the development of individuals' visions along what has been identified in the literature as critical dimensions of high-quality mathematics instruction. The results of this study indicated that participants' articulation of their instructional vision became more sophisticated over the four years in which the project was conducted.

This study defines instructional vision similar to Hammerness (2001) as "a set of images of ideal classroom practice," and takes the position that mathematics educators' images or visions to which they aspire can help us understand how they make meaning of new ideas. Just like Hammerness argued, this study's focus is on mathematics educators' visions of instruction. While there are instances where examples of studies (e.g., Sherin,

2001) will be mentioned and provide insights about the type of learning paths that may result when individuals learn new ideas, this study focuses on developing an instrument to assess vision of instruction. Stated this way, vision is considered as an aspect of prior knowledge that all educators possess, i.e., our ways of seeing the world depend to a great extent on the types of previous experience that one holds. These visions can be surfaced and productively channeled to bring about the desired changes in practice. With these ideas in mind, efforts at instructional change, especially large-scale, might benefit from research that attends to individuals' visions in reform situations.

Standards-Based Reform Perspective

For many decades, the reform in teaching and learning of mathematics has been a significant focus of policymakers. These reforms, frequently packaged as "standards," calls for fundamental changes in what counts as high levels of mathematical teaching and learning for all students. The standards, such as Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989), called for extensive improvements in the way students were taught mathematics, proposing that students should be guided in, "(1) learning to value mathematics; (2) becoming confident in their ability; (3) becoming mathematical problem solver; (4) learning to communicate mathematically, and; (5) learning to reason mathematically" (NCTM, 1989, p. 2). These standards emphasized the use of problem-solving, communication, reasoning, and making mathematical connections as the basis for teaching and learning. Following a decade of focused and coherent efforts to improve mathematics education, NCTM released Principles and Standards for School Mathematics (2000) as an update of its original standards document that proposed five process standards. These standards which refer to the mathematical processes through which students should learn mathematics include, building new mathematical knowledge through problem-solving, selecting and using various types of reasoning and methods of proof, organizing and consolidating mathematical thinking through communication, recognizing and using connections among mathematical ideas, and creating and using representations to organize, record, and communicate mathematical ideas.

Following such proposals, many state and local agencies have exercised control of instruction by devising and implementing their own rigorous and ambitious standards for mathematics teaching and learning and deploying a range of innovations to support the implementation of those standards. Given these ambitious aspirations, significant knowledge is required on the part of teachers and other instructional leaders tasked with helping them. Spillane and Thompson (1997) argued that standards-based mathematics instruction “departs fundamentally from modal practice and notions about teaching, learning . . . and the teaching they envision require . . . pedagogical decision-making that is more complex” (p. 185). To attain ambitious goals, many scholars have outlined new forms of instruction (Franke, Kazemi, & Battey, 2007; Stein, Engle, Smith, & Hughes, 2008; Stein, Smith, Henningsen, & Silver, 2009) and the development of fundamentally different forms of mathematical knowledge (Boston, 2012; Boston & Smith, 2009; Stein et al., 2009; Stein, Smith, & Silver, 1999; Thompson & Zeuli, 1999) that may support the implementation of standards in every classroom.

Knowledge, for example, is no longer viewed as procedural to be delivered in a lecture format; instead, it is something to be developed through productive discussions with other students (Smith & Stein, 2011; Spillane, 2000; NCTM, 2014). According to Hiebert and Grouws (2007), this in-depth knowledge is needed both as teachers set goals for learning, and as they facilitate classroom learning. A typical lesson involves teachers selecting cognitively, demanding tasks that ensure that students understand the context (Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013). Teachers are expected to assess whether the task can be solved in multiple ways and can lead to the kinds of discussions that are projected in given standards (Smith & Stein, 2011). Students collaborate to develop solutions, and through reflection, teachers help students to think through the embedded mathematical concepts (Hufferd-Ackles, Fuson, & Sherin, 2004).

On top of the new knowledge that teachers must develop, they must reassess both their roles and those of their students. As Hill (2001) explained, "teachers should relinquish their role as explainers of mathematical topics and adopt a stance akin to that of a coach who helps guide students through difficult intellectual terrain" (p. 293). Stated this way, classrooms are no longer passive environments but require fundamental changes in typical patterns of teaching and learning (Hiebert & Grouws, 2007).

Despite their soundness, popularity, and motivations, standards-based efforts have faced enormous implementation challenges, revealing that individuals engaged in reform activities have ended up with multiple interpretations of change (e.g., Cohen, 1990; Spillane, 2000; Wiemers, 1990; Wilson, 1990). Because implementing standards depends considerably on learning new forms of instruction, Spillane, Reiser, and Reimer (2002) argued that the challenges witnessed emanate from individuals' perceptions that influence how they notice, then frame, interpret, and construct meanings of given practices. After briefly reviewing the nature of changes in instruction that are called for by standards-based reform, the next section examines the ways that implementation has played out in the context of mathematics education reform

Impact of Vision on Implementation

The role of individuals' conceptions has been a recurring theme in implementation studies. For the last 20 years, this scholarship has examined the ways that individuals' perceptions have produced challenges in implementation (e.g., Spillane, 2000), while also highlighting links of these perceptions to progress in reform ideas (Munter, 2015; Wilhelm, 2014). At its core, the goals of standards-based reform are built on the fact that improvements to learning will require new ways of teaching. As Hill (2001) stated,

> When policies like these standards arrive on the doorstep of local districts, reformers hope, the words will be taken, debated, understood, and, if necessary, transformed into a more usable form . . . Teachers bring to the table practical wisdom, discuss the new ideas from the state, learn from one another, begin the process of personal change, and ultimately provide feedback. (p. 293)

In many cases, implementation researchers have documented that practitioners have faithfully attempted to implement reform ideas, but things have not worked out exactly as anticipated (Cohen, 1990; Spillane & Zeulli, 1999). Studies on implementation of standards have focused not only on assessing the ways that implementation has occurred, but also on deciphering the reason as to why execution happened in the ways that it did

(Spillane et al., 2002). At the forefront are questions concerning the prior knowledge that mathematics educators possess and how this knowledge helps them navigate the complex demands of reform (Sherin, 2002).

Research on the implementation of standards in mathematics education has primarily focused on practitioners' understanding of instructional reform (Hill, 2001; Spillane, 2000; Spillane et al., 2002). Not only are practitioners called on to construct understandings of the new instruction, but they are also expected to build a perspective on how their practice may need to change in the context of those reforms (Spillane et al., 2002). This is a challenging process due in part to the amount of learning that must take place (Hill, 2001). Spillane et al. (2002) argued that prior knowledge is a central factor influencing the understandings that practitioners construct from any given policy. Research has consistently shown that learning to teach in new ways and embracing new ideas is challenging and at times, confusing for those who have to implement them. While some researchers have explained the challenges witnessed in terms of the influence of organizational structures, policy ambiguities, and educators' unwillingness and limited capacity to change (McLaughlin, 1987; Wiemers, 1990), other scholars (Spillane, 2000; Spillane et al., 2002) contend that such views overlook the role that individuals' prior knowledge and organizational contexts play in learning of new ideas. Spillane et al. (2002), for example, argued that "a policy message about changing implementing agents' [teachers and instructional leaders] behavior is not a given that resides in the policy signal . . . rather, the agents must first notice, then frame, interpret, and construct meaning for policy" (p. 393). Spillane et al. (2002) explained three obstacles to the implementation of new ideas. The next sections examine not only these three aspects but also outlines how targeting individuals' perceptions support improvement in instruction.

Interpret the Same Messages in Different Ways

First, Spillane et al. (2002) postulated that every individual has "schemas of knowledge structures" that help in guiding the processing of information. These schemas enable people to construct intuitive models from their past experiences, and then use those models to make predictions about a phenomenon. According to Spillane et al. (2002), what is deemed as a new idea is seen "in terms of what is already known and believed" (p. 395). For example, if a teacher were asked to interpret a practice like "allowing students to discover mathematical insights," a teacher would construct a model of this practice based on their previous experiences or what they envision, and then conclude how effective that practice would be. Some studies have demonstrated that individuals do interpret the same messages about reform in different ways. Hill (2001), for example, found that one district's mathematics writing committee working to support the implementation of the state's mathematics policy understood the reform ideas in different ways than was intended by the writers of the policy. For instance, one teacher's understanding of the word "test" (as an end-of-unit assessment) diverged from the meaning intended by reformers, for whom "test" also includes instructional activities involving reasoning and proof.

Misunderstanding New Ideas as Familiar

Spillane et al. (2002) contended that accommodating, restructuring, and assimilation of new insights into existing knowledge is not solely a matter of encrypting new ideas, but instead, it is a process that involves the restructuring of existing schemas of knowledge. Accordingly, such a process may lead individuals to see new ideas as familiar. In the case study of an experienced middle school teacher learning in the context of a mathematics reform program, Smith (2000) outlined how Elaine, the teacher in the study, struggled through the process of restructuring her knowledge schemas. Smith noted how Elaine had developed the practice of structuring learning so students could experience success, but when she was confronted with the view that students needed to learn through productive struggle, Elaine struggled to accommodate this new practice. In one instance, she seemed to have made some changes, but just quickly reverted to her previous practice, such as posing of low-level questions designed with the sole purpose of ensuring that students felt successful.

Similar patterns have been observed in other studies. Two case studies of teachers' responses to the California Mathematics Framework similarly showed that teachers selectively took up reform proposals (Cohen, 1990; Wiemers, 1990). These teachers understood the new idea from the frameworks as familiar, using practices such as "group work" or "use of manipulatives" to surface mathematics facts and procedures. These teachers fundamentally transformed reform-oriented instruction to fit their pre-existing procedural orientations, resulting in instruction that contained a blend of both old and new practices.

While these studies help explain the little to no change that has historically resulted from standards-based reform efforts, other studies shed light on the modest changes that have occurred. Spillane and Zeuli (1999) found that practitioners made modest or superficial changes to their instructions. In examining the practices of 25 elementary and middle school teachers who had reported familiarity with national and state standards and belief that they were implementing these standards in the intended ways, they found three distinct levels of teachers' practice based on teacher use of productive tasks and commitment to classroom discourse.

Only four teachers exemplified the ideals of standards by supporting meaningful tasks with productive discussions that exposed fundamental mathematical ideas. Ten teachers showed evidence of instruction that was not closely aligned with the intent of the mathematics reform. For example, they used appropriate cognitively demanding tasks but limited discourse during the lesson that resulted in lowering the cognitive demand to procedural knowledge and correct answers. Eleven teachers understood the mathematics standards in ways that involved no fundamental changes, and their mathematical tasks and discourse norms were firmly designed to surface procedural knowledge. These teachers suggested that problem-solving was an occasion for students to apply and practices procedures. They viewed the standards through the lens of their existing practices, and the understandings they constructed failed to mirror the fundamental changes called for by reformers.

Understanding That Falls Short of Meaningful Relationships

Spillane et al. (2002) argued that individuals access and apply knowledge in a specific way, and such a mechanism to access knowledge depends on the "degree of sophistication in that knowledge" (p. 396). Accordingly, while experts may see the deep, meaningful relationship, novices may only notice superficial aspects of the same case.

Depending on how one accesses and applies this knowledge can have negative and positive effects on implementation. For example, Munter (2014) noted that when one's vision of instruction aligns with the quality advanced in reform recommendations, it could create a drive that can act as the basis for educational reform. However, a vision of instruction that lacks a robust pedagogical basis can result in understandings that focus on superficial features and result in the stagnation of practices that are proposed by reform (Wilhelm, 2014).

Wilhelm (2014) used the VHQMI rubrics to determine how teachers' conceptions of teaching and learning mathematics related to teachers' selection and implementation of cognitively demanding tasks. In an examined of 213 middle school mathematics teachers' enactment of cognitively demanding tasks, Wilhelm found that the level of sophistication of VHQMI was significantly related to a teacher's ability to select and maintain cognitively demanding tasks consistently. Specifically, she found that teachers who had firmly held views of instruction that aligned with high-quality math practices were more likely to engage students in cognitively demanding tasks than teachers who had narrower views of mathematical learning.

Munter and Correnti (2017) found somewhat similar results in their longitudinal study of 200 middle grades mathematics teachers in their investigation of the relationship between their instructional visions and changes in practices. Their analyses revealed different patterns of instructional change depending on teachers' sophistication in instructional visions. For example, by year four of data collection, one teacher whose instructional vision at year one had been assessed as being less sophisticated concerning the ways she pressed students for reasoning and facilitation of classroom discourse, experienced only minor changes in the quality of her instruction. However, another teacher whose instructional vision in year one was assessed as highly sophisticated underwent significant changes in her instructional practices. Their findings indicated "instructional visions act not only as filters but also as reflective tools" (p. 21) and therefore concluded that teachers' levels of sophistication in instructional visions might predict how much growth they experience in instructional quality during large-scale efforts to improve instruction.

In another study analyzing ideas about instruction that local district leaders construct from state and national mathematics policy, Spillane (2000) highlighted how the leaders' understandings tended to focus on surface changes (e.g., use of manipulatives) rather than the central themes of the mathematics reform such as "communication" and "reasoning in mathematics." For example, some leaders argued that hands-on activities were vital because they served the purpose of invoking increased student interest and motivation to learn, rather than that these manipulatives could be used as tools for reasoning.

Supporting Instructional Change Through Visions

While the literature on implementation indicates that individuals' visions are a source of implementation challenges, it also indicates that the likelihood of success is increased when studies are designed to attend to individuals' existing knowledge or perceptions. For example, Smith (2000) argued that teachers could change their practices in ways that conform to reform initiatives if they get opportunities to examine their current views of instruction. Smith traced the evolution of one teacher in the context of an ongoing professional learning experience and found that after three rounds of reflecting on her practice, the case teacher made fundamental changes in her practices to conform to high-quality mathematics instruction. For example, she let students work in pairs or groups, facilitated whole-class discussion, used questioning techniques that allowed students to generate multiple solutions, used drawings and sketches as tools to promote problem-solving, provided more time for collaborative thinking, and utilized cognitively demanding tasks.

Several other studies have revealed similar patterns, demonstrating that attending to individuals' visions may support the implementation of new ideas. Sherin (2001) extended the concept of professional vision to her work documenting the evolution of one mathematics teacher's perspective of classroom instruction in the context of a video club. Her teacher, Mr. Louis, had 5 years of teaching experience and participation in the video club involved, observing and videotaping his classroom and meeting weekly with him to watch excerpts of those video recordings. Throughout their 4-year collaboration, Sherin documented how Mr. Louis's interpretation of classroom events changed from a focus on his pedagogical actions towards a focus on student learning. She concluded that these shifts in visions signified that learning had occurred.

In another study, Sherin and Van Es (2009) uncovered similar patterns as they investigated mathematics teacher learning in the context of a 2-year video club in which teachers met monthly to watch and discuss video excerpts from each other's classrooms. Their goal in the study was to explore whether teachers enhanced their professional vision (the ability to notice and interpret significant features of classroom interactions) as they participated in the video club. Over the two years, they observed that teachers' professional visions changed. In related work, Van Es and Sherin (2008) examined the changes in teachers' perceptions in the context of a video club designed to help them learn to notice and make sense of students' mathematical thinking. Their analysis revealed three paths along which teachers' perceptions evolved: direct (single qualitative shift), cyclical (cycling between a broad and narrow perspective over time), and incremental (gradual step by step development). They suggest that understanding the different ways that teachers' perceptions evolve may inform the design of support offered to teachers.

Summary

In summarizing the role of vision or perceptions in the implementation of reform ideas, research suggests that implementation failures are not the result of individuals rejecting new ideas, but result from envisioning such ideas differently. Meaningful changes in instruction might be advanced if individuals' perceptions are surfaced, and perturbed, and become the object of reflecting and thinking differently in ways that align with high-quality education.

Tools Used in Assessing Visions of Instruction

While researchers continue to document evidence the role of vision in instructional improvements and implementation of reform ideas, an examination of this literature suggests a need for more efficient analytical tools for exploring these visions for large-scale efforts. A wide range of tools has been employed to study vision, many of which are researcher intensive. As summarized in Table 1, most researchers have used a review of video records, observations, and interviews to assess individuals' visions of math instruction (e.g., Ho & Tan, 2013; Jackson, Gibbons, & Sharpe, 2017). Very rarely were surveys employed to study vision, and in those cases, the studies were of small scale.

In looking at the summary outlined in Table 1, particular patterns emerge. First, most authors mainly focused on studying small groups of participants, and when this was the case, the authors predominantly used case studies and investigated how teachers' perspectives of instruction changed (e.g., Ho & Tan, 2013; Sherin, Russ, Sherin, & Colestock, 2008). Second, of the large-scale studies, the authors mainly focused on investigating the links between instructional vision and classroom practice (e.g., Jackson et al., 2017; Munter & Correnti, 2017; Wilhelm, 2014). While these results were significant, the methods used to collect the data, especially for the large-scale studies, required substantial resources and extended periods to assess visions.

Table 1

Approaches Used to Characterize Visions of Instruction

Tools	Aspects investigated	References
Review of video records, interviews and classroom observations	Mathematics teachers' development of professional vision, as they participate in a video club (11 participants).	Sherin and Van Es (2009)
Review of video records	The development of a professional vision of classroom practices (two participants).	Ho and Tan (2013)

Table 1

Cont.

Tools	Aspects investigated	References
Interviews	Role of vision in understanding teachers' decisions and experiences in the classroom (two participants)	Hammerness (2003)
Semi-structured interviews	Teachers' views of their students' mathematical capabilities (122 participants).	Jackson et al. (2017)
Review of video record	How teachers apply professional vision in their teaching (one participant).	Sherin et al. (2008)
Survey, interviews & observations	How teachers' vision may be used to better understand the development of their work and careers (16 participants).	Hammerness (2001)
Interviews	Development of teachers' visions of mathematics instruction (18 participants)	Walkowiak, Lee, and Whitehead (2015)
Review of video records, and interviews	How teachers' knowledge and conceptions are related to their enactment of cognitively demanding tasks (213 participants).	Wilhelm (2014)
Observations, interviews, and paper and pencil assessment	Relations between mathematics teachers' instructional vision and knowledge and change in practice	Munter and Correnti (2017)

In contrast, one study used a combination of surveys and interviews to study vision. In this two-year study, Hammerness (1999) collected data through a survey of eighty teachers regarding the content and character of their visions. The eighty teachers were graduates of Hammerness's program or current preservice teachers from two teacher education programs. Hammerness conducted a series of interviews with 16 teachers selected to represent a range of teaching experience, subject matter, gender, ethnicity, and described to have particularly powerful visions. The survey involved responding to a set of prompts in order to express one's vision. For example, one prompt asked teachers to envision and explain in writing, "What you are doing in your ideal class? What is your role? Why?" (p. 93). Although Hammerness was able to show how vision can serve as a guide for directing practice as well as a means of measuring how far teachers may be from their ideals, this approach relied on a survey with all open-responses items, and thus not efficient to be deployed on a large scale. However, this instrument gave some ideas for items but did not precisely fulfill this current study's purposes.

Towards a Survey Methodology

While the previous sections show that the use of tools like interviews and review of videos provide useful avenues for assessing visions of instruction, surveys offer unique and efficient opportunities to gather information, especially for large populations. According to Groves et al. (2009), surveys typically collect data for the purposes of constructing descriptions of attributes that cannot be observed, or that are internalized within persons in a population. This argument focuses attention on the need to ensure that surveys meet specific levels of quality. First, surveys in contrast to interviews do not present opportunities to probe and clarify participants' responses. Second, since surveys are normally conducted in uncontrolled settings of the real world, the results can be affected by these settings (Groves et al., 2009). Given these concerns, part of the task of survey development involves making a broad set of decisions about individual features of the survey with a goal of improving it. According to Groves et al. (2009), considerations have to be made about the types of questions asked, questions and item wordings, sampling methods, and data collection and analysis methods. Since such decisions have the potential to affect the quality of results obtained, it is not surprising that survey methodologists have followed rigorous validation steps to ensure that high-quality surveys are produced (Hinkin, 1998; Onwuegbuzie et al., 2010). For example, Ross et al. (2003) developed a 20-item survey in a study measuring elementary teachers' implementation of standards-based mathematics teaching. In their research, they documented evidence that enabled them to claim that their instrument was reliable and valid.

The first step of this study involved generating dimensions for standards-based teaching based on a review of NCTM policy documents and several empirical studies, resulting in nine dimensions. Teachers were then interviewed to determine which items within each dimension were unclear, and revisions were made to the items. Next, the revised elements were subjected to a measure of reliability in two separate administrations, resulting in quantitative scores that met the threshold for concluding that items within the survey were consistent. To further ensure that the survey was valid, the authors conducted three validity tests.

The first test involved demonstrating predictive validity by showing that survey scores correlated positively with a students' performance in a standards-based assessment. Second, concurrent validity was demonstrated by comparing a small sample of teachers' survey responses to observations in their classrooms. The result suggested a correlation between survey scores and levels of engagement in standards-based practices. Lastly, the authors looked for evidence of construct validity. They identified teachers who reported that they were high users of textbooks that supported standards-based instruction and projected that some teachers would have high survey scores, and others would have low scores. They hypothesized that these two groups of teachers would differ in how they used the standards-based textbook. Findings showed that teachers who scored high on the survey used the text to support standards-based teaching, while those who obtained low scores on the survey transformed the text to promote traditional teaching practices.

The process followed to design and develop the survey in this current study was influenced by the need to develop high-quality surveys. Given that a review of the literature as indicated in previous paragraphs did not find existing instruments appropriate for examining visions on large-scale and with a short timeline, the alternative was to examine literature further to find a theoretical basis and supporting framework for informing the creation of the survey.

Conceptual Framework

The goal of this study was to develop an instrument that can be deployed on a large scale to assess practitioners' visions of mathematics instruction to complement existing instruments and methods. To fulfill this purpose, there was a need to review the literature on mathematics education to find a framework that captures the critical aspects of mathematics classroom teaching and learning, such as those outlined in the introduction of this study. Such a framework would need to address these aspects of math instruction within the survey as a means of assessing the visions that practitioners have for these practices. Though there are a number of syntheses describing high-quality mathematics practices (e.g., Franke, Carpenter, Levi, & Fennema, 2001; Franke et al., 2007; NCTM, 2014; Stein, Smith, Henningsen, & Silver, 2000), only Munter's (2014) Visions of High-Quality Mathematics Instruction (VHQMI) framework captured comprehensive dimensions of mathematics instruction.

Vision of High-Quality Mathematics Instruction (VHQMI)

The VHQMI rubrics were introduced as a framework for the sophistication of individuals' visions of high-quality mathematics instruction (Munter, 2014). The rubric captures the vision of high-quality instruction that can support the successful implementation of new reform ideas or policies that can lead to improvements in students' learning (Chance, & Segura, 2009). The VHQMI framework builds on Hammerness's (2001) research and standard-based practices in mathematics education literature. It delineates levels of sophistication of multiple aspects of high-quality instruction described as critical dimensions of mathematics classroom practice.

As previously mentioned, the VHQMI framework was developed as part of a longitudinal study of four large districts working to promote instructional practices aligned with mathematics education research. Characterizing VHQMI, Munter (2014) developed and tested a set of leveled rubrics to assess educators' descriptions of instruction and their alignment toward research-based descriptions of HQMI.

The VHQMI rubrics assess visions of ideal mathematics classroom practice along four dimensions: the role of the teacher, the role of mathematical tasks, nature of classroom discourse, and student engagement in a classroom activity. Each dimension consists of numbered levels that characterize the increasing sophistication of individuals' instructional visions. Though developed, Munter (2014) reported not using the dimension "student engagement" because information obtained from its interview prompts were similar to those of "tasks" and "discourse" aspects of the rubric. The three dimensions of the rubric are briefly described.

Role of the Teacher

The VHQMI "role of teacher" dimension describes the forms of guidance that teachers provide during instruction. It synthesizes research on teaching practice concerning mathematical activity, classroom norms, and teacher roles in classroom discourse (Clarke, 1997; Franke et al., 2001; Hiebert et al., 1997; Staples, 2007) and contains five levels of sophistication. In moving up the levels in the framework, the descriptors for the role of the teacher are more aligned with a vision of standards-based teaching practices (NCTM, 2000, 2014). Responses categorized at Level 0: teacher as "motivator" describes the role of a teacher as encouraging students to engage in classroom activities. Level 1: "deliverer of knowledge" suggests that the teacher has knowledge that must be imparted unto students through traditional instruction. Level 2: teacher as "monitor" suggests that the teacher monitors as students collaborate to reproduce what they have been taught. Level 3: teacher as "facilitator" suggests that the teacher facilitates student discovery but does not engage students in proactive participation. Level 4: teacher as "more knowledgeable other" suggests that the teachers are proactive, co-participant with students, guiding mathematics learning in meaningful ways.

Role of Mathematical Tasks

A second dimension of the VHQMI rubrics addresses the role of mathematics tasks in teaching. It builds from Smith and Stein's (1998) and Boston's (2012) research on cognitively demanding instructional tasks, and its levels vary in terms of individuals' understanding of the form and function of mathematics tasks in teaching. This rubric contains five levels of sophistication. At Level 0, one does not view mathematical tasks in terms of quality or as a manipulatable aspect of classroom instruction. At Level 1, one describes mathematical tasks in terms of quality, but note that they serve the purpose of reproducing previously learned facts. At level 2, one describes mathematical tasks that are being reform-oriented (e.g., real-world connections, hands-on) but fails to elaborate on their function in terms of high student engagement. At Level 3, one describes mathematical tasks in terms of engaging students in complex thinking, building on prior knowledge, and conceptual understanding, but lacks aspects of generalizations and making connections between ideas. At Level 4, one describes mathematical tasks in terms of its ability to lead to "doing mathematics," i.e., engaging students in explaining their reasoning behind a solution, allowing for insight into the structure of the mathematics, and comparing whether a strategy will work in all cases.

Role of Classroom Discourse

The third dimension of the VHQMI rubric assesses visions that relate to the forms of dialogue in the mathematics classroom. It synthesizes research on classroom discourse in terms of patterns, structure, and nature of talk (Hufferd-Ackles et al., 2004; Thompson, Philipp, Thompson, & Boyd, 1994) and contains four levels of sophistication. Level 1 responses suggest classroom discourse that promotes teacher-to-student patterns of talk or traditional lecturing. Level 2 responses suggest classroom discourse that supports student-student talk regarding small groups. Level 3 responses indicate classroom discourse that promotes whole class discussion with a teacher at the center of talk, and Level 4 responses imply decentralized classroom discourse, i.e., one that invites and orchestrates student-initiated discussion, with the teacher not at the center of conversations.

Summary

The belief that the implementation of reform ideas depends to some extent on vision suggests the importance of considering it when carrying out large-scale reform efforts. The VHQMI framework presents a useful tool for thinking about the critical dimensions of mathematics classroom practices and devising teaching and learning opportunities. The framework has been used successfully in some studies to investigate practitioners' visions of math instruction (e.g., Munter & Correnti, 2017; Wilhelm, 2014). Among one of the suggestions for the application of the VHQMI framework was adapting the prompts and rubric in an electronic survey (Munter, 2014). This study extends the work of Munter through the development of an efficient, reliable instrument that builds from his vision interview and rubric. This rubric constitutes the framework that guided the language of the survey items for this study.

CHAPTER III

RESEARCH DESIGN AND METHODS

The previous chapters introduced the proposed area of research, a description of the study's research problem and research questions, several research purposes, contributions of this study, and a definition of several terms key to the study. A review of relevant literature and the conceptual framework that anchors the study was also outlined. This chapter describes the research design and methodology used to address the following question: To what extent does the collection of survey items based on an established interview rubric create a reliable and valid measure of VHQMI? To answer this question, the next section describes the research context for the study, provides justification for combining quantitative and qualitative methods, and describes the overall instrument development and validation processes.

Research Context

As previously mentioned, the current investigation is a part of an ongoing research and development efforts of a research-practice partnership, including six campuses of a State university system, the State educational agency, and leaders from State school districts. Through this partnership, mathematics educators are working together to advance math teaching and learning across the State by examining the resources and structures that promote systemic coherence when implementing state academic standards.

As part of the research plan, the research team constructed a survey to collect data on teachers and instructional leaders' visions of mathematics instruction, the broader goal is to promote a shared vision of high-quality mathematics instruction (VHQMI) which was identified as one of the paramount strategies to address implementation challenges. The study outlined here reports on the processes that were used to transform the initial set of preliminary items towards a reliable and valid instrument for assessing VHQMI. The instrument will be useful as one component of a successful large-scale implementation agenda.

Mixed Methods Research

Typically, survey instruments are designed for purposes that are tied directly to the specific context in which their usage may be considered valid (Woolley, Benjamin, & Woolley, 2004). Thus, to develop a reliable instrument that assesses VHQMI in other contexts, this study used multiple forms of data that are linked to mixed methods.

Mixed methods research is an approach that allows for the collecting, analyzing, and integrating of both qualitative and quantitative measures at some stage in the research process (Creswell & Clark, 2011; Onwuegbuzie et al., 2010; Teddlie & Tashakkori, 2009). Mixing qualitative and quantitative data within the same study is advantageous since neither quantitative nor qualitative methods by themselves are necessarily sufficient to gain a full understanding of a given situation (Ivankova et al., 2006). A mixed methods approach involves collecting quantitative and qualitative evidence, allowing for a robust analysis of data and collection of validity evidence (Creswell & Creswell, 2017; Nicholas, Hess, & Purzer, 2015).

According to Groves et al. (2009), to claim that a psychometric measure is valid, a researcher must demonstrate a well-articulated theoretical foundation and a thorough development process. This process takes into account the meanings attributed to a measurement instrument within a particular context. This study built on literature about professional vision (Goodwin, 1994), and more specifically, instructional vision (Hammerness, 2001; Munter, 2014; Sherin, 2001), while at the same time accounting for the meanings that participants attributed to the survey they were encountering. The survey was developed iteratively considering quantitative and qualitative evidence available with a goal of making improvements to the instrument being developed and also provide a means of gather evidence of the survey's validity in the given context of the study.

The specific design for this study was dialectic data integration to mixed methods (Nicholas et al., 2015). Dialectic data integration is a method in which the researcher uses both sources of information from qualitative and quantitative approaches to directly inform the other in a spiraling process since each data collection technique acts as a carrier for unique information (Greene & Caracelli, 1997; Nicholas et al., 2015). For example, in this study, by use of conversations, one could ascertain participants' interpretations of their scores, i.e., figure out why they responded in the survey in the way they did. Thus, following a dialectic approach, dialogue became the basis for survey validation, as the survey development proceeded through a process of negotiation between quantitative evidence and qualitative evidence. Interpretations of both quantitative and qualitative data were considered at the concept level (whether the concept as framed was applicable in the educators' context) and item level (whether the survey items were representative of instructional visions and accurately interpreted in the participant's contexts; Nicholas et al., 2015). Additionally, various aspects of validity were considered during data integration in order to give a more holistic understanding of the survey in the participants' context and make further recommendations for improvements.

The Overall Research Approach to Instrument Development

There are several proposals for assessing the psychometric soundness of survey instruments. Several studies on survey instrument development show that thorough validation processes are necessary to establish that survey items measure the construct they purport to measure (e.g., Ross et al., 2003; Schmidt et al., 2009; Woolley et al., 2004). These studies show that creating valid surveys typically follow three general phases: (a) defining the construct, and specifying its domain, (b) testing the scale by designing and conducting studies to examine the extent to which items measure the domain empirically, and (c) using confirmatory studies to finalize the scale. Though these phases do not address all aspects of survey design and are not the only approaches to develop high-quality surveys, they synthesize multiple survey development techniques, organizing them into a cohesive process for developing reliable and valid surveys.

These three phases formed the basis for the approach used in this study. The instrument development process reported here involved the use of literature on visions and the use of local experts concerning the content and language within the survey to generate content. Additionally, the process involved pilot testing, and multiple cycles of development that included feedback analysis, field-testing, validation strategies that included the use of appropriate statistical analysis to confirm the factor structure and reliability of the survey, and revisions. Figure 1 provides an overview of the entire process and the purpose of undertaking each phase. The sections that follow describe in greater detail the processes that were used to develop the instrument, including the aspects of validity evidence that each phase utilized.

Pilot Phase: Pilot Item Generation and Pilot Testing (Spring 2017 – Fall 2018)

The pilot phase of instrument development started in Spring 2017 and involved the creation of items to assess VHQMI of mathematics educators. As mentioned earlier, this item creation phase was part of a broader investigation examining strategies for promoting systemic coherence when implementing state academic standards. In Spring 2017, the partnership conducted a statewide survey of mathematics educators' advice networks, uses of research evidence in practice, and instructional vision. A link to the online survey was sent through email to all mathematics educators subscribing to the state education agency's database. Approximately half of the participants were randomly assigned to questions related to instructional vision, while the remaining half were assigned questions about research use. All participants completed questions addressing professional advice networks.

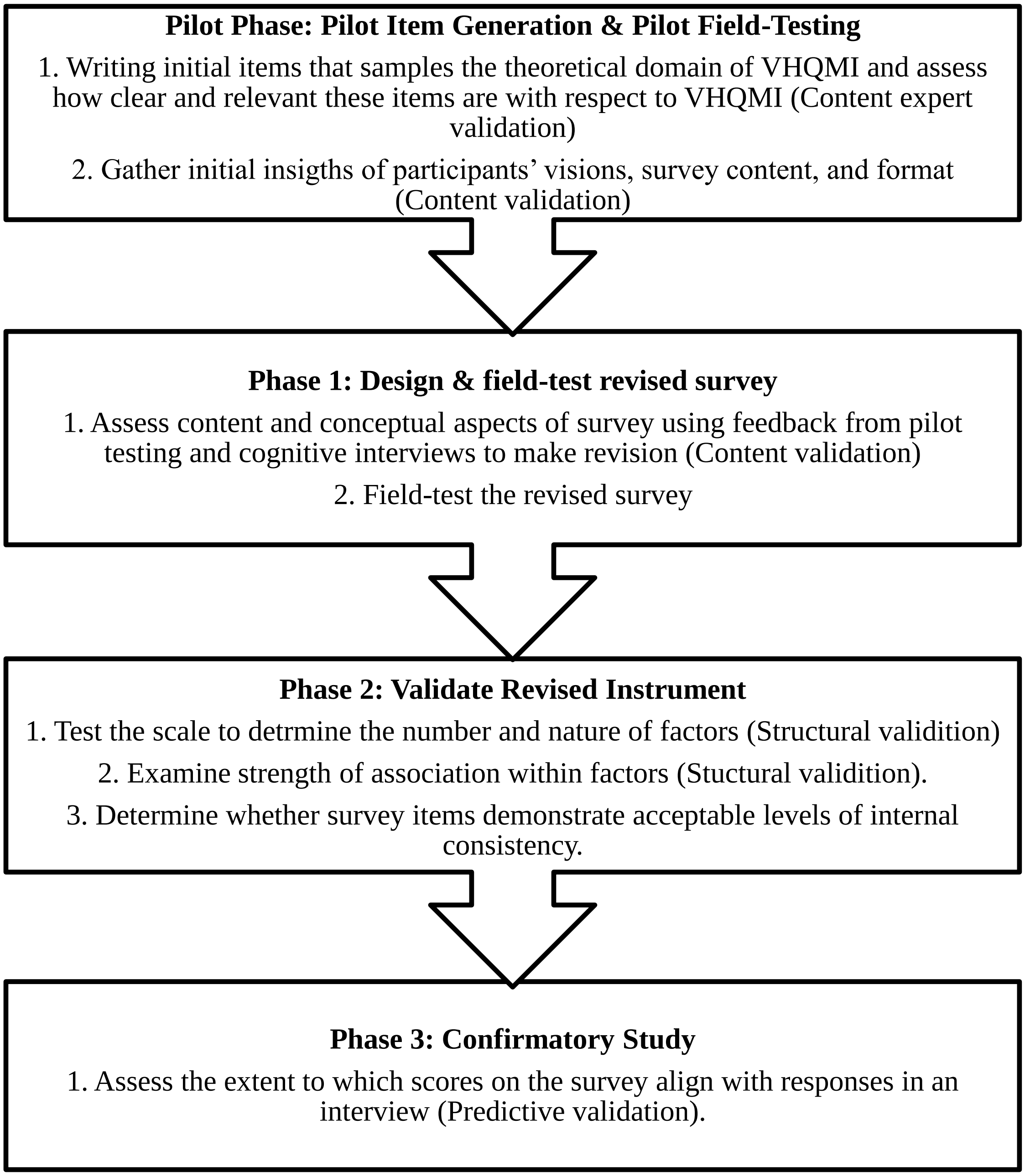

Figure 1. Instrument Development and Validation Process.

Evidence of content validity was assessed in two stages. First, the research team wrote items to reflect different levels of sophistication from Munter's rubric (2014). Following Munter's interview prompts, three survey questions were developed to address each dimension of the VHQMI rubric. The questions asked respondents to choose 2-3 statements from a bank of 8-15 statements that best characterized what they would like to see in an ideal mathematics classroom related to the role of teacher, mathematical tasks, and classroom discourse. The statements were adapted based on the definitions and descriptors of each level across the three dimensions of the VHQMI rubrics. The researchers carefully reviewed each statement in the light of the composition of the VHQMI rubric and then consistently worded the items to more closely align with the language in the VHQMI rubric.

Second, the survey also included an open-ended section asking the respondents to provide any additional insights into their answer choices. This section was necessary to ascertain whether the conceptualization of VHQMI and its content matched how the respondent thought about it. In other words, would respondents identify the practices that were defined in the pilot survey, and would they also use the same language to describe math instructional practices? To answer these questions and ensure that the construct VHQMI was defined from multiple perspectives, this section collected data directly from participants who closely resembled the target population of interest. This section provides an opportunity to get feedback on various aspects of the survey, including the clarity of the instructions, wording of items, format of the survey, and content.

In summary, the data collected provided an initial sense of participants' visions and a primary means of beginning the validation work of the study. Also, respondents were asked to report demographic information, such as educational role, years of experience, currents mathematics assignments, grade bands, tertiary qualifications, and school district. See Appendix A for the pilot survey.

Participants

The pilot survey was administered to a large group of mathematics educators across the state of North Carolina (*N* = 20,000). The reasons for using a large sample was to gather as much insight as possible, especially for the open-ended section, and also be able to get an appropriate number of respondents for a follow-up interview on various aspects of the survey. In total, there were 1,933 survey responses (10% response rate). Seven hundred sixty-nine respondents completed the three instructional vision questions. Of these, there were 641 teachers, 53 instructional coaches, 43 principals, and 32 district curriculum leaders. The response rate is acceptable for a web survey with no pre-notice or remainder follow-up for non-respondents (Kaplowitz, Hadlock, & Levine, 2004). The distribution of responses by educational role, experience, and geographic region gives confidence that the sample participants were representative of math educators across the State.

Data Collection and Analysis

The pilot survey consisting of 35 items addressing VHQMI was sent through email as a link to the State Education Agency's database. The survey consisted of sections for informed consent, and there were assurances that their responses would be confidential. Another section asked participants to indicate whether they would be available for follow-up interviews. After data collection, quantitative analysis was performed on the three dimensions of the survey by generating frequency tables that highlighted the participants' preferences of the instructional practices that they would like to see in their ideal classrooms. These results provide an initial sense of participants' instructional vision and a means of beginning the validation work of the survey.

Phase 1: Design and Field-test Revised Survey (Spring-Summer 2018)

Given that researchers may bring their perspectives and biases in developing a survey, several authors (e.g., Onwuegbuzie et al., 2010) postulate that once these once items have been generated, it is necessary to ascertain whether the conceptualization of the construct matches how respondents think about it. It is imperative to obtain the perspectives of a sample of the target population to identify the semantic and conceptual aspects of a survey that may affect how participants respond (Hinkin, 1998; Ross et al., 2003). A revision of the current survey items was completed in the Spring of 2018 and involved a two-pronged process. First, it involved reviewing respondents' open-ended responses from the pilot survey. Second, cognitive interviews provided an efficient way to get respondents' input about their conceptualization of the pilot survey. The goal of this stage was to gather evidence to inform revisions to the pilot items. The two steps are described next.

Review of Open-ended Responses

One of the first aims of the pilot testing was to gather information from respondents that may not have been captured by the survey questions. Apart from obtaining preliminary quantitative data about participants' VHQMI from their survey responses, the open-ended section included at the end of each dimension of the survey section provided another avenue to gather feedback about the pilot survey. Responses from the open-ended section were reviewed and categorized according to the aspect of instructional vision they addressed. Specifically, a review of these open-ended responses helped in identifying issues with survey mechanics, the wording of items, the format of the survey, and content. The goal of reviewing the open-ended responses was ascertaining the extent to which the responses in the survey represented respondents' VHQMI, collect evidence for revisions to the pilot survey, and inform the design of the cognitive interviews.

Participants

All participants who completed the vision section of the survey elected to participate in the open-ended section. This section invoked an excellent response with over 250 statements made in the role of teacher dimension and over 150 responses both in the role of tasks and discourse dimensions, respectively.

Data Collection and Analysis

Data were analyzed using summative content analysis, which involved counting and comparing keywords or content, followed by the interpretation of the underlying context (Finfgeld-Connett, 2014). The open-ended sections were therefore coded for (a) instances when respondents validated the content in the survey, (b) instances when participants mentioned issues with the structure/format of the survey, and (c) instances when it was directly evident or could be implied from the participants feedback that they were suggesting new items or were restating the survey items using a language that made more sense to them. Responses coded for this last characteristic were used at this stage. The remaining coded responses were saved for use at an advanced stage of the instrument validation process because there was a need to test and revise the pilot items that had been created by the research team before field-testing. The analysis was used to make some revisions to the pilot items and also informed the development of the cognitive interview protocol.

Cognitive Interviews

After pilot testing a survey, it is important to examine potential respondents' understandings of the items and if their interpretation is consistent with what the survey designer intended (Karabenick et al., 2007). Cognitive interviews have proven to be an effective and efficient method of collecting such evidence (Ross et al., 2003; Schwarz & Sudman, 2012; Willis, 2004). The primary goal of cognitive interviews is to give researchers a window into respondents' cognitive processes when answering survey questions, thus assisting in identifying problems with wordings, terminology, comprehension, and potential solutions at each step in this process.

In the present study, cognitive interviews were useful in testing whether the entire pilot survey was working as intended. The interviews were not only crucial for flushing out any ambiguous items and revising them but were also used to thoroughly examine and improve on other aspects of the survey that were identified in the open-ended section as problematic, such as its design. Several cognitive techniques, including a combination of think-aloud and verbal probing (Groves et al., 2009), were used as tools for testing these aspects of the developing survey. These interviews, combined with the findings from the open-ended responses, were the beginnings of the critical work of ensuring that a valid and reliable instrument would result.

A cognitive interview protocol (see Appendix B for the protocol) was developed to address in a more detailed manner what could not be captured by the open-ended responses. These interviews targeted both the issues that emanated from the review of the open-ended responses as well as the soundness of the instrument as a whole. The interviews included a broad range of foci, including evaluating individual sections of the survey, analyzing items, and assessing ease of responding to the survey. Since this instrument under development with intentions for a large scale to evaluate instructional vision, these interviews were particularly important, primarily to ensure that as many issues as possible with the survey were identified and addressed.

Participants. A sample of 12 participants for the cognitive testing was purposefully drawn from teachers and instructional leaders who were known to have high-quality visions. Additionally, they were selected to form a representative section of the mathematics education community, including grade band (K-5, 6-8, and, 9-12), different role groups (teachers, principals, instructional coaches, and district curriculum leaders), and geographic region. The diversity of this sample based on an educational role, experience, and geographic region give confidence that the sample participants reflected the target population of math educators across the State. This sample was assembled in conjunction with the leaders of the broader project.

Data collection and analysis. A cognitive interview protocol (see Appendix B) was used for data to collect data. The cognitive interviews lasted approximately 10-20 minutes and were video recorded. Some of these interviews were conducted through face-to-face meetings, while some were performed using the online video conferencing service. A variety of techniques were used in the interviews. In the first instance, the participants were asked to read the introductory part of the survey and then asked to comment on how they understood that section, i.e., describe in their own words what they understood the survey was trying to assess. Asking participants to explain how they understood the three parts of the survey followed. Verbal probes were used to ask participants to verbalize their thoughts as much as they could as they provided feedback about the specific questions identified in the protocol. The participants were further instructed to identify any items where the vocabulary or wording of the items was not sufficient enough to provide them with all the necessary information required to make a sound judgment.

Analytic memos were created based on detailed notes and video recordings from each cognitive interview. These memos contained a summary of each participant's responses to the questions in the interview protocol, as well as overall impressions or insights of the pilot survey. The memos also included a section to comment on critical items to look at, and even parts of the survey to consider for revisions carefully. See Appendix C for an example of an analytical memo.

Data analysis involved looking at these memos and coding them into the following categories: (a) clarity noted in the items, (b) minor problems in the items, and (c) significant problems in items, and d) feedback on survey design. A summary of these codes was developed. More specifically, a review of responses was done item by item across respondents, documenting items that appeared to work well, and highlighting any problems identified through the interviews. Any issues were coded and explored to determine their potential sources (e.g., item clarity, comprehension, the format of the survey). In concert with the research team, items were improved based on suggestions from the respondents. Recommendations about the items to drop were also made at this stage of the study. Findings from this phase of data collection informed revision to the pilot survey before the field-testing of the revised instrument.

Participants for the Field-testing of the Revised Instrument

The revised survey was administered to mathematics educators across the state, and a total of 189 complete responses were obtained from the respondents. This group of respondents consisted of 140 teachers, four principals, 22 district curriculum personnel, and 23 school-based coaches. The survey was completed in 15-20 minutes. Statistical analyses of the data obtained from the field test are described in Phase 2 of the study.

Phase 2: Validate Revised Instrument (Summer-Fall 2018)

This phase involved the analysis of quantitative data. The goal here was to gather construct-related validity (i.e., structural validity), the nature of the association, and the reliability of the items. To accomplish these goals, scholars recommend exploring the factor structure, examining correlations within the factors, and assessing the consistency of the items (Groves et al., 2009; Tabachnick, Fidell, & Ullman, 2013). The data collected from the field-tested items were examined for statistical properties. In the next section, how the statistical methods that were utilized to gather validity and reliability evidence to improve the instrument are discussed.

Exploratory Factor Analysis (EFA)

Exploratory factor analysis is a tool for empirically examining the interrelationships among variable items, and determining meaningful clusters of items that form a coherent set (Ross et al., 2003; Schmidt et al., 2009; Woolley et al., 2004). This process summarizes patterns of correlations among observed variables, allowing for items with low inter-item correlations to be redacted from the set. EFA is based on the fundamental assumption that holds that responses to all items belonging to the same factor will have similar average correlations. Low correlations indicate items that may not be drawn from the appropriate domain and are usually deleted. Items that simultaneously measure multiple factors, in which case they could be poor indicators of the desired factor, are also customarily considered for elimination (Tabachnick et al., 2013). EFA was an essential process in establishing the number and nature of factors that underlay the given set of items. Although one of the main goals of performing EFA is to reduce the number of items in a survey so that the remaining items in the examination can best explain each of the factors. The primary goal for performing EFA in this study was to identify useful items and those that were questionable items so that they could be evaluated and revised.

EFA utilizing principal axis factoring (PAF) with quartimin rotation and Kaiser normalization was conducted on the data obtained from each dimension of the revised instrument. PAF makes no assumptions about given data, such as multivariate normality, which is almost impossible to meet in practice. Furthermore, the quartimin rotation was utilized because of the assumption that the factors generated would be correlated (Tabachnick et al., 2013; Worthington & Whittaker, 2006).

Next, the rotated factors were examined for items that did not load strongly on any of the factors or cross-loaded into both factors or that theoretically fell into unintended places. At this stage, inappropriately loading items were deleted, the factor analysis repeated until a definite factor structure matrix that explained a high percentage of total item variance was obtained. The choice for the cutoff for the size of loading was based on Comrey and Lee (1992), who suggested that loadings above 0.71 were considered excellent, 0.63 very good, 0.55 good, 0.45 fair, and 0.32 poor. Thus, for this analysis, the minimally acceptable factor loading was 0.45. After determining the factors, each of the items was carefully examined based on the original intent of the items. That is, the instructional quality of the items was considered to determine whether the item corresponded to the category suggested by the factor analysis. Once the significance of the factors was established, items within the same factor were examined in comparison to each other in terms of instructional quality, and to the loadings, and named accordingly. Additionally, each factor was either assigned another label of "low vision" or "high vision," respectively, because the items within them theoretically fitted the criteria.

Assessing the Correlation Within the Sub-dimensions (Factors)

Following the EFA, Spearman correlation analyses were used to assess the strength of association between the items in each sub-dimension (factor). For example, the goal was to determine the strengths of association between an item like "teachers are asking questions that lead to further investigations or revised conjectures," and "teachers are asking conceptually oriented questions to uncover students' thinking," which loaded into factor 1 of the role of discourse dimension. A positive correlation coefficient would indicate a positive relationship between the two items, and specifically, the higher the value of correlation, the stronger the strength of association between items in the same sub-dimension (de Winter, Gosling, & Potter, 2016).

Assessing Internal Consistency (Reliability) of Items

In addition to factor analysis and correlation tests, the remaining items in each dimension were subjected to tests of reliability to establish the extent to which each item correlates with the rest of the items being measured. A widely acceptable technique of establishing reliability is using a statistical measure known as Cronbach's alpha (Cronbach, 1951). Cronbach's alpha reliability coefficient typically ranges between 0 and 1, with values closer to 1.0, indicating more internal consistency of the items a given domain (Groves et al., 2009; Hinkin, 2003; Ross et al., 2003).

George and Mallery (2003) provided the following rules of thumb for the making decisions about alpha values to use in a study: "_ > .90 – Excellent, _ > .80 – Good, _ > .70 – Acceptable, _ > .60 – Questionable, _ > .50 – Poor, and _ < .50 – Unacceptable." The accepted level of reliability depends on the purpose of the research project. For the early stages of a study, alpha reliabilities of 0.50 to 0.60 are acceptable, but the ultimate goal is to increase the reliabilities to beyond the 0.70 levels (Nunnally, 1967). Since instrument development is at an infancy stage, the target level of minimum reliability was set at 0.60 for assessing the internal consistency among the items within each domain

Survey development research has demonstrated that the size of alpha is dependent upon the number of items in a given domain (Cortina, 1993; Gliem & Gliem, 2003; Hinkin, 2003). Although a rigorous validation process was pursued in this study, the low alpha value was anticipated in the role of task dimension due to the presence of very few items. To determine the internal consistency, the items that survived the EFA were subjected to a Cronbach's alpha test. Items that showed lower correlations with other items within a given domain were considered as candidates for deletion and revision, only if, by deleting them, the value of Alpha would increase.

Phase 3: Confirmatory Study (Spring 2019)

The primary purpose of this phase was to increase confidence in the soundness of the measure being developed by gathering further evidence of validity (i.e., predictive validity). Survey methodology shows that comparing the result from an administration of a survey to subsequent analysis of data derived from another source is an excellent means of evaluating the validity of a survey (Groves et al., 2009; Hinkin, 1998; Ross et al., 2003). To achieve the goal of establishing the relationship between respondent responses on two different methods of data collection, a hypothesis was generated to test whether participants' responses from the survey were similar to sentiments expressed in another form of the data collection procedure. It was projected, for example, that such a comparison would show that educators' expressions of their vision on the survey would possibly match their articulations in the interview. The follow-up interview (see Appendix D) was adapted from Munter's interview protocol (Munter, 2014).

Participants

The sample participants for this stage (*n*=5) were purposefully selected to represents participants with known high-quality visions, different grade bands (K-5, 6-8, and 9-12), and various role groups (teachers, principals, instructional coaches, and district curriculum leaders). The reason for selecting these participants was to reflect the intended target population.

Data and Analysis

Both the responses to the revised items an interview protocol were used for data collection for this part of the confirmation study. Some of the data used in the confirmation study were already collected during the field-testing of the revised survey, while others were obtained during this phase. The interviews ranged from 45-60 minutes and were video/audio recorded. The analytical notes that were created contained a summary of participants' responses to the interviews and researcher reflections. Developing these memos involved a focus on statements that participants made about (a) actions a teacher should be doing for instruction to be considered to be high quality, (b) The types of tasks that the teacher would be using for teaching to be regarded as high quality, and (c) what classroom discussion would look and sound like for instruction to be of high quality. These statements were mapped onto the dimension of the VHQMI rubric (2014) and examined to understand the visions that the participants were articulating.

The data analysis involved comparing participants' scores from the field-testing and with analytical memos that were developed for the interviews. If respondents articulated similar sentiments on the survey and the interview, I conclude that the instrument had appropriate levels of predictive validity.

CHAPTER IV

FINDINGS

This study was designed to address the question, to what extent does the collection of survey items based on an established interview rubric create a reliable and valid measure of VHQMI? To investigate this question, the development and validation of the instrument in each phase of the study were explored. Different types of validities were examined in each phase, including content validity, construct validity, predictive validity, and reliability that are recommended in developing psychometrically sound instruments (Groves et al., 2009; Onwuegbuzie et al., 2010). In this chapter, I describe the findings relative to these aspects, while employing both quantitative and qualitative approaches in analyses of the data.

Pilot Results

While a claim can be made that the results of the pilot phase provided an initial sense of participant's views on the three dimensions of mathematics instruction, these results are mainly included for descriptive purposes. Given the fact that participants were able to make selections of their preferential instructional practices implied some level of understanding of the survey content in the participants' context. The pilot results suggest similarities and differences in how different educational role groups envisioned the pilot items. Table 2 summarizes the frequency of statements that fall into each level of the VHQMI rubrics.

Table 2

Frequency of Statements for Each Dimension of the VHQMI Rubric

	SBIC (*n* = 53)		DCL (*n* = 32)		Principals (*n* = 43)		Teacher (*n* = 641)	
	N	%	*N*	%	*N*	%	*N*	%
Role of the Teacher								
Level 0	29	18.24%	20	20.83%	37	28.68%	478	24.86%
Level 1	19	11.95%	8	8.33%	10	7.75%	321	16.69%
Level 2	42	26.42%	30	31.25%	39	30.23%	544	28.29%
Level 3	38	23.90%	14	14.58%	26	20.16%	286	14.87%
Level 4	31	19.50%	24	25.00%	17	13.18%	294	15.29%
Mathematical Tasks								
Level 1	10	9.43%	1	1.56%	6	7.14%	154	12.18%
Level 2	23	21.70%	14	21.88%	27	32.14%	455	36.00%
Level 3	53	50.00%	34	53.13%	37	44.05%	432	34.18%
Level 4	20	18.87%	15	23.44%	14	16.67%	223	17.64%
Classroom Discourse								
Level 1	9	5.70%	7	7.29%	9	7.50%	143	7.66%
Level 2	31	19.62%	23	23.96%	24	20.00%	405	21.69%
Level 3	59	37.34%	32	33.33%	41	34.17%	682	36.53%
Level 4	59	37.34%	34	35.42%	46	38.33%	637	34.12%

Note. The N column is the number of responses selected at least once by each participant.

Analysis indicates that under the “role of teacher” dimension, the majority of statements picked by all participants represented Level 2, “teacher as a monitor,” with coaches making 26.42% of selections, DCLs 31.25%, principals 30.23% and teachers 28.29%. For the “mathematical tasks” dimension, SBICs, DCLs, and principals made most selections that fell into a level 3 category, representing 50.00%, 53.13%, and 44.05%, respectively, whereas most selections (36.00%) made by teachers were at level 2. In terms of the “classroom discourse” dimension, the majority of all participants’ selections across all educator roles fell into a Level 3 and Level 4 category. This represented 37.34% of statements from coaches, 33.33%, and 35.42% from district curriculum leaders, 34.17% and 38.33% from principals and 36.53%, and 34.12% from teachers. These results began to paint a portrait of what mathematics educators envisioned as representative of practices they deemed as high quality. These findings suggested that the contents in the initial survey reflected the instructional practices that participants encounter in mathematics learning environments. The results that follow also talked to this aspect.

Table 3 shows the mean VHQMI scores for each mathematics educator role concerning the three dimensions of the VHQMI rubric. As indicated in the table, mean scores for responding State mathematics educators across dimensions are between 2 and 3, regardless of educators' role. DCLs had both the highest mean role of the teacher scores (2.78) and mean mathematical task scores (2.41), while principals had the highest mean classroom discourse scores (2.49).

Table 3

Mean VHQMI Scores by Educators' Roles

	Role of the Teacher			Mathematical Tasks			Classroom Discourse		
Role	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
SBIC	53	2.49	(0.91)	53	2.32	(0.61)	53	2.45	(0.46)
DCL	32	2.78	(0.72)	32	2.41	(0.56)	32	2.42	(0.49)
Principals	43	2.43	(0.81)	42	2.21	(0.61)	40	2.49	(0.46)
Teachers	641	2.33	(0.85)	632	2.12	(0.75)	625	2.39	(0.49)

Teachers had the lowest mean scores across all dimensions of the VHQMI rubric. Taken together, the mean dimension scores indicated that mathematics educators in different educational roles might share a vision for the role of the teachers as a “monitor” who pays close attention to students as they collaborate to reproduce what they have been taught. On the one hand, educators may share views that tasks serve to make mathematics more relevant and engaging to students through the use of hands-on activities and connections to a real-world context. However, on the other hand, educators may fail to realize the functions of the task in providing other meaningful contexts, such as complex thinking, or making connections between strategies. These mean scores also suggest that educators may share views that classroom talk involves practices that engage students in learning by sharing their strategies in small groups or involves prompting and pressing by teachers to show the mathematical steps in their work.

Analysis of Variance Tests of Significance indicated a statistically significant difference in the mean scores for both “role of the teacher” dimension (F (3, 765) = 3.31, p = 0.02), and “mathematical tasks” dimension (F (3, 755) = 2.78, p = 0.04). No significant differences among means scores were found for the “classroom discourse” dimension (F (3, 746) = 0.73, p = 0.54). A follow-up, pairwise comparisons of mean scores showed statistically significant differences between the groups of DCLs’ and teachers’ in ways in which they view the “role of a teacher” (p = 0.02) (see Table 4). Other comparisons showed no statistically significant differences in mean scores.
A closer look at the data may indicate the extent to which vision is shared across educational roles. Except for the “role of the teacher” dimension that shows a broader

range of differences between the mean scores, that “mathematical tasks” dimension shows much smaller differences between the means scores, and the “classroom discourse” dimension shows almost negligible (or minimal) differences in mean scores.

Table 4

Multiple Comparisons of Mean VHQMI Scores

		Role of the Teacher		Mathematical Tasks		Classroom Discourse	
Role	Role	Diff.	*p*-Value	Diff.	*p*-Value	Diff.	*p*-Value
DCL	Teachers	0.45	(0.02*)	0.29	(0.13)	0.03	(0.62)
DCL	Principals	0.35	(0.29)	0.19	(0.21)	0.07	(0.94)
DCL	SBIC	0.29	(0.42)	0.09	(0.68)	0.03	(0.81)
SBIC	Teachers	0.16	(0.57)	0.20	(0.89)	0.06	(0.99)
Principals	Teachers	0.10	(0.89)	0.10	(0.84)	0.10	(0.99)
SBIC	Principals	0.06	(0.99)	0.12	(0.95)	0.04	(0.99)

Note. $^{*}p < 0.05$. Tukey’s HSD test showing differences between VHQMI mean scores

In summary, the pilot phase was instrumental in capturing how teachers currently articulate perceptions of high-quality practice. On average, VHQMI scores appeared to differ in predictable ways, given the specific role that each educator plays in the broader educational system. There would be a great cause for concern if those tasked with supporting the work of teachers would end up with lower mean VHQMI scores than teachers. The results showed that this was not the case since the mean VHQMI scores for DCLs, SBICs, and principals were higher than teachers’ scores across all dimensions. Based on the findings that the items sampled the initial participant’s visions, it was safe to proceed to other phases of the study to continue to get a better understanding of the overall instrument.

Phase 1: Design and Field-testing of the Revised Instrument

The analysis of the open-ended section and the use of cognitive interviews marked the beginning of revisions to the pilot survey. These two analyses played a crucial role in the examination of content validity, permitting for revision of items that were deemed to be conceptually inconsistent, or did not necessarily sample the content domain of VHQMI.

Open-ended Section

The results of the open-ended section included in the pilot survey produced a wide range of information. However, for this research, there was a focus only on the statements that referred to specific math instructional practices and not on statements about other aspects that support classroom teaching, like behavior management, and grading. Following this line of analysis, the statements from the participants indicated that the survey had evidence of validity and also provided vital information for the instrument development processes.

First, the open-ended section provided some initial assessment of content validity (i.e., face validity & item validity). Specifically, the responses were used to judge whether the items captured their understanding of ideal mathematics classroom instruction. The majority of respondents validated the contents of the pilot survey as representative of the forms of math instruction that they use in their classrooms. This was evident from the numerous ways in which respondents connected the contents of the survey to their views of instruction. Some responses emphasized the teacher's role in facilitating learning. For example, "A teacher should facilitate learning rather than spoon-feed it. Students must learn to communicate, debate, their ideas while working towards consensus." Other responses suggested that the role of the teacher was to provide direct instruction then monitor students' progress. For example, "I believe you should present the information and then allow the students to work collaboratively on problem-solving skills." For responses about tasks, some respondents noted that the role of tasks was to create opportunities for students to learn through hands-on activities. For example, one respondent indicated, "as a teacher, I would like my students to have a hands-on experience/activities of what they have learned since I know learners love to get a hands-on experience to better understand the concept." Other responses under this category indicated that tasks should support multiple solutions paths. For example, one respondent noted that "Since they are many different ways to go about solving a math problem, it is important to validate these approaches and give students meaningful tasks to be engaged in." For the statements that fell under the role of classroom discourse, some responses focused exclusively on the teacher's explanation in terms of clarity and mathematical correctness. For example, one participant stated,

> I believe that before the students are questioned about math strategies, they need to be taught how to solve problems first. After they have been taught how to solve problems, questioning them on their ability to solve problems correctly is a great idea. But, it is difficult to question students and expect them to come to a reasonable solution if they have never been taught the information. We are called teachers because we are meant to teach, not question.

Other responses to discourse addressed several aspects of the teacher's role in facilitating student-to-student discourse, eliciting, and scaffolding students' ideas. For example, a participant noted, "Students should be talking about math, questioning the ideas of others as well as their own, critiquing strategies, and making generalizations and revising when necessary."

Including an open-ended section at the end of the survey was vital in ensuring the items that were created were not only linked to the theoretical content knowledge base on VHQMI (Munter, 2015) but also took into account emic perspectives. Apart from the few issues that are addressed in the next paragraphs, the feedback obtained suggested that the individual items as written were relatively developed enough and satisfied a broad range of the perspectives on math classroom teaching and learning. Consequently, there was no need to rewrite a new set of items that underlie VHQMI. This is not to say that the survey could not be improved by adding additional items, and, as will be explained later, there was a need to add some additional items. The evidence above supported the efforts at proceeding with the next steps in the development of this new instrument.

Second, the responses indicated a weakness in the overall design of the pilot survey. The main issue here was that picking only two or three statements from a list of 8-15 classroom practices was somewhat limiting. There were several instances where respondents talked about the overall design of the survey. For example, one participant noted that "there are other strategies in the list that I would have picked too," and another participant stated, "I think a number of the above choices would be seen in an ideal mathematics classroom, but I was limited to 3." It was evident from these findings that respondents valued more than the 2-3 statements the survey was asking them to choose from the bank of instructional practices. Following these findings, changes to the survey design were undertaken to introduce a Likert scale, making it possible to improve the quality of the survey by gathering data about each item in the survey.

Lastly, the findings from the open-ended section also addressed individual items. As mentioned earlier, only the specific suggestions that addressed changes to individual items were entertained at this phase of instrument development, while any suggestions for new or additional items were saved for a later part of this study. It was in only one instance than a responded directly stated that the word choice was an issue, and suggested improvements. The respondent indicated that the item “demonstrating/leading discussions on how to solve problems” should instead read as “facilitate discussions.” The change in word choice here did make sense in terms of the language that teachers encounter. Moreover, the words “demonstrating” and “leading” could mean two different things. This suggestion was implemented, but we, however, opted to use “engage students in discussions of how to solve problems.”

Cognitive Interviews

Cognitive interviews were used as another assessment of content validity and to inform revision to the pilot instrument. The analytical memos that were created from the cognitive interviews disclosed important details, demonstrating that participants interpreted the meanings of the survey items in the intended ways and understood the purpose of the survey to be about assessing one’s VHQMI. This was an important discovery, given that one of the ways of assessing the quality of a survey is to evaluate the content standards of the survey, i.e., evaluate whether the questions and content are aligned to what the survey is trying to measure and how well respondents comprehend the survey (Groves et al., 2009). While there was significant progress made in establishing the extent to which respondents understood the content, these memos revealed some areas in need of improvements: (a) format of the survey, (b) inclusion/exclusion of items—determining whether certain items warranted to be included or excluded in the survey, and (c) language or wording of the items.

The first type of issue that was identified involved in the format of the survey. Similar to open-ended responses, the interviewees noted that it was a daunting task to pick 2-3 statements that represented their visions of ideal practices. Therefore, respondents had to develop unique strategies to respond to the survey. For example, some respondents viewed the survey as a lesson plan and hence used that perception to make their selection. One respondent saw the survey mostly in terms of what goes into planning the lesson then executing it, noting that while there are a place and time for each practice, his selections were based on what he thought must be evident in every lesson. He pointed out that every lesson must have a plan (role of the teacher), within the plan or what students are doing (role of tasks), and what students are saying (role of discourse). Another respondent shared similar sentiments, eventually making selections under the role of teacher dimension that reflected the typical launch-explore-discuss lesson format (Van de Walle, Karp, & Bay-Williams, 2012). Her three choices were, “creating and introducing interesting math tasks (launch),” “allowing students to work collaboratively (explore),” and “giving students authority to develop their strategies/solutions (discuss).”

The second type of issue that emerged had to do with the inclusion/exclusion of items from the survey. The majority of the respondents implied that it was okay to eliminate some items that described practices that were the necessary foundations needed for learning. The three practices that they specifically mentioned were all under the role of the teacher dimension and included “making connections with students,” “motivating students by captivation their attention,” and “creating and introducing interesting math tasks.” An in-depth analysis of the first two practices revealed that they did not represent practices that were specific to math, and the last item was a practice that every educator must do to prepare for any learning experience. Based on these findings, a decision was made to completely exclude these three items from any future survey developmental work. As a result of removing these three items, the role of the teacher dimension reduced from 15 to 12 items.

Another issue that respondents commented on that had to do with inclusion/ exclusion of items was evident in the role of discourse dimension. At least 5 out of 12 respondents mentioned the need to add more items in this section. The ideas that some items specifically only talked about the teacher while others only talked about students were problematic to these participants. In addressing this issue, one respondent suggested including items that mentioned both teacher action and student action in the same sentence. For example, this participant suggested including the following piece, “students are sharing their ideas and strategies, while teachers are facilitating using questioning.” Another respondent suggested adding items that address student-to-student interactions, recommending the following, “students are listening and responding to others’ ideas.”

Other respondents also mentioned issues with having way more items that address teacher actions as compared to items specifying students' actions. As mentioned earlier, although additional new items were developed based on such findings, they were not included in the modified survey since the pilot items were yet to be thoroughly tested.

The last type of issue that emerged from the analysis involved respondents not understanding some of the terminology used in the survey. For example, a term used by researchers in higher education, such as "problematize" and "mathematical structure" was confusing. Other terms were also indicated as problematic, for example, "demonstrating/leading," "solution," "facts," and "mathematical path." The word "fact" was described as incorrectly used in the phrase, "refraining from providing mathematical facts, formulas, or ideas." To the respondents, "mathematical facts" represents information that needs to be provided, unlike formulas or ideas that students need opportunities to discover. There were also issues with the words "engage" versus "foster" in the role of task dimension. According to two respondents, the word "engage" would be a better fit since it implies sharing authority with students. To combat these issues, several revisions were made to the wordings so that items were more understandable and clearer. For example, "engage students with mathematical structure of strategies" was modification to read "engage students in generalizing mathematical relationships or strategies." Apart from the changes that were made to some wordings, there were also minor changes made to the entire sentences so they could be clearer. For example, the item "problematizing student's prior conceptions of mathematical ideas" was changed to "engaging students in reflecting on their prior conceptions of mathematical ideas." The details in the analytical memos provided valuable information for guiding how to make the necessary changes. Table 5 summarizes the item status based on the modifications that resulted from the interviews.

Table 5

Item Disposition Based on the Cognitive Interviews

		Item disposition	
Dimensions	No. of items	Dropped	Modified
Role of teacher	15	3	8
Role of tasks	8	0	5
Role of discourse	12	0	8

The changes that were made resulted in a revised survey consisting of 32 items for assessing how participants rated their confidence that an instructional practice was suggestive of high quality. There were 12 items for the role of teacher dimension, eight items for the role of task dimension, and 12 items for the role of discourse dimension. Additionally, based on the suggestions from the respondents, and good survey principles (Groves et al., 2009; Onwuegbuzie et al., 2010), the format of the pilot instrument survey was revised from asking respondents to choose two to three statements (from the bank of 8-15 statements in each section) that best characterize what they would like to see in an ideal mathematics classroom to having respondents answer questions based on a 4-point Likert response scale:

1. Does not suggest
2. Mildly suggests

3. Moderately suggests
4. Strongly suggests

This revised survey represented a reasonable measure of the VHQMI and reduced the need for substantial revisions of this instrument in the future. The revised instrument also included similar demographic variables as the initial survey. Table 6 contains the text of the modified survey and the aspects of vision they match. This revised survey was field-tested and quantitatively analyzed.

Table 6

A Revised Survey of Math Educators' VHQMI

Below are sets of observations collected from a variety of math classrooms across North Carolina. If you were to observe these practices independent of the other, how would you rate your confidence that the observation is suggestive of high-quality math instruction?
Role of the Teacher
I see the teacher . . . engaging students in reflecting on their prior conceptions of mathematical ideas. allowing students to discover mathematical insights. monitoring to ensure that students are on task and productive providing opportunities for students to ask questions supporting students in investigation without providing mathematical formulas or ideas encouraging students to explain how to solve problems to each other. monitoring to ensure that students are on task and productive, ensuring student's work is thorough and correct. providing time for students to practice solving mathematical problems. giving students the authority to develop their own strategies/solutions. eliciting and scaffolding students' ideas providing time for students to collaborate and solve assigned mathematical problems engaging students in discussions of how to solve problems

Table 6

Cont.

Role of Mathematical Tasks
I see the teacher using math tasks that engage students in . . .
generalizing mathematical relationships or strategies.
complex thinking and allow for multiple solution paths
making connections between strategies or representations.
creating meaning of the use of mathematics
practicing procedures
practicing procedures prior to engaging in more conceptually challenging tasks.
showing how math is applicable in a “real-world” context.
hands-on activities.

Role of Classroom Discourse
I see and hear classroom discussions where . . .
teachers are asking questions that are carefully ordered to support students in meeting the lesson goals.
students are sharing their ideas with partners or in small groups.
teachers are evaluating students’ answers to ordered questions focused on lesson goals.
teachers are asking questions to uncover students’ mathematical steps.
students are responding to carefully ordered questions focused on the lesson goals.
teachers are prompting and pressing students on the mathematical steps in their work.
teachers are asking conceptually oriented questions that encourage students to explain their approaches, procedures, or calculations.
students are explaining and justifying their ideas in whole-class conversations.
students are explaining their mathematical approaches, procedures, or calculations
in whole-class conversations or small groups/partners.
teachers are asking conceptually oriented questions to uncover students’ thinking.
teachers are supporting students in understanding their procedures or calculations.
teachers are asking questions that lead to further investigations or revised conjectures.

Phase 2: Validating Revised Instrument

The data that was collected from the field-testing of the instrument was subjected to three forms of statistical analyses: factor analysis, correlation tests, and reliability tests. The goal of this phase was to assess construct related validity (i.e., structural validity) and also confirm the reliability (the degree to which items within each dimension were interrelated) of the revised instrument. Structural validity was examined employing exploratory factor analysis, and also by calculating the Spearman correlation of the items within each sub-dimension, while reliability was established with the Cronbach Alpha. Since the construct VHQMI was based on a theoretically sound instrument (Munter, 2015), these statistical analyses provided a means of assessing the extent to which the items in the modified survey were developed to fit the construct's conceptual meanings. The factor analysis results are presented, followed by the results for correlation tests within each sub-dimension and, lastly, the results of the reliability of the items within each dimension.

Factor Analysis Results

The results of the EFA were very promising. The following are the results for each dimension of the survey.

Role of teacher dimension. The initial factor analysis of the role of teacher dimension extracted three factors that were evident in the scree plot and had eigenvalues greater than one. Following this analysis, one questionable item was eliminated because it theoretically loaded into an unintended factor. This item "engaging students in discussions of how to solve problems" was theoretically anticipated to load in either factor 2 or factor 3, but instead loaded into factor 1. A closer look at this item indicates that participants could interpret this statement as a lower-level item because of no mention of whether these discussions were conceptually oriented or procedurally oriented. Another item, “providing time for students to collaborate and solve assigned mathematical problems,” which cross-loaded on both factor 1 (0.68) and factor 2 (0.45), was included in the analysis since it loaded more strongly on factor 1, and the difference between the loadings was huge. Moreover, this item, as written theoretically, fits a general procedural aspect of instruction, hence justified to be included in factor 1. The remaining 11 items were again factor analyzed.

A second factor analysis calling for three factors was run. These remaining items were found to load strongly on respective factors, and cross-loadings were not observed above the 0.45 cutoff. As shown in Table 2, the factor analysis results achieved the minimum reliability scores targeted for this study. The factor pattern showed that most loadings were in the “very good (>0.63)” to “excellent (>0.71)” range (9 out of 11 items), with the exception of 2 items in the “good (>0.55)” and “fair (> 0.45)” range respectively (Tabachnick et al., 2013).

The three factors that were generated were labeled according to the type of instructional practices they depicted. Items loading on to factor one appear to relate to the role of the teacher as the sole source of knowledge. These items included a reference to practicing procedures, focus on getting correct answers, making sense of work that the teacher has demonstrated, and encouraging students to work together. Therefore, this factor has been labeled as “direct instruction” (Munter, Stein, & Smith, 2015). Items loading on to factor two appear to relate to the nature of the classroom environment created by the teacher. These items reference ways in which the teacher shares authority with the students, positioning them as thinkers and decision-makers. Therefore, the second factor was labeled “position students as sense-makers.” The third factor includes items that appear to relate to how the teacher influences classroom discourse by proactively supporting and scaffolding students’ ideas. The third factor was labeled as “dialogic instruction” (Munter et al., 2015). These three extracted factors accounted for 61.6% of the variance in the data set and had the following eigenvalues: 3.22 (factor 1), 2.50 (factor 2), and 1.05 (factor 3).

In addressing the decisions about what type of factors to retain after a factor analysis has been performed, Worthington and Whittaker (2006) suggested that having factors that consist of many items is always desirable. However, these authors recommended that in instances when only two items fall into a given factor after factor analysis, item-generation procedures should be utilized at later stages of study to produce additional items that load into that retained factor. This is the line of reasoning that was followed to decide on retaining factor 3 (dialogic instruction), which ended up with only two items.

To check for structural validity, the three-factor structure was compared to the structure within the VHQMI rubrics to determine the extent of alignment. The results of the EFA demonstrated close alignment between the two instruments, especially in terms of the items that are classified in the VHQMI rubric as either as representative of either less sophisticated or highly sophisticated visions. For example, the items that loaded into direct instruction (factor 1) matched the levels 1 and 2 of the VHQMI rubric, and the items that loaded into the other two factors that are denoted high vision matched the levels 3 and 4 of the VHQMI rubric. Thus, this instrument appeared to be closely aligned in terms of structure to the existing VHQMI rubric.

Table 7

Factor Matrix for Role of Teacher Dimension

Items	Factors
Direct instruction (low vision)	Factor 1
providing time for students to practice solving mathematical problems (1)	0.8684
monitoring to ensure that students are on task and productive (2)	0.7003
providing time for students to collaborate and solve assigned mathematical problems (2)	0.6847
ensuring student’s work is thorough and correct (1)	0.6440
providing opportunities for students to ask questions (1)	0.6219
encouraging students to explain how to solve problems with each other (2)	0.5083
Position students as sense-makers (high vision)	Factor 2
giving students authority to develop their own strategies/solutions (4)	0.8530
supporting students in the investigation without providing mathematical formulas or ideas (3)	0.7337
allowing students to discover mathematical insights (3)	0.7010
Dialogic instruction (high vision)	Factor 3
engaging students in reflecting on their prior conceptions of mathematical ideas (4)	0.8786
eliciting and scaffolding students’ ideas (4)	0.5311

Note. Extraction method: PAF. Rotation method: Varimax with Kaiser normalization.

Role of tasks dimension. The factor analysis of the role of task dimension suggested that a two-factor solution was appropriate. This was evident in the scree plot that showed eigenvalues greater than one. All the items loaded above 0.45, the minimum cutoff set for this study, and there were also no cross-loadings observed. Inspection of the items' loadings on each factor suggested tasks that promote application on the one hand and tasks that promote deep engagement on the other. The two factors that were generated were labeled (a) emphasizing skills (low vision), and (b) emphasizing deep engagement (high vision).

The factor pattern showed that most loadings were in the "very good (>0.63)" to "excellent (>0.71)" range (5 out of 8 items), with the exception of 3 items, 1 in the "good (>0.55)" range and 2 in the "fair (> 0.45)" range, respectively. The factor analysis results showed that the two factors captured 51.7% of the variance in the data set and had the following eigenvalues: 2.37 (emphasize skills) and 1.76 (emphasize deep engagement). Table 8 summarizes the results of the factor matrix.

Table 8

Factor Matrix for Role of Tasks Dimension

Item	Factors
Emphasize skills (low vision)	Factor 1
practicing procedures (1)	0.8726
practicing procedures prior to engaging in more conceptually challenging tasks (1)	0.8336
showing how math is applicable in a "real-world" context (2)	0.5625
hands-on activities (2)	0.4536
Emphasize deep engagement (high vision)	Factor 2
complex thinking and allow for multiple solution paths (3)	0.7157
making connections between strategies or representations (4)	0.7057
creating meaning of the use of mathematics (3)	0.6457
generalizing mathematical relationships or strategies (4)	0.5310

An examination of the structural validity of the extracted factors revealed similar results to the results of the role of the teacher dimension. The two factors appeared to be closely aligned in structure to that found in the VHQMI rubric.

Role of discourse dimension. The initial factor analysis of the role of discourse dimension extracted three factors that were evident in the scree plot and had eigenvalues greater than one. Three questionable items were eliminated because they loaded on unintended factors. The items “students are sharing their ideas with partners or in small groups” and “students are explaining their mathematical approaches, procedures, or calculations in whole-class conversations or small groups/partners” loaded into the same factors although it was anticipated that they would load into different factors. The drawback here might be the words “small groups/partners” that appear in both statements, and therefore, because of this redundancy, these items were culled off. The other item that was eventually discarded “teachers are supporting students in understanding their procedures or calculations” loaded into factor 3 (Procedurally oriented discourse), yet it was written to ascribe to a slightly high quality of practice than the other items that loaded into this sub-dimension.

Following these deletions, second factor analysis was run on the remaining nine items. All the remaining items loaded strongly on only one factor, and each item loaded above 0.45, the minimum cut-off for this study. No cross loading was found above the 0.45 cutoffs. Furthermore, as the factor pattern shows, most loadings are in the “excellent (>0.71)” range (8 out of 9 items), and only one item is in the “very good (>0.63)” range. The three factors that were generated were labeled: (1) procedurally-oriented discourse (focus is on the teacher) (low vision), (2) procedurally-oriented discourse (some focus on the student) and, (3) conceptually-oriented discourse (high vision). These three factors accounted for 67.7% of the variance in the data set and had the following eigenvalues: 3.2536 (factor 1), 1.8259 (factor 2), and, 1.01 (factor 3).

Table 9

Factor Matrix for Role of Discourse Dimension

Items	Factors
Conceptually-oriented discourse (high vision)	Factor 1
teachers are asking questions that lead to further investigations or revised conjectures (4)	0.8029
teachers are asking conceptually oriented questions to uncover students' thinking (4)	0.7393
students are explaining and justifying their ideas in whole-class conversations (4)	0.7384
teachers are asking conceptually oriented questions that encourage students to explain their approaches, procedures, or calculations (3)	0.7147
Procedurally-oriented discourse (focus on answers) (low vision)	Factor 2
students are responding to carefully ordered questions focused on the lesson goals (1)	0.8945
teachers are asking questions that are carefully ordered to support students in meeting the lesson goals (1)	0.8252
teachers are evaluating students' answers to ordered questions focused on lesson goals (1)	0.7865
Procedurally-oriented discourse (some focus on the student) (low vision)	Factor 3
teachers are asking questions to uncover students' mathematical steps (2)	0.8323
teachers are prompting and pressing students on the mathematical steps in their work (2)	0.6919

An examination of the structural validity of the extracted factors revealed similar results to the results of the other two dimensions. The three factors were closely aligned with the pattern that is found in the VHQMI rubric. However, unlike in the other two dimensions, there seemed to be a slightly higher degree of alignment with the VHQMI rubric. For example, the items that loaded into factor 2 (procedurally oriented discourse with a focus on answers) were all developed based on level 1 of the VHQMI rubric, and the items that loaded into factor 3 (procedurally oriented discourse with some focus on the student) were developed based on Level 2 of the rubric.

Evidence of Correlation within Sub-dimensions of the Instrument

The correlation results within the sub-dimensions (factors) of each dimension of the survey are displayed in Tables 10–12.

Table 10

Correlation Results for Role of Teacher Dimension

Variable	by Variable	Spearman ρ	Prob>|ρ|
Direct Instruction (Factor 1)			
R.TE_WORKCORRECT (1)	R.TE_ON_TASK (2)	0.5561	<.0001*
R.TE_PRACTICE (1)	R.TE_ON_TASK (2)	0.4867	<.0001*
R.TE_PRACTICE (1)	R.TE_WORKCORRECT (1)	0.4621	<.0001*
R.TE_COLLAB (2)	R.TE_ON_TASK (2)	0.2232	0.0020*
R.TE_COLLAB (2)	R.TE_WORKCORRECT (1)	0.2149	0.0030*
R.TE_COLLAB (2)	R.TE_PRACTICE (1)	0.4559	<.0001*
R.TE_ASK_QUEST (1)	R.TE_ON_TASK (2)	0.3204	<.0001*
R.TE_ASK_QUEST (1)	R.TE_WORKCORRECT (1)	0.3087	<.0001*
R.TE_ASK_QUEST (1)	R.TE_PRACTICE (1)	0.4802	<.0001*
R.TE_ASK_QUEST (1)	R.TE_COLLAB (2)	0.3322	<.0001*
R.TE_EXPLAIN (2)	R.TE_ON_TASK (2)	0.2156	0.0029*
R.TE_EXPLAIN (2)	R.TE_WORKCORRECT (1)	0.1491	0.0405**
R.TE_EXPLAIN (2)	R.TE_PRACTICE (1)	0.3468	<.0001*
R.TE_EXPLAIN (2)	R.TE_COLLAB (2)	0.4017	<.0001*
R.TE_EXPLAIN (2)	R.TE_ASK_QUEST (1)	0.5004	<.0001*
Position students as decision-makers (factor 2)			
R.TE_DISCOVER (3)	R.TE_AUTHORITY (4)	0.4865	<.0001*
R.TE_INVESTIG (3)	R.TE_AUTHORITY (4)	0.4763	<.0001*
R.TE_INVESTIG (3)	R.TE_DISCOVER (3)	0.3812	<.0001*
Dialogic Instruction (factor 3)			
R.TE_PRIOR_CONC (4)	R.TE_SCAFFOLD (4)	0.2701	0.0002*

Note. *p < .01; **p < .05

Based on the results of the Spearman's correlation, as shown in Table 10, there was a statistically significant correlation between items in the same sub-dimensions. The value of the correlations ranged from 0.14 to 0.55, and were all statistically significant.

Based on the results of the Spearman's correlation, as shown in Table 11, there was a positive correlation between items in the same sub-dimensions. The value of the correlations ranged from 0.13 – 0.64, and except for the association between the items that emphasize that role of tasks is for "complex thinking and allow for multiple solution paths" and "generalizing mathematics relationships or strategies," all other associations were statistically significant.

Table 11

Correlation Results for the Role of Tasks Dimension

Variable	by Variable	Spearman ρ	Prob>|ρ|
Emphasize skills			
R.TA_HANDS_ON (2)	R.TA_PROCEDURES (1)	0.3178	<.0001*
R.TA_PROC_B_CONC (1)	R.TA_PROCEDURES (1)	0.6459	<.0001*
R.TA_PROC_B_CONC (1)	R.TA_HANDS_ON (2)	0.1566	0.0314**
R.TA_REAL_WORLD (2)	R.TA_PROCEDURES (1)	0.3069	<.0001*
R.TA_REAL_WORLD (2)	R.TA_HANDS_ON (2)	0.3884	<.0001*
R.TA_REAL_WORLD (2)	R.TA_PROC_B_CONC (1)	0.3466	<.0001*
Emphasize deep engagement			
R.TA_MULT_SOLN (3)	R.TA_GENERALIZE (4)	0.1379	0.0584
R.TA_MEANING (3)	R.TA_GENERALIZE (4)	0.2014	0.0055*
R.TA_MEANING (3)	R.TA_MULT_SOLN (3)	0.2311	0.0014*
R.TA_CON_REPS (4)	R.TA_GENERALIZE (4)	0.2628	0.0003*
R.TA_CON_REPS (4)	R.TA_MULT_SOLN (3)	0.2806	<.0001*
R.TA_CON_REPS (4)	R.TA_MEANING (3)	0.3373	<.0001*

Note. $^{*}p < .01$; $^{**}p < .05$

Based on the results of the Spearman's correlation, as shown in Table 12, there was a statistically significant correlation between items in the same sub-dimensions. The value of the correlations ranged from 0.37 to 0.69, and were all statistically significant. There was a stronger association between the items in the sub-dimensions of the role of discourse dimension as compared to the sub-dimensions within the role of teacher and task dimensions. In conclusion, the reliability findings that suggested consistency of items within each dimension were supported by these correlation results.

Table 12

Correlation Results for the Role of Tasks Dimension

Variable	by Variable	Spearman ρ	Prob>|ρ|
Conceptually-oriented discourse			
R.TD_REV+CONJ (4)	R.SD_JUSTIFY (4)	0.3770	<.0001*
R.TD_CONC_STU (4)	R.SD_JUSTIFY (4)	0.3909	<.0001*
R.TD_CONC_STU (4)	R.TD_REV+CONJ (4)	0.4449	<.0001*
R.TD_CONC_EXPLAIN (3)	R.SD_JUSTIFY (4)	0.3835	<.0001*
R.TD_CONC_EXPLAIN (3)	R.TD_REV+CONJ (4)	0.4987	<.0001*
R.TD_CONC_EXPLAIN (3)	R.TD_CONC_STU (4)	0.4284	<.0001*
Procedurally oriented discourse (focus on answers)			
R.SD_RESPOND (1)	R.TD_EVALUATE (1)	0.5884	<.0001*
R.TD_SEQ_QUEST (1)	R.TD_EVALUATE (1)	0.5743	<.0001*
R.TD_SEQ_QUEST (1)	R.SD_RESPOND (1)	0.6926	<.0001*
Procedurally-oriented discourse (some focus on the student)			
R.TD_PROBING (2)	R.TD_UNCOVER_STEPS (2)	0.4128	<.0001*

Note. $^{*}p < .01$

Evidence of Reliability

As earlier outlined, new survey instruments should be validated by employing statistical techniques such as reliability tests to confirm the internal consistency of measures (Ross et al., 2003). To estimate the reliability of the instrument, Cronbach's α test was employed to measure the internal consistency of the items that survived in each dimension after the factor analysis. Literature suggests that if the number of retained items at this stage is sufficiently large, the researcher may want to eliminate those items that will improve the internal consistency (Hinkin, 2003). This would not be the case because, by this time in the study, it was already noted that there was a need to increase the number of items to capture the VHQMI within each dimension of the instrument adequately. However, there were several checks performed on each dimension to establish whether the deletion of some items with lower correlations would improve the value of Alpha, but each time an item was deleted, the value of alpha did not change significantly. Because there was a concern that the current level of content validity would be threatened, all the items that survived the factor analysis were retained. Table 13 illustrates the Cronbach's α values that were estimated to examine the internal consistency within each dimension.

Table 13

Reliability of Measurements

VHQMI Dimension	Number of items	Cronbach's α
Role of Teacher	11	0.7298
Role of Tasks	8	0.6415
Role of Discourse	9	0.7752

As shown in Table 13, the Cronbach's α values for the role of teacher and role of discourse dimensions are above the 0.70 level and the role of tasks dimension near the

0.60 thresholds set for this study. The somewhat high Cronbach's α value, for example, for the role of discourse dimension implies higher reliability that the items in this dimension measure the same content universe. Ideally, the desired alpha values are above the 0.80 level (George & Mallery, 2003). Therefore, while the Alpha values appeared to demonstrate that items within each dimension were internally consistent or are were measuring the same constructs, these results indicated that there was some room for improvement. The logical conclusion was to devise means of adding more items in future studies. Thus, in the last section of this study, some items were recommended for addition across all dimensions to improve the reliability of the instrument further.

Phase 3: Confirmatory Study

The confirmation study presented an opportunity to gather further evidence of validity. After examining content validity and construct validity, Phase 3 involved testing for predictive validity. It was predicted that a comparison of participants' scores in the survey instrument to a subsequent assessment of their visions in an interview, would reveal a particular pattern. Given that I had access to some participants with known sophisticated visions based on prior engagements with our research team, a projection was made that the developing survey was able to predict their behavior in particular ways. The predictive validity test was designed to assess the extent to which participants' articulations on the survey were related to articulations of their visions using Munter's interview protocol (Munter, 2015). These interviews were administered following the field-testing of the revised instrument, and it was anticipated that participants would articulate somewhat similar visions in both the survey and interview. In the next section, three case studies were used to demonstrate how participants' articulations of their visions aligned with respective dimensions in the survey. The three cases were drawn from five who participated in the confirmation phase. The reason for only using three participants was because the main goal was to see how each section of the developing survey was performing. Hence, although data were collected for all five participants, these three cases were randomly picked in order for each to examine specific dimensions of the developing survey. The survey data that is used here includes data obtained only from the items that were retained after the field-testing phase.

Lilian

The first participant, Lilian, was used to examine the role of the teacher dimension. Lilian started her teaching career as a fourth-grade teacher, where she taught for 8 years. For the past 13 years, she has played the role of a math and science lead teacher for the district. Her current assignment involves going into schools to model lessons, doing walk-throughs and providing feedback to teachers, and also delivering professional development to teachers. Table 14 captures Lilian's survey results on the role of teacher dimension.

In the survey, Lilian's consistently rated practices that were considered high quality in the "moderately suggests" or "strongly suggests" range. For example, she believed that practices that fell under "positioning students as sense-makers" represented the highest quality of practices that teachers should promote in learning environments. Consistent with these views, Lilian held the perception that certain practices "do not suggest" or "mildly suggest" high-quality math instruction. Such practices that contrasted the high-quality practices involved traditional teacher-directed instruction, with students dutifully listening, and then collaborating to practice demonstrated procedures.

Table 14

Lilian's Survey Results for Role of Teacher Dimension

If you were to observe these practices independent of the other, how would you rate your confidence that the observation is suggestive of high-quality math instruction? I see the teacher . . .				
Direct instruction (low vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
providing time for students to practice solving mathematical problems	X			
monitoring to ensure that students are on task and productive	X			
providing time for students to collaborate and solve assigned mathematical problems			X	
ensuring student's work is thorough and correct	X			
providing opportunities for students to ask questions		X		
encouraging students to explain how to solve problems to each other			X	
Position students as sense-makers (high vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
giving students the authority to develop their own strategies/solutions				X
supporting students in the investigation without providing mathematical formulas or ideas				X
allowing students to discover mathematical insights				X
Dialogic instruction (high vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
engaging students in reflecting on their prior conceptions of mathematical ideas			X	
eliciting and scaffolding students' ideas			X	

During the interview, Lilian expressed the views about the role of the teacher that stressed on high-quality instruction. She consistently talked about designing learning environments to support learning through co-participation. For example, she envisioned teachers “using high-quality tasks that are engaging to students and encourage students’ discussion and reasoning about mathematics,” and while students worked, the teacher would not only be “observing students as they work—listening to student discussions, purposefully selecting which students and in what order they will share during the whole group discussion,” but would also “asks students questions to help uncover their thinking, extend their thinking, or make connections to important mathematics.” Additionally, the teacher would be “allowing students to productively struggle when solving tasks, not rescuing or taking over the students’ thinking,” and would also “facilitate[s] whole group discussion—asking questions, using talk moves, helping students see and make connections to the important mathematics, and orchestrating the discussion to lead to the mathematical goal and learning target of the lesson.”

Summary of Lilian’s Survey Responses and Interview

The comparison of the views expressed by Lilian for the role of teacher dimension in the survey indicated some degree of alignment with her views in the interview, but also some minor divergent views in two instances for the role of the teacher. In comparison to her views in the interview, and based on the knowledge that Lilian held sophisticated views about mathematics instruction, it was anticipated that she would rate the items “providing time for students to collaborate and solve assigned mathematical problems,” and “encouraging students to explain how to solve problems to each other" as either "does not suggest" or "mildly suggest high quality." While this was not the case, since the majority of the responses in the survey corresponded to the responses in the interview, this dimension of the survey seemed to have performed in the general direction that it was expected to perform, thereby indicating some elements of predictive behavior.

Leah

The second participant, Leah, was used to examine the role of the task dimension. Leah had taught mostly at the high school level, and some at the university level as an instructor. Her focus has been on being a teacher leader, helping teachers adopt hands-on types of problems solving strategies, modeling problems so students can see the relevance using technology and data-driven aspects of teaching. She is currently serving as a math teacher, teaching advanced math classed at the high school level. Table 15 captures Leah's survey results on the role of task dimension.

In the survey, Leah consistently rated practices that were considered high quality in the "strongly suggests" range. For example, she believed that tasks that emphasize deep engagement represented the highest quality of practice that teachers should engage students within learning environments. Consistent with these views, Leah held the views that tasks that emphasized application skills "mildly" or "moderately" suggested high-quality math instruction. She specifically rated the low cognitive demanding tasks that promoted practicing procedures practices without connections to meanings as "mildly suggests" high quality but rated the items that focused on application and hands-on activities as "moderately suggests" high quality.

Table 15

Leah's Survey Results for Role of Tasks Dimension

If you were to observe these practices independent of the other, how would you rate your confidence that the observation is suggestive of high-quality math instruction? I see the teacher using math tasks that engage students in . . .				
Emphasize skills (low vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
practicing procedures		X		
practicing procedures prior to engaging in more conceptually challenging tasks		X		
showing how math is applicable in a "real-world" context			X	
hands-on activities			X	
Emphasize deep engagement (high vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
complex thinking and allow for multiple solution paths				X
making connections between strategies or representations				X
creating meaning of the use of mathematics				X
generalizing mathematical relationships or strategies				X

During the interview, Leah expressed the views that good math tasks should have the potential to engage students in complex thinking, multiple solutions paths and insights into problem-solving strategies, and support learning through hands-on activities and connections to the real-world context. Accordingly in the interview, Leah stated that tasks should "be more open-ended, not closed," "have entry points for a variety of levels," "give students opportunities to think and make sense of the mathematics," "use activities that enable students to see how math is connected to the real world," "not be clear-cut so that it opens up a variety of solution methods," and "provide students with opportunities to share their reasoning so that a teacher could build on that reasoning."

Summary of Leah's Survey Responses and Interview

A comparison of the views expressed by Leah in the survey and interview seems to indicate some degree of alignment but also two instances of misalignment, just like in the case of Lilian for the role of the teacher dimension. For example, while the views about the practice "showing how math is applicable in a 'real-world' context" in the survey matched precisely with the views Leah expressed in the interview, this view represents a misalignment with the anticipated views of someone known to have a sophisticated vision. For the most part, the role of task dimension of the survey seemed to have performed in the general direction that it was expected to perform, thereby demonstrating some elements of predictive behavior

Jenny

The third participant, Jenny, was used to examine the role of discourse dimension. Jenny had been teaching for 30 years. She started as a fifth- and eighth-grade math teacher and later worked as a mathematics lead teacher in for her districts. She had maintained that math lead teacher position to date of the interview. Table 16 captures Jenny's survey results on the role of discourse dimension.

Table 16

Jenny’s Survey Results for Role of Discourse Dimension

If you were to observe these practices independent of the other, how would you rate your confidence that the observation is suggestive of high-quality math instruction?
I see and hear classroom discussions where . . .

Conceptually-oriented discourse (high vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
teachers are asking questions that lead to further investigations or revised conjectures				X
teachers are asking conceptually oriented questions to uncover students’ thinking				X
students are explaining and justifying their ideas in whole-class conversations				X
teachers are asking conceptually oriented questions that encourage students to explain their approaches, procedures, or calculations		X		
Procedurally-oriented discourse (focus on answers) (low vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
students are responding to carefully ordered questions focused on the lesson goals	X			
teachers are asking questions that are carefully ordered to support students in meeting the lesson goals	X			
teachers are evaluating students’ answers to ordered questions focused on lesson goals		X		
Procedurally-oriented discourse (some focus on the student) (low vision) items	Does not suggest	Mildly suggests	Moderately suggests	Strongly suggests
teachers are asking questions to uncover students’ mathematical steps				X
teachers are prompting and pressing students on the mathematical steps in their work		X		

Jenny’s survey responses consistently showed that she had confidence that conceptually-oriented classroom discourse “strongly suggested” high-quality practice. It was only in one instance that she rated a conceptually oriented discourse practice as “mildly suggestive” of high-quality practice. Consistent with these observations, Jenny rated procedurally-oriented discourse as either “does not suggest” or “mildly suggests” high-quality discourse, and only on one occasion did she rate a low-quality practice as “strongly suggests” high-quality discourse.

During the interview, Jenny espoused classroom discourse that emphasized developing a supporting a “mathematical discourse community” (Lampert, 1990), expressing the views that productive learning happens when the forms of discourse in the classroom promote shared authority, conceptual reasoning, and sharing of ideas. Accordingly in the interview, Jenny stated that classroom discussions should involve teacher, “facilitating and really looking for student reasoning,” “ not lecturing or teaching per se,” “using unique ways of questioning that are not leading and telling students answers, but trying to elicit information: what does the student know, and how to get them to a deeper understanding,” and “using those five practices” (Smith & Stein, 2011). While the teacher facilitates, the students would be “engaged in the mathematics, they are talking to each other, they are questioning themselves and each other, they are thinking deeply,” and are “critiquing each other, questioning, responding to each other . . . explaining their answers . . . persevering in problem-solving.”

Summary of Jenny’s Survey Responses and Interviews

The findings of these dimensions of the survey also seemed to support the findings of the other two dimensions. A comparison of the views expressed by Jenny in the survey and the interview seemed to suggest some degree of alignment, except for the two instances that have been mentioned. The role of discourse dimensions of the survey seemed to have performed in expected ways, thereby indicating that dimension had some predictive validity.

Summary of the Three Cases

In general, by using these three cases as samples, essential pieces of evidence were used to assess the validity of each dimension of the survey. Overall, based on the results that show participants' views expressed in the interview are reasonably close to the actual views expressed in the survey, it could be inferred that the survey has some degree of predictive validity. However, findings of divergent results warrant some further analysis.

Summary of Instrument Development Process

This study employed survey validation techniques to answer the question, to what extent does the collection of survey items create a reliable and valid measure of VHQMI? The recommended areas of validity evidence include content-related validity, construct validity, criterion-related validity, and reliability tests (Groves et al., 2009; Onwuegbuzie et al., 2010). Table 17 shows how the various aspects of each type of validities were utilized towards creating a reliable and valid measure of VHQMI.

The first type of validity that was employed in this investigation was content validity, which utilized a quantitative and qualitative approach. Both the results from the field-testing of pilot items, feedback form open-ended responses, and cognitive interviews suggested that the instrument possessed an appropriate level of content validity.

Table 17

Summary of the Instrument Validation Process

Areas of validity evidence	Techniques applied	Remarks
Content (Face validity & Item validity)	Use VHQMI rubric as the blueprint to generate content; Inclusion of an open-ended section in initial instrument to generate feedback; Cognitive interviews to assess initial instrument	Responses from the pilot survey indicated an understanding of content. Participants confirmed understanding the meanings of items and purpose of the survey; Participants confirmed that items represented their content universe; feedback informed the revision of items and survey format. This suggests that content validity exists in the final instrument.
Construct		
Structural validity	Factor analysis; The eigenvalue of or above 1; loadings above 0.50; no cross-loading above 0.45; Items that do not load properly are removed from the instrument.	All eigenvalues of 1 and above All items loaded above 0.50; most loadings were in the very good (> 0.63) to excellent range (>0.71); There was no cross-loading of items above 0.45; Total explained variances ranged from 51% - 68%. This confirms evidence of structural validity.
Spearman Correlation	Correlations test within each factor to test the strength of association between items	A positive correlation between items in the same factor; All correlations statistically significant except only one instance. This further verifies the evidence of structural validity
Reliability		
Internal consistency	Cronbach's Alpha value above 0.60	Cronbach's α for all dimensions found above 0.60 threshold which suggests that internal consistency exist in the instrument
Criterion		
Predictive validity	Case study; Compare responses in the survey with responses in a subsequent interview to determine how well the survey could predict views in the interview.	There was a close alignment between views in the survey and interviews, which confirms evidence of predictive validity.

To further answer the research question, this research employed a construct validity approach. Construct validity (i.e., structural validity) was investigated by utilizing factor analysis and then quantitative analysis of the data. The results of Phase 2 (validating revised instrument) demonstrated that this research meets the standard recommendations for factor analysis, which suggests that factors should have eigenvalues of or above 1. Ideally, all loadings should be above 0.40 with no cross-loading above 0.45, and items that do not load properly are removed from the instrument. Additionally, the correlation tests in Phase 2 were quantitatively analyzed, revealing positive associations between items in the same factor, and all correlations were statistically significant, except for only one instance. These results imply that the validated instrument may provide an efficient measure of the theoretical construct of VHQMI.

Third, the internal consistency of the instrument was examined by utilizing Cronbach's α test. The Cronbach's α value for all dimensions of the survey ranges between 0.64 – 0.77, which is well above the 0.60 acceptable thresholds for early stages of research, thereby suggesting internal consistency exists in the instrument.

Finally, Phase 3 blended both quantitative and qualitative data analysis to establish aspects of predictive validity in the instrument. Elements of predictive validity were demonstrated by comparing selections in the survey with participants' articulations in an interview intended to assess instructional visions for the role of the teacher, the nature of tasks used for teaching and learning, and the role of the classroom discourse. The results of Phase 3 indicate a close alignment between views of participants with known high visions in the survey and views in the interviews, but with few instances of divergent results.

In summary, using a variety of validation techniques (content, construct, predictive, and reliability testing) for survey instrument development resulted in a 28-item instrument, comprising three dimensions. The role of the teacher dimension consists of 11 items; the role of task dimensions consists of eight items, while the role of discourse dimension consists of nine items. While the methods in this study have yielded an instrument with aspects of reliability and validity, there is room for improvement.

CHAPTER V

DISCUSSIONS, CONCLUSIONS, AND IMPLICATIONS

The purpose of this study was to develop an instrument that can be used to assess math educators' visions of instruction in the context of implementing new state mathematics standards. The recommended survey development involves either utilizing previously validated instruments or creating new instruments based on the literature (Hinkin, 2003). A review of the literature did not yield any established survey instruments to address the needs of this study adequately. Thus, a new instrument was developed by relying on the existing VHQMI rubric (Munter, 2014). The goal was not to look for a perfect one-to-one match with the VHQMI rubric, but to use the rubric to guide the generation of statements that were representative of the significant dimensions of mathematics classroom practice.

The results of this study were a 28-item instrument, comprising three dimensions, each with 2-3 sub-dimensions. Additionally, the set of items suggests that they may reliably, and validity assess low and high vision. The role of teacher dimensions consists of 11 items in three hypothetical constructs, namely, direct instruction (six items), position students as sense-makers (three items), and dialogic instruction (two items). The role of tasks dimension consists of eight items in a two-factor structure, namely, emphasize skills (four items) and emphasize deep engagement (four items). Lastly, the role of discourse dimension consists of nine items in a three-factor structure, namely, conceptually oriented discourse (four items), procedurally oriented discourse (focus on the answer) (three items), and procedurally oriented discourse (some focus on students) (two items).

An analysis of the factor structure of the final instrument indicates that participants viewed items through the lens of particular instructional quality themes. For example, under the role of teacher dimension, all the items that had a task-focused emphasis ended up in the same factor, the items that emphasize attribution of mathematical authority loaded onto a separate factor, while items involving teachers' role in the influence of high-quality discourse loaded together. This was not a surprising finding given that Munter (2014) also identified three dominant ways that participants characterized the role of the teacher, i.e., conception of typical activity structure, attribution of mathematical authority, and influence of classroom discourse. Under the role of teacher dimensions, two themes emerged with participants viewing the statements about tasks in terms of the cognitive demand or level of student engagement involved. In the lower end were descriptions of tasks that required lower cognitive demand, while the other items described more sophisticated tasks that required higher cognitive demand. Under the role of discourse dimension, the three themes that emerged as evident from the factor structure were similar to Hufferd-Ackles et al.'s (2004) framework for levels of the math-talk community. In the final survey, there were clear distinctions between the procedurally oriented discourse items, with discourse items that focused on answers loading together, and items that described teacher's questions with a focus on student loading in another factor. In terms of the items describing a higher level of discourse, the participants in this study did not distinguish the two levels like Hufferd-Ackles et al. did, but viewed these items together as representing conceptually oriented discourse. The results of this study suggest that the final survey instrument can be applied with sufficient reliability to assess math educators' vision of instruction.

Research Contributions

This study offers several contributions that are summarized in the next section.

Conceptual Contribution

Spillane et al. (2002) concluded that perceptions were the best indicators of the decisions that individuals make concerning the implementation of instruction. Consequently, teachers and instructional leaders' visions about math instruction are critical predictors of their interactions with instructional practices within their context. One of the issues that emerged from the cognitive interviews involved the awareness that some forms of instruction are considered by mathematics educators as non-math specific, and therefore should be used by that all teachers. For example, one participant noted that teaching would not be possible if teachers could not "make connections with students." Such observations imply that the work on studying vision must specify whether items are specific to mathematics or not.

Instrument Use

This instrument is potentially useful for a number of purposes. First, the survey may be a cost-effective instrument to collect large-scale data to explore teachers' and instructional leaders' visions of mathematics instruction to support implementation efforts. Knowing the visions held by teachers and instructional leaders can inform the resources developed to support their work. It may also help to align multiple interpretations and messages across school systems to bring more coherence. Second, the instrument may be useful in monitoring the progress of implementation of reform efforts. Whether the efforts involve helping teachers to improve on their role, or whether they want to improve on their use of complex math tasks, or improve on the use of high-quality discourse, this instrument can be used to monitor progress in the ways that any individuals describe practices associated with these aspects. Third, while the survey is useful in a collective sense, it may also be used at an individual level. For example, if the focus of the effort is on supporting standards-based practices like attending to student thinking or using cognitively demanding tasks, then collecting survey data at several benchmarks may provide an understanding of whether there any notable changes in participants' visions.

While there are several uses for the survey, there are some caveats about what the survey does not tell us about instructional vision. First, this survey gives us general ideas about teachers' perceptions of instructional practices and leaders' expectations for engagement with these practices but does not provide us is with information about the enactment of instructional practices. This type of information can only be gathered through observations or interviews. Second, there are aspects of the vision that this survey may not be able to capture. Hammerness (2001) noted that vision specifically varies across three critical dimensions, focus, range, and distance, of which surveys cannot explicitly assess. For example, focus, which is the area that enjoys the bulk of one's attention, and range, which is the extent of that focus, cannot be captured by the responses in a survey. Additionally, the distance, which refers to how close or far vision is relative to what one is doing, can only be investigated through a combination of observations and interviews, or interviews alone. Lastly, assessing individuals' visions based on a survey does not help to distinguish instances in which participants mean what they are saying from those in which they are merely regurgitating what they have learned to say but do not value (Munter, 2014). What the survey does is give us glimpses of the ways that individuals currently articulate their views about the quality of instructional practices.

Methodological Insights

The use of a dialectic approach for survey development helped to maximize the use of both quantitative and qualitative data at the item and concept level. First, since the survey items were constructed to capture participants' visions of mathematics instruction, the dialectic approach provided a means of examining whether the survey items were representative math practices that participants are aware of. Additionally, it provided a means of examining how accurately the items were being interpreted in the participants' context.

Second, the use of a dialectic approach allowed exploring whether the concept as framed in the survey was applicable in the participants' context. One notable finding was the view that the format of the pilot survey needed to be changed. After revisions were made, the results of the factor analysis seemed to align in anticipated ways in the participants' views. Generally, the use of a dialectic approach provides great utility for developing surveys.

Another methodological insight has to do with considerations that have to be made in terms of theoretical versus field-tested results. The fact that this survey was developed from an existing interview rubric does not mean that the factors obtained after developing the survey would match perfectly with the interview rubric. While Munter's (2014) VHQMI rubric has four levels of instructional sophistication, the results of this study show only 2-3 levels of sophistication.

Process of Instrument Development and Validation

This study extends survey instrument development methodology and is useful to researchers attempting to develop similar instruments. Although the theory on visions as used in this study has been taken from an existing interview rubric, diverse forms of instrumentation are still lacking to study vision has not been developed. Therefore, it was considered essential to create and validate this new instrument. This study offers a promising starting point and first research efforts towards understanding and operationalizing visions in terms of a survey. The process involved several deliberate phases to create, test and revise vision items, and the results showed that these phases yielded desired results. The open-ended section and cognitive interviews seemed to have been particularly helpful in testing how respondents understood the survey. The factor analysis results confirmed distinct factors within each dimension of the survey.

The reliability and validation processes used in each phase of this study presents advancements in methods used to collect data on visions. Reliability and validity are necessary entities of instrument development if one is to report with confidence the results obtained from a survey (Groves et al., 2009). Since this study created a new research instrument, the utmost care was undertaken to demonstrate evidence of reliability and validity in every phase of the instrument's development. The instrument development process involved creating appropriate items and survey prompts, and then undertaking an extensive scale development process, including using respondents' feedback, cognitive interviews, interviews, and statistical analyses of data. Various forms of validities were assessed in each phase of the study. First, given that the pilot items were written by local experts and were based on an existing VHQMI Rubric (Munter, 2014), increased confidence that the pilot items possessed appropriate levels of content validity (i.e., face & item validity). Additionally, the inclusion of an open-ended section in the pilot survey to gather feedback provided an extra layer of assessing the validity of the pilot survey. The suggestions from this section, coupled with responses from follow up cognitive interviews, led to revisions of items, thereby increasing the face and item validities of the instrument.

Second, the use of factor analyses, correlation tests, reliability analysis, and a confirmatory study further demonstrated evidence of validity. Specifically, the form of validity assessment pursued in this section was construct validity (i.e., structural validity), to examine the extent to which the instrument followed a pattern predicted by theory (Onwuegbuzie et al., 2010). As described in previous paragraphs, the results of the factor analyses revealed a factor pattern that is closely aligned to the levels of instructional quality found in other research studies (e.g., Hufferd-Ackles et al., 2004; Munter, 2014). Additionally, the correlation tests revealed that there is an appropriate degree of association between the items within the same factor/subdimension.

Third, the reliability of the survey demonstrated using Cronbach's Alpha revealed that all the dimensions of the survey were internally consistent. While all three sub-dimensions demonstrated appropriate adequacy, the lower Alpha value for the role of task dimension might be attributed to a lack of enough items to measure participants' views on the role of tasks. For example, the presence of only 8 items in the pilot survey did not leave any wiggle room to eliminate any items during the reliability analysis phase. Any attempts to delete any items during this phase to increase the value of Alpha instead served to lower this value. Given that the reliability test for this dimension met the acceptable 0.60 thresholds (Nunnally, 1967) recommended in the early stages of research, it could be argued that the items in this dimension are capable of consistently assessing math educators' vision of the role of tasks.

Lastly, showing that some of the results of the survey correlated with the findings from an interview assessed the predictive validity of the survey. However, the evidence of predictive validity is limited by a concern that when participants articulated their visions in the interview, it was not possible to compare their articulations to every aspect of the survey. Thus, general comparisons were made between the views expressed in the survey and the interviews. The divergent results obtained in Phase 3 could have resulted from a number of likely possibilities. One, it is possible that while the participants are known to have high visions, there are elements of their visions that are still developing. Two of the items could have been worded in such a way that led the participants to make those specific references. Third, some of the practices, for example, "use of hands-on activities," are often assumed to imply high quality.

For the most part, the methodology employed in this study met the acceptable thresholds proposed for survey instrument development. Following the validation techniques outlined, I believe that a roadmap has been provided for any researcher seeking to make similar instruments. Albeit, I believe that one more iteration of refinement to the instrument could strengthen its validity.

Cautions for Use of Findings in the Study

In recommending this instrument to researchers investigating perceptions of math instruction, and conducting future research, a few points of caution are appropriate. First, methodologically, developing this instrument followed multiple steps that involved getting insights directly from the participants. Hence, the reliance on self-reported data assumes a participant's honesty and awareness, which would be viewed by some as a limitation.

Second, the survey items related to instructional practices that participants were assumed to be conversant. Therefore, their responses should be considered with caution. Although cognitive interviews work well in pretesting survey items (Collins, 2003; Groves et al., 2009), it would be illuminating to see what other researchers would find after pretesting similar survey items.

Third, this research is limited in that the comparison of responses in the survey and revealed responses in the interviews in the confirmatory Phase 3 presents discrete choices, and there is a possibility that the case for predictive validity is being overstated. Future research is needed on the predictive validity of the instrument. Maybe the use of classroom observation to accompany the interviews might provide a more compelling context for predictive validity.

Fourth, the sample size in the validation Phase 2 could be a source of concern. In addressing the number of participants needed for performing factor analysis, MacCallum, Widaman, Zhang, and Hong (1999) found that the minimum sample size necessary to obtain factor solutions that was adequately stable and that corresponded closely to population factors was not constant across many studies that they reviewed. First, they found that as sample size increased, sampling error was reduced, and factor analysis solutions were more stable and more accurately represented the correct population structure. Second, they found that when factor loadings were consistently high (>0.60), sample size had relatively little impact on the quality of factor solutions, meaning that accurate description of population structure could be obtained using a reasonably small sample. One of their key conclusions was that when factor loadings are in the range of 0.50, then the use of a sample size of 100-200 participants is acceptable, but not preferred. While this current study on visions indicates that these criteria may have been met, MacCallum et al. (1999) advocated for using sample sizes that result in the mean level of loadings to be at least 0.7, preferably higher. This latter argument may lead to questions about the veracity of the sample size used in this current study.

Future Directions

Results from this study show several directions for future research. First, further work is needed in the establishment and continued development and maintenance of an item bank. An ongoing concern to survey researchers is the length of survey instruments

(Hinkin, 1998; Wilkerson et al., 2018; Worthington, & Whittaker, 2006). Artino, La Rochelle, Dezee, and Gehlbach (2014) argue that although no hard rules guide the decision, it is essential to ensure that all aspects of the domain of interest have been adequately covered. A good practice is to develop more items that will ultimately be needed in the final instrument because some items will likely be deleted or culled off for revisions during the design process (Gehlbach & Brinkworth, 2011). Given that the subdimensions in the final survey have ended up with relatively few items, writing additional items is warranted. For example, the sub-dimensions/subscale “procedurally oriented discourse (some focus on the student)” under the role of teacher dimension has only two items, and therefore, writing additional items for this sub-dimension might provide more examples of practices that sample that domain. Also, given that the design and validation process in most cases leads to the deletion of some items, further refinement of adding more items to the overall survey will ensure that deletions that are made will not threaten the reliability and validity of the instrument.

To begin the work of establishing a large item pool, there was a return to Phase 1 to examine the feedback from the open-ended section and analytical memos from cognitive interviews for instances where new items were being implied or suggested. Additionally, other criteria for adding items were based on overall insights gained from the entire survey development.

Several key aspects emerged. First, in the open-ended sections under the role of the teacher, there were several mentions of instruction that have to do with “reinforcing basic concepts” or “giving basic knowledge,” progression (path) of lessons, providing formulas and strategies to students, providing students with different tools for problem-solving, and an emphasis on differentiated learning. Another idea that was also floated involved "selecting and sequencing students' ideas for whole-class discussion." Other notions emphasized creating items that talked about students as "leaders of their own learning," "debating and critiquing ideas, collaborating in groups, and learning through discovery and mistakes."

Second, in terms of the role of tasks, several features of tasks were described. Tasks were envisioned in terms of their role in reinforcing previous learning, learning foundational concepts, making connections to real-world situations and situations that are of interest, having a hands-on experience, and use of various tools. Tasks were also envisioned in terms of their ability to promote multiple approaches of learning, building and extending prior knowledge, justifying and validating ideas, and not explicitly having a defined way of solving. One participant had an enlightening view of the role of tasks, noting that tasks function to create opportunities for students to grapple with new ideas while applying their own prior knowledge, create mathematical agency within all students (low floor, high ceiling), force students to explain their reasoning, stir up debates and result in learning that can be applied to future situations.

Third, under the role of classroom discourse, several aspects emerged that could be used to guide the addition of items. During the cognitive interviews, two participants mentioned the need to add specific items that referenced both teacher action and student action simultaneously in the same sentence. In the open-ended section, participants mentioned discourse mainly in terms of emphasizing lesson goals, positioning teachers as leaders of instruction (direct instruction), and highlighting correctness of work. Others envisioned discourse in terms of the interactions between students for exploring and sharing thinking, questioning own ideas, as well as that of others, critiquing strategies, and making connections and generalizations, and extending and applying ideas to new situations. Others imagined discourse in terms of teacher facilitation of all the interactions that make learning possible, mainly conceiving it in terms of the use of questioning techniques to support learning, i.e., learning happens through conversations.

These findings from the open-ended and cognitive interviews have been used to begin part of the future work of refining the final survey in this study. Expressly, these ideas represent the first step of generating a larger pool of items, such that the relatively lower than commonly desired reliabilities of 0.80 could be significantly improved. Additionally, the larger pool will present opportunities to perform reliability and factor analysis tests without worrying that the deletion of some items would threaten the validity of the survey.

Given that the cognitive interviews and confirmatory study relied on validating this instrument solely based on sample participants with known high visions, future research should collect data from a larger sample of diverse participants (i.e., with high, medium, and low visions). The inclusion of such diverse participants in terms of vision can present opportunities to analyze what the consequential validity of the process was in designing this instrument using only subjects with sophisticated visions.

Apart from adding more items, and further validating this instrument, the next stage in this research would be developing a means of "scoring" and reporting the survey.

Holistic scoring will be the preferred method of making sense of participants' responses because it provides more in-depth information that does traditional analytical scoring.

Conclusions

This study described the development process for the instrument, which included a pilot phase and three other phases. Evidence of validity and reliability of the instrument was collected at each phase, including content validity, construct validity, predictive validity, correlation, and reliability testing. Feedback analysis and cognitive interviews were used at the content validity phase to improve on the pilot items. Utilizing the revised items, the factor structural within the dimensions was established, and it was found that all dimensions demonstrated acceptable levels of reliability. The outcome of the factor analysis and reliability test was subject to a confirmation study that demonstrated some aspects of predictive validity. In spite of some of the weaknesses mentioned in this study, the results of this study developed through a series of phases that employed various qualitative and quantitative approaches, indicate that the final survey 28-item instrument may be a useful tool for exploring mathematics educators' visions of math instruction.

REFERENCES

Artino Jr., A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher, 36*(6), 463–474.

Bell, C. A., Wilson, S. M., Higgins, T., & McCoach, D. B. (2010). Measuring the effects of professional development on teacher knowledge: The case of developing mathematical ideas. *Journal for Research in Mathematics Education,* 479–512.

Boston, M. D. (2012). Assessing instructional quality in mathematics. *Elementary School Journal, 113,* 76–104.

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education,* 119–156.

Bryk, A. S. (2010). Organizing schools for improvement. *Phi Delta Kappan, 91*(7), 23–30.

Burić, I., Slišković, A., & Macuka, I. (2018). A mixed-method approach to the assessment of teachers' emotions: development and validation of the Teacher Emotion Questionnaire. *Educational Psychology, 38*(3), 325–349.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal, 97*(1), 3–20.

Chance, P. L., & Segura, S. N. (2009). A rural high school's collaborative approach to school improvement. *Journal of Research in Rural Education (Online), 24*(5), 1–12.

Clarke, D. M. (1997). The changing role of the mathematics teacher. *Journal for Research in Mathematics Education, 28*(3), 278–308. doi:10.2307/749782

Cobb, P., & Jackson, K. (2011). Towards an empirically grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education and Development, 13*(1), 6–33.

Cobb, P., & Smith, T. (2008). District development as a means of improving mathematics teaching and learning at scale. *International handbook of mathematics teacher education, 3*, 231–254.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145–170.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher, 32,* 3–12.

Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The Common Core State Standards and implementation research. *Educational Researcher, 45*(4), 243–251

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*(3), 311–329.

Collins, D. (2003). Pretesting survey instruments: an overview of cognitive methods. *Quality of Life Research, 12*(3), 229–238.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98.

Creswell, J. W., & Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Cross, D. I., Hudson, R. A., Adefope, O., Lee, M. Y., Rapacki, L., & Perez, A. (2012). Success made probable: Creating equitable mathematical experiences through project-based learning. *Journal of Urban Mathematics Education, 5*(2), 55–86.

Darling-Hammond, L., Banks, J., Zumwalt, K., Gomez, L., Sherin, M. G., Griesdorn, J., & Finn, L. (2005). Educational goals and purposes: Developing a curricular vision for teaching. *Preparing teachers for a changing world: What teachers should learn and be able to do,* 169–200.

Daugherty, S. D., Harris-Kojetin, L., Squire, C., & Jael, E. (2001). *Maximizing the quality of cognitive interviewing data: An exploration of three approaches and their informational contributions.* Proceedings of the Annual Meeting of the American Statistical Association.

de Winter, J. C., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods, 21*(3), 273.

Drake, C., & Sherin, M. G. (2009). Developing curriculum vision and trust. *Mathematics teachers at work: Connecting curriculum materials and classroom instruction,* 321–337.

Fennema, E., Franke, M. L., Carpenter, T. P., & Carey, D. A. (1993). Using children's mathematical knowledge in instruction. *American Educational Research Journal, 30*(3), 555–583.

Fila, N. D., Hess, J. L., & Purzer, S. (2015). *A dialectic data integration approach for mixed methods survey validation.* In 122th ASEE Annual Conference and Exposition (14-17 June), American Society for Engineering Education, Seattle, WA.

Finfgeld-Connett, D. (2014). Use of content analysis to conduct knowledge-building and theory-generating qualitative systematic reviews. *Qualitative Research, 14*(3), 341–352.

Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., & Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National Society for the Study of Education, 112*(2), 136–156.

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal, 38*(3), 653–689.

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). Greenwich, CT: Information Age Publishing.

Gamoran, A. (Ed.). (2003). *Transforming teaching in math and science: How schools and districts can support change.* Teachers College Press.

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology, 15*(4), 380–387.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston. MA: Allyn & Bacon.

Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales.* Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.

Greene, J. C. (2007). *Mixed methods in social inquiry.* San Francisco, CA: Jossey-Bass.

Greene, J. C., & Caracelli, V. (1997). Defining and describing the paradigm issue in mixed method evaluation. *New Directions for Evaluation, 74,* 5–17.

Greene, J. C., Caracelli, V., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255–274.

Greene, J. C., & McClintock, C. (1985). Triangulation in evaluation: design and analysis issues. *Evaluation Review, 9*(5), 523–545.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25*(1), 1–29.

Hammerness, K. (2001). Teachers' visions: The role of personal ideals in school reform. *Journal of Educational Change, 2*(2), 143–163.

Hammerness, K. (2003). Learning to hope, or hoping to learn? The role of vision in the early professional lives of teachers. *Journal of Teacher Education, 54*(1), 43–56.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte, NC: Information Age.

Hill, H. C. (2001). Policy is not enough: Language and the interpretation of state standards. *American Educational Research Journal, 38*(2), 289–318.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, *1*(1), 104–121.

Hippler, H. J., Schwarz, N., & Sudman, S. (Eds.). (1987). *Social information processing and survey methodology.* New York, NY: Springer Verlag.

Ho, K. F., & Tan, P. (2013). Developing a professional vision of classroom practices of a mathematics teacher: views from a researcher and a teacher. *Teaching Education, 24*(4), 415–426.

Honig, M., & Hatch, T. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher, 33*(8), 16–30.

Hufferd-Ackles, K., Fuson, K. C., & Sherin, M. G. (2004). Describing levels and components of a math-talk learning community. *Journal for Research in Mathematics Education, 35*(2), 81–116. doi:10.2307/30034933

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods, 18*(1), 3.

Jabine, T. B., Straf, M. L., Tanur, J. M., & Tourangeau, R. (Eds.). (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines.* Washington, DC: National Academy Press.

Jackson, K., & Cobb, P. (2013). Coordinating professional development across contexts and role groups. In M. Evans (Ed.)*, Teacher education and pedagogy: Theory, Policy and Practice* (pp. 80–99). Cambridge, UK: Cambridge University Press.

Jackson, K., Cobb, P., Wilson, J., Webster, M., Dunlap, C., & Appelgate, M. (2015). Investigating the development of mathematics leaders' capacity to support teachers' learning on a large scale. *ZDM, 47*(1), 93–104.

Jackson, K., Garrison, A., Wilson, J., Gibbons, L., & Shahan, E. (2013). Exploring relationships between setting up complex tasks and opportunities to learn in concluding whole-class discussions in middle-grades mathematics instruction. *Journal for Research in Mathematics Education, 44*(4), 646–682.

Jackson, K., Gibbons, L., & Sharpe, C. J. (2017). Teachers' views of students' mathematical capabilities: Challenges and possibilities for ambitious reform. *Teachers College Record, 119*(7), 1–43. Retrieved from https://www.tcrecord.org/Content.asp?ContentId=21791

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly, 68*(1), 94–101.

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., . . . Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*(3), 139–151.

Kersting, N. B., Givvin, K., Sotelo, F., & Stigler, J. W. (2010). Teacher's analysis of classroom video predicts student learning of mathematics: Further explorations of a novel measure of teacher knowledge. *Journal of Teacher Education, 61*, 172–181.

Kersting, N. B., Givvin, K. B., Thompson, B. J., Santagata, R. & Stigler, J. W. (2012). Measuring usable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–589.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*(1), 29–63.

Lampert, M., Beasley, H., Ghousseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In *Instructional explanations in the disciplines* (pp. 129–141). Springer US.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84.

McDonald, M., Kazemi, E., & Kavanagh, S. S. (2013). Core practices and pedagogies of teacher education: A call for a common language and collective activity. *Journal of Teacher Education, 64*(5), 378–386.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*(2), 171–178.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741.

Midgley, C., Anderman, E., & Hicks, L. (1995). Differences between elementary and middle school teachers and students: A goal theory approach. *The Journal of Early Adolescence, 15*(1), 90–113.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017.

Munter, C. (2014). Developing visions of high-quality mathematics instruction. *Journal for Research in Mathematics Education, 45*(5), 584–635.

Munter, C. (2015). Envisioning the role of the mathematics teacher. *NCSM Journal of Mathematics Education Leadership, 16*(1), 29–40.

Munter, C., & Correnti, R. (2017). Examining relations between mathematics teachers' instructional vision and knowledge and change in practice. *American Journal of Education, 123*(2).

Munter, C., Stein, M. K., & Smith, M. A. (2015). Dialogic and direct instruction: Two distinct models of mathematics instruction and the debate(s) surrounding them. *Teachers College Record, 117*(11).

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2014). *Principles to actions: Ensuring mathematical success for all.* Reston, VA: The National Council of Teachers of Mathematics.

National Governor's Association, Common Core State Standards Initiative. (2011). *Common Core State Standards in Mathematics.* Retrieved from http://www.corestandards.org/Math

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education.

National Research Council. (1996). *National Science Education Standards.* Washington, DC: National Academy Press.

Nicholas, N., Hess, J., & Purzer, S. (2015). *A dialectic data integration approach for mixed-method survey validation.* 122nd ASEE Annual Conference & exposition.

O'Day, J. A., & Smith, M. S. (1993). Systemic reform and educational opportunity. In S. Fuhrman (Ed.), *Designing coherent education policy: Improving the system* (pp. 250–312). San Francisco, CA: Jossey-Bass.

Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research, 4*(1), 56–78.

Parrish, C. (2016). *Supporting the development of teachers' attributes needed for the selection and implementation of cognitively demanding tasks through engagement with the MathTwitterBlogosphere* (doctoral dissertation, Auburn University). Retrieved from https://etd.auburn.edu/bitstream/handle/10415/5330/Parrish_ Dissertation.pdf?sequence=2&isAllowed=y

Penuel, W. R., Coburn, C. E., & Gallagher, D. J. (2013). Negotiating problems of practice in research-practice design partnerships. *Yearbook of the National Society for the Study of Education, 112*(2), 237–255.

Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology, 30*(4), 59–64.

Redline, C., Smiley, R., Lee, M., & DeMaio, T. (2001, August 5-9). *Beyond concurrent interviews: An evaluation of cognitive interviewing techniques for self-administered questionnaires.* Proceedings of the Annual Meeting of the American Statistical Association.

Ross, J. A., Hogaboam-Gray, A., & McDougall, D. (2002). Research on reform in mathematics education, 1993-2000. *Alberta Journal of Educational Research, 48*(2), 122.

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. doi:10.2307/30034787

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education, 42*(2), 123–149.

Schwarz, N., & Sudman, S. (Eds.). (2012). *Context effects in social and psychological research.* Springer Science & Business Media.

Schuman, H., & Presser, S. (1981). *Questions and answers: Experiments on question form, wording, and context in attitude surveys.* New York, NY: Academic.

Sherin, M. G. (2001). Developing a professional vision of classroom events. Beyond classical pedagogy: Teaching Elementary School Mathematics, 75–93.

Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction, 20*(2), 119–150.

Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 383–395). Hillsdale, NJ: Erlbaum.

Sherin, M. G., & Han, S. (2004). Teacher learning in the context of a video club. *Teaching and Teacher Education, 20,* 163–183.

Sherin, M. G., Russ, R. S., Sherin, B. L., & Colestock, A. (2008). Professional vision in action: An exploratory study. *Issues in Teacher Education, 17*(2), 27–46.

Sherin, M. G., & Van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60*(1), 20–37.

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

Silver, E. A., Mesa, V. M., Morris, K. A., Star, J. R., & Benken, B. M. (2009). *Teaching mathematics for understanding: An analysis of lessons submitted by teachers seeking NBPTS certification.*

Smith, M. S. (2000). Balancing old and new: An experienced middle school teacher's learning in the context of mathematics instructional reform. *The Elementary School Journal, 100*(4), 351–375.

Smith, M. S., & Stein, M. K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School, 3*(5), 344–350.

Spillane, J. P. (1998). State policy and the non-monolithic nature of the local school district: Organizational and professional considerations. *American Educational Research Journal, 35*(1), 33–63.

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, *18*(2), 141–179.

Spillane, J. P., Reiser, B., & Gomez, L. M. (2006). Policy implementation and cognition: The role of human, social, and distributed cognition in framing policy implementation. In M. I. Honig (Ed.), *New directions in educational policy implementation: Confronting complexity* (pp. 47–64). Albany: State University of New York Press.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*(3), 387–431.

Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, *19*(2), 185–203.

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis, 21*(1), 1–27

Staples, M. (2007). Supporting whole-class collaborative inquiry in a secondary mathematics classroom. *Cognition and Instruction*, *25*(2–3), 161–217. doi:10.1080/07370000701301125

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning, 10*(4), 313–340.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488.

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development* (1st ed.). New York, NY: Teachers College Press.

Stein, M. K., Smith, M. S., Henningsen, M., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development* (2nd ed.). New York, NY: Teachers College Press.

Stein, M. K., Smith, M. S., & Silver, E. A. (1999). The development of professional developers: Learning to assist teachers in new settings in new ways. *Harvard Educational Review, 69*, 237–269.

Sullivan, P., Clarke, D., & Clarke, B. (2009). Converting mathematics tasks to learning opportunities: An important aspect of knowledge for mathematics teaching. *Mathematics Education Research Journal, 21*(1), 85–105.

Supporting the Common Core State Standards in Mathematics. NCTM position Statement, 2013. http://www.nctm.org/Standards-and-Positions/Position-Statements/Supporting-the-Common-Core-State-Standards-for-Mathematics/

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics* (Vol. 6). Boston, MA: Pearson.

Tashakkori, A. (2006). *Mixed methodology: Combining qualitative and quantitative approaches.* Sage.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences.* Sage Publications, Inc.

Thompson, C. L., & Zeuli, J. S. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond & G. Sykes

(Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 341–375). San Francisco, CA: Jossey-Bass.

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2012). *Elementary and middle school mathematics: Teaching developmentally* (8th ed.). Upper Saddle River, NJ: Pearson Education.

Van Es, E. A. (2012). Using video to collaborate around problems of practice. *Teacher Education Quarterly, 39*(2), 103–116.

Van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24,* 244–276.

Wager, A. A. (2012). Incorporating out-of-school mathematics: From cultural context to embedded practice. *Journal of Mathematics Teacher Education, 15*(1), 9–23.

Walkowiak, T., Lee, C., & Whitehead, A. (2015, April 15). *The development of preservice teachers' visions of mathematics instruction.* Paper presented at the Research Conference of the National Council of Teachers of Mathematics, Boston, MA.

Weick, K. E. (1995). *Sensemaking in organizations.* Thousand Oaks, CA: Sage.

Wiemers, N. J. (1990). Transformation and accommodation: A case study of Joe Scott. *Educational Evaluation and Policy Analysis, 12*(3), 281–292.

Wilhelm, A. G. (2014). Mathematics teachers' enactment of cognitively demanding tasks: Investigating links to teachers' knowledge and conceptions. *Journal for Research in Mathematics Education, 45*(5), 636–674.

Wilkerson, T. L., Eddy, C. M., Quebec Fuentes, S., Sorto, M. A., Gupta, D., Ward, E. K., . . . Kerschen, K. (2018). Development and validation of the algebra teachers' self-efficacy instrument: Assessment of algebra teachers' knowledge and personal teaching efficacy. *School Science and Mathematics, 118*(6), 206–217.

Willis, G. B. (1999). Cognitive interviewing: A "how to" guide. Research Triangle Institute. 1999 Meeting of the American Statistical Association. Research Triangle Park, NC: Research Triangle Institute.

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design.* Thousand Oaks, CA: Sage.

Willis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design.* Thousand Oaks, CA: Sage.

Woolley, S. L., Benjamin, W. J. J., & Woolley, A. W. (2004). Construct validity of a self-report measure of teacher beliefs related to constructivist and traditional approaches to teaching and learning. *Educational and Psychological Measurement, 64*(2), 319–331.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838.

APPENDIX A

SURVEY OF MATH INSTRUCTIONAL VISION [PILOT]

To help us better understand your vision for ideal mathematics teaching and the ways it relates to other NC mathematics education stakeholders, please answer the following three sets of questions related to your ideal vision of what teachers are doing, what students are engaging in, and what forms of classroom discourse you see in the ideal mathematics classroom.

Your ideal classroom may be somewhat—or very— different from current mathematics classrooms. Imagine what you see as you look around this ideal mathematics classroom as you answer each question:

1. If you were to observe an ideal mathematics classroom for more than one day, what are teachers doing in the ideal mathematics classroom? (Pick the three statements that best characterize what you would want to see in the ideal classroom you have in mind)

In an ideal mathematics classroom, I most want to see that teachers are . . .
☐ Making connections with students.
☐ Demonstrating/leading discussions on how to solve problems.
☐ Allowing students to work collaboratively.
☐ Allowing students to discover mathematical insights.
☐ Problematizing student’s prior conceptions of mathematical ideas.
☐ Creating and introducing interesting math tasks.
☐ Providing time for students to practice.
☐ Giving students authority to develop their own strategies/solutions.
☐ Motivating students by captivating their attention.
☐ Ensuring student’s work is thorough and correct.
☐ Encouraging students to explain to one another.
☐ Eliciting and scaffolding students’ ideas.
☐ Refraining from providing mathematical facts, formulas, or ideas.
☐ Ensuring that students stay on productive mathematical paths.
☐ Allowing students opportunities to ask questions.

2. If you were to observe an ideal mathematics classroom for more than one day, which of the following would best characterize the ideal tasks, you would see students engaging with in this ideal classroom? (Pick the 2 statements that best characterize the ideal classroom you have in mind)

In an ideal mathematics classroom, I most often want to see tasks that . . .

☐ Engage students with mathematical structure or strategies.

☐ Engage students in hands-on activities.

☐ Foster opportunities for students to create meaning of the use of mathematics.

☐ Engage students in complex thinking and allow for multiple solution paths.

☐ Engage students in practicing procedures.

☐ Foster opportunities for students to practice procedures prior to engaging in more conceptually challenging tasks.

☐ Foster opportunities for students to show how math is applicable in a “real-world” context.

☐ Foster opportunities for students to make connections between strategies or representations.

3. If you were to observe an ideal mathematics classroom for more than one day, what are teachers and/or students saying in the ideal mathematics classroom? (Pick 3 statements that best characterize the ideal classroom you have in mind)
In an ideal mathematics classroom, I would most like to hear that . . .

☐ Teachers are inviting students to respond to carefully sequenced questions focused on lesson goals.

☐ Teachers are prompting and pressing students on the mathematical steps in their work.

☐ Teachers are asking questions to uncover students’ mathematical steps.

☐ Teachers are asking conceptually oriented questions to uncover students’ thinking.

☐ Teachers are supporting students in understanding procedures or calculations.

☐ Teachers are asking questions that lead to further investigations or revised conjectures.

☐ Teachers are asking conceptually oriented questions that encourage students to explain their approaches, procedures, or calculations.

☐ Students are sharing their solutions in whole class conversations and/or small groups.

☐ Students are sharing their mathematical strategies or ideas in small groups/partners.

☐ Students are explaining their mathematical approaches, procedures, or calculations.

☐ Teachers are evaluating students’ answers to sequenced questions focused on lesson goals.

☐ Students are responding to carefully sequenced questions focused on the lesson goals.

APPENDIX B

COGNITIVE INTERVIEW PROTOCOL—THINK ALOUDS

MAKE SURE YOU HIT "RECORD" BEFORE YOU BEGIN THE INTERVIEW

Thank you for agreeing to have a conversation with me today. The purpose of this interview is to make sure that the items we have got written in the survey are eliciting the things that we think they are. This interview is voluntary, any time you want to stop, please stop. I won't share this information with anybody else, and when I do share in terms of research, it will be anonymous. You'll never be identified. At any point that you do not want this to be used as data even after this, please let me know. So to start, do you have any questions about what we are doing?

Background Information – To make the interviewee comfortable

- Can you share with me a little bit about your background as a math educator?
  - What is your role?
  - When did you start to teach/work at this school/district?
  - Which grades and courses have you taught? What roles have you performed?
  - How many courses do you typically teach every semester? How about this semester?
- How would you describe your context?
  - How would you describe the students you work with?
  - Tell me about the circumstances that led you to choose to teach as a career?
  - What did you do to prepare yourself for it? (Courses, study programs, workshops, etc.)

Survey Questions

I have got three questions, and I'm going to share with you the text that comes before the questions, and I'm going to let you read it, and if anything about it is unclear, or you have some questions, let me know, then we will take each question at a time, and I will let you read it, and then as you are responding to it, I will ask you questions like: what are you thinking about, why did you pick that? I'm not asking you questions because you are wrong; there is no right or wrong answer. I want to know how you are thinking about it, so if I ask a lot of questions, know that I'm interested in what you have to share, and it's not because you are picking the wrong things. Is that okay? Does that make sense?

Introduction paragraph to the survey

I'm going to let you read the first two paragraphs of the survey, and then we will talk about it in a little bit.

- So, tell me about the first two paragraphs, were they clear?
  - In your own words, tell me what the survey is trying to find out?
  - Do you know what we are trying to do?

Question 1, 2 & 3 (Use similar probes for all items)
So here is the first question. I will let you read it out loud.

- So in your own words, what is the question asking you to do?
- I'm going to let you talk and respond to this question as though you were taking the survey. Talk as much as you want. I would like to know what you are thinking about as you make your selections.

Use the following probes as the participant makes selections

- What are you thinking?
- What about each of these statements isn't clear?
  - Were there words or phrases that were unclear?
  - What is troubling or confusing?
- So if you are to pick <u>three</u> statements about what characterizes what you want to see, can you talk about your process.
  - How did you go about making your selections? What strategy did you use to make your selections?
  - Which ones would you pick? Can you tell me why you chose the ones that you did?
  - What made you not pick particular statements? Can you talk about what you eliminated?
  - So these others that you did not choose were there any that you deliberated about? Were they clearly not the choices you wanted to make?
  - How confident are you with your selections?
- Is there any feedback you would like to give about any item?
  - How might we pause the item differently?
  - Is there a better way to write that item?
  - How might it be written differently, so it makes it easy for you to share your vision of the ideal classroom?
- How easy or difficult did you find the questions in this survey? Why do you say that?

After the participant reads questions 2 and 3, ask the following question below. Then use the probes above.

- What is the difference between question 2 and question 1?
- What is the difference between question 3 and the other questions?

Closing

- Is there anything else you would like to tell me about the survey?
- Is there anything you would like to ask me?

Again, thank you for your time and sharing with me. After I leave today if you think of anything else that you wish you had said, please don't hesitate to contact me.

APPENDIX C

ANALYTICAL MEMO: MATHEMATICS INSTRUCTIONAL [VISION] SURVEY

Participant Name: Mercy
Participant Role: Mathematics Coach

Overall impressions or insights from the interview: Mercy has been in education for 25 years, has taught elementary, participated in many math workshops and so claims to have a strong math background. Later move into curriculum, and later in Math coaching or specialist (Elementary Mathematics). She does understand the whole essence of the survey. It was hard to narrow down to three or two questions. The strategy she used to make her selection involved analyzing to see which items could be grouped together. For example, she stated that under the teacher role, she thought about what the teacher’s job is. So creating and introducing interesting mathematical facts is important because in order for the class to look the way she envisions, the task has to look a certain way, she would be able to make students do the other things if the task was not right. So by looking at all the items under “role of teacher,” it seems like she is picking those items that are key components to any lesson. She claims that the teacher should spend time thinking about the mathematics She is offering to the students. She does group allowing students to discover mathematical insights with giving students authority. She also groups the item on “collaborate” with “encouraging students to explain to one another.” She also stated that eliciting and scaffolding students’ ideas was important to her. Seems like if she was given an opportunity she would have picked more items that represent what her vision of the role of teacher is. Her three choices under role of teacher seem to reflect the typical launch-explore-discuss lesson format. She picks the Task (launch), collaborate (Explore), Discuss (Give authority). She was not confused about the words “engage” and “foster.” Under classroom discourse Mercy hesitated a lot when she was reading the items that talked about the students, probably because of the idea that some items said that students were working in groups, while some did not explicitly say anything. She states that what is missing from this section is the student-to-student interactions. According to her, this part of the survey seen to only capture what is happening between teachers and students and misses to capture explicitly the discourse amongst students. She gives an example that of what could be included: “students are listening and responding to others ideas.” This was a good exercise for her to reflect on what she values. The first question was the hardest, because there is a lot of things she would like to see teachers doing, but she had to focus on the items that talked about “the math.” This survey is an opportunity for her to as a coach, knowing that a class was not ideal, and then she would know where to begin. According to Mercy, this survey could be used in PD with

teachers to gather information about where to begin, to think about what help teachers need, to prioritize which teachers to work with.	
Participant stated instructional vision: Believes in students making sense of math and not being taught procedures	
Survey Items	Feedback
Role of teacher	
Making connections with students	We should be doing this all day long. She wants to see this all the time. Its important overall. Connections is making connections both personally and probably mathematically
Demonstrating/leading discussions on how to solve problems	"This sounds teacher focused to me." The word "demonstrate" takes authority from the students and gives it to the teacher.
Allowing students to work collaboratively	This is a good practice
Allowing students to discover mathematical insights	Wants students to take ownership of their learning.
Problematizing student's prior conceptions of mathematical ideas.	Not sure what this means.
Creating and introducing interesting math tasks.	Yes
Providing time for students to practice.	Not what she wants to see, but understands what it means
Giving students authority to develop their own strategies/solutions	She wants to see this
Motivating students by captivating their attention.	This should be happening all the time.
Ensuring student's work is thorough and correct	Not all the time
Encouraging students to explain to one another.	Yes
Eliciting and scaffolding students' ideas.	Yes
Refraining from providing mathematical facts, formulas, or ideas.	If it is things students cannot conceptually figure out then the teacher needs to provide it for them. Facts are different from formulas or ideas. She wants to give students facts. She suggests taking out the word "facts" from the item.
Ensuring that students stay on productive mathematical paths.	Views mathematical path in terms of the path followed in solving a problem. She would ask questions to ensure that students stay on a productive path. What might look like an

	unproductive path could really turn out to lead to deep learning.
Allowing students opportunities to ask questions.	Definitely
Instructional Tasks	
Engage students with mathematical structure or strategies.	It implies more practice with particular strategies, but not what they should always do
Engage students in hands-on activities.	Good, they need to have their hands-on things.
Foster opportunities for students to create meaning of the use of mathematics.	Yes
Engage students in complex thinking and allow for multiple solution paths.	Yes, math is about solving problems and making sense
Engage students in practicing procedures.	Important at times to build fluency
Foster opportunities for students to practice procedures prior to engaging in more conceptually challenging tasks.	No, conceptual is important then followed by procedural fluency
Foster opportunities for students to show how math is applicable in a “real-world” context	Yes
Foster opportunities for students to make connections between strategies or representations.	Yes
Classroom Discourse	
Teachers are inviting students to respond to carefully sequenced questions focused on lesson goals.	We are not to force students into a particular pathway. It implies that the teacher is looking for specific answers
Teachers are prompting and pressing students on the mathematical steps in their work.	Maybe
Teachers are asking questions to uncover students’ mathematical steps.	Like this than the one before
Teachers are asking conceptually oriented questions to uncover students’ thinking.	Yes, wants to see this
Teachers are supporting students in understanding procedures or calculations.	No

Teachers are asking questions that lead to further investigations or revised conjectures.	Yes
Teachers are asking conceptually oriented questions that encourage students to explain their approaches, procedures, or calculations.	Yes, but not really.
Students are sharing their solutions in whole class conversations and/or small groups.	Has an issue with the word “solution.” There is a difference between sharing “solutions” and “strategies.” Solution means we are focused on the answers. Strategy means we are focused on what is happening. She gives an example of her classes where if she notices that the students are focused on the solution, then she would write it on the board, and the ask the students to figure out how to get to that solution.
Students are sharing their mathematical strategies or ideas in small groups/partners.	This is what she wants. She likes this one more than the one above because the focus should be on the strategies and not the solution. She suggests that we should add the term “whole group” on this item too.
Students are explaining their mathematical approaches, procedures, or calculations.	This one does not specify whether it is small group/whole class, etc.
Teachers are evaluating students’ answers to sequenced questions focused on lesson goals.	She did not comment on this.
Students are responding to carefully sequenced questions focused on the lesson goals.	Teacher is not getting at students’ thinking.
Key ideas for items for revision: We need to change the format	
Parts of the interview to look at carefully: Look at the items about discourse (that mention student)	
Anything else that needs to be shared:	

APPENDIX D

CONFIRMATORY STUDY FOLLOW UP INTERVIEW PROTOCOL

MAKE SURE YOU HIT “RECORD” BEFORE YOU BEGIN THE INTERVIEW

Thank you for agreeing to have a conversation with me today. The purpose of this interview is to gain an insight of how we can improve on the survey that you participated in. This interview is totally voluntary, any time you want to stop, please stop. Specifically, the goal is to capture mathematics educators’ instructional vision. There are no right or wrong answers, or desirable or undesirable answers. I would like you to feel comfortable saying exactly what you are thinking.
With your permission, I would like to record this interview so that I can concentrate on what you are saying rather than focus on writing everything down. I won’t share this information with anybody else and when I do share in terms of research it will be anonymous. You’ll never be identified. At any point that you do not want this to be used as data even after this, please let me know. So to start, do you have any questions about what we are doing?

Background Information – To make interviewee comfortable

- Can you share with me a little bit about your background as a math educator?
  - How did you get into the teaching field?
  - What is your role?
  - When did you start to teach/work at this school/district?
  - Which grades and courses have you taught? What roles have you performed?
  - How many courses do you typically teach every semester? How about this semester?
- How would you describe your context?
  - How would you describe the students you work with?
  -

Capturing Vision – through Interview
If you were asked to observe a variety of math classroom for one or more lesson, what are some important things you would look for to decide whether the practices were suggestive of high-quality instruction?

- What can you describe as “good teaching” and “productive mathematics learning”?
    - Can you give specific/some examples of what this will look like?
- What are some of the things the teacher should actually be doing?
- Can you please describe what classroom discussion would look and sound like?
- What types of task do you think the teacher should be using?

If now you were provided with a set of observations collected from these classrooms, sort them out to decide which ones suggest higher quality than the others. This part will be divided into three sections

Role of Teacher (provide cards for each item)

- How do you make sense of the following practices? Do you see any distinctions between the following practices? Which ones suggest higher quality of instruction?
  "I see the teacher engaging students in discussions of how to solve problems" and "I see the teacher encouraging students to explain how to solve problems to each other"

How would you make sense of the following practices?

- "I see the teacher providing time for the students to collaborate and solve assigned mathematical problems."
- "I see the teacher engaging students in reflecting on their prior conceptions of mathematical ideas."

I'm now going to give you a pile of cardstocks with instructional practice on them. Is it possible to make piles according to practices that you think have the same level of quality? Talk as much as you want. I would like to know what you are thinking about (Provide the participant with card stock of level 1 and 2 items, then level 3 and 4 items, and then have them do sorting –Do both "open sorting" and "closed sorting")

- Sort the cards according to which practices you think belong together (open sorting).
- Now sort the cards into exactly two groups (closed sorting).

Use the following probes as the participant makes selections

- What are you thinking?
- What about each of these statements ins't clear?
  - Were there words or phrases that are unclear?
  - What is troubling or confusing?
- How did you decide which item belongs to a particular group?
  - How did you go about making your selections? What strategy did you use to make your selections?
  - How confident are you with your selections?
- Which group suggests a higher quality of instruction than the other?
  - How confident are you with your selections?

Role of Tasks (Use similar probes)

Role of Tasks (provide cards for each item)

- How do you make sense of the following practices? Do you see any distinctions between the following practices? Which ones suggest higher quality of instruction?

“I see the teacher engaging using math tasks that engage student in hands on activities”
“I see the teacher engaging using math tasks that engage student in creating meaning of the use of mathematics”

- How would you make sense of the following practices?
  “I see the teacher using tasks that engage the student in complex thinking and allow for multiple solution paths.”

I’m now going to give you a pile of cardstocks with instructional practice on them. Is it possible to make piles according to practices that you think have the same level of quality? Talk as much as you want. I would like to know what you are thinking about (Provide the participant with card stock of level 1 and 2 items, then level 3 and 4 items, and then have them do sorting – Do both “open sorting” and “closed sorting”)

- Sort the cards according to which practices you think belong together (open sorting).
- Now sort the cards into exactly two groups (closed sorting).

Use the following probes as the participant makes selections

- What are you thinking?
- What about each of these statements isn’t clear?
  - Were there words or phrases that are unclear?
  - What is troubling or confusing?
- How did you decide which item belongs to a particular group?
  - How did you go about making your selections? What strategy did you use to make your selections?
  - How confident are you with your selections?
- Which group suggests a higher quality of instruction than the other?
  - How confident are you with your selections?

Role of Discourse (Use similar probes) ****Level 2 and 3 also have some overlap
Role of Discourse (provide cards for each item)

- How do you make sense of the following practices? Do you see any distinctions between the following practices? Which ones suggest higher quality of instruction?
  “I see and hear classroom discourse where students are sharing their ideas with partners in or in small groups”
  “I see and hear classroom discourse where teachers are supporting students in understanding their procedures or calculations”
  “I see and hear classroom discourse where teachers are asking questions to uncover students’ mathematical steps”

I’m now going to give you a pile of cardstocks with instructional practice on them. Is it possible to make piles according to practices that you think have the same level of quality? Talk as much as you want. I would like to know what you are thinking about (Provide the participant with card stock of level 1 and 2 items, then level 3 and 4 items, then level 2 and 3 items, and then have them do sorting –Do both "open sorting" and "closed sorting")

- Sort the cards according to which practices you think belong together (open sorting).
- Now sort the cards into exactly two groups (closed sorting).

Use similar probes as other sections

CLOSING

- Is there any feedback you would like to give about any items mentioned above?
    - How might we pause the item differently? Is there a better way to write that item?
    - How might it be written diffently so it makes it easy for you to share your vision of the ideal classroom?
    - Are there other items (practices) that you think should be added that would make it easy for you to share your vision of ideal classroon instruction?
- Is there anything you would like to ask me?

Again, thank you for your time and sharing with me. After I leave today, if you think of anything else that you wish you had said, please don't hesitate to contact me.