Investigating the Dimensionality of Teachers' Mathematical Knowledge for Teaching Secondary Mathematics Using Item Factor Analyses and Diagnostic Classification Models

by

Inah Ko

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Educational Studies)
in the University of Michigan
2019

Doctoral Committee:

Professor Patricio G. Herbst, Chair
Professor Ben B. Hansen
Associate Research Scientist Mark Hoover
Associate Research Scientist Stephen Gerard Schilling
Professor Edward A. Silver

Inah Ko

inahko@umich.edu

ORCID iD: 0000-0002-5348-5879

Dedication

To my family

Acknowledgements

Throughout this dissertation-writing journey, I have benefited from extraordinary help and support. First, I would like to express my deepest thanks to my advisor, Professor Patricio Herbst, who patiently provided guidance and encouragement through each stage of the process. His expertise and insights into the research topic and methodology have been invaluable in the development and completion of my dissertation. All that he has shared with me as my advisor, mentor, and an outstanding researcher has been a great inspiration.

I would also like to thank other committee members, Professor Edward Silver, Dr. Mark Hoover, and Professor Ben Hansen. Their feedback and suggestions have helped me improve my dissertation immensely. In addition, I owe a special thank you to Dr. Stephen Schilling, also a committee member, for generously sharing his valuable time and for all the suggestions he provided that enabled me to develop the dissertation's methodological strength.

My thanks also go to Christine Feak, who was always willing to help me with my writing for the dissertation. Her constructive advice and astute suggestions have been extremely helpful and allowed me to understand the genre of academic writing and the elements I need to be aware of when writing a research paper.

My colleagues and friends also deserve my sincere thanks. They offered consistent academic and emotional support as I made this journey. I am grateful to the current and prior members of the GRIP research group, Amanda Milewski, Nicolas Boileau, Mollee Shultz, Emanuele Bardelli, Michael Ion, Claudine Margolis, Jacqueline Benczarski, Srikanth Lavu, Umut Gursel, Vu Minh Chieu, Kristi Hanby, Ander Erickson, and Justin Dimmel. Their continuing support and encouragement from my first day in the GRIP have been a great help throughout my PhD program. I would also like to recognize the support and friendship of Sara Ryoo, Minsung Kwon, Seoyoon Woo, and Yeon Kim, whose empathy and caring when I needed emotional support helped me reach my goal. I am blessed to have great friends in both the US and Korea who supported me not only through difficult days but also sleepless nights.

Last but not least, my heartfelt appreciation and gratitude go to my family. My mother, Jinyeong Hong, has given me unconditional love and made sacrifices to allow me to follow my dreams. My sister, Inyeong Ko, and brother-in-law, Geunmin Ryu, have offered me consistent encouragement, academic and emotional support, and wisdom since the day I decided to undertake this journey. I must also acknowledge my wonderful nephew, Aiden Ryu, who has kept my spirits up with the greatest fun and smiles. Finally, I wish to thank my late father, Thongsung Ko, with all my heart. I finished this journey because I know he is always watching over me and believing in me, and I would like to tell him that, now, I understand his joy in research and I share it.

Table of Contents

Dedication **ii**

Acknowledgements **iii**

List of Tables **ix**

List of Figures **xii**

List of Abbreviations **xv**

Abstract **xvii**

Chapter 1 Objective of the Study **1**

1.1 Introduction **1**

1.1.1 A descriptive and large scale study on teachers' knowledge 2

1.1.2 The need of an alternative knowledge framework 3

1.1.3 Organization of the dissertation 5

1.2 Background **6**

1.2.1 Conceptualizing mathematical knowledge for teaching as a multidimensional construct 12

1.2.2 The role of multidimensional understanding in teacher education 17

1.2.3 The need for re-conceptualizing the dimensionality of mathematical knowledge for teaching 20

1.2.4 The need for re-operationalizing the dimensionality of mathematical knowledge for teaching 23

1.3 Guiding research question **26**

1.4 Chapter summary **27**

Chapter 2 Review of Concepts and Methods Used in Literature **28**

2.1 Operationalization of teachers' mathematical knowledge 28

2.1.1 Frameworks distinguishing CK and PCK 28

2.1.2 Differences among the frameworks 30

2.1.3 Dimensionality of secondary school teachers' mathematical knowledge 35

2.1.4 Dimensionality of knowledge in other professions 37

2.2 Dimensionality analysis in the studies of teachers' mathematical knowledge 40

2.2.1 IFA within SEM and (unidimensional) IRT frameworks 41

2.2.2 Bi-factor model 43

2.2.3 Multidimensional IRT 44

2.2.4 Item vector plot 46

2.2.5 Diagnostic Classification Models 47

2.3 Chapter summary 49

2.4 Research Questions 50

Chapter 3 Proposed Item Blueprint 52

3.1 The foundations of the hypothesis on the framework 52

3.2 Operational definition of the knowledge construct 55

3.3 The two organizers and the conceptual rationale of the blueprint 56

3.3.1 Organizer One: Task of Teaching 59

3.3.2 Organizer Two: Instructional Situation 66

3.4 Chapter summary 70

Chapter 4 Method 71

4.1 Instruments 71

4.1.1 MKT-G items 73

4.1.2 SMK-G items 74

4.1.3 MKT-A items 74

4.2 Participants 75

4.3 Data Analysis 76

4.3.1 Missing values 78
4.3.2 Item content analysis 79
4.3.3 Tasks of Teaching 81
4.3.4 Instructional Situations 86
4.3.5 Dimensionality of item scores 95

4.4 Chapter summary 98

Chapter 5 Results: Validation of the Dimensions Proposed in the Framework 99

5.1 Dimensionality within the same task of teaching 101

5.1.1 Organization of the items within the task USW 101
5.1.2 Organization of the items within the task CGP 123

5.2 Dimensionality within the same instructional situation 138

5.2.1 Organization of the items within the instructional situation CG 139
5.2.2 Organization of the items within the instructional situation DP 148
5.2.3 Organization of the items within the instructional situation EF 157

5.3 Dimensionality within the same course of studies 165

5.3.1 Organization of the items within geometry 165
5.3.2 Organization of the items within algebra 176

5.4 Differences in geometry knowledge profile proportions 186

5.4.1 Knowledge profile proportions within the same task of teaching 186
5.4.2 Knowledge profile proportions within the same instructional situation 189

5.5 Differences in algebra knowledge profile proportions 193

5.6 Differences in a relationship with educational and teaching experience 194

5.7 Chapter summary 198

Chapter 6 Discussion and Conclusion 200

6.1 Conclusion and summary of findings 201

6.1.1 The role of the two organizers in measuring dimensions of the knowledge 201
6.1.2 Interpretations on the indistinguishable dimensions 208

6.1.3 Different relationships with teachers' background among dimensions 215

6.2 Implications 218

6.2.1 Implications for the development of theory of mathematics teaching 218

6.2.2 Implications for the practice of teacher education 219

6.2.3 Implications for the theory of measurement 221

6.2.4 Implications for item development 223

6.3 Limitations and future research 225

6.4 Concluding remarks 230

Appendix 232

Bibliography 241

List of Tables

Table 3.1 LMT group's MKT framework 58

Table 3.2 Framework proposed in this study 58

Table 4.1. Organization of administered items 73

Table 4.2. Descriptive statistics for teachers' background 76

Table 4.3. Organization of items 81

Table 5.1 Summary of findings 100

Table 5.2 Items in a task of USW 102

Table 5.3. Estimated standardized factor loadings within USW 107

Table 5.4. Estimated correlations among dimensions within USW 109

Table 5.5. Comparison of fit among MIRTs within USW 110

Table 5.6. Q-matrix for items within USW 111

Table 5.7. Estimated item parameters (logit) within USW 115

Table 5.8. Items in the task CGP 124

Table 5.9. Estimated correlations among dimensions within CGP 125

Table 5.10. Comparison of fit among MIRTs within CGP 127

Table 5.11. Estimated standardized factor loading within CGP 128

Table 5.12. Q-matrix for items within CGP 130

Table 5.13. Estimated item parameters (logit) within CGP 132

Table 5.14. Items in the instruction situation CG 139

Table 5.15. Comparison of fit among the IFA models within CG 141

Table 5.16. Estimated standardized factor loadings within CG..................................... 142

Table 5.17. Q-matrix for the items within CG.. 143

Table 5.18. Estimated item parameters (logit) within CG... 146

Table 5.19. Items within DP 149

Table 5.20. Comparison of fit among IFAs within DP.. 151

Table 5.21. Estimated standardized factor loadings within DP 151

Table 5.22. Q-matrix for items within DP 153

Table 5.23. Estimated item parameters (logit) within DP 155

Table 5.24. Items in the task EF 158

Table 5.25. Comparison of fit among IFAs within EF 159

Table 5.26. Estimated standardized factor loadings within EF 160

Table 5.27. Q-matric for the items within EF... 161

Table 5.28. Estimated item parameters (logit) within EF.. 163

Table 5.29. Items within the course of high school geometry....................................... 166

Table 5.30. Estimated correlations among dimensions within geometry 167

Table 5.31. Estimated standardized factor loadings within geometry............................ 168

Table 5.32. Comparison of tit among multidimensional models within geometry 169

Table 5.33. Q-matrix for the items within geometry 170

Table 5.34. Estimated item parameters within geometry 173

Table 5.35. Items within algebra... 177

Table 5.36. Estimated correlations among dimensions within algebra 178

Table 5.37. Estimated standardized factor loadings within algebra 178

Table 5.38. Comparison of tit among multidimensional models within algebra 179

Table 5.39. Q-matrix for the items within algebra 180

Table 5.40. Estimated item parameters (logit) within algebra 183

Table 5.41. Effects of teachers' background on knowledge in doing USW 194

Table 5.42. Effects of teachers' background on knowledge in doing CGP 195

Table 5.43. Effects of teachers' background on knowledge in algebra 1 195

List of Figures

Figure 1.1. Domains of MKT (Ball et al., 2008, p.403) 9

Figure 1.2. Difference in a framework between MKT and SMK-G. a. Unidimensional (left) b. MKT (middle) c. SMK-G (right) 22

Figure 1.3 Example of SCK problem (Ball et al., 2008, p.400) 25

Figure 2.1. Knowledge/Skill framework for clinical nurse specialists (Royal College of Nursing, 2005, p.9) 40

Figure 2.2. Dimensional models (TEDS-M) a. IRT (left) b. Between-item MIRT (middle) c. Within-item MIRT (right) (Blömeke et al., 2014, p. 485-487).................................... 46

Figure 2.3. CCK/SCK item vectors in MIRT (Schilling, 2007, p. 103) 47

Figure 3.1 The relationship between student task and instructional situation 68

Figure 4.1. Sample GCN task 88

Figure 5.1. Five-dimensional model within the task USW.. 105

Figure 5.2. Four-dimensional model within the task USW 105

Figure 5.3. Three-dimensional model within the task USW 106

Figure 5.4. Item characteristic bar chart for USW_CG items... 117

Figure 5.5. Item characteristic bar chart for USW_DP items 117

Figure 5.6. Item characteristic bar chart for USW_EF items 118

Figure 5.7. Item characteristic bar chart for USW_SR items 118

Figure 5.8. Item characteristic bar chart for USW_SE items 119

Figure 5.9. Test information function of USW 122

Figure 5.10. Four-dimensional model within the task CGP 126

Figure 5.11. Three-dimensional model within the task CGP... 127

Figure 5.12. Item characteristic bar charts (ICBCs) for CGP_CG items........................ 133

Figure 5.13. Item characteristic bar charts (ICBCs) for CGP_DP.................................. 134

Figure 5.14. Item characteristic bar charts (ICBCs) for CGP_EF 134

Figure 5.15. Item characteristic bar charts (ICBCs) for CGP_CN 135

Figure 5.16. Test information function of CGP under 4-dimensional MIRT................. 137

Figure 5.17. Test information function of CGP under 3-dimensional MIRT................. 137

Figure 5.18. Two-dimensional model within the instructional situation CG 140

Figure 5.19. Unidimensional model within the instructional situation CG 141

Figure 5.20. Item characteristics bar chart for USW_CG items 147

Figure 5.21. Item characteristics bar charts for CGP_CG items..................................... 147

Figure 5.22. Two-dimensional model within the instructional situation DP.................. 150

Figure 5.23. Unidimensional model within the instructional situation DP..................... 150

Figure 5.24. Item characteristic bar chart for USW_DP items 156

Figure 5.25. Item characteristic bar chart for CGP_DP items 156

Figure 5.26. Two-dimensional model within EF 159

Figure 5.27. Unidimensional model within EF.. 159

Figure 5.28. Item characteristic bar chart for USW_EF items 164

Figure 5.29. Item characteristic bar chart for CGP_EF items... 164

Figure 5.30. Four-dimensional model within U.S. high school geometry...................... 167

Figure 5.31. Item characteristic bar chart for USW_CG items....................................... 174

Figure 5.32. Item characteristic bar chart for CGP_CG items....................................... 174

Figure 5.33. Item characteristic bar chart for USW_DP items...................................... 175

Figure 5.34. Item characteristic bar chart for CGP_DP items 175

Figure 5.35. Three-dimensional model within algebra.. 177

Figure 5.36. Item characteristic bar chart for USW_SR items 184

Figure 5.37. Item characteristic bar chart for USW_SE items 184

Figure 5.38. Item characteristic bar chart for CGP_CN items....................................... 185

Figure 5.39. Profile proportions within the task USW 187

Figure 5.40. Profile proportions within the task CGP 189

Figure 5.41. Profile proportions within CG within geometry.. 190

Figure 5.42. Profile proportions within DP within geometry 191

Figure 5.43. Profile proportions within EF within geometry... 191

Figure 5.44. Marginal attribute mastery within algebra 193

Figure 5.45. Effect sizes of years of teaching geometry within geometry attributes...... 196

List of Abbreviations

CCK: Common Content Knowledge

CFA: Confirmatory Factor Analysis

CFI: Comparative Fit Index

CG: Calculation in Geometry

CGP: Choosing the Givens for a Problem

CK: Content Knowledge

CN: Calculating with Numbers

CTT: Classical Test Theory

DCM: Diagnostic Classification Model

DP: Doing Proofs

EF: Exploring a Figure

FA: Factor Analysis

FIML: Full Information via Maximum Likelihood

GPK: General Pedagogical Knowledge

HCK: Horizon Content Knowledge

ICBC: Item Characteristics Bar Charts

IFA: Item Factor Analysis

IRT: Item Response Theory

KCC: Knowledge of Content and Curriculum

KCS: Knowledge of Content an Students

KCT:Knowledge of Content and Teaching

MCK: Mathematics Content Knowledge

MIRT: Multidimensional Item Response Theory

MKT: Mathematical Knowledge for Teaching

MKT-A: Mathematical Knowledge for Teaching Algebra

MKT-G: Mathematical Knowledge for Teaching Geometry

MML: Marginal Maximum Likelihood

MPCK: Mathematics Pedagogical Content Knowledge

PCK: Pedagogical Content Knowledge

RMSEA: Root Mean Square Error of Approximation

SCK: Specialized Content Knowledge

SEM: Structural Equation Modeling

SMK: Subject Matter Knowledge

SMK_G: Subject Matter Knowledge for Teaching Geometry

SMK_USW: Subject Matter Knowledge for Understanding Students' Work

SMK-G: Subject Matter Knowledge for teaching Geometry

SR: Simplifying Rational expressions

TIF: Test Informatoin Function

TLI: Tucker-Lewis Index

USW: Understanding Students' Work

WLSMV: Weighted Least Squares Mean and Variance adjusted estimator

Abstract

This study proposes a way of organizing mathematical knowledge for teaching that permits to reveal its multidimensionality. Scholars concerned with teachers' mathematical knowledge have traditionally distinguished knowledge dimensions by knowledge types, such as mathematical content knowledge or pedagogical content knowledge (e.g., the MKT framework). This approach has been widely adopted in studies that measure teachers' knowledge using assessment items. But it remains an open question whether these conceptualizations can lead to precise measures of the different domains, as it is highly likely teachers simultaneously use multiple knowledge types when teaching mathematics. This creates challenges in measuring only mathematical content knowledge not mixed with any pedagogical aspects but still used in the work of teaching. While this way of conceptualizing knowledge dimensions has allowed researchers to develop measures that reflect professional knowledge, it has been less adept to documenting whether and how the knowledge varies depending on the specific teaching assignments teachers have experience with.

The challenge in developing distinct measures has motivated me to propose a new way to organize assessment items. I describe this new way in terms of an item blueprint that specifies the correspondence between the organization of the items and the dimensions of the knowledge purported to be measured by the items. The proposed item blueprint is then evaluated regarding its purposes: 1) to capture multiple aspects of

teachers' mathematical knowledge used in teaching; 2) to develop precise multiple measures reflecting the identified dimensions of knowledge. Ultimately, the developed measures were designed to allow a fine-grained description of the knowledge used in the work of teaching secondary mathematics.

The proposed item blueprint uses two organizers: task of teaching and instructional situation. *Task of teaching* alludes to each of the activities that comprise the practice of a mathematics teacher (e.g., understanding students' work). *Instructional situation* alludes to each of the types of mathematical work students are assigned within a course of study (e.g., doing proofs in geometry). Following the blueprint, I assigned each set of items to measure one knowledge dimension associated with one task of teaching and one instructional situation. By organizing the knowledge using these two organizers, the item blueprint allows a description of teachers' knowledge with respect to the characteristics of the components of the work of teaching. With this conceptual rationale, the methodological feasibility of the item blueprint was evaluated by fitting item-factor models to the item responses collected from a nationally distributed sample of 602 U.S. practicing mathematics teachers. The distinctions among the factors were examined using model-comparison tests conducted under three different measurement models: structural equation modeling, item response theory, and diagnostic classification models.

The results consistently showed that the majority of the hypothesized dimensions are statistically distinguishable by either or both of the organizers within and across both geometry and algebra courses of study. This distinction was further supported by different relationships with teachers' educational background and teaching experience across the identified knowledge dimensions.

By presenting an innovative item blueprint that is theoretically warranted and methodologically feasible, this study shows great promise for measuring multiple dimensions of teachers' mathematical knowledge used in the work of teaching. It contributes to developing theory of mathematics teaching and to future item development for measuring knowledge used in professional tasks and instructional situations.

Chapter 1

Objective of the Study

1.1 Introduction

This study explores the feasibility of measures of mathematical knowledge for teaching that attend to the knowledge management tasks the teacher needs to engage in and the specificity of the mathematical work they need to manage. In that context the study explores how assessment items can be used to identify distinguishable dimensions of teachers' mathematical knowledge for teaching, which is defined as a profession-specific mathematical knowledge used in the work that teachers do to help students learn mathematics (Ball, Thames, & Phelps, 2008; Thames, 2009). Teachers' mathematical knowledge for teaching has been hypothesized to be a critical factor for students' learning of mathematics and this hypothesis has been supported by empirical studies that closely examine teachers' instructional practice and their mathematical knowledge.

For example, compared to less knowledgeable teachers, knowledgeable teachers are more likely to present problems properly, considering students' prior knowledge (Sherin, 2002), use appropriate mathematical representations (Lehrer & Frank, 1992; Heid, Blume, Zbiek, & Edwards, 1998), or competently implement mathematics curriculum materials (Manouchehri & Goodman, 2000). Given these important associations with teaching quality, attempts have been made to examine teachers' mathematical knowledge using assessment instruments that allow for empirical investigations.

1.1.1 A descriptive and large scale[1] study on teachers' knowledge

Compared to other studies that empirically examine teachers' mathematical knowledge, the current study has two particular characteristics: it is a descriptive (rather prescriptive) and a large-scale (rather small scale) study. First, this study is descriptive in that its purpose is to obtain fine-grained description of teachers' mathematical knowledge used in the work of teaching. To obtain this fine-grained description, this study aims to establish multiple measures that represent distinguishable dimensions of teachers' knowledge. To establish the multiple measures, this study proposes a way to develop and structure assessment items by referring to different components of actual teaching practices (i.e., according to kinds of work teachers do and kinds of subject matter being transacted between students and a teacher). This approach is different from studies that aim to evaluate individual teachers' knowledge against a prescribed knowledge or a list of elements of desirable knowledge. As this study aims to provide a way to describe multidimensional knowledge dimensions rather than to suggest desirable dimensions composing the whole knowledge construct, this study does not attempt to obtain an exhaustive list of dimensions of teachers' knowledge required in the work of teaching.

Second, this study is a large-scale study in that it aims to investigate the dimensionality of teachers' knowledge through measures established by using item responses collected from a large number of participants, such as 602 nationally representative U.S high school teachers. The reason for using assessment items measuring teachers' knowledge instead of observations or interviews, which are generally used in small-scale studies, is that this study aims to investigate the

[1] The sample of this study can be considered large in mathematics education scholarship

dimensionality of teachers' mathematical knowledge generalizable to the U.S. secondary mathematics teacher population. To empirically examine teachers' knowledge, this knowledge needs to be operationalized through an instrument. This is because knowledge is a latent construct that cannot be directly observed. In particular, to measure multiple dimensions of knowledge, the items that purport to measure those dimensions need to be structured according to the hypothesized structure of the knowledge. How to structure items to establish multiple knowledge measures allowing for a fine-grained description of the knowledge is the focus of this study.

1.1.2 The need of an alternative knowledge framework

Through theoretical arguments and case studies, scholars have argued that mathematical knowledge for teaching consists of multiple dimensions. Many researchers have attempted to measure distinguishable dimensions of teachers' knowledge for teaching, but they have not yet been successful in developing precise measures that reflect the different dimensions they hypothesized. This raises the question of whether and how the multidimensionality of teachers' mathematical knowledge for teaching can be operationalized by using assessment items. The question includes whether we do indeed have a knowledge framework that is conceptually plausible and methodologically viable to create an item blueprint that operationalizes the descriptive investigation of teachers' mathematical knowledge in which different dimensions can be examined. If the framework is plausible, can we design items measuring distinguishable aspects of teachers' mathematical knowledge? If such items can be developed, which measurement model can be used to evaluate whether the items function as intended? Furthermore, which measurement model can be used to yield measures reflecting teachers' competencies in multiple dimensions? To address these important questions, this study therefore addresses both conceptual and methodological challenges the field has had understanding the multidimensionality of teachers' mathematical knowledge.

To explore these questions, first, this study proposes a framework for structuring assessment items that is hypothesized to allow for the measurement of multiple distinguishable dimensions of teachers' mathematical knowledge for teaching secondary mathematics. Also, the established measures are intended to capture variations in teachers' knowledge that can be explained by the variations in the components of the work of teaching. The framework is informed by the analyses of teachers' performance on knowledge instruments and by an understanding of the work of teaching as well as by prior approaches to measuring teachers' mathematical knowledge. Specifically, it is organized according to *tasks of teaching* (e.g., understanding students' work; discussed in Chapters 3 & 4) and *instructional situations* (e.g., doing proofs; discussed in Chapters 3 & 4).

To develop distinguishable measures of teachers' knowledge, which other research groups have not been able or attempted to do with a multidimensional item blueprint (Hill, Schilling, & Ball, 2004; Gitomer, Phelps, Weren, Howell, Croft, 2014; Etkina et al., 2018), this study distinguishes assessment items on the basis of tasks of teaching and instructional situations. The rationale for doing this is that if any distinguishable dimensions of teachers' knowledge are identifiable in terms of differences in the tasks of teaching or instructional situations that call for that knowledge, it would be reasonable to hypothesize that the task of teaching and (or) instructional situations can be classifiers of the dimensions of teachers' knowledge. This classification mechanism for classifying the dimensions of the knowledge can also shed light on the specificity of the work of teaching by identifying the specific knowledge dimensions involved in the components of the work.

With this conceptual rationale of the proposed framework, this study evaluates the methodological feasibility of the suggested knowledge framework using participants' performance on assessment items on geometry and algebra 1 content. The participants are 602 teachers who represent national in-service high school teachers in the U.S. Informed by psychometric theory, this study evaluates the extent to which item-score interrelationships support the hypothesis of the multidimensional item structure I propose. Not only does the evaluation using multiple measurement models aim to validate the proposed framework, it also provides guidance for future item development that can allow for the measurement of multiple dimensions of teachers' mathematical knowledge. Again, ultimately, the results of this study contribute to fine-grained descriptions of teachers' mathematical knowledge used in the work of teaching secondary geometry and algebra 1.

1.1.3 Organization of the dissertation

I begin with Chapter 1 that discusses the theoretical and practical benefits of a multidimensional understanding of teachers' mathematical knowledge, with an emphasis on the advantages of a multidimensional measure over a unidimensional measure of knowledge. Next, I discuss how the need for re-conceptualization and operationalization of the dimensionality of teachers' knowledge leads to the research questions guiding this dissertation study. In Chapter 2, I review how different research groups have defined teachers' mathematical knowledge in ways that could benefit from multidimensional analysis and how they have attempted to measure the constructs empirically. I also review the methodological approaches including structural equation modeling, item response theory, and diagnostic classification modeling in the use in the teacher knowledge literature. In Chapter 3, I suggest a new item blueprint hypothesized to allow the identification of multiple dimensions of teachers' mathematical knowledge for teaching high school geometry and algebra 1. The discussion of the foundations of the proposed knowledge framework is followed by the definitions of the two organizers (i.e., task of teaching and instructional situation) that characterize teachers' knowledge dimensions. In Chapter 4, I describe empirical procedures such as item type selection, defining hypothesized knowledge dimensions, scaling responses, and data analysis methods. I also describe the definitions of particular knowledge dimensions examined in this study and the rationale for using them. In Chapter 5, I describe the results. In Chapter 6, I discuss findings and implications of the results. I conclude this dissertation with a discussion of the limitations and some directions for future study.

1.2 Background

There has been a long-standing interest in finding attributes of teachers' mathematical knowledge that is specific to the work of teaching (Ball, Lubienski, & Mewborn, 2001). It is widely accepted that in addition to subject matter knowledge, teachers have mathematical knowledge that is pedagogical in nature and is used to help students learn mathematics. In an effort to understand the unique knowledge that teachers need and use, Shulman (1986) suggested three distinct types of content knowledge (subject matter content knowledge, pedagogical content knowledge, and curricular knowledge), which can be defined in the context of teaching. In particular, among these three categories, the conceptualization of pedagogical content knowledge (PCK) has been instrumental in enabling the research community to think about a type of knowledge that blends subject matter and pedagogical knowledge. In relation to mathematics teaching, Shulman's (1986) notion of PCK emphasizes the role of teachers' mathematical knowledge in relation to representing and formulating mathematical content in a way that is comprehensible to students, rather than merely presenting the knowledge about the content itself (Shulman, 1986, p. 9). Shulman's multiple categories of teacher knowledge has been extended to further develop a variety of multidimensional structures of teachers' mathematical knowledge for teaching, all of which see mathematical knowledge for teaching as composed of at least two sub-constructs, e.g., subject matter knowledge and pedagogical knowledge (e.g., Ball et al., 2008; Herbst & Kosko, 2014; Krauss, Brunner, Kunter, Baumert, Blum, Neubrand, & Jordan, 2008; McCrory, Floden, Ferrini-Mundy, Reckase, & Senk, 2012; Mohr-Schroeder, Ronau, Peters, Lee, & Bush, 2017; Saderholm, Ronau, Brown, & Collins, 2010; Tatto, Schwille, Senk, Ingvarson, Peck, & Rowley, 2008).

Among the many research groups that have endeavored to conceptualize teachers' mathematical knowledge in a multidimensional manner, the LMT group at the University of Michigan has defined mathematical knowledge for teaching (hereafter MKT) as a profession-specific mathematical knowledge used in the work of teaching (Ball, Lubienski, & Mewborn, 2001; Ball et al., 2008). In this framework, MKT is conceptualized as a construct composed of subject matter knowledge (hereafter SMK) and pedagogical content knowledge (hereafter PCK) (Figure 1.1) (Ball et al., 2008, p.403). SMK is described as the mathematical content knowledge used in the tasks of teaching, while PCK is a blend of content knowledge and pedagogical knowledge (p.392). In the MKT framework, SMK and PCK are subsumed into one overarching construct of mathematical knowledge for teaching (MKT), but each of them is also divided into sub-domains. SMK is decomposed into three domains[2] of content knowledge. The first is common content knowledge (CCK), "the mathematical knowledge known in common with others who know and use mathematics" (p. 403). The second is specialized content knowledge (SCK), which is conceptualized as "the mathematical knowledge and skill unique to teaching" (p. 400), and the third is horizon content knowledge (HCK), defined as "an awareness of how mathematical topics are related over the span of mathematics included in the curriculum" (p.403). PCK is decomposed into knowledge of content and student (KCS), defined as "knowledge that combines knowing about students and knowing about mathematics" (p. 401), knowledge of content and student (KCT), defined as knowledge that "combines knowing about teaching and knowing about mathematics" (p. 401), and knowledge of content and curriculum. Among these domains, horizon content knowledge and knowledge of content and curriculum have been reconceptualized by some researchers (Fernandez, Figueiras, Deulofeu, and Martinez, 2011; Zazkis and Mamolo, 2011; Koponen, Asikainen, Viholainen, & Hirvonen, 2016). However, to the best of my knowledge, these domains have not yet been operationalized by test items.

[2] The term "domain" is used interchangeably with the term "dimension" for the MKT framework.

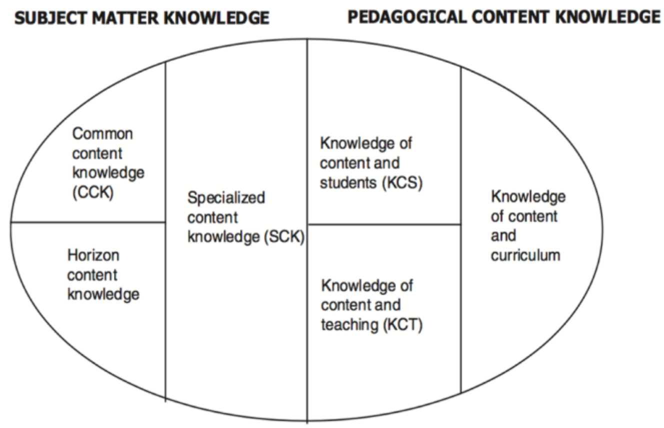


Figure 1.1. Domains of MKT (Ball et al., 2008, p.403)

Other research groups have also conceptualized teachers' mathematical knowledge as a multidimensional construct in different ways (Baumert, Kunter, Werner, Brunner, Voss, Jordan, …, & Tsai, 2010; Herbst & Kosko, 2014; McCrory et al., 2012; Saderholm et al., 2010; Tatto et al., 2008). Some scholars have focused on a specific subject: McCrory et al. (2012) and Phelps et al. (2014) worked in algebra, Herbst & Kosko (2014) and Mohr-Schroeder, Ronau, Peters, Lee, & Bush (2017) worked in geometry. Some scholars have focused on a level of schooling: Hill, Schilling, and Ball (2004), Ball et al., (2008), Hill (2010) worked with elementary school teachers, Saderholm et al., (2010) and Hill (2007) with middle school teachers, and Herbst & Kosko (2014), Mohr-Schroeder et al. (2017), and Wilson & Heid (2015) with high school teachers. These diverse ways that research programs have conceptualized the structure of teachers' mathematical knowledge will be discussed in more detail in 2.1.2.

The diverse ways of conceptualizing the structure of teachers' mathematical knowledge have often been criticized for the failure to produce one single theoretical framework that the mathematics education community could agree on (Tirosh & Even, 2007). This lack of a universal agreed upon framework has also been cited as a main factor explaining the unorganized approach to designing teacher education courses (Petrou & Goulding, 2011, p. 9). Some frameworks are criticized for the lack of attention to factors that other frameworks consider. For example, the MKT framework is criticized for not acknowledging teachers' beliefs (Rowland & Ruthven, 2011).

Admittedly, the existence of divergent frameworks for the same construct – teachers' mathematical knowledge – could be problematic if a framework is indiscriminately applied to different situations, contexts, or purposes without an understanding of the unique features of each framework. For example, applying the MKT knowledge framework, which emphasizes the distinction between CCK and SCK, may not be useful if the purpose of using the framework is to understand distinguishable aspects of mathematical knowledge among college mathematics professors, considering that the framework is based on the work of elementary teachers. On the other hand, the same framework may be appropriate for the purpose of capturing the unique aspects of mathematical knowledge that elementary teachers use in the work of teaching, as compared to the mathematical knowledge that mathematicians use in their work.

The divergent conceptualizations of teachers' multidimensional knowledge have contributed to challenges in understanding the nature of teachers' knowledge, but, nonetheless, the existence of diverse frameworks can be a blessing in that they provide diverse means to measure dimensions of teachers' knowledge in terms of purpose. Put in other words, if the purpose and the foundation of each framework are correctly understood, one does not need to default to a competitive perspective on the relationship among frameworks (i.e., which framework is better to explain the dimensionality of teachers' knowledge than others). Instead, one can consider a complementary perspective among frameworks. In such a perspective, one might, for example, consider the LMT group's MKT framework as complementary to other frameworks that might be organized according to the level of cognitive demand (e.g., Tschoshanov, 2011) rather than knowledge type, in that the former provides a blueprint for capturing a unique dimension of teachers' mathematical content knowledge (i.e., SCK) that the latter framework does not allow. Similarly, a framework organized by mathematical topics can complement the MKT framework by providing a way to examine a knowledge dimension specific to a certain mathematical topic that the MKT framework, where each dimension is composed of more than one mathematical topic, cannot provide. The importance of complementarity was also alluded to by Petrou and Gouldin (2011), who state that "the conceptualizations of teacher knowledge proposed are not inconsistent; rather, they build on each other. Even though the researchers have stressed different domains of teacher knowledge, all focus on the importance of seeing the content to be taught as an important part of teaching" (p.20).

Given this perspective on complementarity, one framework is not necessarily better than another in explaining variance in teachers' knowledge. However, one framework might be more suitable than another depending on the specific goals of the framework. If the framework is to be used as a blueprint of an assessment measuring teachers' knowledge, the framework needs to be valid not only from a theoretical perspective but also from a measurement perspective. The validity of the assessment is then evaluated based on the purpose of the assessment, which is determined by the intended use and interpretation of the results (Kane, 2006). Therefore, it is crucial to clarify the intended use and interpretation of the results of an assessment in order to suggest a framework that guides the development of the assessment instrument.

In this study, the intended use and interpretation of the results from the focused assessment are to describe and interpret multidimensional teachers' mathematical knowledge with respect to a particular task of teaching and a particular instructional situation. Therefore, the suggested knowledge framework is expected to attain following two goals. First, the framework allows measuring multiple distinguishable dimensions of teachers' mathematical knowledge. Second, each of the identified dimensions reflects teachers' knowledge in doing a specific work of teaching in a specific instructional situation. To investigate how the multiple dimensions of teachers' knowledge are organized, this study examined the structure of a set of assessment items, which had been designed to measure teachers' mathematical knowledge for teaching secondary geometry and algebra 1.

1.2.1 Conceptualizing mathematical knowledge for teaching as a multidimensional construct

To recognize the benefits of a multidimensional conceptualization of teachers' mathematical knowledge, it is important to understand first the limitations of a unidimensional conceptualization on teachers' knowledge, and then understand how a multidimensional conceptualization could overcome those limitations. Before discussing the advantages of a multidimensional conceptualization over a unidimensional conceptualization, it is necessary to make clear the difference in the meaning between unidimensional and multidimensional conceptualizations of teacher knowledge. Edwards

(2001) defined *multidimensional construct* and contrasted it from that of *unidimensional construct* and from what he called “multiple dimensions regarded as distinct” as follows.

> A construct is multidimensional when it refers to several distinct but related dimensions treated as a single theoretical concept. Multidimensional constructs may be distinguished from unidimensional constructs, which refer to a single theoretical concept, and from multiple dimensions regarded as distinct but related concepts rather than a single overall concept (Edwards, 2001, p.144).

Following this definition, teacher knowledge could be said to be multidimensional if it is composed of several distinct but related dimensions subsumed within a larger theoretical concept of teachers’ mathematical knowledge for teaching. In contrast, teachers’ knowledge is unidimensional if it is conceptualized as a single theoretical concept without any distinct knowledge dimensions under it. Thus, the unidimensional perspective on teacher knowledge refers to the assumption that teachers’ knowledge can be represented by a single attribute. Consequently, this assumption may lead to the use of a unidimensional measure (e.g., a single score) as a proxy for the amount of teachers’ knowledge. However, the use of a single score for the amount of a teacher’s knowledge can be valid only if the teacher’s responses (or actions) to the instrument (a written assessment or an observation tool) are evidence of only one type of knowledge. If indeed a single measure is used for a multidimensional construct – teachers’ mathematical knowledge – this may foreclose the opportunity for capturing differences in how individuals know different aspects of the mathematics they use in teaching. Furthermore, this reduction of a multidimensional knowledge construct to a unidimensional measure might cause inconsistent results on the relationship between teachers' mathematical knowledge, which is indicated by subject matter preparation, and student achievement. In other words, a unidimensional measure of teachers' knowledge could cause threats to internal validity (Shadish, Cook, & Campbell, 2002, p.38) by distorting a causal relationship from teachers' subject matter preparation to teachers' mathematical knowledge.

In some early studies, teachers' mathematical knowledge has been conceptualized as equivalent to the amount of mathematics studied, so it was operationalized by the number of courses taken in college. These studies then reported that the effect of more than five mathematics courses was negatively correlated with 11^{th} grade students' performance (Monk, 1994). Similarly, Begle (1979) reported that the number of teachers' course credits at the level of calculus or beyond was negatively associated with student (4th, 7th, and 10th grade) achievement in 43 percent of cases which had significant main effects from teacher credits on student achievement (Begle, 1979, p. 41). However, teachers' mathematics content preparation, as measured by the number of mathematics courses taken, was positively related to 10^{th} grade student performance in mathematics (Monk, 1994; Rowan et al., 1997). Hence, the evidence is inconsistent. I suggest this inconsistency may be due to the lack of means to capture the multidimensional nature of mathematical knowledge. I do not argue that teachers' mathematics course credits or number of courses taken are inappropriate proxies for teachers' mathematical knowledge. Rather, I admit that those proxies may account for a portion of teachers' mathematical knowledge, which can be developed through mathematics courses. However, these proxies may not reflect other dimensions of knowledge which could not be developed solely by taking courses or studying advanced mathematics, but which probably influence student achievement. In other words, the attempt to understand the overall effect of teachers' knowledge on student achievement with a single measure might prevent us from observing a plausible relationship whose observation might be enabled if we had a better conceptualization of other dimensions of teacher knowledge, perhaps more centrally related to the work teachers do in classrooms.

For example, mathematical content knowledge gained from pure mathematics courses can be important in one aspect of the work of teaching (e.g., evaluating the correctness of student mathematical work), but it may not be equally important in other aspects of the work (e.g., anticipating students' misconception). While the depth of teachers' mathematical knowledge may partially account for teachers' overall mathematical knowledge, it is likely that other key components of the knowledge used in the work of teaching also need to be taken into account. Specifically, considering the finding of a negative effect of more than five mathematics courses on students' performance, it is possible that what seems to be negative from a unidimensional perspective might not be negative from a multidimensional perspective.

Consider the following thought experiment. Suppose that teachers' mathematical content knowledge was validly measured by the number of mathematics courses taken, and pedagogical content knowledge was validly measured by the time spent in field experiences, but only the number of courses was used as a proxy in examining the relationship between teachers' mathematical knowledge and student achievement. The proxy used to measure teacher knowledge might not capture the knowledge learned by those who made good use of field experiences and would not account for the lack of this knowledge among the teachers who took many classes beyond requirements at the expense of their field experience. In this case, the relationship between teacher knowledge and student achievement could be shown as negative, even though the negative relationship depended on the lack of the knowledge that could have been learned from field experience.

With this example, we can see the limitations of considering the teachers' mathematical knowledge as a unidimensional construct. But what does a multidimensional perspective entail? We would expect that a multidimensional perspective would allow access to multiple measures reflecting distinct aspects of teachers' knowledge. If we assume that teachers' knowledge is a two-dimensional construct instead of a unidimensional one, then student achievement could be modeled by a function of two variables. For example, one variable might represent teachers' mathematical content knowledge (presumably gained through content coursework) and the other variable might represent teachers' pedagogical content knowledge (presumably gained through mathematics education courses and field experiences). In this example of two-dimensional teacher knowledge, a student taught by a teacher with moderate levels of knowledge in both content and pedagogical domains could show higher performance than a student taught by a teacher with a high level of content knowledge but a low level of pedagogical knowledge. The unidimensional perspective on teachers' knowledge might not allow us to detect this phenomenon and could lead to an incorrect interpretation of the effect of coursework (e.g., that knowing more mathematics is somehow detrimental to teachers).

Similarly, a multidimensional perspective on teacher knowledge might explain the source of variation in teacher effectiveness at different stages of their teaching career, particularly early in their career, as shown in previous studies (Harris & Sass, 2011; Clotfelter, Ladd, & Vigdor, 2006). For example, an inconsistent effect of teaching experience on teachers' knowledge could be dependent on the types of knowledge that might be acquired through experience. Specifically, gains in teachers' effectiveness might rise sharply in teachers' initial years if the effectiveness is associated with teachers' ability in understanding students' work, whereas sharp gains can be shown in later years of teachers' career if the measured effectiveness is associated with teachers' ability in creating problems. This conjecture is supported by a previous empirical study showing different patterns of relationships between different dimensions of mathematical knowledge for teaching geometry (MKT-G) and years of teaching geometry (Ko & Herbst, under review). As such, a multidimensional perspective can provide a lens for explaining some inconsistent results reported in early studies. Furthermore, it allows identifying multiple key components of teachers' knowledge required in multiple demands of the work of teaching.

1.2.2 The role of multidimensional understanding in teacher education

A better theoretical understanding of the nature of teachers' mathematical knowledge through the lens of a multidimensional conceptualization may offer several benefits to teacher education; for example, in teacher knowledge diagnosis, teacher training, and teacher certification or badging.

Identifying and understanding multiple dimensions composing teachers' mathematical knowledge may enable the development of assessments measuring multiple domains of knowledge, which are important in teaching. These assessments can be used to diagnose the status of individual teachers' knowledge, and this diagnosis can therefore contribute to the development of tailored teacher training programs. For example, if an assessment finds that teachers have the most difficulty on items asking them to understand student work, the teacher education program could focus on providing learning opportunities to improve mathematical knowledge required for this task of teaching. On the other hand, if teachers are shown to be weak in basic mathematical content knowledge, the program could be organized to require teachers to take more content courses before taking a methods course, or field work. This highly specific information about teachers' knowledge can provide teacher educators with more actionable guidance to customize their instruction for their students.

Furthermore, teacher training programs could provide differentiated training for beginning teachers and experienced teachers by using the most common knowledge profiles that are driven by multidimensional assessments within each group. For example, some studies have shown that beginning teachers need the most help in areas of content-specific pedagogies or in the task of planning lessons (Reynolds, 1995). A multi-dimensional conceptualization of teacher knowledge might enable the identification of the related knowledge dimensions on which beginning teachers are less knowledgeable than experienced teachers. Accordingly, a teacher development program could be designed to focus on these areas for beginning teachers.

Another potential benefit for teacher education is that a multidimensional conceptualization might support the exploration of multiple ways of organizing the field of mathematical knowledge for teaching, hence impact the curriculum for teacher preparation. Based on a multidimensional conceptualization of the knowledge needed for teaching, different types of courses could be designed that do not reproduce the classical division between content courses and methods courses. Instead, these courses could be designed to attend to specific dimensions of teacher knowledge. In this study, I investigate two organizers as crucial for finding dimensions of teacher knowledge. These organizers could also provide alternative ways of designing courses; for example, by types of tasks of teaching or by types of instructional situations, instead of simply by mathematical topics. As such, a well-defined knowledge structure could be mapped to the teacher education curriculum.

Lastly, a multidimensional understanding of teacher knowledge could help improve teacher certification or badging systems. The need for tools providing a comprehensive view of teachers' competence has been discussed and Reynolds (1995) suggested a portfolio model for teacher assessment which would "display badges of competency in important teaching tasks and knowledge domains" (Reynolds, 1995, p. 218). Along with the idea of the portfolio model, a multidimensional understanding of teacher knowledge could offer a map laying out multiple knowledge dimensions with related teaching tasks. Teachers and teacher educators, then, can use the map to track their mastery of the important tasks of teaching along the path to be competent in all areas/dimensions. For example, a map based on the multidimensional structure of teacher knowledge would allow teachers to plan their professional development based on their mastery status or interests in a specific teaching skill. Furthermore, a map composed of multiple knowledge dimensions could help scaffold teachers' learning during and after teacher education.

1.2.3 The need for re-conceptualizing the dimensionality of mathematical knowledge for teaching

To empirically investigate a hypothesis about the dimensionality of teachers' mathematical knowledge for teaching, researchers have developed instruments consisting of items measuring hypothesized sub-constructs (i.e., each construct corresponds to one dimension of teachers' knowledge in this study). In general, the sub-constructs operationalized in those instruments included at least two knowledge domains, CK and PCK, which were originally introduced by Shulman (1986, 1987). This categorization that distinguishes CK and PCK, however, has run into challenges in operationalizing the knowledge domains.

To develop an instrument, the targeted knowledge domains need to be articulated (Furr, 2011, p. 12), and the articulation necessitates clarifying 1) the similarities and differences with other relevant domains as well as 2) the contexts that engage the construct (Furr, 2011, p.12). Wilson (2005) uses the expression "variable clarification" to refer to this process, where "the construct to be measured is distinguished from other closely related constructs" (Wilson, 2005, p. 38). In accordance with Furr's (2011) and Wilson's (2005) guidelines, there seems to be a lack of construct clarification in current frameworks that follow Shulman's knowledge categorization. I argue that this is one of the main limitations that need to be improved to empirically measure teachers' mathematical knowledge as a multidimensional construct. Specifically, regarding the relationship among domains, even though multiple subdomains of teacher knowledge have been conceptualized to be subsumed in an overarching domain of mathematical knowledge for teaching in most studies, the relationship among subdomains tends to be conceptualized as disjoint or is not clearly defined. For example, items are assigned to either CK or PCK without consideration that items may contribute to more than one domain.

To allow interactions between the two or more domains may be important, given that the nature of teachers' knowledge is complex and may not be so clearly divided into disjoint sub-dimensions defined by a domain of knowledge (CK, PCK, etc.) used in teaching. Teachers may simultaneously use more than one knowledge type when they teach. In other words, it would be more sensible to assume that domains are interrelated. For example, an item may require teachers to use multiple relevant skills and knowledge domains when responding to the item rather than require only one knowledge domain. Given this possibility of using more than one knowledge domain at the same time, it is worthwhile to re-conceptualize how knowledge is organized. Specifically, re-conceptualization of how we organize knowledge that can accommodate the intersection of sub-domains is needed. To incorporate intersections between domains, this study proposes to consider more than one organizer for a knowledge structure.

Comparatively speaking, the MKT framework is conceptualized with one organizer such as knowledge domain, which distinguishes six knowledge dimensions (Figure 1.2.b), whereas the framework suggested in this study is conceptualized with two organizers: task of teaching and instructional situation, each of which might distinguish different numbers (e.g., M and N) of knowledge dimensions, respectively. Therefore, the proposed framework suggests a total of M*N distinguishable dimensions (Figure 1.2.c). Figure 1.2.c is described with an example of the SMK-G (subject matter knowledge for teaching geometry) instrument where two tasks of teaching (understanding students' work, hereafter USW and choosing appropriate givens for a problem, hereafter CGP) and two instructional situations (doing proofs, hereafter DP) and calculation in geometry (hereafter, CG) are hypothesized as distinguishable dimensions for each organizer (Ko & Herbst, 2017).

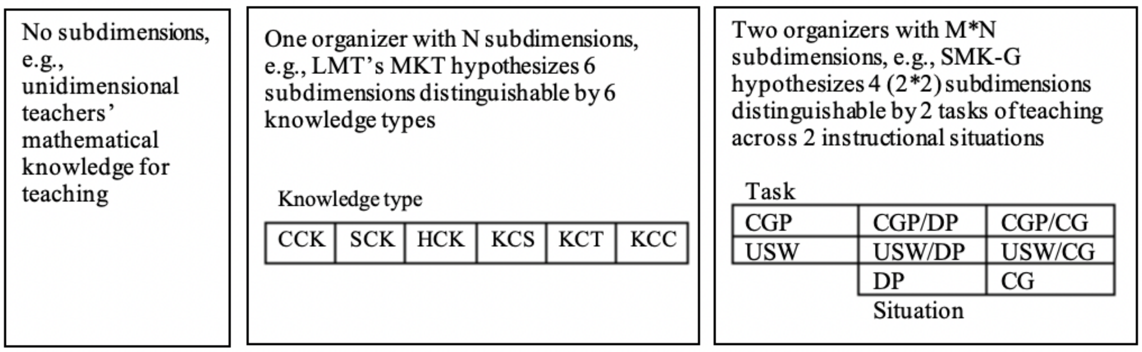


Figure 1.2. Difference in a framework between MKT and SMK-G.
a. Unidimensional (left) b. MKT (middle) c. SMK-G (right)

Furthermore, as mentioned previously, the contexts that engage a construct (knowledge) also need to be clarified to enable a measured knowledge component to reflect how aspects of mathematical knowledge are involved in specific teaching contexts. Considering the difference in teacher knowledge between different grade levels, it also seems reasonable to conjecture that there are differences in teachers' mathematical knowledge among different instructional situations even within the secondary school level. However, there are very few research groups that developed a teacher' mathematical knowledge framework that accounts for this aspect of situational knowledge. The meaning of instructional situation and its hypothesized role in teacher knowledge structure is described more in detail in Chapter 3.

1.2.4 The need for re-operationalizing the dimensionality of mathematical knowledge for teaching

After defining theoretical constructs (e.g., overall MKT and the various domains of MKT identified by Ball et al., 2008), researchers have developed instruments as operational tools to measure such constructs (viz., sets of items, each of which taps into each of the domains of teachers' mathematical knowledge). Researchers, then, infer the characteristics of the underlying construct from participants' responses on the instrument. Furthermore, the inferred nature of teachers' knowledge can then be used to refine the initial conceptualization of teachers' knowledge. This cyclic measurement process has been used in the studies of measuring teachers' mathematical knowledge (Baumert et al., 2010; Blömeke et al., 2014; Hill et al, 2004; Wilson, 2005).

In the process of making inferences on the knowledge domains in terms of teachers' responses to items, the plausibility of the theory (conceptualization of dimensionality of teachers' knowledge) needs to be evaluated empirically (Kane, 2013) to accurately relate teachers' responses (observable attribute) to a theoretical construct (dimensions of teachers' knowledge). However, many empirical studies have failed to provide empirical support to their hypothesized theory of teachers' knowledge. Limitations in the theory itself or a limitation in the empirical study (including the affordances of the sample and the measurement model used) can be the reason for the discrepancy between what the theory posits and its empirical findings. One of the challenges in an empirical check is to choose an appropriate measurement model that establishes a relationship between the item score and the construct. Measurement models used in the teacher knowledge literature commonly relate one item response to a single construct, one domain of teachers' knowledge such as CCK or SCK, relying on unidimensional item response models. If, however, we consider that constructs such as SCK are highly complex, it is reasonable to consider a measurement model allowing multiple domains to intervene in an item's response. This is because each item may activate multiple knowledge domains. For example, Figure 1.3 is one of the MKT items designed to measure teachers' SCK (mathematical knowledge uniquely used by teachers). However, this item raises the question of whether SCK is the only type of knowledge that a teacher may use to evaluate the given story problems. Consider the item shown in Figure 1.3 where the respondent needs to determine, for each of three story-problems, whether the problem represents 1 ¼ divided by ½. A participant may compare the numerical answer for each problem, which is ⅝ in a) and 5/2 in b) and c), with the result of calculating 1 ¼ divided by ½ (= 5/2). If a given problem leads to a different numerical answer as that for 1 ¼ divided by ½, as story problem (a) does, a mathematically knowledgeable participant would likely choose an answer "No" if he or she were using their common content knowledge. Yet, the correct answer for story problem (b) is also "No," and to come to that answer a participant may need to assess the story problem on the basis of the mathematical conception of division being represented in the problem as well as in the calculation used to model the problem. Knowledge of different conceptions of division is knowledge special to teachers. While, mathematically speaking, the only correct problem is problem (c), even mathematically knowledgeable participants could choose "No" for problem (c), if they relied on knowledge of students' difficulties with long sentences or some students' (e.g., ELLs) difficulties with particular vocabulary (viz, taffy). In sum, a participant may need to activate various types of knowledge unique to a teaching context, in addition to common mathematical knowledge that non-teachers may use in their job, in order to answer an item.

Story Problem That Represents 1¼ divided by ½

Which of the following story problems can be used to represent 1 $^1/_4$ divided by $^1/_2$?

	Yes	No
a) You want to split 1¼ pies evenly between two families. How much should each family get?	1	2
b) You have $1.25 and may soon double your money. How much money would you end up with?	1	2
c) You are making some homemade taffy and the recipe calls for 1¼ cups of butter. How many sticks of butter (each stick = ½ cup) will you need?	1	2

Figure 1.3 Example of SCK problem (Ball et al., 2008, p.400)

In regard to the item shown in Figure 1.3, the issue is whether the type of knowledge involved in answering the sub-items b) and c) relies on knowledge from a single domain such as SCK or whether it involves both CCK and SCK, or even KCS. The type of measurement model allowing the latter situation ("the situation in which some items on a multidimensional assessment measure more than one attribute") is called within-item multidimensionality (Rupp, Templin, & Henson, 2010, p. 330), an assumption in measurement modeling that stands in contrast to between-item multidimensionality in which "each item on an assessment measures a distinct latent attribute" (Rupp, Templin, & Henson, 2010, p. 317). Measurement models can be differentiated in regard to what assumption they make.

Dimensionality analysis is important not only for conceptualizing a complex construct, but also for methodological reasons. Specifically, if any additional dimensions are measured by any of the items in the set (assumed to be unidimensional), this presence of multiple dimensions can threaten the accuracy of item parameter estimates when using a unidimensional model (e.g., IRT) as a measurement model. Given the possibility that more than one knowledge domain is required to solve an item, this study relaxes the assumption that one item needs to measure one construct or that domains and constructs have one-to-one relationships. With this unconstrained assumption, this study explores an alternative measurement model that increases a feasibility of multidimensional evaluation with respect to the complexity of the construct – mathematical knowledge called for in the work of teaching (the concept will be further discussed in Chapter 3).

1.3 Guiding research question

As mathematical knowledge for teaching is an unobservable construct that is generally operationalized through assessment items, this study takes a fresh look at these items to understand how they can be used to operationalize the dimensionality of in-service teachers' mathematical knowledge for teaching used in the work of teaching.

This study asks the following research question:

What kind of item blueprint can guide the development and structuring of items that 1) capture the variations in teachers' knowledge used in different components of the work of teaching and 2) allow for establishing multiple distinguishable measures of teacher' mathematical knowledge for teaching high school geometry and algebra 1.

To answer this question, I propose to structure teachers' knowledge with two organizers – task of teaching and instructional situation. I expect that this proposal can embody a potential framework for developing sets of measures reflecting teachers' mathematical knowledge used in the work of teaching. At this point, only a general, research question guiding this study is posed. The specific research questions including both the conceptual and methodological questions are introduced after the literature review which discusses relevant prior empirical studies and the need for additional research.

1.4 Chapter summary

In this chapter, I introduce the purpose of this study, which is to obtain a fine-grained description of teachers' mathematical knowledge for teaching by proposing a knowledge framework that allows for the measurement of multiple distinguishable dimensions of teachers' mathematical knowledge for teaching secondary mathematics. The purpose is supported by the theoretical and methodological rationale of this study. The rationale is discussed with consideration of benefits of a multidimensional understanding of teachers' mathematical knowledge. The purpose and the rationale lead to the research questions guiding this dissertation study.

Chapter 2

Review of Concepts and Methods Used in Literature

2.1 Operationalization of teachers' mathematical knowledge

As researchers have acknowledged the benefits of a multidimensional view of teachers' mathematical knowledge for teaching, considerable efforts have been made to operationalize the multiple dimensions of teachers' mathematical knowledge in terms of observable or measurable variables. This has led to the development of research frameworks using empirical instruments (e.g., observational tools or assessment items) that allow for the operationalization of multiple dimensions of teachers' knowledge. In this section, I review studies developing a multidimensional framework of teachers' mathematical knowledge and attempting to empirically identify the distinctions among hypothesized knowledge dimensions.

2.1.1 Frameworks distinguishing CK and PCK

Most researchers, who empirically measured teachers' mathematical knowledge, have refined Shulman's conceptualization of teachers' knowledge based on the idea that knowledge for teaching is distinguishable according to knowledge domains, such as CK and PCK. In other words, these researchers have attempted to measure teachers' professional knowledge assuming conceptual distinctions among knowledge dimensions.

Among them, the LMT group conceptualized teachers' mathematical knowledge as a multidimensional construct composed of six sub-domains: common content knowledge (CCK− “the mathematical knowledge used in settings other than teaching”, Ball et al., 2008, p.399), specialized content knowledge (SCK – “the mathematical knowledge that allows teachers to engage in particular teaching tasks”, Hill, Ball, & Schilling, 2008, p.377), horizon content knowledge (HCK), knowledge of content and student (KCS), knowledge of content and teaching (KCT), and knowledge of content and curriculum (KCC) (Ball et al., 2008). Similarly, the TEDS-M[3] project conceptualized teachers’ knowledge as consisting of three knowledge domains (mathematics content knowledge: MCK; mathematics pedagogical content knowledge: MPCK; and general pedagogical knowledge: GPK) across three difficulty levels of curriculum (novice, intermediate, and advanced; Senk et al, 2012). They further classified the Mathematical Content Knowledge (MCK) into three cognitive knowledge dimensions such as knowing, applying, and reasoning, and also classified the domain of Mathematics Pedagogical Content Knowledge (MPCK) into three dimensions of teaching practices such as mathematical curricular knowledge, knowledge of planning for mathematics teaching and learning, and enacting mathematics for teaching and learning. Similar to the TEDS-M group, Bush and his colleagues at the University of Louisville developed the DTAMS (Diagnostic Teacher Assessment in Mathematics and Science) assessment. They categorized dimensions of teachers’ mathematical knowledge by the depth of cognitive knowledge: memorized knowledge, conceptual knowledge, higher-order thinking (problem solving/reasoning), and pedagogical content knowledge (Saderholm et al., 2010). The Cognitively Activating Instruction, and the Development of Students’ Mathematical Literacy (COACTIV) group (Baumert et al., 2010; Krauss et al, 2008) also

[3] IEA’s Teacher Education and Development Study in Mathematics

conceptualized teachers' knowledge as having two components: Content Knowledge (CK) – "a conceptual understanding of the mathematical knowledge taught"– and Pedagogical Knowledge (PCK) – "the area of knowledge relating specifically to the main activity of teachers." Within the domain of PCK, they defined three aspects of teachers' knowledge: "the ability to identify multiple solution paths" (tasks); the "ability to recognize students' misconceptions, difficulties, and solution strategies" (student); and "knowledge of different representations and explanations of standard mathematics problems (instruction)" (Baumert et al., 2010, p. 1).

2.1.2 Differences among the frameworks

Even though researchers have generally agreed that teachers' mathematical knowledge is multidimensional in that it is composed of at least two conceptually distinguishable dimensions such as CK and PCK, hypothesized construct maps representing the organization of subdimensions of knowledge vary considerably. There has been ongoing discussion about what characteristics of knowledge need to be used to identify sub-dimensions or how those subdimensions are internally organized. Diverse hypotheses on the principles used to distinguish subdimensions have generated many divergent frameworks. Accordingly, the issue of how to integrate different frameworks to have a better understanding of the dimensionality of teachers' knowledge remains a focus of debate. Among the frameworks reviewed in the previous section, three main sources of differences, which may cause challenges for an integrated understanding, could be identified with respect to the 1) terminology; 2) level of specification; 3) principles in organizing knowledge dimensions. Each of these sources of differences is discussed in the following paragraphs.

First, one of the sources of challenges for integrated understanding is the terminology used to define the components of teachers' mathematical knowledge. Consider that according to the LMT group, teachers' mathematical content knowledge that is not pedagogical is subject matter knowledge (SMK); but, in the view of the German research group COACTIV (Cognitive Activation in the Mathematics Classroom and Professional Competence of Teachers; Krauss et al., 2008), a similar concept is named *Content Knowledge* (CK), and for the project TEDS-M project (Teacher Education and Development Study in Mathematics; Tatto et al., 2008) the same concept is referred to using the term *Mathematical Content Knowledge* (MCK). Moreover, researchers have different perspectives on the characteristics of sub-dimensions, even when sub-dimensions are defined similarly using similar terms. Consider different perspectives on *subject matter knowledge* or *content knowledge*. While subject matter knowledge is typically conceptualized without consideration of the practice of teaching, Ball's group (Ball et al., 2008) introduced the concept of specialized content knowledge (SCK) within the domain of subject matter knowledge to emphasize this unique part of teachers' subject matter knowledge that cannot exist independently from the work of teaching. In contrast to Ball's group, other research groups (TEDS-M: Senk et al., 2012; COACTIV: Krauss et al., 2008) have merged this specialized content knowledge with subject matter knowledge or pedagogical knowledge rather than define it as a stand-alone domain.

Second, there are differences among the frameworks in the level of specification used to organize teachers' knowledge. Regarding the number of dimensions needed to represent teachers' mathematical knowledge, different research groups hypothesize a different number of dimensions even within a dimension of mathematical content knowledge. For example, the LMT group conceptualizes SMK as an aggregate of more than one knowledge subdomain (including CCK, SCK, and HCK), whereas the COACTIV group sees CK as unidimensional. Another example of this is the test used by Tchoshanov (2011), who divides knowledge into three kinds of content knowledge such as "*knowledge of facts and procedures (Type 1 knowledge)*, *knowledge of concepts and connections (Type 2 knowledge)*, and *knowledge of models and generalizations (Type 3 knowledge)*" (Tchoshanov, 2011, p. 142).

Third, different research groups have used different organizing principles, which in this study I call *organizers* (i.e., the principles structuring knowledge dimensions), in organizing their hypothesized knowledge dimensions. For example, the LMT group uses *knowledge domain* (e.g., CCK, SCK, etc.) as a main organizer for their framework. In doing so, Ball et al.'s (2008) distinctions contribute to unpack the diffuse category of general teacher knowledge into domain-specific teacher knowledge (de Jong & Ferguson-Hessler, 1996). Each dimension (domain) is distinguished from other dimensions by the object of knowledge in each domain. A similar approach is taken by TEDS-M (Tatto et al., 2008) at their first level of differentiation between mathematics content knowledge (MCK) and mathematical pedagogical content knowledge (MPCK). Other groups use curriculum levels (e.g., KAT[4] project uses school algebra vs. college-level mathematics; McCrory et al., 2012), mathematical domain distinctions (DTAMS uses whole numbers vs. rational numbers; Saderholm et al., 2010), the level or type of cognitive demand (e.g., Tchoshanov, 2011, uses knowledge of facts and procedures vs. knowledge of concepts

[4] Knowledge of Algebra for Teaching

and connections vs. knowledge of models and generalizations). The classification effort continues at lower levels in some of those frameworks, often combining distinctions of domain and cognitive demand with what de Jong and Ferguson-Hessler (1996) called epistemological distinctions, or distinctions based on the tasks in which knowledge is used (e.g., KAT's distinction among knowledge used in the practices of trimming, bridging, and decompressing, or in TED-S's distinction of MPCK among knowledge of planning for mathematics teaching and learning, and knowledge of enacting mathematics for teaching and learning).

As described, there are several differences among the frameworks used for conceptualizing teachers' mathematical knowledge as multidimensional. A common characteristic of the research groups reviewed thus far is that they assume that teachers' mathematical knowledge is something teachers may or may not possess. In contrast to these groups, Thompson (2015) focused on "teachers' mathematical meaning" to refer to the possibly tacit mathematical understandings that could be inferred in teachers' handling of the mathematics they teach. He emphasized the advantage of a focus on "teachers' mathematical meaning" over a focus on mathematical knowledge which he termed as "declarative knowledge" as he considers that teachers' mathematical meaning is more productive for understanding sources of teachers' instructional decisions and actions (Thompson, 2015, p. 438). Within these meanings, one can see developmental distinctions similar to those that have been observed in children's knowledge. Similar to how children develop their mathematical knowledge by constructing different schemes on a learning trajectory (Steffe, 2004), teachers may develop their mathematical knowledge for teaching by transforming their content knowledge into the mathematical knowledge that is pedagogical (Silverman & Thompson, 2008).

As those examples demonstrate, different groups of researchers have used a variety of organizers to conceptualize their frameworks for mathematics teachers' knowledge. Those organizers are generally chosen to express what each research group considers crucial in defining teachers' knowledge a priori. This hypothesized structure is then further operationalized by test items, which are administered to teachers and possibly others (e.g., non-teachers for comparison's sake; Reckase, McCrory, Floden, Ferrini-Mundy, & Senk, 2015; Krauss et al., 2008), and the data collected from responses to the test items is examined to determine whether the hypothesized structure fits well the data or not. If the data fit the hypothesized structure statistically well, this lends credibility to the hypothesis. However, that a certain multidimensional framework accounts for the structure of teachers' item responses well with respect to a certain purpose (e.g., identifying a mathematical topic in which teachers have difficulty) does not necessary mean that the framework is good for another purpose (e.g., identifying what knowledge dimension is related to understanding student misconceptions). Again, I would like to argue that the existence of diverse frameworks is not problematic by itself, but it is problematic if those frameworks are not associated with the use to which they may and may not be put. Efforts to validate a dimensionalization of teachers' knowledge need to be understood and applied to an appropriate context according to the purpose of the intended use of the framework (Kane, 2013).

2.1.3 Dimensionality of secondary school teachers' mathematical knowledge

Even though some research groups used curriculum distinctions to organize their framework (e.g., the KAT group distinguished the knowledge of college level algebra for secondary teachers from their knowledge of high school algebra; McCrory et al., 2012), they used the same dimensionality framework for teachers teaching different school levels. For example, the LMT group used the same MKT framework (Ball et al., 2008) to structure items for elementary and middle school teachers, though the items for two teacher groups were different regarding the curricular level of mathematics contents (Hill, 2007). Likewise, the TEDS-M project used the same conceptual framework to develop items, which are different in mathematics contents, for future primary and lower secondary teachers (Tatto, 2014, p. 37).

However, considering the differences among the groups of teachers who teach different grade-levels of students, the wisdom of using the same knowledge structures for the different teacher groups is not something that should be taken for granted. In particular, given that secondary mathematics teachers take more advanced courses in the discipline of mathematics than elementary mathematics teachers, teachers' content knowledge may need to be conceptualized differently between two groups. For example, what the KAT group calls "college-level mathematics" in the category of high school teachers' algebra content knowledge might be close to "horizon content knowledge", which is distinguished from elementary teachers' common content knowledge in the LMT group's MKT framework.

As such, the issue of whether the same structure or the same definitions of sub-domains of MKT, which is grounded in the analysis of elementary teachers' work, apply to teacher populations in different grade levels such as secondary or post-secondary gives rise to the need for research. Speer, King, and Howell (2015) discussed this concern stating, in particular for CCK and SCK, that "…distinctions between CCK and SCK for elementary teachers have been recognized and accepted in the mathematics education community, but these distinctions may be less compelling and clear at higher levels, as few examples have been presented at the secondary and post-secondary levels" (Speer et al., 2015, p. 107). Other researchers also have emphasized the need for efforts to examine the similarities and differences in teachers' mathematical knowledge between different grade level contexts in which teachers' mathematical work is called for (Hill, 2007; Rowland & Ruthven, 2011).

The need to take differences among teacher populations into account is also important given that the dimensionality of teachers' mathematical knowledge has been generally investigated through the dimensionality of a test designed to measure multiple domains of knowledge. According to Reckase's definition of test dimensionality - "a sample-specific characteristic of the data matrix" (Reckase, 2009, p. 194) - , the dimensionality of the test is dependent not only on "the dimensions of sensitivity for test items," but also on "the amount of variation that is present in the sample of examinees on the constructs that are the target of the test" (Reckase, 2009, p. 201). In other words, even when the test items are sensitive enough to capture two distinguishable dimensions of CCK and SCK, the secondary teachers' response data may not show distinction between CCK and SCK if the correlation between sample secondary teachers' CCK and SCK knowledge is very high. Given that secondary teachers receive more content preparation in mathematics than elementary teachers (Speer et al., 2015), the variance among secondary teachers' amount of knowledge in mathematics' content knowledge may be less than it is at the elementary level. This possibility of difference in the amount of variation between samples of elementary and secondary teachers on each construct implies a need for a unique dimensionality hypothesis for each population.

In short, the dimensionality of mathematical knowledge for teaching for elementary school teachers can reasonably be expected to be different from that for high school teachers. For example, sets of items designed to measure CCK and SCK separately could detect those two dimensions in the sample of elementary population but not necessarily in the sample of secondary population, or vice versa. To handle this possibility, the present study proposes a framework for the study of dimensions of teacher knowledge that eschews knowledge domain distinctions (CCK, SCK, etc.) and instead is solely based on aspects of teachers' work.

2.1.4 Dimensionality of knowledge in other professions

Given that a multidimensional view on a construct is useful in that it "provides holistic representations of complex phenomena" (Edwards, 2001, p.145), it is worth looking at whether and how scholars have attempted to conceptualize the dimensionality of knowledge and skills associated with professions other than teaching. Although this study focuses on teachers' professional knowledge used in teaching, it is worth exploring how other professional fields conceptualize the dimensionality of knowledge or skills needed in doing their professional work. To know how dimensionality has been discussed in other professions may help this study enhance its applicability, in that we could draw inspiration from the ways in which professions other than teaching have conceptualized dimensions of knowledge. With regard to this purpose, I reviewed literature from two professions – nursing and engineering – that have made efforts to establish frameworks for the dimensionality of knowledge, competence, and skills.

In a number of professions, dimensionality of the professional knowledge is framed in terms of intrinsic tasks or activities carried out by each professional, or vice versa (i.e., dimensionality of the tasks is framed in terms of the dimensionality of the types of professional knowledge). For example, Lee (1986) identified dimensions of engineering tasks using an exploratory factor analysis on survey items that represent different types of requisite job knowledge. Specifically, the task factor reflecting "work with hands" was distinguished from the factor reflecting "work with other people", according to the distinction in the required skill/knowledge for the two tasks. The former task requires "knowledge of machine shop and model building skills" or "skills to work with specific instrument or test machines," whereas the latter requires "to interact with many people" or "to work with others in a team effort" (Lee, 1986, p.129-130).

Similar to the use of correspondence between a task and required knowledge, the knowledge needed for a nursing career is characterized as being composed of four levels of knowledge and skills required for doing a specific nursing task. The knowledge and skills framework (KSF) for a UK clinical specialist is organized by "dimensions which describe different aspects of work" (Royal College of Nursing, 2005, p.2). Similarly, the framework for nurses involved in cancer care is defined by the focus of their practice (New Zealand Ministry of Health, 2014). For example, each competency dimension corresponds to a dimension of cancer care (e.g., competency for disease and treatment-related care, ability in supportive care). In consonance with this convergence of attention to the characteristics of professional tasks as a major distinction in structuring professional knowledge, the knowledge framework of the present study also considers professional tasks in looking for a way to understand the dimensionality of teachers' mathematical knowledge.

Another similarity in frameworks among professionals across professions is that the level of knowledge is conceptualized as a scale, whereby the characteristics of knowledge a professional has can be ranked according to their performance in professional practice. For example, Lee's (1986) study, examining knowledge attributes of U.S. engineers, used multidimensional performance rating scales assessed by the engineers' supervisors to identify distinguishable types of abilities demanded of young engineers (Lee, 1986). Similarly, in the UK, each dimension of the National Health Service (NHS) staff's knowledge has four levels, where the higher number represents the higher level of knowledge (Figure 2.1). This multidimensional framework delineating types of knowledge dimensions and levels of each dimension is adjustable for different NHS related roles and at different career levels. For example, some dimensions could be dropped for senior level positions, if the dimensions are no longer required knowledge. Figure 2.1 is an example framework for clinical nurse specialists.

NHS KSF DIMENSIONS	Required for post	Level for post 1	2	3	4	Notes
CORE DIMENSIONS relates to all NHS posts						
1 Communication	✓				✓	
2 Personal and people development	✓				✓	
3 Health, safety and security	✓			✓		
4 Service improvement	✓			✓		
5 Quality	✓			✓		
6 Equality and diversity	✓			✓		
SPECIFIC DIMENSIONS						
IK2 Information collection and analysis	✓			✓		
IK3 Knowledge and information resources	✓			✓		
GI Learning and development	✓				✓	
G7 Capacity and capability	✓		✓			

Figure 2.1. Knowledge/Skill framework for clinical nurse specialists (Royal College of Nursing, 2005, p.9)

With regard to levels of knowledge or skill, the present study also considers each dimension of teachers' knowledge as a scale where the level of knowledge can be assessed. The idea of establishing a framework consisting of core dimensions and specific dimensions (such as the NHS KSF framework above) would be worth considering when modifying subdimensions appropriately for the specificity of teachers' work (by subject or by grade level).

2.2 Dimensionality analysis in the studies of teachers' mathematical knowledge

In general, the dimensionality of teachers' knowledge has been investigated through examining the dimensionality of item scores. Methods for doing so include "identifying the number of dimensions reflected by a test, the meaning of those dimensions, and the degree to which the dimensions are associated with each other" (Furr

& Bacharach, 2013, p. 72). This section describes the methods that researchers have used to identify distinguishable aspects or components of teachers' mathematical knowledge for teaching. As mentioned above in 1.2, identifying the dimensionality of teachers' mathematical knowledge has important implications both theoretically and practically. In what follows, various statistical methods such as item factor analysis and bi-factor model with scored teachers' response data will be described with example studies.

2.2.1 IFA within SEM and (unidimensional) IRT frameworks

Item factor analysis (hereafter, IFA; factor analysis for categorical data) is the most common method used in examining the dimensionality of a test. To examine the dimensionality can be understood as identifying the interpretable factors that account for the correlations among the set of items (Furr & Bacharach, 2013). By using factor analysis (hereafter, FA), researchers "identify sets of items that are relatively strongly correlated with each other but weakly correlated with other items" (Furr & Bacharach, 2013, p. 80). In general, research groups develop each item to reflect one of the theoretically hypothesized dimensions in advance and conduct confirmatory factor analysis to provide evidence confirming the conjectured structure of the item responses. In the teacher knowledge literature, factor analysis has been conducted within the structural equation modeling (SEM) or within the item response theory (IRT) frameworks. The main difference between IFA within the SEM and IRT frameworks is that IFA within SEM uses limited information such as correlation or covariance matrices to assess the structure of items, whereas IFA within IRT generally uses full information such as the pattern of raw response data (Wirth & Edwards, 2007, p. 66).

Regarding the use of factor analysis in the SEM framework, the COACTIV group used a confirmatory factor analysis (CFA) with continuous parcel scores, which is defined as "an aggregate-level indicator comprised of the sum (or average) of two or more items, responses, or behaviors" (Little, Cunningham, & Shahar, 2002, p. 152). In the study, the parcel scores were developed for CK and PCK separately, based on the hypothesis that CK and PCK are the two distinguishable domains of teachers' mathematical knowledge. A confirmatory factor analysis (CFA) with these parcel scores showed good fit statistics (e.g., RMSEA values below 0.05 and CFI and TLI values above 0.95), and the researchers concluded that their conjectured two-factor structure of teacher knowledge (CK and PCK) fits the teachers' responses well (Krauss et al, 2008). Similarly, the DTAMS group conducted item factor analysis (IFA), using polychoric correlations among items (SEM-based), to assess how well their items fit into the hypothesized factor model. As a result, the DTAMS group confirmed that all their structural models adequately described the sample data, except for the probability and statistics test (Saderholm, 2010, p. 185). In addition to the SEM-based IFA, the DTAMS group conducted unidimensional IRT (one-factor IFA within the IRT framework) for each hypothesized dimension to evaluate the unidimensionality of a set of items and assess item quality, such as item difficulty, discrimination, and item misfit (Saderholm et al., 2010, p.184). Similarly, the LMT group used unidimensional IRT to scale teacher knowledge as a single score and to evaluate the score reliability (Hill, 2007, 2010).

As such, some research groups evaluated the unidimensionality of a set of items reflecting one dimension, and they used subscale scores, each of which was estimated from a unidimensional model, to represent the amount of knowledge for one knowledge dimension (Krauss et al., 2008; Baumert et al., 2010). However, separately tested unidimensional models do not provide evidence that items across dimensions are separable. Again, to empirically confirm a hypothesized structure of a set of items, the number of final factors should be the same as the hypothesized number of domains, and each domain should consist of the intended items. However, this is very challenging to obtain, not only because large sample sizes are needed, but also because it is possible that factors (knowledge domains) are highly correlated. For example, it could be more difficult to distinguish CCK from SCK than to distinguish SMK from PCK, as CCK and SCK are likely to be highly correlated, being both included in SMK. Furthermore, as unidimensional IFA forces an item to load on only one dimension, it excludes the possibility that an item evokes more than one knowledge domain when solving the item. To accommodate the possibility that one item is associated with more than one domain, or those domains are correlated to each other, some research groups have applied more sophisticated methods, such as bi-factor model, item vector plot, or multidimensional IRT (MIRT).

2.2.2 Bi-factor model

An alternative to FA is the bi-factor model. If the factors identified among items are substantially correlated, allowing the possibility that an item response would contribute information about two factors would be a reasonable approach: a general factor taking into account the level of performance over all dimensions and a factor reflecting a specific domain. The bi-factor model assigns items onto two factors: a general factor explaining a common variance across items and a specific factor explaining only the variance in a specific category. The main difference between FA and a bi-factor model is that the bi-factor model allows each item to load in two places: one for a general factor that explains an overall variance across all items, and the other for a specific factor that explains variance within a specific dimension other than a general factor. On the other hand, FA or unidimensional IRT allow each item to load only on one factor.

The LMT group (Hill et al., 2004) used a bi-factor model to examine whether their MKT items represent one general domain of MKT or multiple distinct domains of MKT (Hill et al, 2004). The items used in the analysis were designed to reflect the domain of SMK (consisting of CCK and SCK) and KCS, across two mathematics topics (number concepts and operations). They found evidence of multidimensionality in measures using bi-factor analysis. Their bi-factor model supported the existence of SCK as distinguishable from CCK, and SCK items were better explained by a specific factor. This specific factor excludes the influence of the general factor, which reflects the influence of common factor across all items.

2.2.3 Multidimensional IRT

One variation of item response theory (IRT) that allows for multidimensionality within a construct is called multi-dimensional IRT (MIRT). The need for a model allowing more than one dimension in determining the level of a construct has led researchers to pay attention to a multidimensional IRT model (MIRT) (Ayala, 2009). MIRT allows an item to reflect more than one construct and domain to associate, so it can represent the level of the teacher's knowledge as a vector, each of whose coordinates represent the location of the item in one knowledge dimension, in a multidimensional space. For example, an item could actually require both CCK and SCK. In this case, MIRT could be used to simultaneously measure both domains from the item.

In MIRT, items that have item vectors pointing in a similar direction (i.e., these items are composed with similar kinds of knowledge) are clustered as one-dimensional (Reckase, 2009). In this way, MIRT can detect item clusters that cannot be detected by unidimensional IFA or IRT by using more comprehensive item characteristic information. Among the research groups studying teacher knowledge, the TEDS-M group estimated teachers' knowledge using MIRT. They used both unidimensional IRT (Figure 2.2.a) and two MIRT approaches (Figure 2.2.b & 2.2.c). Their unidimensional IRT model treated teachers' knowledge as a single construct combining Mathematics Content Knowledge (MCK) and Mathematics Pedagogical Knowledge (MPCK). In contrast to this unidimensional IRT model, multidimensional models allow a correlation between MCK and MPCK. One of TEDS-M's MIRT models treated MCK or MPCK as a separable unidimensional construct, but the model assumed the two domains are correlated (Figure 2.2b). The result of fitting this model supported their assumption that a correlation between the two domains exists (MCK and MPCK), but the domains are separable (Blömeke et al., 2014). In the second MIRT (Figure 2.2.c), the researchers relaxed the assumption that one item measures only one construct; instead, they allowed an item's response to load not only on MCK but also on MPCK. The two-multidimensional IRT models showed a significantly better fit than a unidimensional model, which supports a claim on the multidimensional nature of teachers' mathematical knowledge.

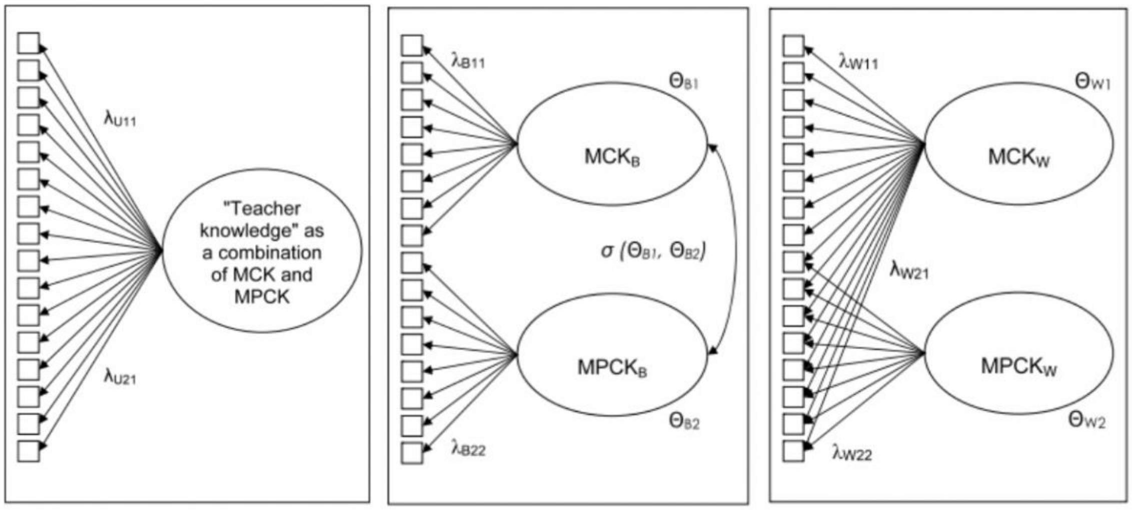


Figure 2.2. Dimensional models (TEDS-M)
a. IRT (left) b. Between-item MIRT (middle) c. Within-item MIRT (right)
(Blömeke et al., 2014, p. 485-487)

2.2.4 Item vector plot

Another method for examining what constructs an item response reflects is the item vector plot. In Schilling's (2007) study that validates the hypothesized structure of the MKT instrument, Schilling used item vector plots, a graphical technique used for examining the multiple features of MIRT. The plot was drawn in two-dimensional space, where each axis represents the scale of item difficulty with respect to one of two factors (Figure 2.3). An item vector plot can be considered as a scatterplot of two-dimensional item difficulties with respect to an orthogonal (independent) factor analysis solution with two factors (Reckase, 2009). As shown in Figure 2.3, the SCK items varied widely in their orientation, whereas all of the CCK items were tightly clustered in a narrow sector (Schilling, 2007, p. 104). This indicates that CCK items have similar direction with respect to two factors, whereas SCK items are not. This suggests that SCK items contributed much to the multidimensionality within content knowledge for teaching. This multidimensionality of SCK may warrant an inquiry into the potential factors that can explain variation within a domain of SMK (subject matter knowledge), given that all the items involved number concepts and operations.

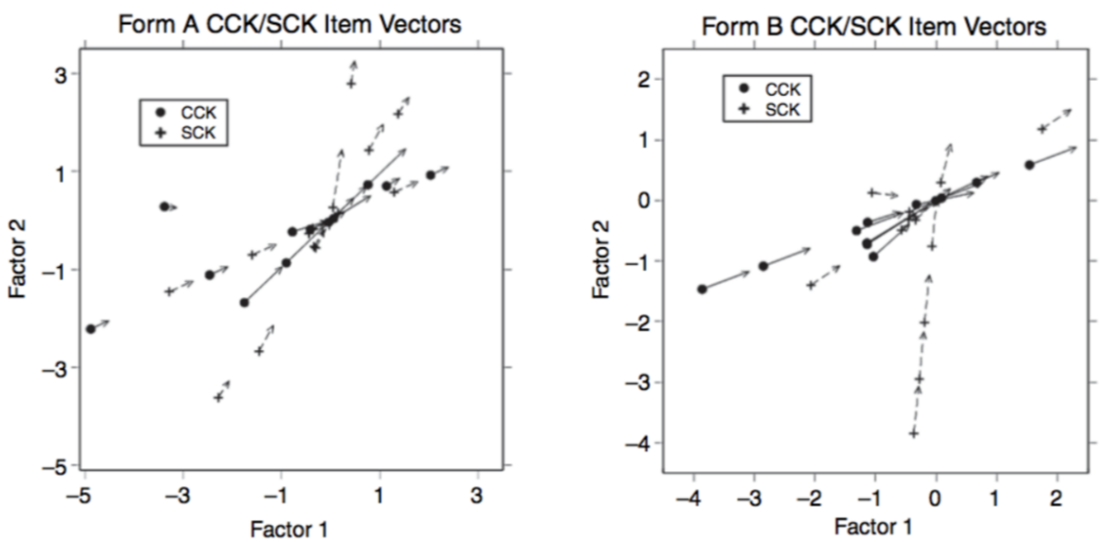


Figure 2.3. CCK/SCK item vectors in MIRT (Schilling, 2007, p. 103)

2.2.5 Diagnostic Classification Models

Diagnostic Classification Models (hereafter, DCMs) are also models that may be used to estimate multiple dimensions of a construct using multiple item responses. However, in contrast to the above models that consider the knowledge construct as continuous, DCMs consider the latent construct (here, teachers' knowledge) being measured as binary: 0 when a respondent (here, teacher) has not achieved the mastery level of skill/knowledge; 1 when a respondent has achieved the mastery level of skill/knowledge. Accordingly, respondents are classified by a list of binary numbers specifying knowledge *attributes* that individual respondents have mastered or not. The term *attribute* refers to "a latent characteristic of respondents", and a list of binary numbers is an *attribute profile* that refers to "the particular pattern of values on the latent attribute variables that is assigned to a respondent" (Rupp, Templin, & Henson, 2010, p. 316). In comparison with IRT and SEM framework, one knowledge attribute in DCM corresponds to one knowledge dimension in those other models and a list of 1s and 0s in DCM corresponds to a list of real-numbered scores representing the amounts of knowledge on multiple dimensions.

As such, DCMs simplify the level of knowledge for each attribute/dimension into two categories (e.g., masters and non-masters) based on statistically estimated cut-scores. By setting those cut-scores to maximize the reliability of classifying respondents, DCMs provide a relatively higher reliability for their attribute estimates with fewer items than other models (e.g., MIRT or SEM) that locate respondents on continuous scales. Classifying respondents based on their continuous scores might also be possible in MIRT or SEM frameworks by setting cut-scores. However, the cut-scores may need to be determined by human judges, which might create multiple sources of error in estimates. This advantage of DCMs in classifying respondents is referred to as one of the most important characteristics of DCMs that distinguishes them from other multidimensional models, such as MIRT or SEM (Rupp et al., 2010).

In the context of assessing teacher knowledge, the attribute profile would provide information about which knowledge domains individual teachers have mastered or have not mastered. However, in spite of the advantage of DCMs described above, the application of DCM models to study teachers' knowledge is only emerging. I only found one study examining middle grade teachers' understanding of fraction arithmetic (Bradshaw, Izsák, Templin, & Jacobson, 2014). The study developed an instrument measuring four attributes of teachers' understanding of rational numbers and analyzed the item responses using the log-linear cognitive diagnosis model (a general form of DCM models). The authors illustrated how this DCM model can be used to detect distinct patterns of attribute mastery and how the model can provide detailed feedback with respect to multiple facets of hypothesized knowledge subdomain.

Motivated by the advantages of DCMs and the process of applying DCMs illustrated in Bradshaw et al.'s study (2014), I decided to apply the log-linear cognitive classification (LCDM) model (Henson, Templin, & Willse, 2009) in this dissertation study. Specifically, I applied the LCDM model, a general measurement model within DCMs, to examine groups of teachers classified by their knowledge levels on multiple knowledge attributes. Moreover, considering the small number of items (minimum three items) for some of the hypothesized knowledge dimensions, the LCDM model promises to be an effective way of estimating teachers' knowledge attributes with acceptable reliabilities. The specific models used in this study will be described in more detail in Chapter 5.

2.3 Chapter summary

In this chapter, I introduce the literature about conceptualizations of teachers' mathematical knowledge. In this introduction, I compared and contrasted how different research groups (in particular, those who have developed tests to measure teachers' knowledge for large-scale studies) have conceptualized the dimensionality of teachers' mathematical knowledge for teaching. I also introduce measurement models that have been used for investigating the dimensionality of teachers' mathematical knowledge. As discussed in this chapter, there might not be one best dimensionality framework or one best methodology that can reveal the true structure of teachers' knowledge. In other words, each framework has its own advantages with respect to its purpose. However, if the purpose is to distinguish multiple aspects of teachers' mathematical knowledge, which are specific to their professional work, we may need to consider an alternative framework that allows us to do so. This alternative framework needs to be devised not only based on theoretical but also based on methodological considerations, in particular, if the conceptualized framework is to be operationalized by test items.

2.4 Research Questions

The knowledge framework that I suggest for examining teachers' distinguishable aspects of mathematical knowledge is described by making use of two organizers: tasks of teaching and instructional situations. In terms of these two organizers, the guiding research questions posed in Section 1.3 can be articulated as follows.

Given a nationally representative sample of in-service U.S. high school teachers, can we make valid inferences about the dimensionality of teachers' mathematical knowledge based on the responses on the items measuring the knowledge? Here, the knowledge is specifically what is needed to teach geometry and algebra 1 in the context of U.S. secondary school.

The research question in this dissertation consists of two kinds: conceptual plausibility and methodological feasibility

1. Conceptually, how well do the organizers – task of teaching and instructional situation – capture multiple aspects of teachers' mathematical knowledge used in the work of teaching?
2. Methodologically, are the knowledge scales estimated by items measuring teachers' mathematical knowledge used in either or both different tasks of teaching and different instructional situations statistically distinguishable?

In an effort to better understand the results from the methodological questions (identified differences and similarities among the hypothesized dimensions), the relationships between the dimensions and teachers' background information was further investigated. Thus, the subsequent research question is that

3. Are there differences among the hypothesized dimensions of teachers' mathematical knowledge for teaching geometry and algebra 1 in terms of their relationships with teachers' self-reported background (number of mathematics courses taken in college, number of geometry (or algebra 1) teaching years, number of non-geometry (or non-algebra) teaching years)?

The inquiry processes conducted to investigate these research questions and related concepts is described in the following chapters.

Chapter 3

Proposed Item Blueprint

To conceptualize and operationalize dimensions of teachers' mathematical knowledge, this study suggests a framework organized by tasks teachers do and instructional situations where those tasks are employed. The purpose of the framework is to structure test items so that they can measure distinguishable (but related) dimensions of teachers' mathematical knowledge. The developed measures are expected to allow obtaining a fine-grained description of teachers' mathematical knowledge used in the work of teaching.

Given that the focus of this dissertation is on the dimensionality of mathematical knowledge for teaching secondary mathematics, this study limits the population to U.S. high school in-service teachers and considers the teaching context of U.S. high school mathematics courses (geometry and algebra 1). The focus on this narrowed teacher population should highlight the specificity of knowledge used in particular courses of mathematical study. Moreover, as discussed in 2.1.3, this teacher population (high school teachers) differs from the one used to develop the MKT framework (elementary teachers) and this difference also warrants the consideration of an alternative knowledge framework.

3.1 The foundations of the hypothesis on the framework

The focal construct of the present study is teachers' mathematical knowledge for teaching high school geometry and algebra 1. However, broad definitions of constructs are generally not sufficient to operationalize multiple dimensions of those constructs. Given that ambiguous definitions of constructs of interest have sometimes resulted in divergent measures, which in turn have led to scattered and inconsistent understandings, it is essential to provide a clear theoretical conceptualization and operational definition of the construct this study intends to measure.

The theoretical conceptualization of the knowledge organization suggested in this chapter builds on preliminary studies investigating teachers' mathematical knowledge for teaching high school geometry (Herbst & Kosko, 2014; Ko & Herbst, 2017; Ko & Herbst, 2019). One of the studies (Ko & Herbst, 2019) operationalized the construct of SMK-G (subject matter knowledge for teaching geometry) and showed evidence supporting the argument that teachers' SMK-G is distinguishable depending on the tasks of teaching where mathematical work is required. Here, the tasks of teaching refer to the "professionally recognizable components" of the work of teaching (Hoover et al., 2014, p. 8) that teachers routinely do in the course of their daily work (Gitomer, Phelps, Weren, Howell, & Croft, 2015, p. 501).

Building on this evidence, a second preliminary study (Ko & Herbst, 2017) investigated the dimensionality of an MKT-G instrument consisting of items designed to measure not only SMK-G, but also pedagogical content knowledge for teaching geometry (PCK-G). The items of PCK-G were designed to attend to what Ball and colleagues (2008) had called KCS or KCT. The underlying assumption was that there would be more than one distinguishable item cluster, where each cluster could be characterized by item features reflecting the type of task of teaching and instructional situation. The consideration of instructional situation in the structure of teachers' mathematical knowledge is based on a study by Herbst & Kosko (2014). In that study, teachers' performance on the items (designed to measure teachers' MKT-G) were observed to be dependent on whether the items were set in the context of a recurrent instructional situation. Specifically, there were differences in teachers' performance between the items tapping into knowledge needed to manage a customary instructional situation and the items tapping into knowledge needed to manage students' work on a task that was novel (not customary).

Observations on these previous results provided the foundational hypothesis of the framework used in this study, namely that task of teaching and instructional situation are two main knowledge organizers that can be used to create an instrument that represents the knowledge needed by high school mathematics teachers to teach their courses. In this present study, I do not intend to cluster items in terms of MKT domains (e.g., CCK, SCK, KCS, KCT). Rather, I propose to define each dimension of the knowledge construct by using the product of simultaneous consideration of dimensions from two organizers (task of teaching and instructional situation) to identify distinguishable dimensions of teachers' knowledge. By *task of teaching* I mean a mathematical but general (that is, not specific to particular mathematical content) description of what the teacher is doing: creating problems for students to solve, observing students' work on problems, responding to students' work on problems, explaining new ideas, etc. By *instructional situation* (Herbst, 2006) I refer to the types of mathematical work that students do in particular courses of study: In algebra 1, these include solving, graphing, simplifying, calculating; in geometry, these include constructing, doing proofs, calculating, exploring figures (see Herbst et al., 2010). The two concepts, task of teaching and instructional situation, are described in greater detail in 3.3.

3.2 Operational definition of the knowledge construct

In addition to the conceptualizing the construct of interest, it is necessary to provide an operational definition, clarifying how the unobservable theoretical construct will be measured (Furr & Bacharach, 2013, p. 5). The main construct of this study is operationally defined as the pattern in teachers' responses on multiple-choice items, which are represented as numbers (binary for multiple-choice items and ordinal for testlet[5] items). The theoretical and operational definitions of constructs also necessitate an articulation of the context where those constructs are present (Furr, 2011, p.12). The context here is instruction in geometry and algebra 1, and the items used in this are presented in written-scenarios of instruction in those courses of studies. For example, a situation where students work on a task of making conjectures from a given geometric figure is hypothesized to be different from a situation where students work on doing a proof for a statement provided by the teacher.

Regarding the need of a description of the similarities and differences among dimensions when clarifying a construct, this study hypothesizes high correlations among different items if those items belong to the same dimension with respect to one of the organizers considered. For example, a significant correlation might exist between an item focusing on the knowledge needed for evaluating student work in a situation of calculation in geometric calculation and an item targeting the knowledge needed for choosing the givens for a problem in the same situation of calculation in geometric

[5]A testlet (Wainer & Kiely, 1987) is defined as an aggregation of items that are based on a single stimulus

calculation, because both items are related to the same instructional situation. Yet, these dimensions are conjectured to be distinct inasmuch as the tasks of teaching that call for the knowledge differ. In other words, the dimensions representing different tasks of teaching or instructional situations are hypothesized to be less correlated with each other than the dimensions of the same task of teaching or instructional situation.

3.3 The two organizers and the conceptual rationale of the blueprint

Keeping in mind these theoretical and operational articulations of teachers' mathematical knowledge and the contexts where the knowledge is used, this study suggests that teachers' mathematical knowledge for teaching is organized around an integrated knowledge trait that can be defined in terms of the types of teaching tasks (e.g., understanding student mathematical work) and instructional situations (e.g., doing a proof in a geometry class). This perspective is different from that of the LMT group's MKT instrument, which categorized items by pairing content areas (e.g., number concepts and operations, patterns, functions, and algebra) and knowledge domains (e.g., CCK, SCK, KCS) (Hill, Schilling, & Ball, 2004; Hill, 2007). The notion of task of teaching is also present in the LMT group's item development. Several tasks of teaching may be used to create items classified within a single MKT domain (e.g., CCK, SCK, etc.) (Table 3.1). But the LMT group does not conceptualize tasks of teaching as an organizer of dimensionality in MKT.

The present study, however, suggests a knowledge map where the type of task of teaching is one organizer that undergirds divisions of the construct (
Table 3.2). I expect that this suggested framework organized by two components (task of teaching and instructional situation) simultaneously would allow us to capture multiple

distinguishable aspects of knowledge specific to teaching secondary mathematics. This proposed way to organize teachers' mathematical knowledge is also warranted by the descriptive purpose of this study. In other words, if the purpose is to describe the knowledge used in the work of teaching, it would be more reasonable to have a framework that distinguishes knowledge dimensions by the natural divisions in the work of teaching than by the object of the knowledge (i.e., knowledge about student misconception).

The distinction between my proposed framework and a knowledge domain framework (e.g., MKT) can also be understood in the context of de Jong and Ferguson-Hessler (1996)'s *knowledge types* that distinguishes situational knowledge and conceptual knowledge. Their "situational knowledge", which refers to the knowledge of problem situations, may correspond to the knowledge dimensions generated by my suggested framework. On the other hand, their "conceptual knowledge", which refers to the static knowledge about facts, concepts, and principles, may correspond to the knowledge domains classified by domain-based knowledge framework that specifies an object of the knowledge (e.g., knowledge about student misconception).

In short, the proposed way that organizes the teachers' mathematical knowledge by tasks of teaching and instructional situation is conceptually plausible regarding the purpose of the measures. Specifically, if the purpose of using measures is to describe the dimensionality of teachers' knowledge used in the work of teaching, the knowledge framework should be embedded in the characteristics of the work of teaching mathematics rather than in the concepts hypothesized to be required for the work of teaching mathematics. In the following sections, I further provide a justification for each of the suggested organizers in terms of empirical and theoretical background.

Table 3.1 LMT group's MKT framework

Knowledge Type				
CCK: Common Content Knowledge (in doing Task A, B, or C, etc.)	SCK: Specialized Content Knowledge (in doing Task A, B, or C, etc.)	KCS: Knowledge of content and student (in doing Task A, C, or D, etc.)	KCT: Knowledge of content and teaching (in doing Task A, B, or C, etc.)	KCC: Knowledge of content and curriculum (in doing Task A, B, or C, etc.)

Table 3.2 Framework proposed in this study

		Task of Teaching				
		Task A, e.g., Understanding students' work (USW)	Task B, e.g., Choosing the givens for a problem (CGP)	• • •	Task F, e.g., Explaining a mathematical concept	• • •
Instructional situation	Instructional Situation A, e.g., Geometric calculation (CG)	Knowledge for doing A (USW) in a situation A (CG)	Knowledge for doing B in a situation A	• • •	Knowledge for doing F in a situation A	• • •
	Instructional Situation B, e.g., Doing proofs (DP)	Knowledge for doing A in a situation B	Knowledge for doing B in a situation B	• • •	Knowledge for doing F in a situation B	• • •
	⋮	⋮	⋮		• • •	• • •
	Instructional Situation F, e.g., Exploring a figure (EF)	Knowledge for doing A in a situation F	Knowledge for doing C in a situation F	• • •	Knowledge for doing F in a situation F	• • •
		⋮	⋮	⋮	⋮	⋮

3.3.1 Organizer One: Task of Teaching

By the definition of mathematical knowledge for teaching, a crucial feature that makes it professional is its specific use in the work of teaching. Thus, it is reasonable to propose that an analysis of this work could serve to unpack the knowledge. Accordingly, I conceptualize the dimensionality of teachers' knowledge based on an operational decomposition of the work of teaching based on what specific work teachers do. These components of the work have been referred to as "tasks" and the descriptions of these tasks and the knowledge associated within them could be described via "task analysis" (Jonassen, Tessmer, & Hannum, 1999; Glaser & Bassok,1989). In the following, the concept of "task analysis" is described in general; then, its potential role in understanding teachers' mathematical knowledge is discussed in more detail.

3.3.1.1 General description of task analysis

The field of task analysis has been developed within the area of instructional design, where tasks referred to the work the learner did. In this dissertation, I apply the concept to what the teacher does. Jonassen et al. (1999) defined "task analysis for instructional design" as the process of analyzing and articulating the kind of learning outcomes that the educator expects from the learners (i.e., persons who perform the tasks). Given that teachers are the persons who perform the tasks, analyzing tasks of teaching includes articulating expected teachers' performance within each task. While this articulation could be used for preparing teachers, this is not a focus of the present work. Rather, the notion of task analysis is used to describe the work in which teachers might be making use of mathematical knowledge. Jonassen et al (1989) described (cognitive) task analysis as including activity analysis, learning analysis, job/procedural analysis, and subject matter content analysis. Given that cognitive task analysis focuses more on the underlying knowledge associated with task performance and other domains of task analysis focus more on the behaviors of task performance, it seems reasonable for the purpose of this study to specify task analysis as "cognitive task analysis." In addition, as my objective is to investigate the multiple dimensions of knowledge associated with multiple components of the work of teaching, I particularly focus on two of the functions of task analysis introduced by Jonassen et al. (1989). The two functions are determining 1) "the operational components of jobs, skills, learning goals or objectives, that is, to describe what task performers do, how they perform a task or apply a skill and how they think before, during, and after learning" and 2) "how to construct performance assessments and evaluation."

In this study, the first function of describing the operational components can be considered as describing tasks of teaching, and the second function of constructing assessments can be considered as constructing assessments measuring teachers' mathematical knowledge for teaching high school mathematics. How the function of task analysis has been used in the studies focusing on the work of teaching is discussed in the next section.

3.3.1.2 Task analysis in the literature of mathematics teaching

The work of teaching has been conceptualized as being represented by multiple tasks of teaching. For example, Hoover et al. (2014) describe a task of teaching as "a decomposition of the work of teaching into professionally recognizable components" (p. 8). Similarly, Ball and Forzani (2009) define the work of teaching as "core tasks that teachers must execute to help pupils learn" (p. 497), which can be identified through an analysis of teachers' work happening both inside and outside the classroom. The process of identifying these tasks, (i.e., task analysis) has generally been undertaken to inform teacher education curricula (Grossman & McDonald, 2008; Ball & Forzani, 2009; Haertel, 1991; Ball, Sleep, Boerst, & Bass, 2009). For example, Ball and Forzani (2009) argued that teaching practice needs to be unpacked so that the core tasks can form the basis of the content of a professional curriculum. Similarly, Grossman and McDonald (2008) emphasized the need of a framework for parsing teaching practices so that a curriculum for professional education can focus on core practices.

A variety of methods and criteria have been used to identify these core tasks or practices. Regarding methods, some studies used previous studies or experts' views on the category of tasks (Reynolds, 1992), whereas some other studies mainly identified tasks through classroom observations (Ball & Bass, 2003b; Ball et al., 2008). Across different studies, some differences could be identified regarding the list of tasks of teaching. One of the differences is the range of tasks. For example, some studies identified tasks which are distinctively mathematical (e.g., "finding an example to make a specific mathematical point;" Ball et al., 2008, p. 400), whereas some other studies included not only mathematical tasks of teaching, but also some institution-centered acts (e.g., attending teachers' meetings; Cooney, Davis, & Henderson, 1983, p. 10). Another difference is seen in the studies that conceptualized tasks of teaching by a time frame such as "preactive" tasks (i.e., tasks for planning a lesson; e.g., comprehend content and materials), "interactive" tasks (i.e., tasks done in the classroom; e.g., implement and adjust plans during instruction), and "postactive" tasks (i.e., tasks of reflecting a lesson;

e.g., seek professional development) (Reynolds, 1992). But other studies ignore such different time frames and instead focus on tasks within lessons (Ball & Bass, 2003b). Some others have described the work of teaching in terms of its purposes. For example, the Connecticut Competency Instrument (CCI), which was developed to assess beginning teachers' lessons, was based on a framework that included ten tasks of teaching categorized into three categories according to their purposes, such as classroom management, instruction, and assessment (Haertel, 1991).

Another main difference among frameworks is the degree of specification in task description. For example, Ball et al. (2008) described tasks that are specifically mathematical, such as adapting the mathematical content of textbooks. Similarly, Etkina et al. (2018), who conceptualized physics teachers' content knowledge for teaching energy, used physics-specific tasks of teaching, e.g., "anticipating student thinking around science ideas" (Etkina et al., 2018, p.010127-3). In contrast, Haertel (1991) described tasks more generally, such as creating a structure for learning. The tasks related to managing the classroom environment, such as "engaging students in activities of the lesson" (Haertel, 1991, p.21), tend to be described more generally than the tasks teachers do with the subject content, such as "making and explaining connections among mathematical ideas" (Ferrini-Mundy & Findell, 2010, p. 34).

While there is a lack of consensus on the core tasks, there is some agreement on the criteria that researchers have used to identify them. First, most tasks identified across studies are those that teachers routinely or frequently do to teach mathematics. For example, Ball et al. (2008) listed 16 mathematical tasks of teaching, each of which is "something teachers routinely do" (Ball et al., 2008, p. 400). Similarly, Ball et al. (2009)

employed the criterion "is done frequently when teaching mathematics" to identify practices essential for beginning mathematics teachers (Ball et al., 2009, p. 461). Kazemi, Lampert, and Ghousseini (2007) also used a set of routines of practice in mathematics teaching to decompose the work so that novice teachers can learn how to do it.

Second, all the lists of core tasks include tasks associated with teachers' interaction with students. Some examples are "engaging students in activities of the lesson" (Haertel, 199, p. 21) , "evaluate student learning" (Reynolds, 1992, p.4), "respond productively to students' mathematical questions and curiosities" (Ball & Bass, 2003b, p. 11), "finding the logic in someone else's (students) argument or the meaning in someone else's representation" (Ferrini-Mundy and Findell, 2010, p. 34), "student-centered acts" (Cooney, Davis, & Henderson, 1983, p. 11), and "stimulating and managing classroom discourse" (NCTM, 1991, p.5). In terms of the time frame, most tasks associated with the interactions with students are expected to be performed during a lesson. Considering that teachers need to do mathematical work not only during the class (interactive task), but also before the class (preactive task), this present study use both an interactive task (namely, understanding students' work) and a preactive task (namely, choosing the givens for a problem for students) in operationalizing teachers' knowledge in doing tasks of teaching.

Among the various criteria that researchers have used to define tasks of teaching, I particularly focus on tasks that are distinctively mathematical. Also, I consider the frequency with which a task is done as possibly related to mathematical knowledge in that the frequency of use of mathematical knowledge for doing the task may matter in how well the mathematics is known. My hypothesis on the effects of those multiple characteristics of tasks of teaching on the organization of teachers' mathematical knowledge and the potential benefits of organizing teachers' knowledge by tasks of teaching are discussed in the next section.

3.3.1.3 The role of task of teaching in the knowledge structure

In addition to the theoretical rationale, the organization scheme of teachers' knowledge by task of teaching can be warranted on account of previous empirical results demonstrating a difference in teachers' knowledge due to a difference among tasks. Regarding the previous studies, in a study investigating the structure of MKT items, Schilling (2007) suggested a factor other than subject matter that could explain the variation within a dimension of SCK, and he alluded to the task of teaching as a potential factor organizing items within a mathematics content area. In addition, the conjecture that teachers' knowledge would differ depending on the tasks of teaching was supported by differentiated effects of experience teaching geometry on teachers' level of knowledge between two different tasks of teaching geometry (Ko & Herbst, under review). Specifically, teachers' experience teaching geometry showed greater effect on their knowledge for more frequently encountered tasks (namely, understanding students' work) than for the tasks encountered less frequently (namely, creating givens of a problem).

The organization scheme of teachers' knowledge by tasks of teaching also has potential methodological benefits. To organize teachers' knowledge around types of teaching tasks will provide a basic structure applicable to other subject areas and grade levels (Phelps et al., 2014), given that many tasks of teaching are common across subject areas and grade levels. Meanwhile, this basic structure could provide a foundation in which a specificity of tasks of teaching for a specific subject area to a grade level can be better elaborated and compared to other areas under the common understanding about the basic structure. Consequently, the blueprint for an instrument could be more conveniently adaptable for different grade levels where the characteristics of tasks of teaching are different or when a new teaching task is identified (Hill, 2016).

Lastly, as a task-focused knowledge structure directly relates teachers' mathematical knowledge to teaching practice, more practical feedback that can be readily applicable to the actual teaching practice could be provided for teachers under task-focused knowledge framework. For example, Hill (2016) emphasized the advantage of the task-oriented map in that it would enable the teacher knowledge measure to better approximate the actual practices that teachers are expected to master. (Hill, 2016, p. 5).

In sum, in accordance with these benefits, this study proposes organizing teachers' mathematical knowledge in terms of tasks of teaching. Specifically, this study asks: if teachers are knowledgeable in one task of teaching for a situation, how likely is it that they would be knowledgeable in another task of teaching for the same situation? This question examines teachers' competence in the knowledge used for different tasks of teaching in the same instructional situation. The specific categories of task of teaching operationalized in this study are defined in 4.3.2. In addition to the task of teaching, this study proposes one more criterion organizing dimensions and this additional organizer is instructional situation. In other words, this study examines the possibility of organizing teachers' mathematical knowledge with two organizers, task of teaching and instructional situation, simultaneously. The concept of instructional situation is described in the next section.

3.3.2 Organizer Two: Instructional Situation

In addition to *task of teaching*, this study introduces *instructional situation* as another organizer structuring the variation in teachers' mathematical knowledge for teaching in addition to *task of teaching*. Herbst (2006) defines "instructional situation" as "any one of the customary ways in which classroom actions are framed into units of [mathematical] work so as to be traded in for (or accounted to) claims over the knowledge at stake (and, reciprocally, any one of the customary ways in which the teaching or learning of objects of knowledge is deployed as classroom [mathematical] work)" (Herbst, 2006, p. 316, square-bracketed text added for clarification). In line with this, different instances of mathematical work framed by the same instructional situation (e.g., different proof problems) are regulated by the same norms (i.e., the same expectations of who is to do what and when; e.g., problems are likely to include a labeled diagram). Instances of student work that are framed by the same instructional situation are likely to be similar to each other in terms of "what mathematical elements they contain, what actions they call forth from the teacher and students, and what their completion is evidence of" (Aaron & Herbst, 2015, p. 4). Also, instances of work framed under different instructional situations (e.g., a proof problem and a geometric calculation problem) respond to different norms regarding the actions students and teachers are expected to do (e.g., while the first case is likely to include a labeled diagram, the second is less likely to do so)

In terms of operationalization, this study conjectures that the degree of relationship among items measuring teachers' knowledge can be explained by how likely the student work represented in those items are framed by the same instructional situation. I speak of the likelihood that an instance of student work be framed by an instructional situation because the specific piece of work may abide by some but maybe not all of the norms of an instructional situation (for example, while a norm of doing proofs is that the teacher provides a labeled diagram, it is possible that a teacher might have their students work on a proof problem whose diagram was given without labels). The conjecture then says that, the likelihood is high that a strong correlation will exist among items measuring teacher knowledge about teacher's management of student work in the same instructional situation. In contrast, a weak correlation will exist among items that call the teacher to manage novel student mathematical work which is less likely to be framed by any instructional situation. Moreover, those items will be less likely correlated with the items framed in customary instructional situations. These hypothesized relationships among items with respect to different instructional situations hold within and across Task of Teaching (hereafter, ToT). Therefore, the likelihood of being framed as one or another situation (or none) is independent of ToT.

Instructional situations group sets of similar instances of student work. Figure 3.1 conceptualizes this relationship between instances of student work and instructional situations proposed in this study. In this figure, the black dots represent instances of student work; they may be situated (framed) in one of the instructional situations (represented by big circles) or otherwise be completely novel student work, outside of situations (e.g., Dot 1). The two dots inside Situation 1 would be instances of student work with the greatest likelihood of being framed by Situation 1, while dots that are outside Situation 1 would have a smaller likelihood. The dots outside any of the

Situations represent novel (student) tasks for which the probability is low that they are framed under any instructional situation.

With this conception of instructional situation in mind, this study asks: if teachers are knowledgeable in a task of teaching for one situation, how likely is it that they would be knowledgeable in that same task of teaching for another situation? This question is central to my study aiming to assess teachers' competence in the knowledge needed for the task of teaching in different instructional situations.

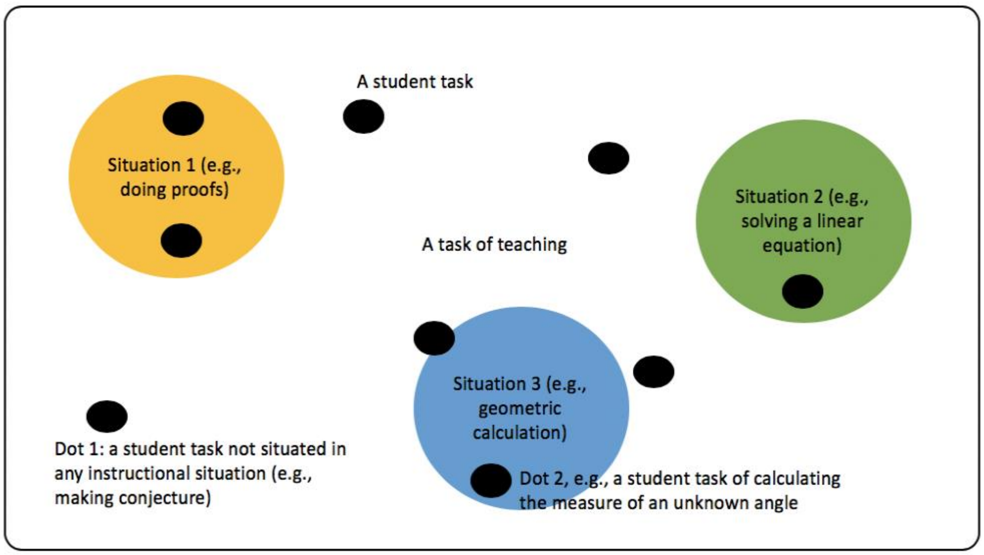


Figure 3.1 The relationship between student task and instructional situation

The hypothesis that teachers' knowledge in a task of teaching for one situation is different from that in the same task of teaching for another situation was built from an earlier study of MKT-G (Herbst & Kosko, 2014). Importantly, that study found that more experienced geometry teachers did better than less experienced teachers in some SCK items contextualized in mathematical work framed by an instructional situation common in the geometry course (e.g., calculating a measure), whereas there was no difference between more experienced teachers and less-experienced teachers for items that involved mathematical work that was novel for students and was not framed by an instructional situation (e.g., generating a construction method for a figure). From this finding, the authors conjectured that the mathematical work a teacher needs to do to support instructional situations that frame familiar work in the geometry course of studies is different from the mathematical work they might need to do to support students' involvement in novel geometry tasks.

As different instructional situations frame the way students undertake mathematical work differently, teachers' mathematical knowledge called for in the same task of teaching (e.g., evaluating students' work) may differ depending on the instructional situation. For example, in an instance of the instructional situation of calculation in geometry, the teacher task of evaluating student work may include verifying whether students used the known properties of a figure, set up the equations matched to those properties, and found an unknown measure of a single geometric object (Herbst, 2010). In an instance of another instructional situation, such as the situation of doing proofs, evaluating student work may include verifying whether students associated each statement with a correct reason and showed all necessary steps for the proof.

Similarly, teachers' mathematical knowledge used to choose appropriate givens (including diagrams) for a problem in a situation of calculation in geometry may differ from the knowledge used to choose appropriate givens (including diagrams) in a situation of exploring a figure. This is especially true in regard to what the teacher needs to anticipate how students would interact with those givens (particularly the given diagrams) in those two situations. Specifically, students may not measure the diagram in a situation of calculation in geometry, because diagrams are typically given with some extra signs (numbers or hash marks) revealing the properties of a figure. In another situation, students may be expected to measure or mark in a diagram and state properties they found when exploring a figure (Herbst, 2010). Thus, teachers' knowledge used to choose an appropriate diagram that can possibly engage students in intended mathematical work will be different between these two instructional situations. In this regard, this study proposes the need to structure teachers' mathematical knowledge not only in terms of specific tasks of teaching but also in terms of the specific instructional situations where the knowledge might be required. This hypothesis is also supported in one preliminary study (Ko & Herbst, 2017) investigating the dimensionality of MKT-G. In that study, two distinguishable item clusters were identified within an entire set of MKT-G items. These two identified item clusters could be characterized by the type of task of teaching and instructional situation, which have been hypothesized as two dimensions explaining the variances among teachers' performance on the items (Ko & Herbst, 2017). The specific categories of instructional situations operationalized in this study are described in more detail in a later 4.3.4.

3.4 Chapter summary

In this chapter, I introduce a new item organization scheme that would allow the identification of the multiple sub-dimensions of teachers' mathematical knowledge for teaching high school geometry and algebra 1. Following this suggested organization, the next chapter describes how the hypothesis on the dimensionality of teachers' knowledge was investigated using the data collected from nationally representative sample of U.S. high school mathematics teachers.

Chapter 4

Method

4.1 Instruments

The data analyzed in this study includes 602 U.S. high school mathematics teachers' responses on three knowledge assessments (MKT-A, MKT-G, SMK-G)[6] and a background survey asking them about their educational background and teaching experiences. The data were collected from March 2015 to January 2016 as a part of a larger research project[7]. Among the three instruments, two geometry instruments (MKT-G and SMK-G) were developed by the GRIP (Herbst's research group) and these instruments were designed to measure teachers' mathematical knowledge for teaching high school geometry. The other instrument MKT-A was developed as a part of the Measures of Effective Teaching (MET) study (Phelps et al., 2014) to measure teachers' content knowledge used in teaching algebra 1. The initial item development framework of these instruments was not the same (MKT-G was framed by the domains of knowledge type proposed by the LMT group; SMK-G included only CCK and SCK items according to knowledge type, but also focused on two tasks of teaching and three instructional situations; MKT-A was framed by task of teaching). Still, most items from these

---

[6] MKT-A (MKT-Algebra), MKT-G (MKT-Geometry), SMK-G (Subject Matter Knowledge-Geometry)

[7] NSF project DRL- 0918425, 2009-2017. (Herbst, PI; Chazan, coPI)

instruments could be described in terms of at least two common organizers – task of teaching (ToT) and instructional situation.

The similarities and differences among the items with respect to a course of studies, task of teaching, and instructional situation have enabled me to evaluate the dimensionality of teachers' mathematical knowledge in a multifaceted way. For example, items across different courses of mathematical study (algebra 1 or high school geometry) could be combined into the same group by task of teaching criterion (e.g., the mathematical knowledge used in the task of evaluating student work), so the hypothesis such as whether the relationship among items designed to measure teachers' mathematical knowledge can be approximated by distinguishing task of teaching regardless of the course of studies (geometry or algebra 1) could be tested. Similarly, the influence of distinct instructional situation on the dimensionality of teachers' mathematical knowledge for teaching geometry or algebra 1 could be assessed. Given that instructional situation is a finer distinction than course of studies, the dimensions organized by instructional situation was assessed within a course of study (geometry or algebra 1).

In addition to the benefits of having items measuring multiple constructs subsumed within the same larger construct the data from the instruments developed by two different research groups (GRIP, MET) also has enabled me to evaluate convergent validity. In other words, a positive relationship among teachers' performance on different instruments supports the validity argument that teachers' mathematical content knowledge for teaching geometry and teachers' mathematical content knowledge for teaching algebra 1, which theoretically should be related, are indeed related. Using multiple measures for a broad theoretical construct – teachers' mathematical knowledge – is also in accord with one of the preceding treatments of construct validity emphasized by Shadish, Cook, & Campbell (2002) that "to use multiple operations to index each construct when possible" (Shadish, Cook, & Campbell, 2002, p. 81).

Regarding the data that will be used for this study, participants completed all three instruments by logging into the LessonSketch platform (http://www.lessonsketch.org). All the items were randomly ordered regardless of which instrument they were drawn from. The organization of items administered in this study is shown below (Table 4.1). More detailed information about each of the instrument will be introduced in the next section.

Table 4.1. Organization of administered items

Instrument	Course of study	The number of items: stem (total)
MKT-G	High School Geometry	28 (42)
SMK-G	High School Geometry	33 (60)
MKT-A	Algebra 1	23 (37)

4.1.1 MKT-G items

One set of items that used in this study is from the MKT-G instrument developed by Herbst's research group (the GRIP; Herbst & Kosko, 2014). As the item development framework was built on Ball et al.'s (2008) MKT theory, the 28 stem items (some of the items include multiple true/false sub-items) were initially developed to measure teachers' MKT-G (mathematical knowledge for teaching geometry) with items eliciting knowledge across four MKT domains − CCK, SCK, KCS, KCT − (Ball et al., 2008). The items represent a variety of tasks of teaching and they were situated in classroom contexts; each item calling the respondent to make a choice on behalf of a teacher involved in a task of teaching. Herbst and Kosko (2014) argued that this instrument captures specific mathematical knowledge for teaching high school geometry on the basis that participants' experience teaching high school geometry significantly correlates with their MKT-G scores, while experience teaching mathematics in general was not significantly correlated with MKT-G scores (Herbst & Kosko, 2014).

In the present study, 28 MKT-G items, each of which was originally developed to be mapped to one of the knowledge domains (CCK, SCK, KCS, KCT), has been re-analyzed and coded with respect to the task of teaching and instructional situation represented in the item.

4.1.2 SMK-G items

The second set of items used in this study was drawn from a second MKT-like instrument developed in the GRIP (SMK-G) which consists of 33 stem items developed to measure teachers' subject matter knowledge for teaching geometry (CCK, SCK). Herbst and his colleagues designed the SMK-G instrument to empirically investigate the role of instructional situation in teachers' MKT-G. Items were organized according to two tasks of teaching (understanding students' work and choosing appropriate givens for a problem) nested in three instructional situations (doing proofs, exploring figures, and geometric calculation). In this present study, 33 SMK-G items have been coded in terms of type of task of teaching and instructional situation.

4.1.3 MKT-A items

The MKT-A items used in this study is a set of 23 items that had been developed, as part of MET project (Phelps et al., 2014), to measure teachers' content knowledge for teaching algebra 1. The items had been developed based on tasks of teaching, which is defined as “the recurrent practices that make up the work of teaching [mathematics]” (p.3). Thus, each item involved one task of teaching (e.g., “creating and adapting resources for instruction”). In this study, 23 MKT-A items have been coded in terms of not only a type of task of teaching but also instructional situation.

4.2 Participants

A total of 84 items drawn from the three instruments (MKT-G, SMK-G, MKT-A) were administered to a nationally distributed sample of U.S. high school mathematics teachers across 47 states. To recruit participants, more than 12,000 public secondary schools in the United States were selected using a stratified systematic probability proportional to size sampling of schools with respect to geographical region and urbanity. One secondary mathematics teacher was randomly selected from each school and then recruited via email.

Table 4.2 includes descriptive statistics of the participating teachers’ educational background and their teaching experience. A total of 602 participants responded to at least one item and a total of 406 participants among them completed all 84 items. On average, participants had been teaching mathematics for 14.2 years (SD=8.7, min=1, max=40), and had taken 14 college-level mathematics courses (SD=7.3, min=2, max=40). In addition, teachers had been teaching geometry for an average of 5.6 years. The way of dealing with missing data is described in the section of data analysis.

Table 4.2. Descriptive statistics for teachers' background

	M	SD	MIN	MAX
Gender	M: 40%, F: 60%	N/A	N/A	N/A
College math coursework	13.9	7.3	2	40[8]
Geometry coursework	1.9	1.8	0	20
Algebra 1 coursework	3.8	3.0	0	21
Total years of teaching	14.2	8.7	1	40
Years of teaching geometry	5.6	5.9	0	35
Years of teaching algebra 1	6.3	5.9	0	32
Currently teaching geometry	45% (Yes)	N/A	N/A	N/A
Currently teaching algebra 1	40% (Yes)	N/A	N/A	N/A

The majority of the participants self-identified as Caucasian (83%), while the rest included 8% identified African American, 2.6% as Asian, and 1.9% as Hispanic or Latinx; and they were mostly female (60%). The participants' schools were diverse with respect to urbanicity (city: 23.9%; suburbs: 30%; towns: 16.8%; rural:29.3%). These figures are consistent with nationally representative data obtained from the National Center for Education Statistics (NCES) database. (Caucasian: 81.5%, female: 57.3%[9], city: 27.0%, suburbs: 31.5%, towns: 13.7%, rural: 27.7%[10]).

4.3 Data Analysis

Teachers' responses to the 84 multiple-choice or true/false questions were scored dichotomously as "1" for correct and "0" for incorrect. 21 items have multiple true/false

[8] The participants were asked to choose one of the options ranged from 0 to 40.

[9] U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public School Teacher Data File," 2011-12, Table 209.50. Percentage of public school teachers of grades 9 through 12, by field of main teaching assignment (here, mathematics) and selected demographic and educational characteristics: 2011-12

[10] U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "Public Elementary/Secondary School Universe Survey," 2013–14 (version 1a), Table A.1.a.-2 Number of public elementary and secondary schools, by school urban-centric 12-category locale and state or jurisdiction: 2013–14.

questions and 63 items are single multiple-choice questions. For testlet questions with 3 or 4 sub true/false items within one stem, the sum of the multiple true/false binary scores within one stem question were used as an item score. Therefore, the maximum score possible is the same as the number of sub-items. This scoring method was applied to control for testlet effects (Phelps & Schilling, 2004) and it provided a single ordinal score for each testlet.

The ordinal scores for 84 items were then analyzed together under classical test theory (CTT) to examine item difficulties and biserial correlations (the correlation between a binary value and a continuous total score). This enabled me to eliminate items with extreme difficulty levels (too easy or too difficult) and items with a negative or negligible correlation with other items[11]. In addition, according to the focus of this study, we used only the items that required participants to do mathematics in the context of teaching work and excluded items that asked the same mathematical questions that students might be asked in class, e.g., an item just asking participants to select, from a set of four statements, the one statement that is not a property of parallelograms. Considering the reliability of a scale, several items were further excluded when they involved a unique task or instructional situation that cannot form a cluster of three items that have common trait of task of teaching or instructional situation.

[11] I eliminated items yielding p-values (difficulty) greater than 0.95 or less than 0.05. In addition, items yielding a negative or lower than or equal to 0.1 item-rest correlation were eliminated.

4.3.1 Missing values

All items from the three instruments (MKT-A, MKT-G, SMK-G) were randomly ordered and divided into eight forms, each of which includes 9~11 items. All participants were also randomly divided into 12 groups, consisting of 30 ~50 participants. Participants were required to complete all items and could not proceed if any item was unanswered. To avoid the possibility that items located at the end have a lower response rate than other items, the order of forms was changed across groups. For example, Group One participants took the forms in order 1, 2, 3, …, 8, whereas Group Two participants took the forms in order 2, 3, 4, …., 8, 1. This instrument design indeed resulted in similar response rate for each item.

Before conducting the dimensionality analyses, missing data were further evaluated to understand patterns and determine the best method for handling missing values. The analysis showed that a total of 602 participants answered at least one item and a total of 406 participants completed all 84 items. In an initial analysis, any participant who had one or more missing items had been removed from the sample (list-wise deletion). But this approach was, recognized as possibly problematic: It was possible that teachers with weak mathematical knowledge might complete fewer items due to difficulty in solving them. Thus, I anticipated that list-wise deletion could result in removing participants who have low mathematical knowledge, and that this could bias the results. To avoid this potential bias, I decided to conduct all the analyses for both cases: a case with the complete data (N = 406) set using listwise deletion and case with the full data set including missing values (N = 602). As the results from complete data and full data set with missing values yielded the same conclusions, I report only the results from full data set. For the full data set with missing responses, dimensional analyses were conducted under *full information maximum likelihood* (FIML), which is known as superior to other conditions such as list-wise deletion, pairwise deletion, and imputation (Enders & Bandalos, 2001). The pairwise deletion approach was also conducted under the SEM framework using WLSMV estimator (discussed in the section 2.2.1). The two results (under FIML and under pairwise deletion) were compared by using different estimators (FIML is used in the IRT framework and pairwise deletion is used in the SEM framework) to confirm the consistency in the evaluation of different factor models.

4.3.2 Item content analysis

As described in the previous section of data description, each item was coded prior to the instrument administration with respect to the task of teaching involved (e.g., formulating a problem, reviewing student work) and the instructional situation involved (e.g., geometric calculation, doing proofs) to capture the components of the items related to the task of teaching and instructional situation. A thorough analysis of the item content with respect to the hypothesized knowledge organizers yielded nine clusters of items (Table 4.3) The items selected for this study have at least two components, tasks of teaching and instructional situations.

Each cluster[12] corresponds to one of the sub-dimensions characterized by one task of teaching and one instructional situation. Among the nine clusters, five clusters include items measuring teachers' mathematical knowledge for understanding students' work and

---

[12] A cluster was formed only if there are more than 3 items that have common trait of task of teaching or instructional situation.

four clusters include items measuring teachers' mathematical knowledge for choosing appropriate givens for a problem (Table 4.3). Within the same task of teaching, clusters were categorized according to instructional situation. Specifically, each of the five item clusters within the task of understanding students' work corresponds to one of the following instructional situations: calculation in geometry, doing proofs in geometry, exploring a figure in geometry, simplifying rational expressions in algebra, and solving equations in algebra. Similarly, each of the four item clusters within the task of choosing appropriate givens for a problem consists of items reflecting instructional situations of calculation in geometry, doing proofs in geometry, exploring a figure in geometry, and calculating with numbers. Sample items for each category are presented in Appendix. The organization of clusters by task of teaching and instructional situation is described in Table 4.3 and each cluster is described in the following section.

Table 4.3. Organization of items

		Task of Teaching	
		Understanding Students' Work (USW)	Choosing appropriate Givens for a Problem (CGP)
Instructional Situation	Calculation in Geometry (CG)	Items measuring mathematical knowledge for doing USW in CG (USW_CG): 6 items	Items measuring mathematical knowledge for doing CGP in CG (CGP_CG): 5 items
	Doing Proofs in geometry (DP)	Items measuring mathematical knowledge for doing USW in DP (USW_DP): 3 items	Items measuring mathematical knowledge for doing CGP in DP (CGP_DP): 4 items
	Exploring a Figure in geometry (EF)	Items measuring mathematical knowledge for doing USW in EF (USW_EF): 3 items	Items measuring mathematical knowledge for doing CGP in EF (CGP_EF): 4 items
	Simplifying Rational expressions in algebra (SR)	Items measuring mathematical knowledge for doing USW in SR (USW_SR): 3 items	No item
	Solving Equations in algebra (SE)	Items measuring mathematical knowledge for doing USW in SE (USW_SE): 5 items	No item
	Calculating with Numbers in algebra (CN)	No item	Items measuring mathematical knowledge for doing CGP in CN (CGP_CN): 3 items

4.3.3 Tasks of Teaching

Of interest in this analysis are the responses to items that measured knowledge associated with two different tasks of teaching: the task of Understanding Students' Work (hereafter USW, defined in the next section), and the task of Choosing appropriate Givens for a Problem (hereafter, CGP), which are defined as follows.

4.3.3.1 Understanding Students' Work (USW)

The task of understanding students' work (USW) includes a teacher's reading, making sense, representing a student's thinking in the teacher's mind, deciding whether the student's answer or inferred process is mathematically correct, identifying specific errors, and generating mathematical commentary on that work through inscriptions or writing.

4.3.3.2 Choosing appropriate Givens for a Problem (CGP)

The task of choosing appropriate givens for a problem (CGP) includes writing text of various kinds of questions such as classroom tasks or test items, choosing particular information (numerical, algebraic, diagrammatic) to include with the text and deciding how to encode it, ascertaining that the text calls for the student to do what the author of the problem expects them to do, and verifying that the information given is consistent and sufficient for students to engage in the work envisioned.

4.3.3.3 Rationale for using the two particular tasks of teaching chosen

As described in 3.3.1.2, researchers have used diverse criteria in identifying core tasks of teaching mathematics. For example, some studies focused on tasks that are distinctively mathematical (Ball et al., 2008; Hoover et al., 2014), whereas some other studies have also dealt with tasks that are more general, such as managing the classroom environment (Haertel, 1991). This study focuses on the teachers' knowledge which is needed to be used in doing the mathematical work of teaching; thus, the tasks of teaching focused in this study are distinctively mathematical. However, these tasks are described as generic insofar as they name work a teacher does across different subjects in mathematics, such as geometry or algebra; and the analysis explores whether hypothesizing a common knowledge dimension for each task of teaching enables us to describe the knowledge needed to respond to items in which the teacher has to do each task across content areas. By focusing on the mathematical tasks, this study provides a way to better understand the organization of the mathematical knowledge needed to be used for teaching mathematics.

The validation of methodological feasibility includes assessing whether the proposed item blueprint allows for developing multiple distinguishable measures representing the hypothesized dimensions of teachers' mathematical knowledge. In this section, I describe a rationale for using two particular tasks of teaching among others for the evaluation of methodological feasibility. The rationale is based on their representativeness in teacher's work managing instructional exchanges and their distinguishable characteristics in the work of teaching.

First, the two particular tasks were hypothesized to produce meaningful variation among the item responses in that they are representative of two key elements of the work that teachers do: creating mathematical work for students and evaluating what students do (Herbst, 2006). If we consider the term "instruction" in the context of the transaction of mathematical knowledge between teacher and students (or student work), the role of a teacher in instruction is to manage the instructional exchanges between students' work on problems and mathematical ideas at stake. To manage the exchange from the mathematical ideas at stake to students' work on problem, a teacher needs to engage in "deploying mathematical objects of study in the form of work for students to do" (Herbst & Chazan, 2012, p.606). This task includes choosing or creating problems for students in support of the knowledge at stake and providing students the problems to work on, which prominently includes choosing the givens for a problem (CGP), one of the tasks of teaching used in this study. Instructional exchanges also require the teacher to do work in the opposite direction, from students' work to the mathematical ideas at stake: a teacher needs to engage in "interpreting work (being) done by students in light of a mathematical object of study" (Herbst & Chazan, 2012, p.606). This work includes, prominently, the task of understanding and interpreting what the students do and think in those problems (USW), which is the other task of teaching used in study. In this regard, the two tasks of teaching, USW and CGP, are reasonably assumed to be different and to represent important, key tasks the teacher needs to engage in when they manage instructional exchanges in mathematics.

Second, not only were these two tasks of teaching (USW and CGP) hypothesized to be different in regard to their role in instructional exchanges, but they were also hypothesized to be different regarding the frequency in which each task might be found in the work of individual teachers. Related to findings from prior research, it was expected to be common for teachers to draw problems from textbooks or websites (Pepin & Haggarty, 2001; Robová, 2013; Schmidt et al., 1999) hence not engage very often in creating or choosing the givens for a problem (CGP). Yet, we surmised that most K-12 teachers in the U.S. have to read and understand student work as they usually grade their students' homework and tests. In this regard, it might be reasonable to conjecture that USW (understanding students' work) is a much more common task of teaching than CGP (choosing appropriate givens for a problem).

This conjecture was supported by a survey of 60 U.S. high school teachers who were selected from the 602 nationally represented sample in a subsequent wave of data collection. This group of teachers included 30 more-experienced and 30 less-experienced (than 5 years of teaching experience) teachers and they were asked to respond to survey-items asking how often in the academic year they had engaged in the tasks USW and CGP. The result derived from the teachers' responses showed that the frequency of doing CGP is significantly lower than the frequency of USW ($t (59) = -0.33$, $p = 0.0001$), which supports my hypothesis on the difference in the frequency of doing the task between CGP and USW.

Third, the two tasks of teaching (USW and CGP) were hypothesized to create meaningful variations among the items regarding the difference in a time frame in which the task is planned to take place. As mentioned in 3.3.1.2, some researchers conceptualized tasks of teaching by a time frame such as "preactive" work, "interactive" work, or "postactive" work, depending on whether the work is expected to happen before interacting with students (e.g., planning a lesson), expected to be done in the classroom, or for reflecting a lesson. According to this time frame criterion, the task CGP represents preactive work, whereas the task USW represents interactive or postactive work (e.g., responding to student work in class, grading student work after class).

Fourth, the two tasks of teaching were hypothesized to be different regarding the type of interaction with student work. As a teacher needs to read a student's work and makes sense of a student's thinking in doing USW, the task USW demands a wider range of literacy skills than CGP. On the other hand, CGP does not require a teacher to interact with students' work which is usually not written in a standard form accepted in a discipline.

The representativeness and differences between these two tasks of teaching warrant their use in testing the hypothesis that teachers' mathematical knowledge used in doing different tasks of teaching (in an instructional situation) is distinguishable. This conjecture on the differences between USW and CGP also led me to further hypothesize differentiated effects of teachers' experience in teaching geometry between two knowledge dimensions. The results are reported in Chapter 5.

4.3.4 Instructional Situations

As described earlier in Chapter 3, this study hypothesizes that the (strength of the) relationships among different items can be explained by how likely the student mathematical work represented in those items are regulated by the same instructional situation (that is, by the same division of labor between student, teacher, and the mathematical objects in the task). Instances of classroom mathematical work are considered to be framed by the same or by different instructional situations depending on the norms that regulate who has to do what and when. These norms specify mathematical elements of student tasks, actions called forth from the teacher and students, and the evidence of the completion of the task (Herbst & Chazan, 2012). In this regard, I describe the six instructional situations operationalized in this study in terms of "the roles and responsibilities of teachers and students in relationship to the mathematics at hand" (Chazan & Lueke, 2009, p.22), a focus of the concept of instructional situation (Herbst, 2006). Situations can be described at various levels of detail; while one might eventually distinguish a situation of construction with straightedge and compass from a situation of construction with dynamic geometry software, one could also speak of a situation of construction that includes the prior two and that could be characterized by norms that apply to both (Herbst, 2010). With this in mind I describe the following.

4.3.4.1 Calculation in Geometry (CG) (GCN and GCA)

In the situation of Calculation in Geometry, a task given to students (by a teacher or a textbook) involves calculating some unknown dimensions of a figure represented by a diagram and by some information (given numerically or algebraically) about other dimensions of the figure (Boileau & Herbst, 2015; Herbst, 2010, Herbst 2004, Hsu, 2010; Hsu and Silver, 2014). Students are required to use their knowledge of properties of figures and the given information included in the diagram to solve for the unknown dimensions and without using measurements of the diagram (Herbst, 2010). The category of Calculation in Geometry includes at least two subordinate situations represented in items used in the study: Geometric Calculation in Algebra (GCA; Boileau & Herbst, 2015) and Geometric Calculation in Number (GCN; Hsu & Silver, 2014). The main difference between the two situations is whether algebraic skills are needed to find the unknown dimensions. GCA requires algebraic skills to find the value of a dimension, whereas GCN does not (Hsu, 2010). The roles of the teacher and students in each situation are described below. The CG category in this study includes 10 GCA items and 1 GCN item, so I categorized both under CG in my analysis without distinction between them.

In GCN, a teacher is expected to

- provide a diagram of a figure in which some of its dimensions are set numerically while others are unknown. The diagram includes sufficient information enabling the students to use the information with their knowledge of properties of the figure and known theorems to find other information required to find the unknown (Herbst, 2010, p.50).

In GCN, students are expected to

- find deductively the unknown dimensions of a figure, using their knowledge of geometric properties and application of related theorems to set up arithmetic calculations.

In GCN, “What is stake is a claim on students’ capacity to use a property they already know” (Herbst, 2010, p.50). Figure 4.1. shows an example of a GCN task.

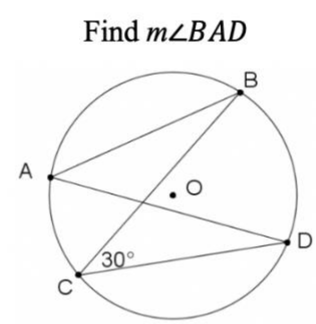


Figure 4.1. Sample GCN task

In this task, students are required to use their knowledge of a theorem (the inscribed angle theorem) and a given numeric value to find an unknown dimension (∠BAD).

Compared to GCN tasks, a GCA task involves algebraic calculation.

In GCA**,** the teacher is expected to

- provide a diagram of a figure in which some of its dimensions are represented using algebraic expressions while others are set numerically or unknown. The diagram includes sufficient information enabling the students to use the information and their knowledge of properties of the figure and known theorems to find other information required to find the unknown (Herbst, 2010, p.53).

In GCA, students are expected to

- Find deductively the numerical value of an unknown dimension of the figure by posing and solving algebraic equations representing relationships among the relevant dimensions of the figure. The algebraic equations are set up using their knowledge of geometric properties and applications of related theorems.

In GCA, "What is stake is a claim on students' capacity to use a property they already know as well as a claim on maintaining knowledge of algebra skills" (Herbst, 2010, p.53). For example, one of the items in GCA provides a parallelogram with some angle measures expressed algebraically. This item is presented in a context where students are asked to find the value of unknown angle *x*. The information given to students must yield a positive value of *x* and be consistent with the other information known about the figure. To find *x*, students are expected to use one of the properties of a parallelogram – that consecutive angles are supplementary – and set up an algebraic equation using the given algebraic expressions. Students solve the equation to find the value of *x*. When the teacher looks at students' work, they examine whether and how students connect the known properties of a parallelogram to the algebraic expressions given so as to set up an equation. The teacher also looks at the students' algebraic solution and how they use that solution to answer the question about geometric quantities.

4.3.4.2 Doing Proofs in geometry (DP)

In DP, a teacher is expected to

- provide students a specified 'given' and 'prove' statements when assigning a problem (given-prove norm) (Herbst, Aaron, Dimmel, & Erickson, 2013).
- provide students with a diagram representing the geometric objects involved in the given and prove statements using the diagrammatic register, that is, using geometric notation that refers to the specific geometric objects (rather than the more general geometric concepts). All points to be used in the proof are labeled in the diagram. The diagram accurately represents the figure of a diagram (diagrammatic-register norm).

In DP, students are expected to

- understand some information implicitly given by the diagram (particularly regarding properties of incidence, collinearity, and separation) and avoid making assumptions about other information given by the diagram (particularly regarding properties of parallelism, perpendicularity, and congruence).
- figure out chains of descriptive statements (about the geometric objects at hand) that connect the 'givens' to the 'prove'.
- justify those statements deductively by reasons drawn from their knowledge of theorems, definitions, and postulates that they learned in class.

  (Herbst, 2002b, Herbst 2004, Herbst & Brach, 2006; Herbst et al. 2009, Herbst et al., 2010).

In DP, what is stake is a claim on “students’ capacity to produce chains of statements and reasons that connect a given statement to a conclusion”, and “students’ capacity to recall and apply known theorems, definitions, and postulates to justify specific statements” (Herbst, 2010, p.54). For example, one of the items in this category provides a diagram of two triangles whose vertices are labeled. The proof problem is presented in the form of givens (the properties of the lines; e.g., $\overline{AE} \parallel \overline{BD}$) and the statement to be proved by students ($\angle EFA \cong \angle DCB$). The given information (diagrammatic) is expected to be sufficient for students to engage in the proof work envisioned. A student organizes a proof using a two-column form: one for statements and one for reasons. The teacher evaluates the student’s work by attending to whether each statement and reason are mathematically correct; whether the student’s proof contains all the required steps in a correct order; and whether the student’s statements are written using correct mathematical symbols and notations.

4.3.4.3 Exploring a Figure (EF)

In EF, a teacher is expected to

- provide students with a geometric object and tools (e.g., protractors, rulers) to use when exploring.
- ask students to explore figures given through diagrams or other concrete materials and to generate conjectures stating properties of the geometric figure represented by the objects at hand.
- not constrain a specific object or property that the students make a conjecture
- accept, revise, or reject the properties conjectured by students

In EF, students are expected to

- look at, measure, mark, or draw in a given diagram.
- state any property for the diagram using a combination of informal language and mathematical notation and as many properties as they wanted for each diagram. (Herbst, 2010, p.41)

- freely choose among a range of material operations to apply on concrete (physical or pictorial) embodiments of the concept depending the tools available to them, their reading of the particular results of those operations, and the translation of those results into general statements made in the conceptual register. The reasoning that students could thus have the opportunity to engage in can be described as abductive, proceeding from particular to general (Herbst, 2010; Aaron & Herbst, 2015).

In EF, what is stake is a claim on students' "knowledge of the definition, names, and properties of geometric figures, knowledge of symbols to express the properties of a particular geometric object, skill manipulating instruments for measuring and other instruments to check on properties, practice identifying various objects after geometric terms, and experiential learning of geometry" (Herbst, 2010, p.37). Students' capacity to reason inductively and abductively in inferring general properties from a particular models and capacity to reason deductively in verifying known properties of a given model are also at stake in EF. One of the items in this category provides students a circle that has a point in the center and has two particular arcs, formed by two intersecting chords, that are set to be the same. The teacher asks students to make conjectures based on the diagram. Each student makes their own conjecture on the diagram. The first student makes a conjecture regarding triangle congruence; the second student makes a conjecture regarding the length of chords; and the third student makes a conjecture regarding an angle of a triangle. As such, the diagram properties/features the students focus on for their conjectures can be diverse.

4.3.4.4 Simplifying Rational expression (SR)

In SR, a teacher is expected to

- provide a rational expression and ask students to simplify the given expression.

In SR, students are expected to

- find the greatest common factor of both numerator and denominator, if possible
- factor completely both the numerator and denominator
- use the Fundamental Principle of Rational Expressions[13] to divide out the common factor from the numerator and denominator and simplify (ElHitti, Bonanome, Carley, Tradler, & Zhou, 2017, p. 132; Cunningham & Yacone, 2013, p.459)

The mathematical work expected from students in SR is based on the methods that algebra textbooks and secondary school algebra teachers most commonly use when illustrating simplifying rational expressions (Cunningham & Yacone, 2013).

One item in this category shows an example of student task simplifying $\frac{10a+4}{2a}$.

4.3.4.5 Solving a linear or quadratic equation in algebra (SE)

In SE, a teacher is expected to

- provide students a linear or quadratic equation for which a solution exists.

In SE, students are expected to

- identify the type of equation

---

[13] For polynomials P, Q and R with $Q \neq 0$ and $R \neq 0$, $\frac{P}{Q} = \frac{P \cdot R}{Q \cdot R}$

- simplify or distribute the algebraic expressions involved, if necessary
- put like terms with like terms (put all the terms containing the unknown on one side of the equals sign with the constants on the other side)
- divide or to do something else in order to simplify
- write a string of equivalent equations that terminates in an equation of the form $x = a$ number (without justification).

(Chazan & Lueke, 2009, p.31-32; Buchbinder, Chazan, & Capozzoli, 2019)

One item in this category presents an example of student work solving the equation $3x^2 - 6x - 24 = 0$, while all the others are linear equations.

4.3.4.6 Calculating with Numbers (CN)

In CN, a teacher is expected to

- provide a problem of number calculation that can be simplified using the basic properties of real numbers (e.g., commutative, associative, distributive)

In CN, students are expected to

- apply the basic properties along with their knowledge of the decimal number system to computational settings.
- choose an appropriate property and apply the property efficiently.

One of the examples shows multiple expressions and asks the teacher to choose the best example to illustrate how using the distributive property can simplify computations. The answer is the one expressed by two terms that have the same factor and add up to 100 when divided by the factor.

According to the characteristics of categories described above, items were categorized into nine clusters and each of the cluster represents one of the hypothesized knowledge dimensions. The next sections describe how the items scores were analyzed to examine distinctions among the hypothesized dimensions.

4.3.5 Dimensionality of item scores

To evaluate the dimensionality of teachers' mathematical knowledge, several statistical models were fitted to item scores and the results from different models were compared. The models that I applied for the dimensional analysis are two prominent models under different assumptions: two models assumed teachers' knowledge as a continuous quantity (IFA) (under SEM & IRT framework) and a model that assumed teachers' knowledge as discrete quantity - mastery level or non-mastery level (diagnostic classification models) (discussed in Chapter 2). The comparison among results derived from different methods was done for the purpose of evaluating the consistency of the results as well as understanding the limitations and benefits of each model.

4.3.5.1 IFA and DCM models

Following the model taxonomy used in Wirth and Edwards (2007), this study classifies the SEM-based and IRT-based item factor analysis as "variants of a general item factor analytic framework" (Wirth & Edwards, p.59). The main difference between IFA within the SEM and IRT frameworks is that IFA within SEM uses limited information such as correlation or covariance matrices to assess the structure of items, whereas IFA within IRT generally uses full information such as raw data (Wirth & Edwards, 2007, p.66). In a comparison of the usage between the two frameworks, Wirth and Edwards (2007) suggested to use an IRT-based item factor analysis (IFA) if the research purpose is to understand individual item characteristics or obtain scores for individual participants. On the other hand, SEM-based IFA is appropriate if the purpose is to explain the relationship between constructs (e.g., number of factors or cross-loadings; Wirth and Edwards, 2007, p. 70). As this current study aims to not only understand the relationship between constructs, but also obtain scores for individual participants, both methods (SEM & IRT[14] – based IFA) were conducted. A comparison of the parameter estimates from the two methods, confirming consistency between results also supported more confidence in the estimates.

A multidimensional model that allows correlations among factors, where one item corresponds to one factor, was conducted for each IFA. In each IFA, the variation due to one of the organizers (e.g., task of teaching) was controlled when examining the dimensionality associated with the other organizer (instructional situation). For example, all the items measuring teachers' mathematical knowledge for doing one task of teaching (e.g., understanding students' work) were used in an IFA model to examine whether the dimensions of teachers' mathematical knowledge could be distinguished by different instructional situations. Similarly, all the items measuring teachers' mathematical knowledge in one instructional situation (e.g., doing proofs) were used in another IFA to examine whether the dimensions of teachers' mathematical knowledge could be distinguished by different tasks of teaching. These IFAs conjectured by different tasks of teaching and instructional situations were conducted both within and across the courses of studies (within geometry or algebra items).

The analysis using the DCM approach followed the same hypothesis on the relationships between item responses and knowledge dimensions, which were tested in IFA. In other words, the DCM model tested a simple structure where each item measures

[14] Specifically, the model is the Graded Responses Model, an extension of the 2PL model for ordinal response data.

only one hypothesized attribute (Rupp et al., 2010, p.328). I used a predefined Q-matrix, that is, a table specifying which attributes are measured by which item in terms of numeric value (1s for an attribute measured by an item and 0s for an attribute not measured by an item; Rupp et al., 2010, p.54). In the analysis, in order to use a DCM model which models dichotomous responses, I dichotomized ordinal item responses by using the mean value of each item as a cut point, following the approach of Templin and Henson (2006). For example, if a mean of a testlet item is 2.5 (sum of 4 True/False items), scores greater than or equal to 2.5 were coded as 1 and scores lower than 2.5 were scored as 0.

4.3.5.2 Relationship with teachers' educational and teaching background

Dimensionality analysis with only item scores is a variable-centered approach in that it attempts to find a dimensionality of knowledge based on the covariance among item scores. As this approach does not account for different person characteristics such as subject matter preparation or experience teaching, it does not capture different patterns in teachers' item scores associated with teachers' educational background or teaching experience. In this regard, I further investigated dimensionality of teachers' mathematical knowledge by examining the differences in the relationships with their educational and teaching experience across identified dimensions. Among other studies on teacher knowledge, the COACTIV group had found differences between CK and PCK in terms of teachers' effect on instruction and on student learning. (CK was not found to affect student learning, whereas PCK significantly affected student learning; Baumert et al., 2010). This result supported their conjecture about PCK's being empirically distinguishable from CK. Meaningful differences in teachers' performance associated with teachers' background variables would inform claims about the structure of teachers' mathematical knowledge. For example, the examination of the correlation between teachers' background variables such as years of teaching experience and the number of coursework that had taken and their test scores was expected to reveal the difference in teachers' performance depending on a set of items.

4.4 Chapter summary

In this chapter, I describe the characteristics of analyzed data, including the types of instruments (MKT-Algebra, MKT-Geometry, SMK-Geometry), item formats (multiple-choice and True-False items), demographic information of participants (602 U.S. high school mathematics teachers). The item contents were described in terms of the characteristics of the item clusters created by the hypothesized item blueprint. The item blueprint was organized by associated tasks of teaching and instructional situations.
The empirical procedures such as scaling responses and data analysis methods were also introduced in this chapter. By using the three different measurement models – structural equation modeling, item response theory, and diagnostic classification model, my study aims to support consistency of the results and provides insights into the functions of each model.

Chapter 5

Results: Validation of the Dimensions Proposed in the Framework

This chapter presents the results of the seven dimensionality analyses conducted to validate the methodological feasibility of the proposed item blueprint. In other words, this chapter examines whether the proposed item blueprint allows establishing distinguishable measures that reflect the hypothesized distinguishable knowledge dimensions. First, I report the results comparing different item factor models in which each dimension is characterized by different instructional situations within the same task of teaching (5.1.1~5.1.2). Second, I report the results comparing different item factor models in which each dimension is characterized by different tasks of teaching within the same instructional situation (5.2.1~5.2.3). Third, I report the models in which each dimension is characterized by different tasks of teaching and different instructional situations within the same course of study (5.3.1~5.3.2). Fourth, I report the differences among hypothesized dimensions in terms of the proportions of knowledge profiles (5.4~5.5). Fifth, I report the results examining the effects of teachers' educational and teaching experience on each of the hypothesized dimensions (5.6). Lastly, I summarize the results validating the suggested dimensions of the framework.

For each dimensional analysis, three different modeling approaches were applied: 1) Item Factor Analysis under SEM framework; 2) Item Factor Analysis under IRT framework; 3) Diagnostic Classification Modeling.

Table 5.1 present the organization of the analyses including a hypothesis of each analysis and the summary of it. The detailed analysis results comparing different factor models are described in the following sections.

Table 5.1 Summary of findings

Section	Hypothesis	Result
5.1.1	Hypothesized knowledge factors associated with the five different instructional situations (CG, DP, EF, SR, SE) **within the task USW** are distinguishable each other	The hypothesized factors are significantly distinguishable, except USW_CG and USW_DP.
5.1.2	Hypothesized knowledge factors associated with the four different instructional situations (CG, DP, EF, CN) **within the task CGP** are distinguishable each other	The hypothesized factors are significantly distinguishable, except CGP_DP and CGP_EF.
5.2.1	Hypothesized knowledge factors associated with the two different tasks of teaching (USW, CGP) **within the instructional situation CG** are distinguishable each other	The hypothesized factors USW_CG and CGP_CG are significantly distinguishable.
5.2.2	Hypothesized knowledge factors associated with the two different tasks of teaching (USW, CGP) **within the instructional situation DP** are distinguishable each other	The hypothesized factors USW_DP and CGP_DP are significantly distinguishable.
5.2.3	Hypothesized knowledge factors associated with the two different tasks of teaching (USW, CGP) **within the instructional situation EF** are distinguishable each other	The hypothesized factors USW_EF and CGP_EF are not significantly distinguishable.
5.3.1	Hypothesized knowledge factors associated with the two different tasks of teaching (USW, CGP) across the two different instructional situations **in geometry** (CG, DP) are distinguishable each other	The four hypothesized factors USW_CG, USW_DP, CGP_CG, and CGP_DP are significantly distinguishable.
5.3.2	Hypothesized knowledge factors associated with the two different tasks of teaching (USW, CGP) across the three different instructional situations **in algebra 1** (SR, SE, CN) are distinguishable each other	The three hypothesized factors USW_SR, USW_SE, and CGP_CN are significantly distinguishable.

5.1 Dimensionality within the same task of teaching

To examine whether multiple dimensions can be identified by the organizer instructional situation, the dimensionality of the item responses was investigated within the same task of teaching. The constraint on task of teaching was set to control any effect from different tasks of teaching on the dimensionality.

The distinction among hypothesized dimensions was examined through Chi-Square Difference Tests. Specifically, the fit of a higher dimensional model (where all the factors are distinguished as hypothesized) was compared to a lower dimensional model (where two factors are combined). The analyses were conducted within the task of understanding students' work (USW) (4.3.3.1) and within the task of choosing the givens for a problem (CGP) (4.3.3.2), respectively.

5.1.1 Organization of the items within the task USW

The list of situations within the task of understanding students' work (USW) was sorted into five separate situational categories, as shown in Table 5.2. As described in 4.3.3.1, the task of USW includes a teacher's reading, making sense, representing students' thinking in the teacher's mind, deciding whether the student's answer or inferred process is mathematically correct, identifying specific errors, and generating mathematical interpretation of that work through inscriptions or writing. In this section, I report the results that can answer the question posed in 3.3.2: if teachers are knowledgeable in a task of teaching for one situation, how likely is it that they would be knowledgeable in that same task of teaching for another situation? In other words, the analyses examine whether the items measuring teachers' knowledge in doing USW can be distinguished by different instructional situations.

The analyses were conducted under two different frameworks 1) a framework in which the knowledge being measured is considered to be continuous; 2) a framework in which the knowledge being measured is considered to be discrete: whether a teacher had achieved a mastery level of knowledge (1: master) or not (0: nonmaster). The framework of continuous latent construct (hereafter, IFA models) is further categorized into two different approaches: SEM-based modeling (use limited information) and IRT-based modeling (use full information) (discussed in Chapter 4).

Table 5.2 Items in a task of USW

Task of teaching	**Instructional situation**	**Knowledge dimension/Attribute**	**# items**
Understanding Students' Work (USW)	Calculation in Geometry (CG)	USW_CG	6
	Doing Proofs in geometry (DP)	USW_DP	3
	Exploring a Figure in geometry (EF)	USW_EF	3
	Simplifying Rational expression in algebra (SR)	USW_SR	3
	Solving a linear or quadratic Equation in algebra (SE)	USW_SE	5

5.1.1.1 Dimensionality analysis of the items within USW using IFA

To examine whether the items commonly measuring teachers' mathematical knowledge for the task of understanding students' work (USW), can be distinguished in terms of the instructional situation, confirmatory item factor analyses were conducted under two different approaches (SEM-based and IRT-based framework) using Mplus version 7.4 (Muthén & Muthén, 1998 – 2015).

5.1.1.1.1 SEM-based modeling (using WLSMV[15] estimator)

The factor structure among the items was first evaluated by a confirmatory item factor analysis using a WLSMV estimator. In all item factor analyses conducted in this study, latent factor variances and means were set to 1 and 0, respectively, and all item factor loadings were freely estimated.

The initial 5-factor model (Figure 5.1) hypothesized based on the five different situations yielded an acceptable fit (RMSEA[16]=0.000; CFI[17]=1.000; TLI[18]=1.000) by every criterion suggested by Hu & Bentler (1999; RMSEA <0.06, TLI > 0.95, CFI > 0.95). However, the factors of Calculation in Geometry (hereafter, CG, 4.3.4.1) and Doing Proofs (hereafter, DP, 4.3.4.2) were highly correlated as 0.95, implying that they may not be distinguishable. Consequently, I compared the 5-factor model with the 4-factor model where DP and CG items formed one common factor (Figure 5.2). To examine the significance of a difference in a model fit between two models, I conducted the differential test called DIFFTEST[19], which adjusts the difference in chi-square values for the WLSMV estimator (Muthén & Muthén, 1998-2015). The DIFFTEST result suggested that the 4-factor model (Figure 5.2) was not significantly worse than the 5-factor ($\chi^2 = 3.296$, $df = 4$, $p = 0.5096$) (Figure 5.1), implying that DP and CG may not be statistically distinguishable. I further compared the 4-factor model where DP and CG are combined with the 3-factor model (Figure 5.3) where all three geometry situation factors (EF, DP, and CG) are combined into one factor. The differential test revealed that the 3-

[15] Diagonally weighted least squares estimator
[16] Root Mean Square Error of Approximation
[17] Comparative Fit Index
[18] Tucker-Lewis Index
[19] As the chi-square value for WLMSV cannot be used for chi-square difference testing in the regular way, the DIFFTEST option of Mplus was used to compare two models (Muthén & Muthén, 1998-2015).

factor model is significantly worse than the 4-factor model ($\chi^2 = 9.115$, $df = 3$, $p = 0.0278$). In other words, the 4-factor model (Figure 5.2) was significantly better than the 3-factor model, but not significantly worse than the 5-factor model.

Table 5.3 presents estimated factor loadings estimated by the 3-factor, 4-factor, and 5-factor models under SEM-based models (column 2 ~ 4) and IRT-based models (column 5 ~ 7). As shown in the first three columns of the table, geometry item loadings estimated from the 5-factor model tend to go up slightly relative to the loadings in the 3-factor model or 4-factor model, where some of the geometry factors are combined.

Algebra item loadings stayed the same between models, as their structure was not modified across models. This result implies that most geometry items can explain more variance when they are loaded on three different dimensions than when they combine together in a single factor. In other words, even though the hypothesized factors are highly correlated, they are better to be distinguished. For example, CG items yielded slightly higher loadings in the 5-factor model where CG items are distinguished from DP items, even though CG and DP are not significantly distinguishable.

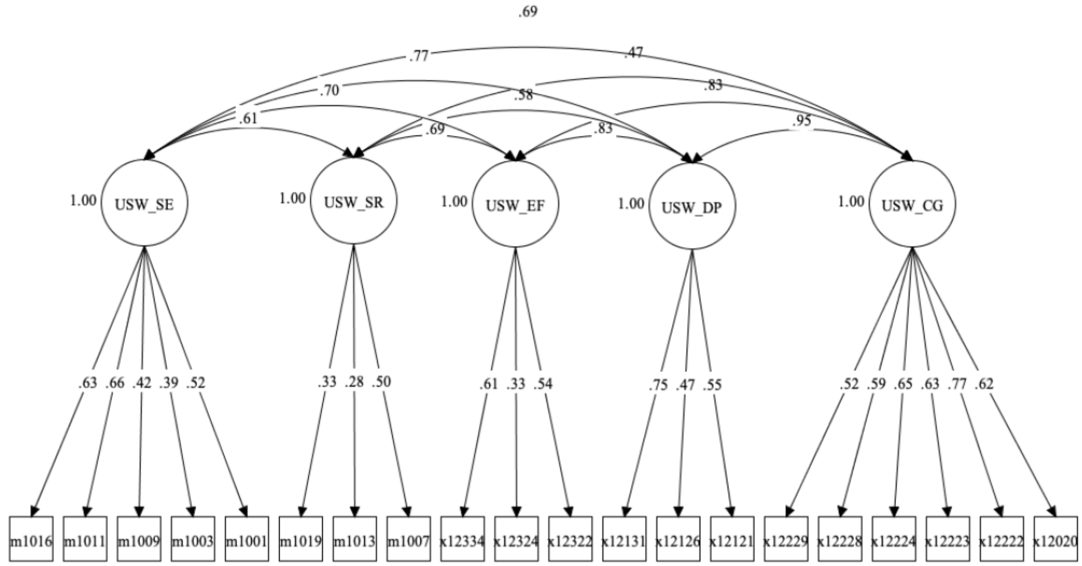


Figure 5.1. Five-dimensional model within the task USW[20]

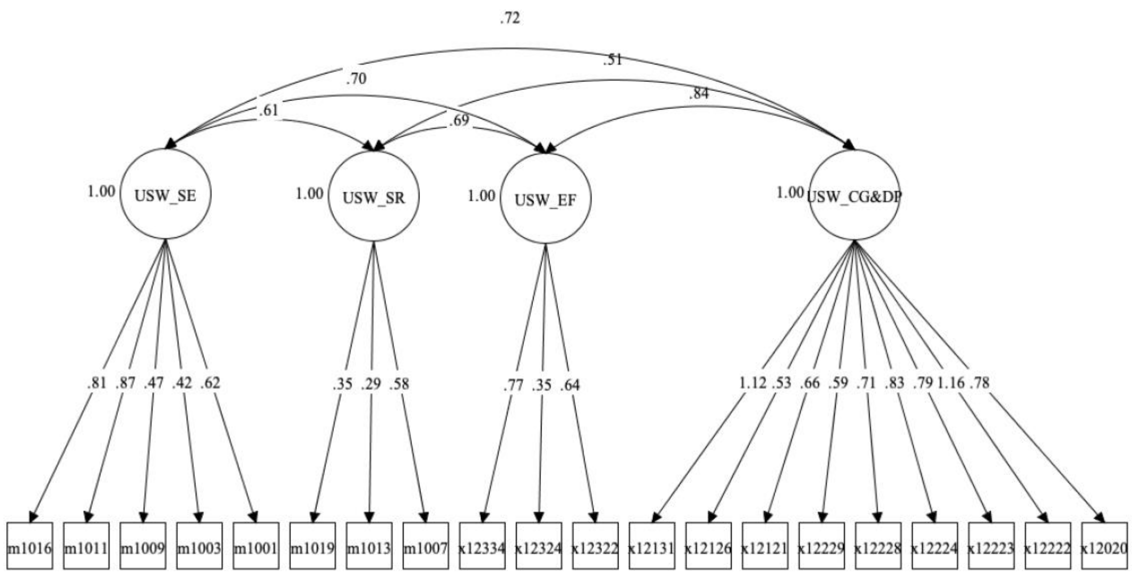


Figure 5.2. Four-dimensional model within the task USW

[20] Standardized factor loadings are presented in the diagrams.

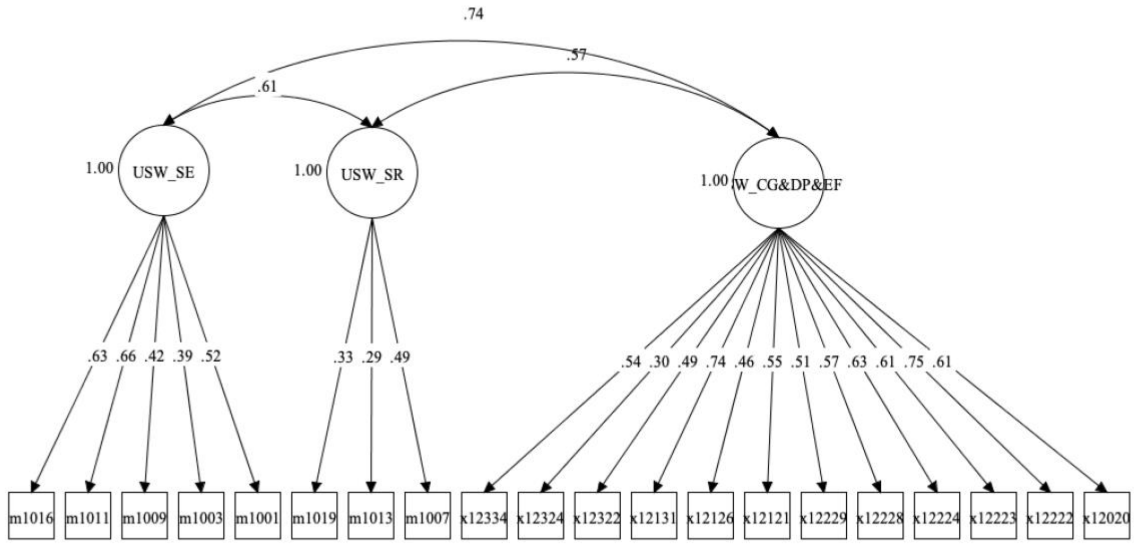


Figure 5.3. Three-dimensional model within the task USW

Table 5.3. Estimated standardized factor loadings within USW

		SEM-based			IRT-based		
Item	Dimension	3 CG+DP +EF	4 CG+DP	5 all separated	3 CG+DP+ EF	4 CG+DP	5 all separated
X12020	USW_CG	0.61 (0.05)	0.62 (0.05)	0.62 (0.05)	0.56 (0.05)	0.56 (0.05)	0.56 (0.05)
X12222		0.75 (0.05)	0.76 (0.05)	0.77 (0.05)	0.73 (0.05)	0.74 (0.05)	0.75 (0.05)
X12223		0.62 (0.05)	0.62 (0.05)	0.63 (0.05)	0.59 (0.05)	0.59 (0.05)	0.60 (0.05)
X12224		0.63 (0.04)	0.64 (0.04)	0.65 (0.04)	0.61 (0.04)	0.62 (0.04)	0.63 (0.04)
X12228		0.57 (0.05)	0.58 (0.05)	0.59 (0.05)	0.55 (0.05)	0.56 (0.05)	0.57 (0.05)
X12229		0.51 (0.06)	0.51 (0.06)	0.52 (0.06)	0.47 (0.06)	0.47 (0.06)	0.47 (0.06)
X12121	USW_DP	0.55 (0.05)	0.55 (0.05)	0.55 (0.06)	0.48 (0.06)	0.49 (0.06)	0.50 (0.06)
X12126		0.46 (0.06)	0.47 (0.06)	0.47 (0.06)	0.41 (0.06)	0.41 (0.06)	0.42 (0.06)
X12131		0.74 (0.05)	0.75 (0.05)	0.75 (0.06)	0.69 (0.05)	0.69 (0.05)	0.70 (0.07)
X12322	USW_EF	0.49 (0.06)	0.54 (0.06)	0.54 (0.06)	0.44 (0.06)	0.48 (0.06)	0.48 (0.06)
X12324		0.30 (0.06)	0.33 (0.06)	0.33 (0.06)	0.26 (0.05)	0.28 (0.06)	0.28 (0.06)
X12334		0.54 (0.05)	0.61 (0.06)	0.61 (0.06)	0.51 (0.05)	0.59 (0.06)	0.59 (0.06)
M1007	USW_SR	0.49 (0.12)	0.50 (0.12)	0.50 (0.12)	0.44 (0.11)	0.49 (0.12)	0.49 (0.12)
M1013		0.29 (0.10)	0.28 (0.10)	0.28 (0.10)	0.24 (0.09)	0.18 (0.09)	0.18 (0.09)
M1019		0.33 (0.10)	0.34 (0.10)	0.33 (0.09)	0.32 (0.10)	0.33 (0.09)	0.32 (0.09)
M1001	USW_SE	0.52 (0.08)	0.52 (0.08)	0.52 (0.08)	0.47 (0.07)	0.47 (0.07)	0.47 (0.07)
M1003		0.39 (0.12)	0.39 (0.12)	0.39 (0.12)	0.37 (0.11)	0.36 (0.11)	0.38 (0.11)
M1009		0.42 (0.07)	0.42 (0.07)	0.42 (0.07)	0.41 (0.07)	0.41 (0.07)	0.41 (0.07)
M1011		0.66 (0.07)	0.66 (0.07)	0.66 (0.07)	0.61 (0.07)	0.61 (0.07)	0.62 (0.07)
M1016		0.63 (0.07)	0.63 (0.07)	0.63 (0.07)	0.60 (0.07)	0.60 (0.07)	0.59 (0.07)

All item loadings are significant at the $p < 0.05$ level

Standardized loading (standard error)

5.1.1.1.2 MIRT-based modeling

IRT-based dimensional analysis was also conducted to assess whether the items commonly measuring teachers' mathematical knowledge for the task of understanding students' work can be distinguished in terms of the instructional situation. The difference[21] between this analysis and the previous analysis is that MIRT-based modeling, (specifically, Multidimensional Graded Responses Model, an extension of the 2PL model for ordinal response data; Samejima, 1997) examines the structure of items using the MML estimator[22] which uses patterns in responses (Schilling & Bock, 2005; Wirth & Edwards, 2007) instead of using WLSMV estimator which uses correlations among item responses.

The estimated correlations among the five knowledge dimensions are presented in Table 5.4. Factors, except for a pair of Calculation in Geometry (CG) and Doing Proof (DP), have correlations significantly different from 1, suggesting that the factors reflecting teachers' mathematical knowledge for teaching geometry or algebra 1 in doing the task of understanding students' work (USW) are statistically distinguishable by different instructional situations.

As shown in Table 5.4, geometry factors (Calculation in Geometry: CG, Doing Proofs: DP, and Exploring a Figure: EF) are more strongly correlated with each other than with algebra factors (Simplifying Rational expressions: SR and Solving Equations:

---

[21] Two different approaches (IRT-based models and SEM-based models) are sometimes termed as IRT models or Factor analytic models (Kamata & Bauer, 2008) instead of using the terms IRT-based and SEM-based frameworks under the broad category of Item factor analysis (IFA). This study classifies IRT-based and SEM-based framework under IFA, following the taxonomy used Wirth & Edwards (2007).

[22] Estimator MML (marginal maximum likelihood) was used with an adaptive integration (Gauss-Hermite integration with a five integration points).

SE). For example, the correlation between USW_CG (understanding students' work in calculation in geometry) and USW_DP (understanding students' work in doing proofs) was as high as 0.94, whereas the correlation between USW_CG (understanding students' work in calculation in geometry) and USW_SR (understanding students' work in simplifying rational expressions) was as low as 0.39.

Table 5.4. Estimated correlations among dimensions within USW

	Dimension				
Dimension	USW_CG	USW_DP	USW_EF	USW_SR	USW_SE
USW_CG					
USW_DP	0.94				
USW_EF	0.83	0.85			
USW_SR	0.39	0.53	0.72		
USW_SE	0.68	0.75	0.72	0.62	

Similar to the case of SEM-based modeling, the high correlation between USW_CG and USW_DP led me to examine any advantage of distinguishing two factors over combining them. To compare the nested IRT-based models, I calculated the difference in the log-likelihoods twice (deviance) with the difference in the number of free parameters.

The result of likelihood ratio test was consistent with that found with the SEM-based modeling described above (which used the WLSMV estimator) in that the 4-dimensional MIRT model fit the data significantly better than the 3-dimensional IRT model, (χ^2=10.56, *df=3, p* =0.014) (Table 5.5), but not significantly worse than the 5-dimensional IRT model (χ^2=2.97, *df=4, p* =0.564). Moreover, all the item loadings of the EF (Exploring a Figure) factor (i.e., standardized item discriminations in the MIRT framework) on the 4-factor model went up slightly relative to the loadings in the 3-factor model (x12322: 0.44 → 0.48; 12324: 0.26→0.28; x12334: 0.51→0.59) (Table 5.3), indicating that these EF items can explain more variance when they are distinguished from the factor of USW_CG&DP. AIC information statistic also suggested that a 4-dimensional model is better than a 3-dimensional model where all the geometry items are loaded on a unidimensional factor (note that lower information value indicates better fit) (Table 5.5).

Table 5.5. Comparison of fit among MIRTs within USW

Model	Deviance (-2log likelihood) statistics	Number of free parameters	Change in Deviance	Change in Degrees of Freedom	*p*	*AIC*
5-Dimension	13096.14	60				13216.143
4-Dimension (CG+DP)	**13099.11**	**56**	**2.97**	**4**	**0.56**	**13211.108**
3-Dimension (CG+DP+MC)	13109.66	53	10.56	3	0.01	13215.660

In summary, tested measurement models supported my hypothesis that the items within the same task of understanding students' work (USW) could be distinguished by different instructional situations, though the distinction between calculation in geometry (CG) and doing proofs (DP) requires preferring a less parsimonious solution for the sake of higher loadings. Moreover, the results were consistent under both SEM and IRT based framework. In the next section, the same hypothesis on the structure of USW items is investigated under a DCM framework.

5.1.1.2 Dimensionality analysis of the items of USW using DCM

In addition to the IFA models in which the latent construct is considered to be continuous, a DCM model was applied to the same data set with the same hypothesis on the structure of items. Specifically, a loglinear cognitive diagnosis model (LCDM), a general DCM model allowing item parameterization to vary at the item level, was retrofitted to the items designed to measure five distinguishable dimensions of teachers' knowledge in doing the task of understanding students' work (USW_CG, USW_DP, USW_EF, USW_SR, USW_SE). Following the correspondence between items and targeted constructs, the LCDM model was set to estimate the main effects of items on each of the five knowledge attributes distinguished by instructional situations within the task of USW. The Q-matrix specifying the item-to-knowledge attribute alignment is presented in Table 5.6.

Table 5.6. Q-matrix for items within USW

Item	Attribute 1: USW_CG	Attribute 2: USW_DP	Attribute 3: USW_EF	Attribute 4: USW_SR	Attribute 5: USW_SE
x12020, x12222, x12223, x12224_b, x12228_b, x12229,	1	0	0	0	0
x12121, x12126, x12131	0	1	0	0	0
x12322, x12324_b, x12334_b	0	0	1	0	0
m1007, m1013, m1019	0	0	0	1	0
m1001, m1003_b, m1009, m1011_b, m1016	0	0	0	0	1

*items with _b are dichotomized (originally ordinal items) responses

As shown in the Q-matrix, each item measures only one attribute. This type of Q-matrix is referred to as a simple structure (Jurich & Bradshaw, 2013). The example formula below expresses the LCDM proposed for estimating teachers' knowledge attribute USW_CG measured by an item *i*, which is designed to measure only USW_CG within the task of USW.

$$ln(\frac{P(X_i = 1|\alpha_{rUSW\_CG})}{P(X_i = 0|\alpha_{rUSW\_CG})}) = \lambda_{i,0} + \lambda_{i,1,(USW\_CG)}\alpha_{rUSW\_CG} \quad (1)$$

In equation (1), (adapting notations from Rupp et al., 2010), X_i indicates a dichotomous item response to item *i* by the respondent group (latent class) who have attained the mastery status α_{rUSW_CG}, which is 1 if the knowledge USW_CG is mastered by the group, and 0 otherwise. Accordingly, the dependent variable of $ln(\frac{P(X_i =1|\alpha_{rUSW\_CG})}{P(X_i=0|\alpha_{rUSW\_CG})})$ is the logit of a correct answer to item *i* by the respondents. Regarding the use of notations, the second subscript (0 or 1) of λ indicates whether the item parameter corresponds to the intercept (0) or main effect (1) for item *i*, and the third subscript, which is within parentheses, indicates the attribute to which the main effect refers. Thus, the parameter $\lambda_{i,0}$ – the intercept – indicates a logit for non-masters of the attribute measured by item *i* and the parameter $\lambda_{i,1,(attribute)}$ – the main effect – indicates the increase in the logit for mastering the attribute measured by item *i*. For another example, the equation could be expressed as below for the item x12121 which measures USW_DP.

$$ln(\frac{P(X_{X12121} =1|\alpha_{rUSW\_DP})}{P(X_{X12121}=0|\alpha_{rUSW\_DP})}) = \lambda_{X12121,0} + \lambda_{X12121,1,(USW\_DP)}\alpha_{rUSW\_DP}$$

Each mastery indicator (α_{rUSW_CG} or α_{rUSW_DP}) is an element of a vector that represents a mastery profile α_r. In this example, where the LCDM model estimates teachers' five knowledge attributes reflecting the five different instructional situations within the task of USW, the mastery profile of a respondent can be expressed as α_r=[$\alpha_{rUSW\_CG}, \alpha_{rUSW\_DP}, \alpha_{rUSW\_EF}, \alpha_{rUSW\_SR}, \alpha_{rUSW\_SE}$] . This indicates which attribute the respondent has mastered ($\alpha_{rattribute}$= 1 if the attribute is mastered, and $\alpha_{rattribute}$=

0 otherwise). For example, the mastery profile of the respondent who has mastered USW_CG and USW_DP, but not for other attributes can be expressed as α_r= [α_{rUSW_CG}, α_{rUSW_DP}, $\alpha_{rUSW\_EF=}$, α_{rUSW_SR}, α_{rUSW_SE}]= [1,1,0,0,0].

Given that one item measures only one knowledge attribute within the same task of teaching, the probability of a correct response for item *i* is estimated by two parameters – an intercept $\lambda_{i,0}$ and a main effect $\lambda_{i,1,(attribute)}$. Using these estimated item parameters (the intercept and the main effect), the proportions of teachers for each of the mastery profiles could be calculated, which are reported in a later section.

5.1.1.2.1 Model fit

Even though the Q-matrix used in estimating the LCDM model was based on a measurement model that had shown good model fit statistics in IFA models, the LCDM model fit also needed to be evaluated before interpreting the estimated item parameters. In regard to model fit indices, there has been no specific reliable test that can evaluate the global model fit of DCMs (Rupp et al., 2010). Alternatively, limited-information methods using item-pair associations have been used in several studies to produce indices of model fit of a DCM. For example, one study using DCM in student assessment (Jurich & Bradshaw, 2013) used a bivariate goodness of fit statistic to evaluate a DCM model's ability in reproducing the data. Following their approach, this present study calculated a bivariate goodness of fit statistic with a $\chi^2(1)$ distribution for each pair of items to examine the pairs of items that demonstrate poor fit. These bivariate fit indices – an index for a pair of items – were calculated by Mplus.

The result suggested that 95.8% (182 pairs) of the 190 (=$\frac{20*19}{2}$) item pairs showed good model fit (i.e., chi-square value is insignificant) when chi-square values were evaluated at a 0.05 significance value and only one item pair showed misfit at a 0.01 significance level. Overall, my LCDM model estimating teachers' knowledge in understanding students' work across five knowledge attributes provided an acceptable model fit for the data used in this study. Therefore, the item parameters estimated from the model were further interpreted as follows.

5.1.1.2.2 Item parameter estimates

Again, as the model estimates a simple structure where one item measures only one knowledge attribute, only main effects ($\lambda_{i,1,(a)}$) and intercept ($\lambda_{i,0}$) parameters were estimated. The estimated item parameters are listed in Table 5.7.

As shown in the last row of the second column in the table, the average intercept across items is -0.72, indicating the average predicted logit of a correct response for teachers who had not mastered the targeted attributes. This means that approximately 0.33% (= $\frac{e^{-0.72}}{(1+e^{-0.72})}$) of teachers, who had not mastered an attribute measured by an item of understanding students' work (USW), answered the items correctly. Next, the average main effect ranged from 1.47 to 2.36. This means that the increase in the logit of a correct response by mastering an attribute ranged from 1.47 to 2.36. At an item level, for example, nonmasters of the understanding students' work in solving equations (USW_SE) measured by item M1016 have a log-odds of correct response -2.29, and masters of the USW_SE measured by item M1016 have a log-odds 0.61, which is equal to the sum of the intercept value –2.29 and the main effect 2.90.

Table 5.7. Estimated item parameters (logit) within USW

Item	Intercept $\lambda_{i,0}$	USW_CG $\lambda_{i,1,(1)}$	USW_DP $\lambda_{i,1,(2)}$	USW_EF $\lambda_{i,1,(3)}$	USW_SR $\lambda_{i,1,(4)}$	USW_SE $\lambda_{i,1,(5)}$
x12020	-1.04 (0.29)	1.98 (0.31)				
x12222	-0.09 (0.22)	3.07 (0.50)				
x12223	-0.60 (0.21)	2.12 (0.29)				
x12229	-0.12 (0.19)	1.69 (0.28)				
x12224_b	-0.71 (0.25)	2.66 (0.34)				
x12228_b	-1.64 (0.33)	2.65 (0.34)				
x12121	-1.58 (0.47)		2.24 (0.48)			
x12126	-1.45 (0.26)		1.43 (0.33)			
x12131	-0.64 (0.38)		2.98 (0.45)			
x12322	-0.69 (0.31)			1.76 (0.3)		
x12324_b	-1.06 (0.19)			1.15 (0.33)		
x12334_b	-0.78 (0.45)			3.67 (1.16)		
M1007	-0.81 (0.46)				2.03 (0.50)	
M1013	-1.07 (0.28)				0.90 (0.37)	
M1019	-1.72 (0.46)				1.48 (0.48)	
M1001	0.29 (0.24)					1.83 (0.46)
M1009	-0.39 (0.33)					1.64 (0.39)
M1016	-2.29 (0.71)					2.90 (0.69)
M1003_b	1.94 (0.34)					1.17 (0.50)
M1011_b	-0.01 (0.28)					2.26 (0.38)
Average	**-0.72 (0.33)**	**2.36 (0.34)**	**2.22 (0.42)**	**2.19 (0.60)**	**1.47 (0.45)**	**1.96 (0.48)**

Standard errors in parentheses

The main effect size of each item on each knowledge attribute was evaluated using the conventional metrics for an odds ratio suggested by Bradshaw et al. (2014, p.6) with small effect sizes for odd ratios between 1.44 and 2.47, medium effect sizes for odd ratios between 2.47 and 4.25, and large effect sizes for odd ratios larger than 4.25. According to this convention, one item had small effect sizes (M1013: 2.46 $=e^{0.90}$), three items had medium effect sizes (x12126, x12324_b, M1003_b), and the remaining 16 items had large effect sizes. This result suggests that all the items measured the intended attributes with significant effect sizes.

5.1.1.2.3 Probability of a correct response between nonmasters and masters

Following the same approach of Bradshaw et al. (2014), the strength of associations between measured attributes and each item were examined by using item characteristics bar charts (ICBCs). Considering that the ICBCs present the probability of a correct response to an item for each of the latent classifications (masters and nonmasters), it can be considered as a DCM version of item characteristics curve in IRT framework.

The ICBCs for each knowledge attributes are displayed in Figures 5.4 ~ 5.8. By examining the difference in the probability of a correct response between nonmasters and masters, the ability of each item in distinguishing masters from nonmasters was evaluated. For example, as shown in Figure 5.8, a difference in the probability of a correct response on the item M1003_b between nonmasters and masters is very small, in that masters of USW_SE answered the item M1003_b correctly with a 0.96 probability and even nonmasters of USW_SE answered the same item correctly with a probability of 0.87. Based on this small increase in probability of a correct response from nonmasters to masters, the item M1003_b may provide little practical information in measuring teachers' knowledge of USW_SE. This item might be too easy to distinguish teachers' mastery level, given that 87% of teachers could get a correct answer even though they had not mastered the knowledge attribute. Consequently, whether to include this item for estimation was evaluated by using a model comparison test, which is described in the next section "Problematic items".

Even though some items were weak in discriminating mastery level, overall, significant differences could be identified in the probability of getting a correct response between nonmasters and masters across items, with respect to the hypothesized constructs. The hypothesized constructs were categorized by instructional situations in the task of understanding students' work.

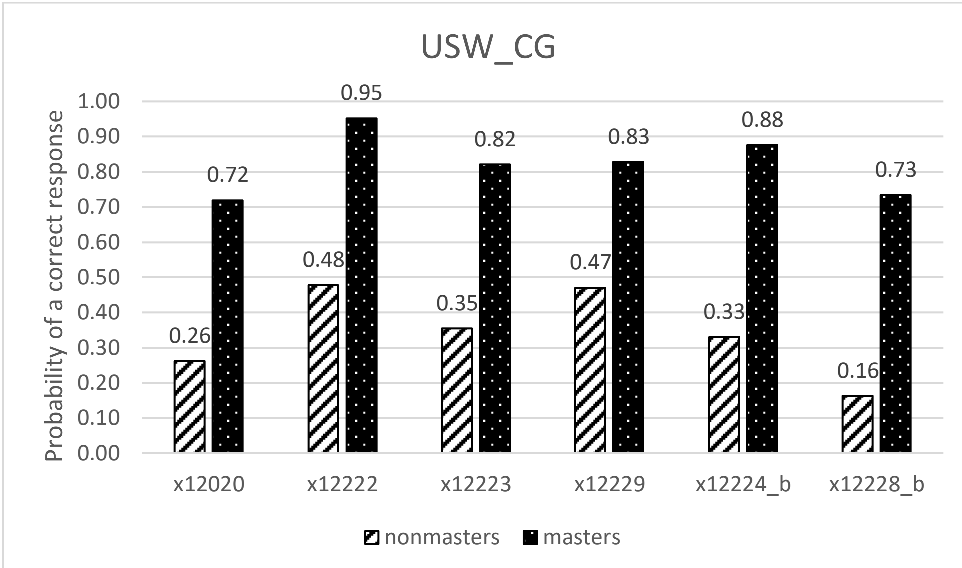


Figure 5.4. Item characteristic bar chart for USW_CG items

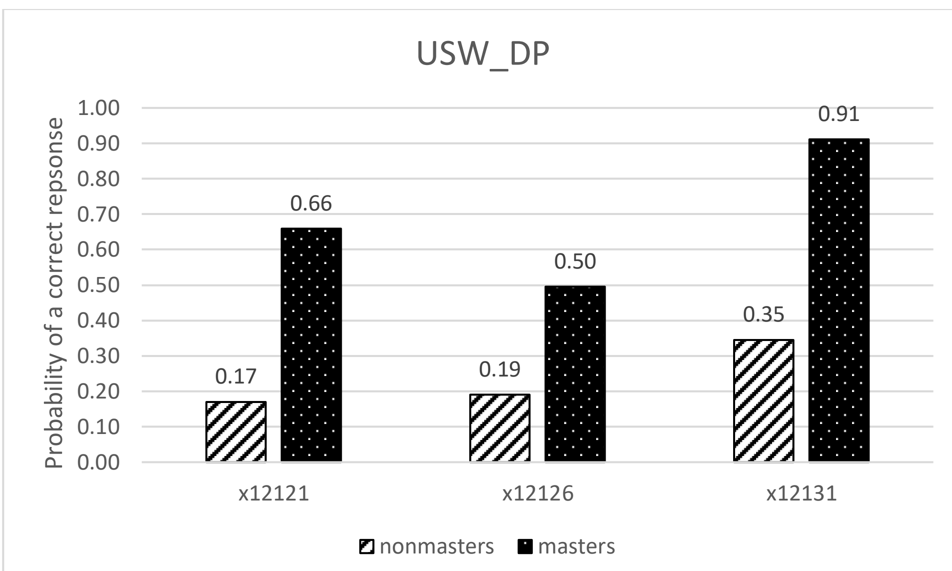


Figure 5.5. Item characteristic bar chart for USW_DP items

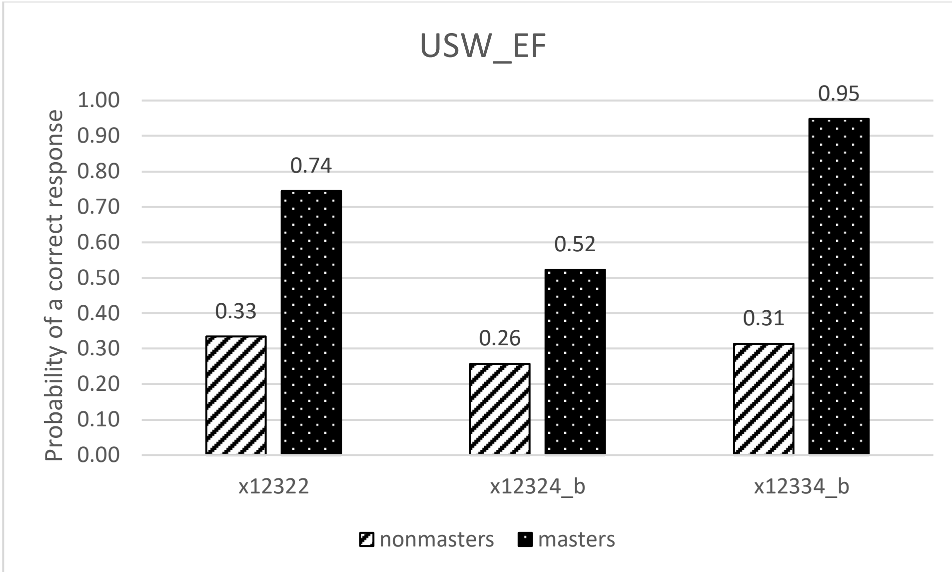


Figure 5.6. Item characteristic bar chart for USW_EF items

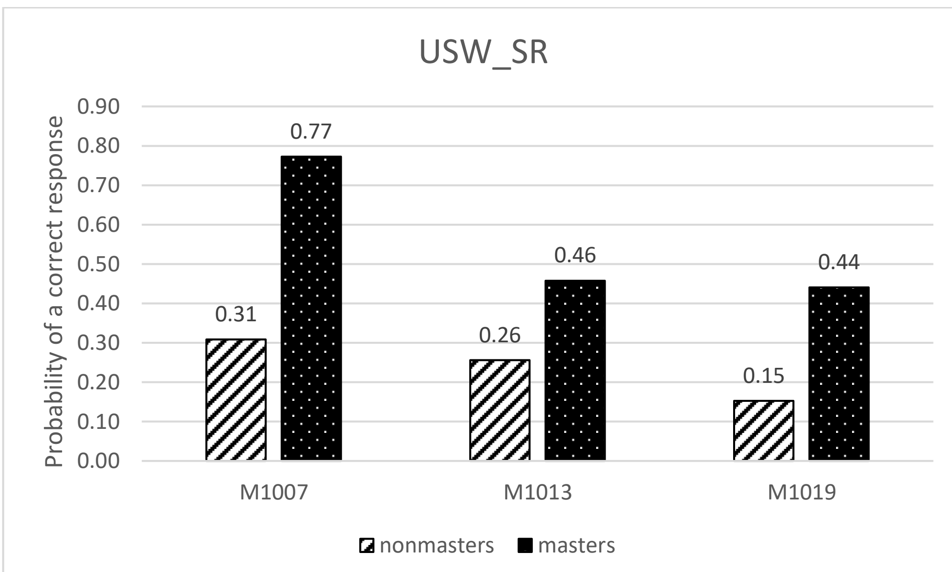


Figure 5.7. Item characteristic bar chart for USW_SR items

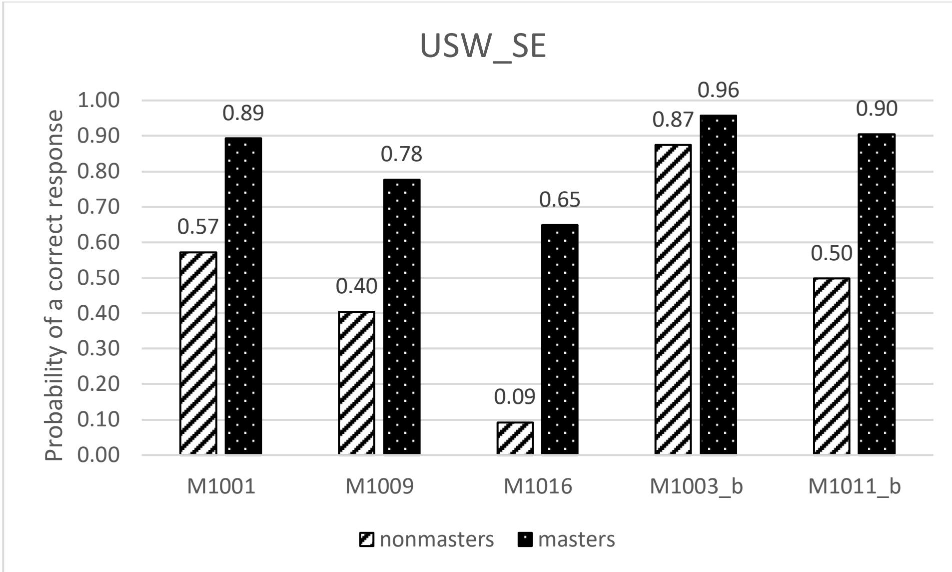


Figure 5.8. Item characteristic bar chart for USW_SE items

5.1.1.2.4 Problematic items

As I mentioned in the previous section, item M1003_b showed the smallest main effect size on USW_SE, and the difference in the probability of a correct response between nonmasters and masters was as small as 0.09. This indicates that its significance (i.e., significantly different from zero) presented in the output may not be substantial, even though the main effect size of M1003_b seems significant. This is because the main effect sizes of all the items in the LCDM model were constrained to be greater than zero in the estimation, so the probability of incorrectly rejecting the null hypothesis (p-value) would never indicate insignificance (Bradshaw et al., 2014). Therefore, instead of using p-value, the likelihood test for nested model comparison was used to examine the significance of the main effect of M1003_b. For the test, the model including the main effect of M1003_b was compared to the model removing the main effect of M1003_b. The comparison test between these models suggested that the model without the main effect of M1003_b is significantly worse than the model with the main effect ($\chi^2(1)$ =

6.89, p <0.009), indicating that the main effect was significant. Therefore, the main effect of M1003_b was not removed in the subsequent analysis.

5.1.1.2.5 Model comparison

Similar to the test conducted under the MIRT framework, the 5-attribute LCDM model was compared to the 4-attribute LCDM model where the factors CG and DP combine together. The result was consistent with that of IFA in that the 4-attribute model was not significantly worse than the 5-attribute model ($\chi^2(16) = 7.70$, p=0.96), but significantly better than the 3-attribute model (($\chi^2(8) = 27.69$, p<0.001). The other fit statistics such as AIC also suggested better fit of the 4-attribute model (9377.053) than the 5-attribute model (9401.349) or 3-attribute model (9387.74).

5.1.1.3 Reliability

The degree of consistency in item scores (i.e., reliability) was evaluated for each hypothesized knowledge dimension under two different frameworks: IRT and DCM. Under the IRT framework, I examined a range of knowledge levels where the items can provide information with acceptable precision using test information functions (TIF) for the USW_CG, USW_DP, USW_EF, USW_SR, and USW_SE, respectively. In contrast to classical test theory in which reliability of the test scores are estimated by a single value, the reliability of the score estimated in IRT is evaluated by a function that varies across ability levels. This function is called test information function, which is the sum of the information functions from the individual items (Bandalos, 2018, p.429).

The test information functions for estimated scores, each of which represents a IRT score for each hypothesized dimension in USW, are shown in Figure 5.9. There were differences in the level of test information among hypothesized USW dimensions. For example, USW_CG items provided precise estimates for the teachers across a wide range of theta (-3.0 ~ 1.2) with information greater than 2, which is equivalent to Cronbach alpha coefficient of 0.67. On the other hand, USW_SR items provided a little precision (less than 1.5) for the entire range of theta. The lowest test information for USW_SR seems to be due to the item difficulty. In other word, considering the small number of participants who answered the USW_SR items correctly, USW_SR items were too difficult to reliably distinguish teachers' knowledge level. For example, even knowledgeable teachers could not answer the items correctly, which is also shown in the DCM model (Figure 5.7) that even masters have a less than 50% probability of answering the USW_SR item correctly (46% for M1007, 44% for M1019).

Overall, all the hypothesized knowledge dimensions of USW, except USW_SR, provided acceptable precisions for the scores ranged from -2.8 to 2.0, using a threshold 1.5(= Cronbach alpha coefficient of 0.60[23]) for an acceptable level of reliability. For improved reliability, more items may need to be added for each of the dimensions. The suggestions for item development are discussed in Chapter 6 in more detail.

[23] If the scale is short, such as less than 10 items, a slightly lower criteria (of about 0.6) can be considered as the criteria of acceptability for reliability coefficients (Lewis & Loewenthal, 2015, p.60)

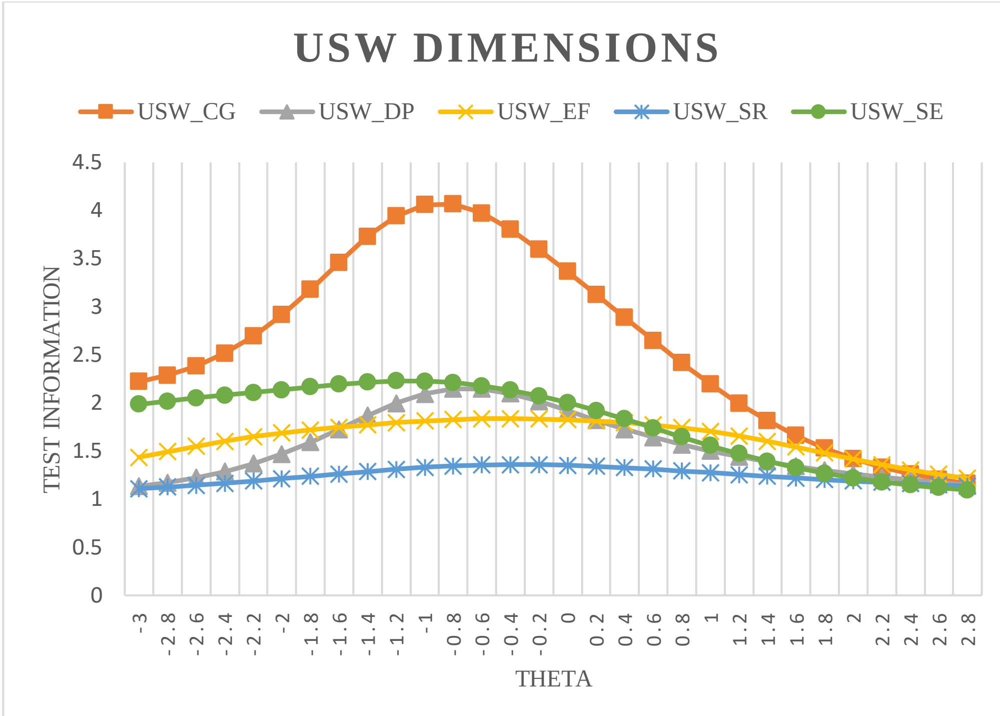


Figure 5.9. Test information function of USW

The reliability of estimates (the reliability of mastery classification, in terms of DCMs) was also evaluated under a DCM framework. Using the DCM measure of reliability from Templin and Bradshaw (2013), reliability of each of the hypothesized knowledge attributes was calculated by using R code created by Templin and Bradshaw (2013). The estimated reliabilities were 0.97, 0.93, 0.91, 0.76, 0.91 for USW_CG, USW_DP, USW_EF, USW_SR, USW_SE, respectively. As expected, the DCM model provided higher reliabilities than the IRT model. The difference in reliability among dimensions was also consistent with the results shown in the IRT model. For example, the reliability for USW_SR (0.76) was much less than that for USW_CG (0.93). Overall, the result that a DCM model (LCDM in this study) yielded acceptable reliabilities for all the dimensions (greater than 0.70), which was not attainable in IRT, suggests that a DCM could be an alternative measurement model for a small number of items.

In summary, tested measurement models supported my hypothesis that multidimensional models fit the item responses better than a unidimensional model. The results supporting the hypothesis were also consistent across three different approaches: SEM, MIRT, and DCM. Furthermore, the items within the same task of USW could be distinguished by different instructional situations. Even though, two geometry factors could not be statistically distinguishable, items yielded higher factor loadings when loaded on the two separated factors than when loaded on the combined factor. The result may imply that two sets of items (USW_CG and USW_DP) are measuring two different dimensions of knowledge reflecting two instructional situations, but they are not statistically distinguishable because of a high similarity between two types of instructional situations.

5.1.2 Organization of the items within the task CGP

The task of choosing the givens for a problem (hereafter, CGP) was conceptualized as involving writing text of various kinds of questions such as classroom tasks or test items, choosing particular information (numerical, algebraic, diagrammatic) to include with the text and deciding how to encode it, ascertaining that the text calls for the student to do what the author of the problem expects them to do, and verifying that the information given is consistent and sufficient for students to engage in the work envisioned. Items that involved the respondent in this task of teaching were classified into four separate instructional situations, as shown in Table 5.8. Compared to the previous section where I examined the dimensionality of USW items, this section examines whether the items measuring teachers' knowledge in doing CGP can be distinguished by different instructional situations. The analyses were also conducted under an IFA framework (assuming continuous constructs) and a DCM framework (assuming discrete attributes); the continuous construct (IFA) framework is further applied under SEM-based modeling and IRT-based modeling.

Table 5.8. Items in the task CGP

Task of teaching	**Instructional situation**	**Knowledge dimension/Attribute**	**# items**
Choosing appropriate Givens for a Problem (CGP)	Calculation in geometry (CG)	CGP_CG	5
	Doing Proofs in geometry (DP)	CGP_DP	4
	Exploring a Figure in geometry (EF)	CGP_EF	4
	Calculation with Numbers (CN) in algebra	CGP_CN	3

5.1.2.1 Dimensionality analysis of the items within CGP using IFA

As all the item factor analyses conducted in this study follow the same design as the models described in the previous section with respect to the estimators and settings used, I hereafter describe the results from SEM-based modeling and MIRT-based modeling together in one section.

Confirmatory item factor analyses using WLSMV estimator and MML estimator were conducted to examine whether the items commonly measuring teachers' mathematical knowledge for the task of choosing the givens for a problem can be distinguished in terms of the instructional situation. The result suggested that the factors of CGP_EF and CGP_DP were highly correlated as 0.92 (Table 5.9). This high correlation between two factors led me to compare this 4-factor model (Figure 5.10.

Four-dimensional model within the task CGP) and 3-factor model where CGP_DP and CGP_EF are combined (Figure 5.11).

Confirmatory item factor analyses using WLSMV estimator and MML estimator were conducted to examine whether the items commonly measuring teachers' mathematical knowledge for the task of choosing the givens for a problem can be distinguished in terms of the instructional situation. The result suggested that the factors of CGP_EF and CGP_DP were highly correlated as 0.92 (Table 5.9. Estimated correlations among dimensions within CGP). This high correlation between two factors led me to compare this 4-factor model (Figure 5.10) and 3-factor model where CGP_DP and CGP_EF are combined (Figure 5.11).

Table 5.9. Estimated correlations among dimensions within CGP

	Dimension			
Dimension	CGP_CG	CGP_DP	CGP_EF	CGP_CN
CGP_CG				
CGP_DP	0.77			
CGP_EF	0.74	**0.92**		
CGP_CN	0.66	0.79	0.68	

The comparison was tested using DIFFTEST and likelihood ratio test for SEM-based and MIRT-based modeling, respectively. The result suggested that the 3-factor model, where EF and DP are combined, was not significantly worse than the initial 4-factor model (DIFFTEST: $\chi^2 = 1.305$, $df = 3$, $p = 0.7279$; LR test: $\chi^2 = 1.48$, $df = 3$, $p = 0.687$). In contrast, another 3-factor model, where DP and CG are combined, was significantly worse than the initial 4-factor model (DIFFTEST: $\chi^2 = 12.200$, $df = 3$, $p = 0.007$; LR test: $\chi^2 = 15.656$, $df = 3$, $p = 0.001$). The 3-factor model, where EF and DP are combined, was further compared to the 2-factor model, where CG, DP, and EF are combined. The comparison results suggested that the 2-factor model was significantly worse than the 3-factor model (DIFFTEST: χ^2 =12.762, *df* =2, *p* =0.0017; LR test: χ^2 =18.54, *df* =2, *p* < 0.001). In sum, the results indicated that all the pairs of hypothesized dimensions are distinguishable except the pair of EF and DP.

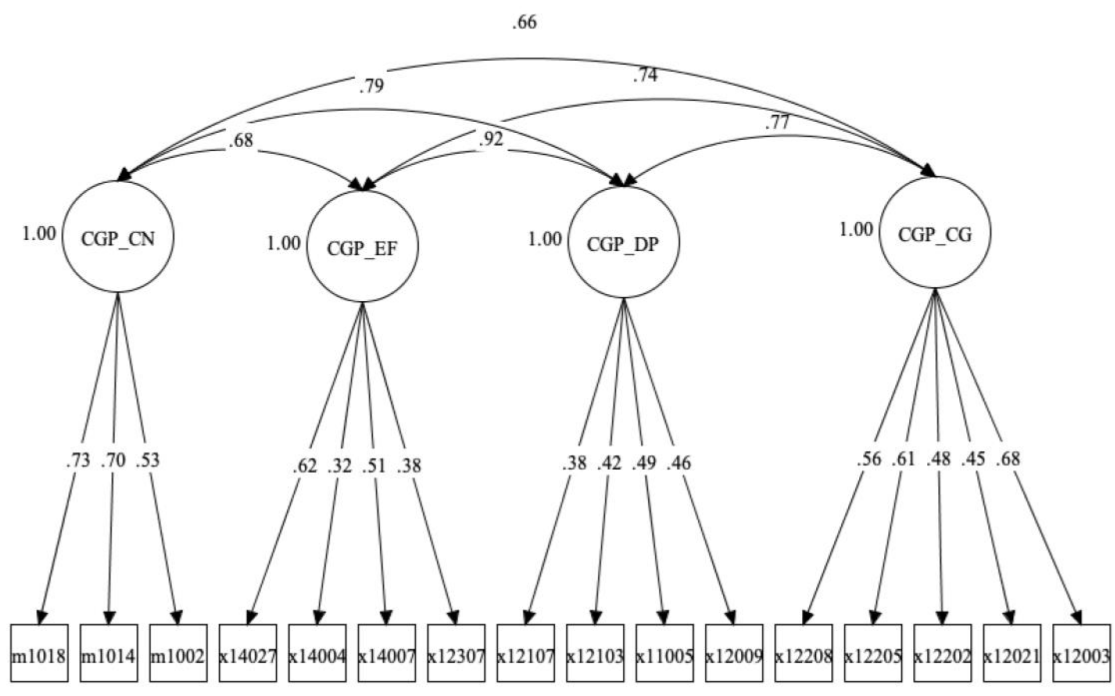


Figure 5.10. Four-dimensional model within the task CGP

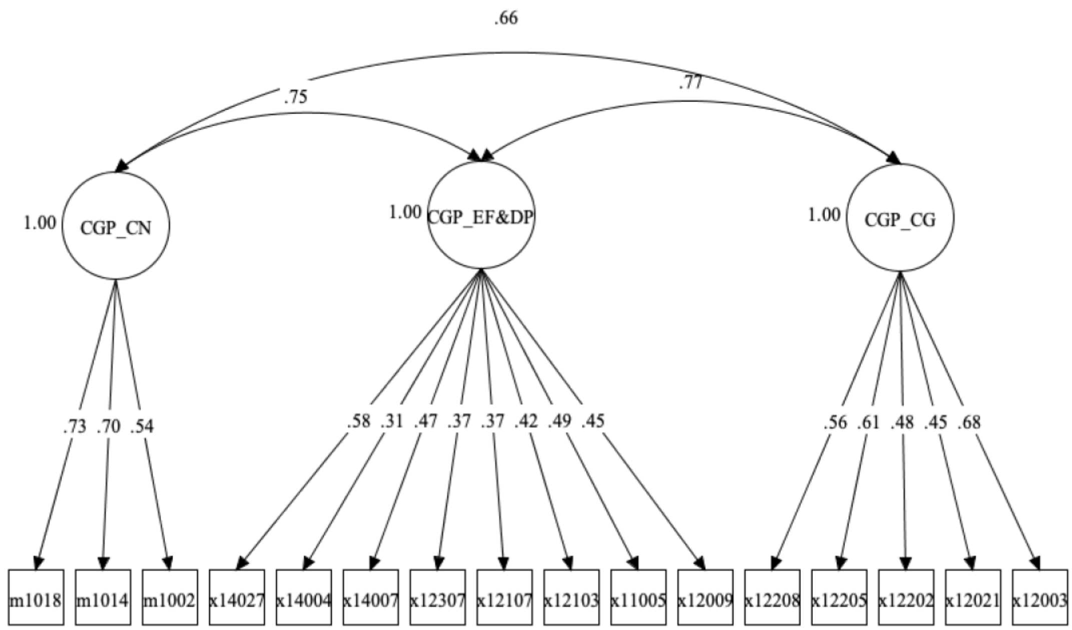


Figure 5.11. Three-dimensional model within the task CGP

Therefore, a 3-factor model, where the factors CGP_EF and CGP_DP are combined (Figure 5.11), was evaluated as the most appropriate model based on parsimony and fit statistics. The fit statistics of the 3-factor model also shows good global fit (RMSEA = 0.011; CFI = 0.994; TLI = 0.993). The fit indices for each model are provided in Table 5.10.

Table 5.10. Comparison of fit among MIRTs within CGP

Model	Deviance (-2log likelihood) statistics	Number of free parameters	Change in Deviance	Change in Degrees of Freedom	*p* (significance)	*AIC*
4-Dimension	10986.49	46				11078.49
3-Dimension (DP+EF)	**10987.97**	**43**	**1.48**	**3**	**0.68**	**11073.97**
3-Dimension (CG+DP)	11002.14	43	15.66	3	0.001	11088.14
2-Dimension (CG+DP+EF)	11006.508	41	18.54	2	< 0.001	11088.51

Table 5.11 below presents estimated factor loadings estimated by the 3-factor and 4-factor models under SEM- and MIRT-based models.

Table 5.11. Estimated standardized factor loading within CGP

		SEM-based			IRT-based		
Item	Dimension	2 (CG+DP +EF)	3 (EF+DP)	4	2 (CG+DP +EF)	3 (EF+DP)	4
X12003	**CGP_CG**	0.63 (0.04)	0.69 (0.04)	0.69 (0.04)	0.62 (0.05)	0.68 (0.05)	0.68 (0.05)
X12021		0.48 (0.06)	0.50 (0.06)	0.50 (0.06)	0.45 (0.06)	0.45 (0.06)	0.45 (0.06)
X12202		0.47 (0.05)	0.51 (0.05)	0.51 (0.05)	0.44 (0.05)	0.48 (0.05)	0.48 (0.05)
X12205		0.59 (0.05)	0.63 (0.06)	0.63 (0.06)	0.57 (0.06)	0.61 (0.06)	0.61 (0.06)
X12208		0.52 (0.06)	0.56 (0.06)	0.56 (0.06)	0.50 (0.07)	0.56 (0.07)	0.56 (0.07)
X12009	**CGP_DP**	0.44 (0.05)	0.45 (0.05)	0.45 (0.05)	0.43 (0.05)	0.45 (0.05)	0.46 (0.06)
X11005		0.52 (0.06)	0.54 (0.06)	0.54 (0.07)	0.50 (0.06)	0.49 (0.06)	0.49 (0.06)
X12103		0.42 (0.06)	0.44 (0.06)	0.44 (0.06)	0.37 (0.06)	0.42 (0.06)	0.42 (0.07)
X12107		0.42 (0.06)	0.43 (0.06)	0.43 (0.07)	0.36 (0.06)	0.37 (0.06)	0.38 (0.06)
X12307	**CGP_EF**	0.38 (0.06)	0.39 (0.07)	0.41 (0.07)	0.33 (0.06)	0.37 (0.07)	0.38 (0.07)
X14007		0.50 (0.06)	0.52 (0.06)	0.54 (0.07)	0.45 (0.06)	0.47 (0.06)	0.51 (0.07)
X14004		0.32 (0.06)	0.34 (0.07)	0.36 (0.07)	0.27 (0.06)	0.31 (0.07)	0.32 (0.07)
X14027		0.59 (0.05)	0.62 (0.06)	0.65 (0.06)	0.56 (0.06)	0.58 (0.06)	0.62 (0.07)
M1002	**CGP_CN**	0.58 (0.07)	0.58 (0.07)	0.58 (0.07)	0.53 (0.07)	0.54 (0.07)	0.53 (0.07)
M1014		0.72 (0.06)	0.72 (0.06)	0.72 (0.06)	0.70 (0.06)	0.70 (0.06)	0.70 (0.06)
M1018		0.74 (0.06)	0.74 (0.06)	0.74 (0.06)	0.73 (0.06)	0.73 (0.06)	0.73 (0.06)

All item loadings are significant at the $p < 0.05$ level
Standardized loading (standard error)

As shown in the table above, all the standardized item loadings of the EF factor on the 4-factor model tend to go up slightly relative to the loadings in the 3-factor model, even though 4 factor model is not significantly better than the 3-factor model. This may indicate that these items can explain more variance when they are distinguished from DP factor. Similarly, all the items of the geometry related factors (CG, DP, EF) have higher item factor loadings in the 3-factor model than the 2-factor model in both SEM and IRT-based models.

In summary, tested measurement models supported my hypothesis that the items within the same task of CGP could be distinguished by different instructional situations, though the distinction between EF and DP requires preferring a less parsimonious solution for the sake of higher loadings. The results were consistent under both SEM and IRT based framework. In the next section, the same hypothesis on the structure of CGP items is investigated under a DCM modeling.

5.1.2.2 Dimensionality analysis of the items within CGP using DCM

A loglinear cognitive diagnosis model (LCDM) was retrofitted to the items measuring teachers' knowledge in doing the task of choosing the givens for a problem (CGP_CG, CGP_DP, CGP_EF, CGP_CN). Following the same 4-factor structure tested in IFA, the LCDM model was set to estimate the main effects of items on each of the four knowledge attributes distinguished by instructional situations within the task CGP. The specified Q-matrix (Table 5.12) and the equation demonstrating the model parameters (Equation 2) are as follows.

Table 5.12. Q-matrix for items within CGP

Item	Attribute 1: CGP _CG	Attribute 2: CGP_DP	Attribute 3: CGP_EF	Attribute 4: CGP_CN
X12003, X12021, X12202_b, X12205, X12208	1	0	0	0
X12009_b, X11005, X12103_b, X12107	0	1	0	0
X12307, X14007, X14004, X14027	0	0	1	0
M1002, M1014, M1018	0	0	0	1

Items with _b are dichotomized (originally ordinal items) responses

Following the same notations used in the previous section, the equation estimating parameters of the CGP items can be represented as follows:

$$\ln\left(\frac{P(X_i = 1|\alpha_{ra})}{P(X_i = 0|\alpha_{ra})}\right) = \lambda_{i,0} + \lambda_{i,1,(a)}\alpha_{ra}, \text{ where a=CGP\_CG, CGP\_DP, CGP\_EF, or CGP\_CN, depending on the attribute item } i \text{ measures} \quad (2)$$

For each item, the equation (2) has one intercept and one main effect. Given that there are four hypothesized dimensions within CGP, a respondent's profile vector is expressed as a vector with four elements α_r=[$\alpha_{rCGP\_CG}, \alpha_{rCGP\_DP,}\alpha_{rCGP\_EF,}\alpha_{rCGP\_CN}$], each of which has 1, when a respondent has mastered the knowledge attribute. Thus, the main effect parameter of an attribute can be included in the calculation of the logit, only if a respondent's mastery profile has "1" for that attribute.

For example, item x12003 measures knowledge attribute CGP_CG and the equation estimating item parameters of the item can be represented as

$$\ln\left(\frac{P\left(X_{x12003} = 1\middle|\alpha_{rCGP\_CG}\right)}{P\left(X_{x12003} = 0\middle|\alpha_{rCGP\_CG}\right)}\right) = \lambda_{x12003,0} + \lambda_{x12003,1,(CGP\_CG)}\alpha_{rCGP\_CG}$$

The parameter $\lambda_{x12003,0}$ is the intercept, indicating the logit of a correct answer to item x12003 by nonmasters of CGP_CG ($\alpha_{rCGP\_CG} = 0$). The parameter $\lambda_{i,1,(CGP\_CG)}$ is the main effect (indicated by the second subscript) for CGP_CG. It represents the increase in the logit of a correct answer to item x12003 for mastering CGP_CG for a respondent (a teacher in this study) who has not mastered the attribute.

5.1.2.2.1 Model fit

Following the same approach used in evaluating the LCDM model with USW items (5.1.1.2), the fit of the LCDM model with CGP items was evaluated by calculating a bivariate goodness of fit statistic with a $\chi^2(1)$ distribution for each pair of CGP items (Rupp et al., 2010).

The result suggested that 95% (115 pairs) of the 120 (=$\frac{16*15}{2}$) item pairs showed good model fit (i.e., chi-square value is insignificant) when chi-square values were evaluated at a 0.05 significance value and no item pair showed misfit at a 0.01 significance level. Overall, my LCDM model estimating teachers' knowledge in choosing the givens for a problem across four knowledge attributes provided an acceptable model fit for the data used in this study. Therefore, the item parameters estimated from the model were further interpreted as follows.

5.1.2.2.2 Item parameter estimates

The item parameters estimated from the LCDM are listed in Table 5.13. As shown in the table, the average intercept across items is -0.73, indicating the average predicted logit of a correct response for teachers who had not mastered any of the attributes. This means that approximately 0.33% (= $\frac{e^{-0.73}}{(1+e^{-0.73})}$) of teachers who had not mastered any of the three CGP knowledge attributes answered the items correctly. The average main effect ranged from 1.65 to 2.58. This means that the increase in the logit of a correct response by mastering one of the attributes ranged from 1.65 to 2.58, which correspond to odds ratios between 5.21 and 13.20. The main effect size of each item on each knowledge attribute was evaluated using the criterion for odds ratio suggested by Bradshaw et al. (2014, p.6), with small effect sizes for odd ratios between 1.44 and 2.47, medium effect sizes for odd ratios between 2.47 and 4.25, and large effect sizes for odd ratios larger than 4.25). According to these suggested thresholds, three items had medium effect sizes (x12107, x12307, x14004) and the remaining 13 items had large effect sizes. This result suggests that all the items measured the intended attributes with significant effect sizes.

Table 5.13. Estimated item parameters (logit) within CGP

Item	Intercept $\lambda_{i,0}$	CGP_CG $\lambda_{i,1,(1)}$	CGP_DP $\lambda_{i,1,(2)}$	CGP_EF $\lambda_{i,1,(3)}$	CGP_CN $\lambda_{i,1,(4)}$
X12003_b	-2.38 (0.37)	2.70 (0.39)			
X12021	-0.27 (0.17)	1.65 (0.31)			
X12202_b	-1.20 (0.25)	1.92 (0.32)			
X12205	-0.12 (0.21)	2.63 (0.48)			
X12208	0.61 (0.20)	2.32 (0.51)			
X12009_b	-1.02 (0.18)		1.65 (0.31)		
X11005	-1.57 (0.26)		1.92 (0.34)		
X12103	-0.66 (0.19)		1.70 (0.32)		
X12107	-0.82 (0.23)		1.33 (0.31)		
X12307	0.31 (0.17)			1.25 (0.33)	
X14004	-1.39 (0.27)			1.21 (0.35)	
X14007	-0.75 (0.24)			2.09 (0.35)	
X14027	-0.90 (0.24)			2.27 (0.37)	
M1002	-0.19 (0.22)				1.92 (0.32)
M1014	-1.14 (0.34)				2.92 (0.40)
M1018	-0.23 (0.23)				2.89 (0.50)
Average	**-0.73 (0.24)**	**2.24 (0.40)**	**1.65 (0.32)**	**1.70 (0.35)**	**2.58 (0.40)**

Standard errors in parentheses

5.1.2.2.3 Probability of a correct response between nonmasters and masters

Following the same approach of Bradshaw et al. (2014), the strength of associations between measured attributes and each item were examined by using item characteristics bar charts (ICBCs). The ICBCs for each knowledge attributes are displayed in Figure 5.12~5.15. By examining the difference in the probability of a correct response between nonmasters and masters, the ability of each item in distinguishing masters from nonmasters was evaluated. Overall, significant differences could be identified in the probability of getting a correct response between nonmasters and masters across items, with respect to the four CGP knowledge attributes. The ICBCs also provide information about item difficulties. For example, x14004 seems to be a difficult item given that even masters choose a correct answer with less than 0.5 probability (Figure 5.14).

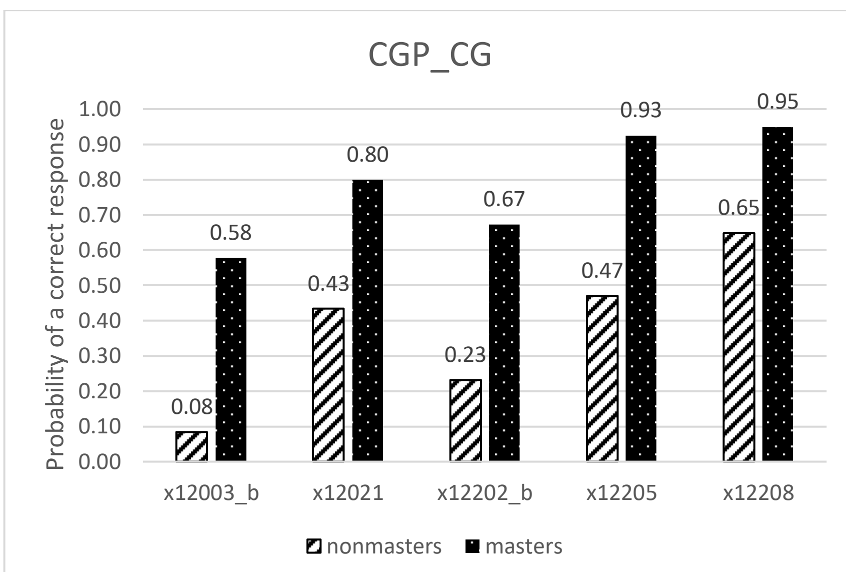


Figure 5.12. Item characteristic bar charts (ICBCs) for CGP_CG items

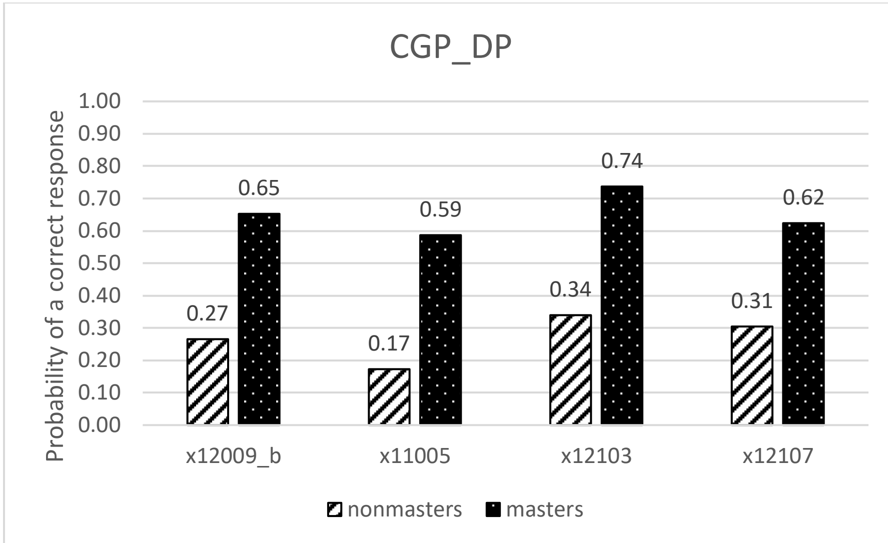


Figure 5.13. Item characteristic bar charts (ICBCs) for CGP_DP

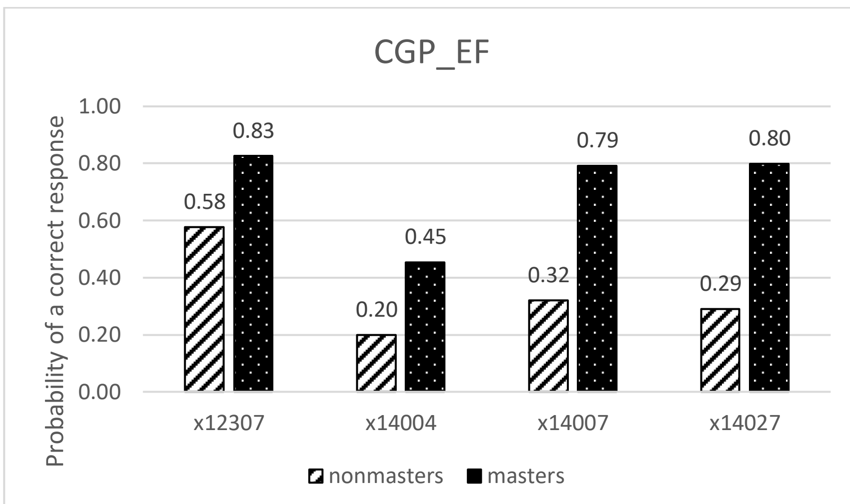


Figure 5.14. Item characteristic bar charts (ICBCs) for CGP_EF

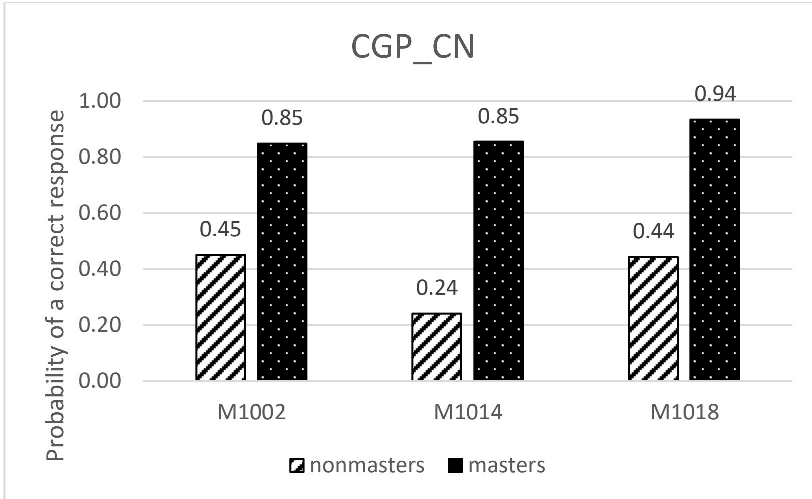


Figure 5.15. Item characteristic bar charts (ICBCs) for CGP_CN

Overall, ICBCs showed noticeable differences in the probability of a correct response between nonmasters and masters for all the CGP items.

5.1.2.2.4 Model comparison

Similar to the test conducted under the MIRT framework, the 4-attribute LCDM model was compared to the 3-attribute LCDM model (where the attributes DP and EF combine together) and the 2-attribute LCDM model (where the attributes CG, DP, and EF combine together). The result was consistent with that of IFA in that the 3-attributes model (DP+EF) was not significantly worse than the 4-attributes model ($\chi^2(8) = 8.89$, $p=0.35$), but significantly better than the 2-attribute model (($\chi^2(4) = 22.31$, $p<0.001$). Other fit statistics, such as AIC, also suggested better fit of the 3-attribute model (7892.451) than the 4-attribute model (7899.558) or 2-attribute model (7906.761).

5.1.2.3 Reliability

The degree of consistency in item scores (i.e., reliability) was evaluated for each hypothesized knowledge dimension under two different modeling: IFAs and DCM. Under the IFA modeling I examined a range of knowledge levels where the items can provide information with acceptable precisions using test information functions (TIF) for the CGP_CG, CGP_DP, CGP_EF, CGP_CN, respectively. In contrast to classical test theory in which reliability of the test scores are estimated by a single value, the reliability of the score estimated in IRT is evaluated by a function that varies across ability levels.

The test information functions for estimated scores are presented in Figure 5.16 and Figure 5.17. As shown in the figure, CGP_CG items provided precise estimates for teachers across a wide range of theta (-3.0 ~ 1.4) with information greater than 2, which is equivalent to Cronbach alpha coefficient of 0.67. The combined factor of CGP_DPEF provided acceptable information (greater than 1.5) for the entire range of theta (Figure 5.17), while CGP_DP and CGP_EF provided less than 2.0 information when they are scaled separately (Figure 5.16). In particular, the combined factor of CGP_DPEF provided the highest information at the higher values of the knowledge scale (Figure 5.17). This suggests that CGP_DP&EF items better discriminate high-scoring teachers than other items (Figure 5.17). CGP_CN items provided the most precise estimates for the teachers whose knowledge scales is between -1.6 and 0.4.

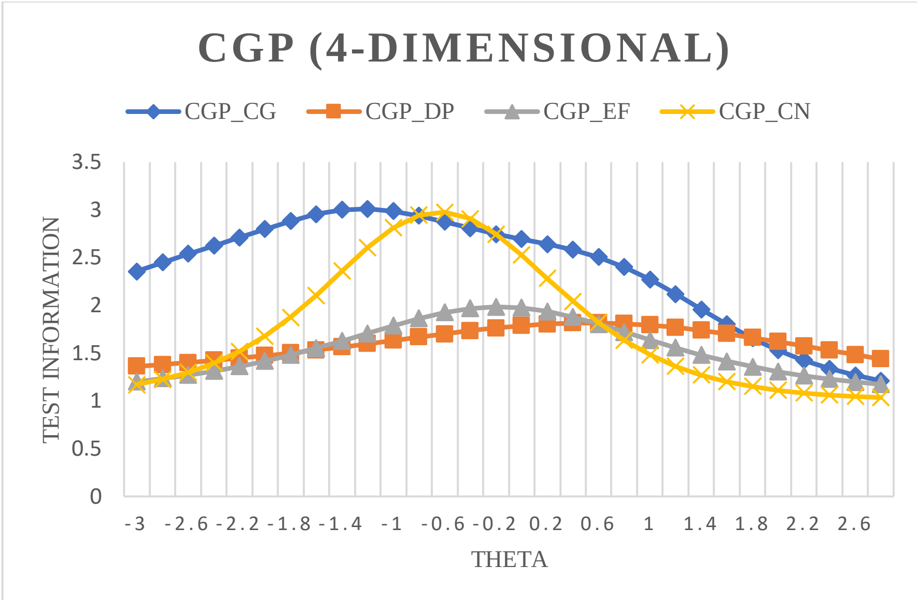


Figure 5.16. Test information function of CGP under 4-dimensional MIRT

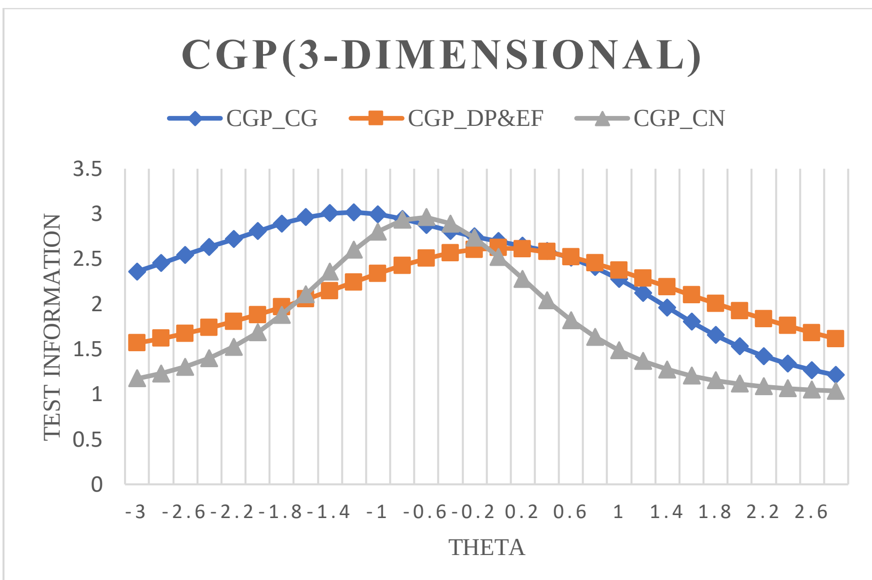


Figure 5.17. Test information function of CGP under 3-dimensional MIRT

The reliability of estimates (the reliability of mastery classification, in terms of DCMs) was also evaluated under a DCM modeling. Using the DCM measure of reliability from Templin and Bradshaw (2013), reliabilities of the identified knowledge attributes were calculated by using an R code created by Templin and Bradshaw (2013).

The estimated reliabilities were 0.87, 0.89, 0.85, and 0.88 for CGP_CG, CGP_DP, CGP_EF, and CGP_CN, respectively. As expected, the DCM model provided much higher reliabilities than the IRT model.

In summary, tested measurement models supported my hypothesis that multidimensional models fit the item responses better than a unidimensional model within the task of CGP. The results supporting the hypothesis were also consistent across three different approaches: SEM, MIRT, and DCM. The items within the same task of CGP could be distinguished by different instructional situations. However, two geometry factors reflecting CGP_DP (knowledge for choosing appropriate givens for a problem in a situation of doing proofs) and CGP_EF (knowledge for choosing appropriate givens for a problem in a situation of exploring figures) could not be statistically distinguishable. This may imply that there are high similarities between two types of instructional situations – doing proofs and exploring figures, which will be discussed in Chapter 6 in more detail.

5.2 Dimensionality within the same instructional situation

To examine whether multiple dimensions can be identified by the organizer task of teaching, the dimensionality of the item responses was investigated within each instructional situation. Given that three item clusters of instructional situations – Doing Proofs (DP), Calculation in Geometry (CG), and Exploring a Figure (EF) – were identified across two different Tasks of Teaching, the distinction was examined through three separate Chi-Square Difference Tests within the situation of DP, CG, and EF respectively. Specifically, the fit of two-dimensional model (where USW is distinguished from CGP) was compared to the unidimensional model (where two tasks are combined) for each of the instructional situations – DP, CG, and EF.

5.2.1 Organization of the items within the instructional situation CG

The items commonly measuring teacher' mathematical knowledge in the instructional situation of Calculation in Geometry (CG) were divided into two categories according to the task of teaching involved: one is the set of items measuring teachers' mathematical knowledge in doing the task of understanding students' work (USW_CG) and another is the set of items measuring teachers' mathematical knowledge in doing the task of choosing appropriate givens for a problem (CGP_CG) (Table 5.14).

Table 5.14. Items in the instruction situation CG

Task of teaching	**Instructional situation**	**Knowledge dimension/Attribute**	**# items**
Calculation in Geometry (CG)	Understanding Students' work (USW)	USW_CG	6
	Choosing appropriate Givens for a Problem (CGP)	CGP_CG	5

With these two sets of items, this section examines whether the items measuring teachers' knowledge used in CG can be distinguished by different tasks of teaching. The analyses were conducted using IFA and DCM models, and the IFA models were implemented under SEM-based and IRT-based framework.

5.2.1.1 Dimensionality analysis of the items within CG using IFAs

Using the same approach applied in the previous sections, confirmatory item factor analyses were conducted under SEM and IRT based framework. In other words, the comparison between the 2-factor model distinguishing CGP_CG and USW_CG (Figure 5.18) and the unidimensional model where all the CG items are loaded on one factor (Figure 5.19) was tested using DIFFTEST (under SEM) and likelihood ratio test (under MIRT). The results suggested that the two factors representing two different tasks of teaching within the same instructional situation of CG are statistically distinguishable (DIFFTEST: χ^2 =8.167, *df* =1, *p* =0.004; LR test: χ^2 =9.894, *df* =1, *p* =0.002).

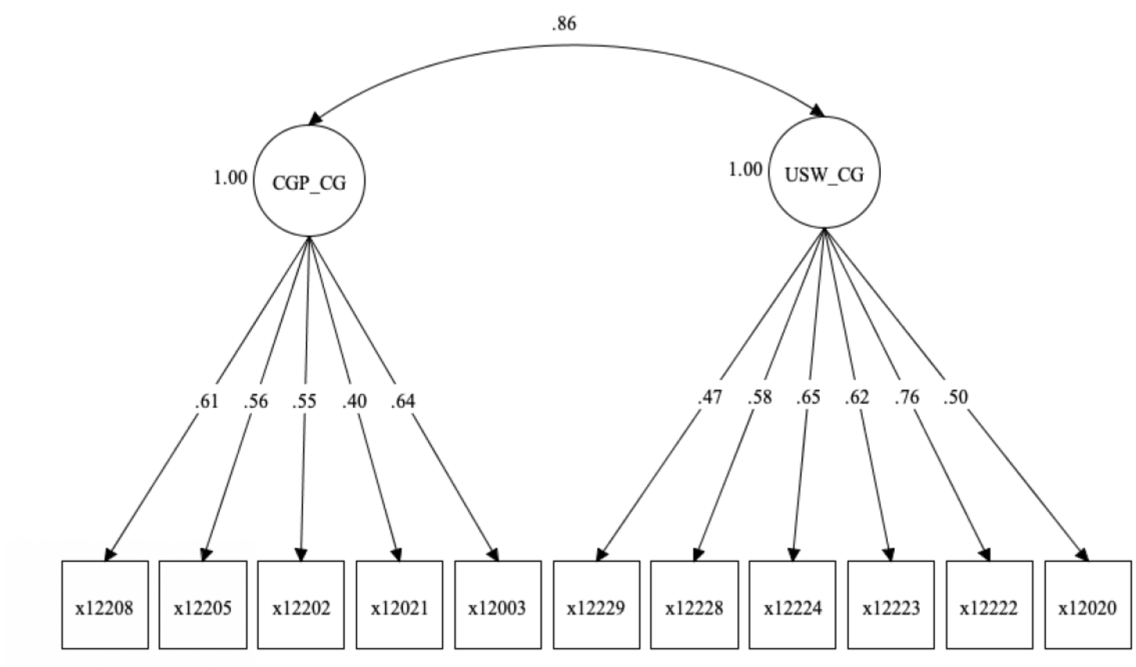


Figure 5.18. Two-dimensional model within the instructional situation CG

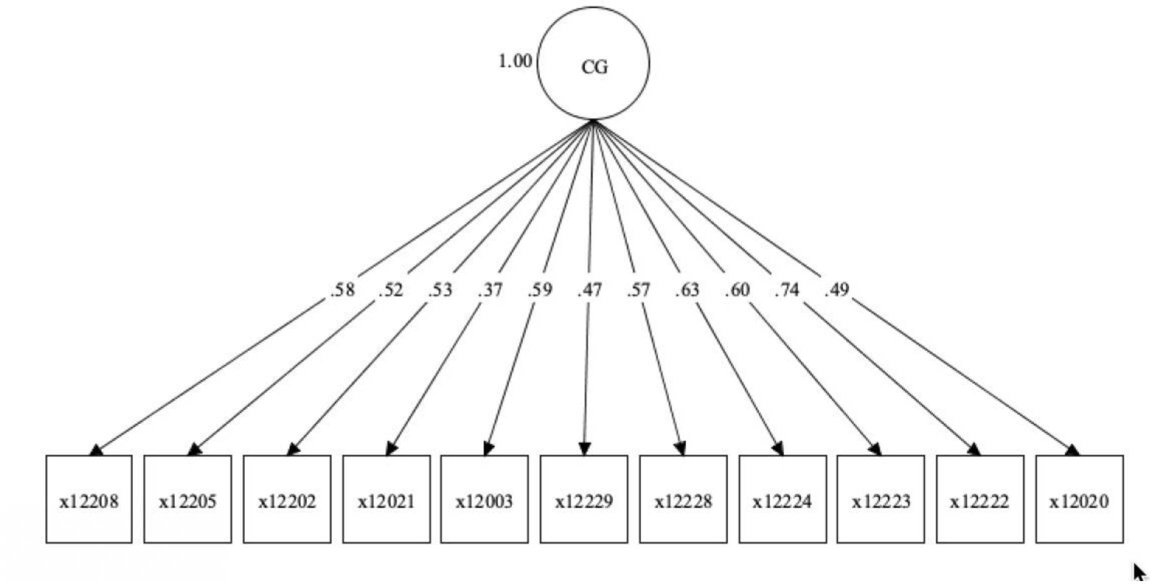


Figure 5.19. Unidimensional model within the instructional situation CG

Both unidimensional and 2-factor model yielded good global fit statistics. The fit indices for each model are provided in Table 5.15. The factor loadings estimated by the 2-factor and 1-factor model under SEM- and MIRT-based models are presented in Table 5.15.

Table 5.15. Comparison of fit among the IFA models within CG

	SEM-based			MIRT-based					
Model	RMSEA	CFI	TLI	Deviance (-2log likelihood) statistics	Number of free parameters	Deviance change	df change	p	AIC
2-Dimension	0.000	1.000	1.000	7582.51	32				7646.50
1-Dimension	0.018	0.994	0.993	7592.40	31	9.89	1	0.002	7654.40

Table 5.16. Estimated standardized factor loadings within CG

		SEM-based		IRT-based	
Item	Dimension	1	2	1	2
X12020	**USW_CG**	0.54 (0.06)	0.55 (0.06)	0.49 (0.06)	0.50 (0.06)
X12222		0.74 (0.05)	0.75 (0.05)	0.74 (0.05)	0.76 (0.05)
X12223		0.64 (0.05)	0.65 (0.05)	0.60 (0.05)	0.62 (0.05)
X12224		0.65 (0.04)	0.67 (0.04)	0.63 (0.04)	0.65 (0.04)
X12228		0.60 (0.05)	0.61 (0.05)	0.58 (0.05)	0.59 (0.05)
X12229		0.50 (0.06)	0.51 (0.06)	0.47 (0.06)	0.47 (0.06)
X12003	**CGP_CG**	0.62 (0.04)	0.65 (0.04)	0.59 (0.05)	0.64 (0.05)
X12021		0.43 (0.06)	0.45 (0.06)	0.37 (0.06)	0.40 (0.06)
X12202		0.55 (0.05)	0.58 (0.05)	0.53 (0.05)	0.55 (0.05)
X12205		0.56 (0.06)	0.59 (0.06)	0.52 (0.06)	0.56 (0.06)
X12208		0.60 (0.06)	0.63 (0.06)	0.58 (0.06)	0.61 (0.06)

All item loadings are significant at the $p < 0.05$ level
Standardized loading (standard error)

As expected from the result of model comparison, all the items showed higher factor loadings when they were loaded onto separated factors than when they were loaded on the same factor of CG (Table 5.16). This indicates that CG items can explain more variance when they are distinguished by two different tasks of teaching. In summary, tested measurement models supported my hypothesis that the items within the same instructional situation CG could be distinguished by different tasks of teaching. Moreover, the results were consistent under both SEM and IRT based framework. In the next section, the same hypothesis on the structure of CG items is investigated using a DCM model.

5.2.1.2 Dimensionality analysis of the items within CG using DCM

A loglinear cognitive diagnosis model (LCDM) was retrofitted to the items designed to measure two distinguishable dimensions of teachers' knowledge in the situation of calculation in geometry (USW_CG and CGP_CG). Following the same 2-factor structure tested in IFA, the LCDM model was set to estimate the main effects of items on each of the two knowledge attributes distinguished by tasks of teaching within the instructional situation CG. The specified Q-matrix (Table 5.17) and the equation demonstrating the model parameters (Equation 3) are as follows.

Table 5.17. Q-matrix for the items within CG

Item	Attribute 1: USW_CG	Attribute 2: CGP_CG
X12020 X12222 X12223 X12224_b X12228_b X12229	1	0
X12003_b X12021 X12202_b X12205 X12208	0	1

Items with _b are dichotomized (originally ordinal item) responses

Following the same notations used in the previous section, the equation estimating parameters of the CG items can be represented as follows:

$$\ln\left(\frac{P(X_i = 1|\alpha_{ra})}{P(X_i = 0|\alpha_{ra})}\right) = \lambda_{i,0} + \lambda_{i,1,(a)}\alpha_{ra}, \text{ where } a = \text{USW\_CG or CGP\_CG,} \tag{3}$$

depending on the attribute item *i* measures

For each item, the equation (3) has one intercept and one main effect. Given that there are two hypothesized dimensions within CG, a respondent's profile is expressed as a vector with two elements α_r=[$\alpha_{rUSW-CG}, \alpha_{rCGP-CG}$], each of which has 1 when a respondent has mastered the knowledge attribute. Thus, the main effect parameter of an attribute is included in the calculation of the logit, only if a respondent's mastery profile has "1" for that attribute ($\lambda_{i,1,(a)}\alpha_{ra}$= $\lambda_{i,1,(a)}$*0 = 0).

For example, item x12003 measures knowledge attribute CGP_CG and the equation estimating item parameters of the item can be represented as

$$\ln\left(\frac{P\left(X_{x12020}=1\middle|\alpha_{rUSW\_CG}\right)}{P\left(X_{x12020}=0\middle|\alpha_{rUSW\_CG}\right)}\right)=\lambda_{x12020,0}+\lambda_{x12020,1,(USW\_CG)}\alpha_{rUSW\_CG}$$

The parameter $\lambda_{x12020,0}$ is the intercept, indicating the logit of a correct answer to item x12020 by non-masters of USW_CG ($\alpha_{rUSW-CG}=0$). The parameter $\lambda_{i,1,(USW\_CG)}$ is the main effect (indicated by the second subscript) for USW_CG. It represents the increase in the logit of a correct answer to item x12020 for mastering USW_CG for a respondent (a teacher in this study) who had not mastered the attribute.

5.2.1.2.1 Model fit

The fit of the LCDM model with CG items was evaluated by calculating a bivariate goodness of fit statistic with a $\chi^2(1)$ distribution for each pair of 11 CG items (Rupp et al., 2010). The result suggested that all the item pairs except one (between x12208 and x12223) showed a good model fit (i.e., chi-square value is insignificant) when chi-square values are evaluated at a 0.05 significance value. Overall, my LCDM model estimating teachers' knowledge in the situation of CG across two knowledge attributes provided an acceptable model fit for the data used in this study. Therefore, the item parameters estimated from the model were further interpreted as follows.

5.2.1.2.2 Item parameters

The item parameters estimated from the LCDM are listed in Table 5.18. As shown in the table, the average intercept across items is -0.76, indicating the average predicted logit of a correct response for teachers who had not mastered any of the attributes. This means that approximately 0.32% (= $\frac{e^{-0.76}}{(1+e^{-0.76})}$) of teachers who had not mastered any of the two CG knowledge attributes answered the items correctly. The average main effect is 2.57 and 2.17 for USW_CG and CGP_CG, respectively. This means that the increase in the logit of a correct response by mastering USW_CG is 2.57, which correspond to odds ratio of 13.07, and the increase in the logit of a correct response by mastering CGP_CG is 2.17, which corresponds to odds ratio of 8.76. The main effect size of each item on each knowledge attribute was evaluated using the criterion for odds ratio suggested by Bradshaw et al. (2014, p. 6), with small effect sizes for odd ratios between 1.44 and 2.47, medium effect sizes for odd ratios between 2.47 and 4.25, and large effect sizes for odd ratios larger than 4.25. According to these suggested thresholds, all 11 CG items had large effect sizes. This result suggests that all the items measured the intended attributes with significant effect sizes.

Table 5.18. Estimated item parameters (logit) within CG

Item	Intercept $\lambda_{i,0}$	USW_CG $\lambda_{i,1,(1)}$	CGP_CG $\lambda_{i,1,(2)}$
USW_CG			
X12020	-0.89 (0.22)	1.86 (0.26)	
X12222	-0.05 (0.19)	3.56 (0.87)	
X12223	-0.60 (0.21)	2.28 (0.29)	
X12224_b	-0.58 (0.23)	2.67 (0.32)	
X12228_b	-1.44 (0.30)	2.48 (0.31)	
X12229	-0.03 (0.18)	1.61 (0.26)	
CGP_CG			
X12003_b	-2.77 (0.51)		2.76 (0.51)
X12021	-0.47 (0.19)		1.55 (0.26)
X12202_b	-1.44 (0.25)		1.89 (0.30)
X12205	-0.36 (0.20)		2.25 (0.31)
X12208	0.29 (0.20)		2.37 (0.37)
Average	**-0.76 (0.24)**	**2.57 (0.41)**	**2.17 (0.35)**

Standard errors in parentheses

5.2.1.2.3 Probability of a correct response between nonmasters and masters

Following the same approach of Bradshaw et al. (2014), the strength of associations between measured attributes and each item were examined by using item characteristics bar charts (ICBCs). The ICBCs for each CG knowledge attributes are displayed in Figure 5.20 and Figure 5.21. As shown in both figures, significant differences in the probability of getting a correct response between nonmasters and masters could be found in all 11 CG items. The probability of getting a correct response for masters was more than 0.5 across all the items, whereas the probability of getting a correct response for nonmasters was less than 0.5 across all the items except one item (57% for the item x12208).

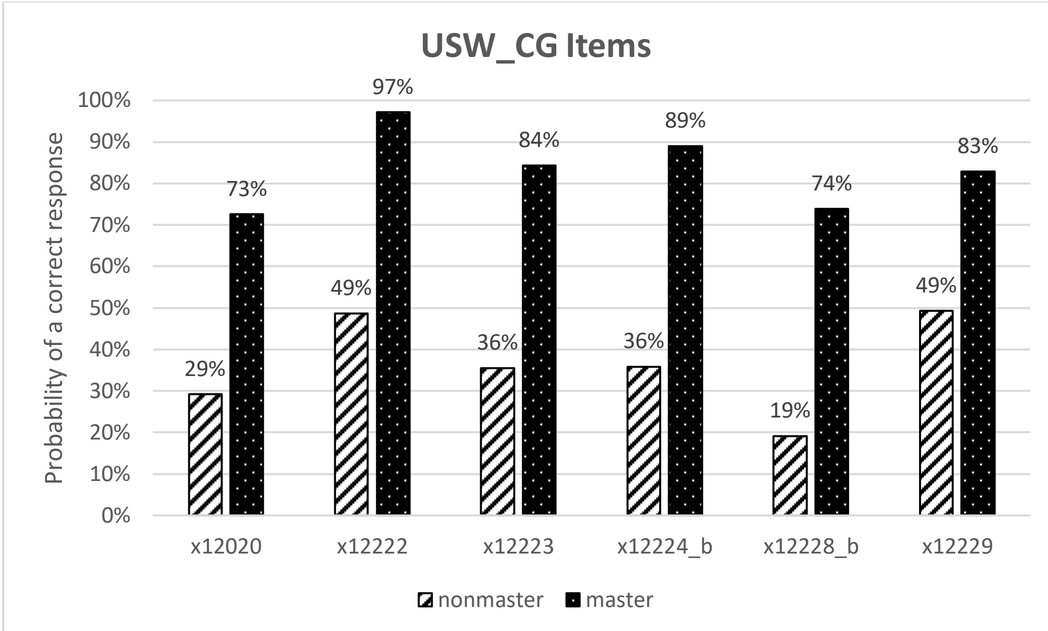


Figure 5.20. Item characteristics bar chart for USW_CG items

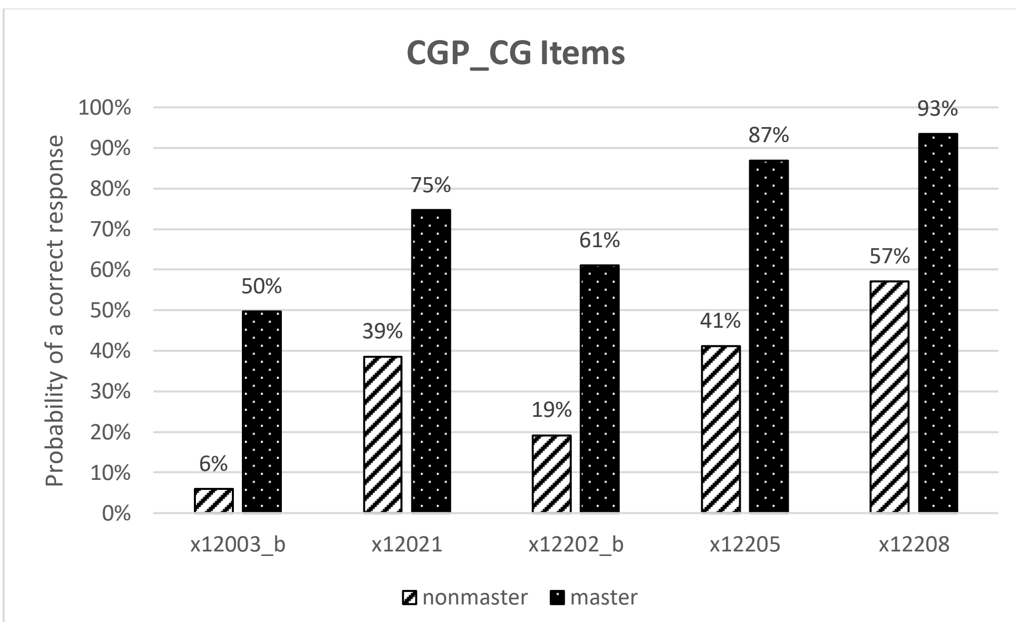


Figure 5.21. Item characteristics bar charts for CGP_CG items

5.2.1.2.4 Model comparison

Similar to the test conducted under the MIRT framework, the 2-attribute LCDM model was compared to the 1-attribute LCDM model (where all the items are assumed to measure the mastery of a single attribute CG). The result was consistent with that of IFA in that the 2- attribute model was significantly better than the 1- attribute model (($\chi^2(2)$ = 8.984, p=0.01). Other fit statistics, such as AIC, also suggested better fit of the 2-attribute model (5312.895) than the 1-attribute model (5317.878).

In summary, tested measurement models supported my hypothesis assuming that the two- factor (or two- attribute) model fit the item responses better than the unidimensional model within the instructional situation of calculation in geometry (CG). The results supporting the hypothesis were also consistent across three different approaches: SEM, MIRT, and DCM. The items within the same situation of CG could be distinguished by different tasks of teaching – understanding students' work (USW) and choosing the givens for a problem (CGP).

5.2.2 Organization of the items within the instructional situation DP

The items that involved the respondents in the instructional situation Doing Proofs in geometry (DP) were classified into two separate tasks of teaching, as shown in Table 5.19. One category is the set of items measuring teachers' mathematical knowledge in doing the task of understanding students' work (USW_DP) and another is the set of items measuring teachers' mathematical knowledge in doing the task of choosing appropriate givens for a problem (CGP_DP).

Table 5.19. Items within DP

Task of teaching	**Instructional situation**	**Knowledge dimension/attribute**	**# items**
Doing Proofs in Geometry (DP)	Understanding Students' Work (USW)	USW_DP	3
	Choosing appropriate Givens for a Problem (CGP)	CGP_DP	4

With these two sets of items, this section examines whether the items measuring teachers' knowledge used in the situation of DP can be distinguished by two different tasks of teaching. The analyses were conducted using IFA (assuming continuous constructs) and DCM models (assuming discrete attributes); the continuous construct (IFA) model was implemented under SEM-based and IRT-based framework.

5.2.2.1.1 Dimensionality analysis of the items within DP using IFAs

To examine whether the two hypothesized dimensions characterized by different tasks of teaching can be distinguished within the same instructional situation of doing proofs (DP), confirmatory item factor analyses were conducted under SEM and IRT based framework. The result suggested that the factors CGP_DP and USW_DP were correlated as 0.634, which is significantly less than 1 (Figure 5.22). The comparison between the 2-factor model (Figure 5.22) and the unidimensional model (Figure 5.23) was further tested using DIFFTEST and likelihood ratio test to confirm that the two factors are statistically distinguishable. The results suggested that the two factors representing two different tasks of teaching within the same instructional situation of DP are statistically distinguishable (DIFFTEST: χ^2 =8.355, *df* =1, *p* =0.004; LR test: χ^2 =10.83, *df* =1, *p* =0.001).

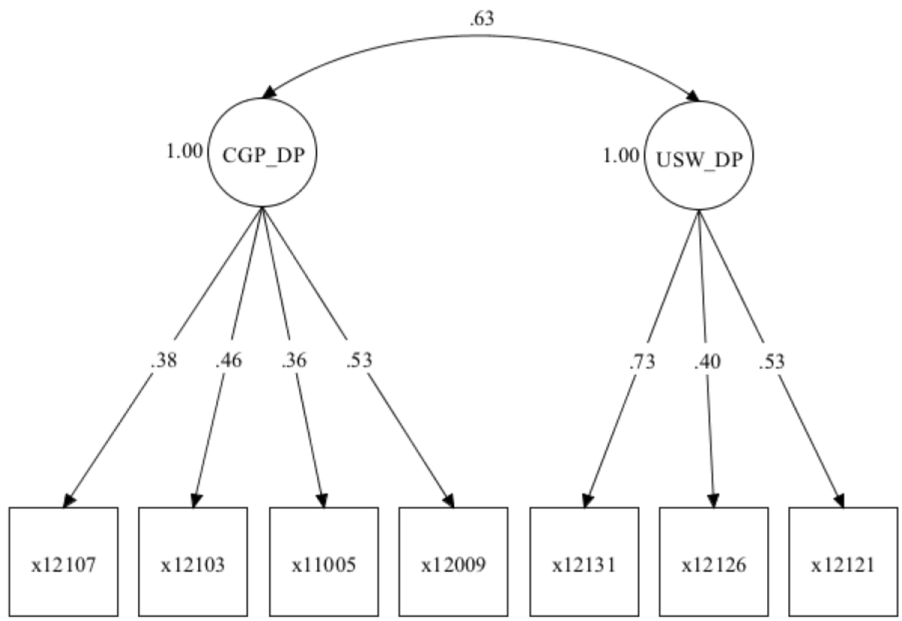


Figure 5.22. Two-dimensional model within the instructional situation DP

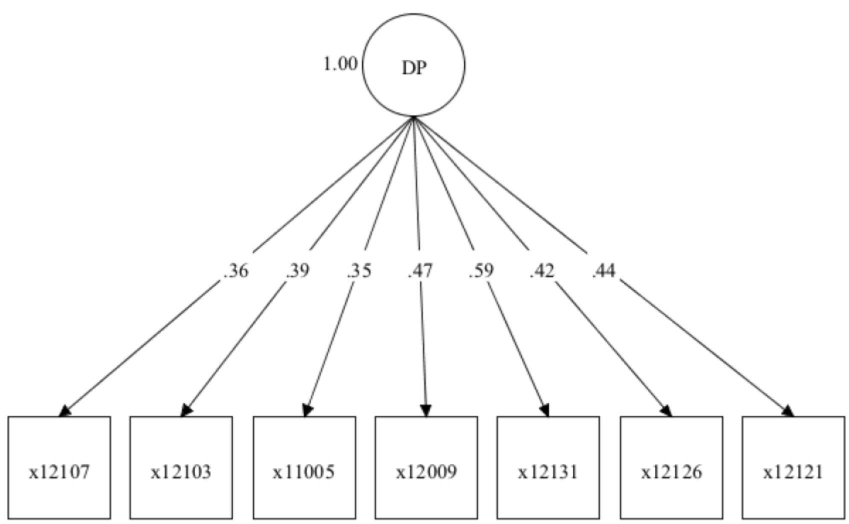


Figure 5.23. Unidimensional model within the instructional situation DP

Table 5.20 presents model fit statistics for each factor model as well as the result of comparison test (2-factor vs. 1-factor). As shown in the table, the 2-dimensional model shows acceptable fit statistics, whereas 1-dimensional model does not (CFI=0.904, TLI=0.857). The information statistics (AIC) also suggested that a 2-dimensional model better fits the data than a unidimensional model (note that lower information value indicates better fit).

Table 5.20. Comparison of fit among IFAs within DP

	SEM-based				MIRT-based				
Model	RMSEA	CFI	TLI	Deviance (-2log likelihood) statistics	Number of free parameters	Deviance change	df change	p	AIC
2-Dimension	0.039	0.946	0.912	5175.41	18				5211.41
1-Dimension	0.050	0.904	0.857	5186.24	17	10.83	1	0.001	5220.24

To examine the advantage of 2-factor model over 1-factor model at an item level, the factor loadings estimated from the 2-factor and 1-factor model under SEM- and MIRT-based models are presented, respectively in Table 5.21.

Table 5.21. Estimated standardized factor loadings within DP

		SEM-based		IRT-based	
Item	Dimension	1	2	1	2
X12121	USW_DP	0.50 (0.07)	0.55 (0.08)	0.44 (0.08)	0.53 (0.08)
X12126		0.46 (0.08)	0.47 (0.08)	0.42 (0.08)	0.40 (0.08)
X12131		0.65 (0.07)	0.74 (0.09)	0.59 (0.09)	0.73 (0.09)
X12009_b	CGP_DP	0.49 (0.06)	0.56 (0.07)	0.47 (0.07)	0.53 (0.08)
X11005		0.38 (0.08)	0.41 (0.08)	0.35 (0.08)	0.37 (0.08)
X12103		0.42 (0.07)	0.48 (0.08)	0.39 (0.07)	0.46 (0.08)
X12107		0.38 (0.08)	0.41 (0.08)	0.36 (0.07)	0.38 (0.08)

All item loadings are significant at the $p < 0.05$ level
Standardized loading (standard error)

The item loadings for each model suggested that the DP items explain more variance when they are distinguished by the task of teaching than when they are loaded on one factor of DP, given higher item loadings on 2-dimensional model than 1-dimensional model under both SEM and IRT based modeling.

In summary, tested measurement models supported my hypothesis that the items within the same instructional situation of doing proofs (DP) could be distinguished by different tasks of teaching – USW and CGP. Moreover, the results were consistent under both SEM and IRT framework. In the next section, the same hypothesis on the structure of USW items was investigated under an assumption that teachers' knowledge level is discrete (DCM model).

5.2.2.2 Dimensionality analysis of the items within DP using DCM

The two-dimensional structure of the DP items categorized by two tasks of teaching (USW and CGP) was estimated by a LCDM model where teachers' knowledge is assumed to be a discrete construct. Compared to the previous analysis which examined the distinction between the items of USW and CGP within the situation of calculation in geometry (CG), this section reports the LCDM analysis examining the distinction between the USW and CGP items within the situation of doing proofs (DP). Using the same 2-dimensional model, which showed a good fit under the IFA, the LCDM model was set to estimate the main effects of items on each of the two knowledge attributes of DP distinguished by tasks of teaching. The Q-matrix and the equation demonstrating the model parameters are presented in Table 5.22 and Equation (4), respectively.

Table 5.22. Q-matrix for items within DP

Item	Attribute 1: USW_DP	Attribute 2: CGP_DP
X12121, X12126, X12131	1	0
X12009_b, X11005, X12103, X12107	0	1

Items with _b are dichotomized (originally ordinal items) responses

The equation estimating parameters of the DP items is represented as follows:

$$\ln\left(\frac{P(X_i = 1|\alpha_{ra})}{P(X_i = 0|\alpha_{ra})}\right) = \lambda_{i,0} + \lambda_{i,1,(a)}\alpha_{ra} \text{ , where } a = \text{USW\_DP or CGP\_DP} \tag{4}$$

depending on the attribute item *i* measures

For each item, the equation (4) has one intercept and one main effect. Given that there are two hypothesized dimensions within DP, a respondent's profile is expressed as a vector with two elements α_r=[$\alpha_{rUSW-DP}, \alpha_{rCGP-DP}$], each of which has 1 when a respondent has mastered the knowledge attribute. Thus, the main effect parameter of an attribute can be included in the calculation of the logit, only if a respondent's mastery profile has "1" for that attribute.

The parameter $\lambda_{i,0}$ indicates the logit of a correct response (to item *i*) for nonmasters of the attribute ($\alpha_{rUSW-DP} = 0 \ or \ \alpha_{rCGP-DP} = 0$). The parameter $\lambda_{i,1,(a)}$ is the main effect (indicated by the second subscript) for USW_DP (or CGP_DP). It represents the increase in the logit of a correct response (to item *i*) for mastering USW_DP (or CGP_DP) for a respondent (a teacher in this study) who has not mastered the attribute.

5.2.2.2.1 Model fit

The fit of the LCDM model with DP items was evaluated by calculating a bivariate goodness of fit statistic with a $\chi^2(1)$ distribution for each pair of 7 DP items (Rupp et al., 2010). The result suggested that all the item pairs except one (between x11005 and x12126) showed a good model fit (i.e., chi-square value is insignificant) when chi-square values are evaluated at a 0.05 significance value. Thus, the LCDM model estimating teachers' knowledge in the situation of DP across two knowledge attributes was considered as acceptable and the item parameters estimated from the model were further interpreted.

5.2.2.2.2 Item parameter estimates (2-attribute model)

The item parameters estimated from the LCDM for DP items are listed in Table 5.23. As shown in the table, the average intercept across items is -1.03, indicating the average predicted logit of a correct response for teachers who had not mastered any of the attributes. This means that approximately 0.26% (= $\frac{e^{-1.03}}{(1+e^{-1.03})}$) of teachers who had not mastered any of the two DP knowledge attributes answered items correctly. The average main effects are 2.29 and 1.67 for USW_DP and CGP_DP, respectively. This means that the increase in the logit of a correct response by mastering USW_DP is 2.20, which correspond to odds ratio of 9.02, and the increase in the logit of a correct response by mastering CGP_DP is 1.67, which correspond to odds ratio of 5.31. The main effect size of each item on each knowledge attribute was evaluated using the criterion for odds ratio suggested by Bradshaw et al. (2014, p. 6), with small effect sizes for odd ratios between 1.44 and 2.47, medium effect sizes for odd ratios between 2.47 and 4.25, and large effect sizes for odd ratios larger than 4.25. According to these suggested thresholds, 2 items had medium effects and the remaining 5 items had large effect sizes. This result suggests that all the items measured the intended attributes with significant effect sizes.

Table 5.23. Estimated item parameters (logit) within DP

Item	Intercept $\lambda_{i,0}$	USW_DP $\lambda_{i,1,(1)}$	CGP_DP $\lambda_{i,1,(2)}$
<u>USW_DP</u>			
X12121	-1.51 (0.59)	2.39 (0.49)	
X12126	-1.33 (0.24)	1.41 (0.34)	
X12131	-0.36 (0.38)	3.06 (0.84)	
<u>CGP_DP</u>			
X12009_b	-1.13 (0.31)		1.96 (0.46)
X11005	-1.34 (0.31)		1.48 (0.36)
X12103	-0.65 (0.32)		1.84 (0.43)
X12107	-0.86 (0.19)		1.40 (0.49)
Average	**-1.03 (0.33)**	**2.29 (0.56)**	**1.67 (0.43)**

Standard errors in parentheses

5.2.2.2.3 Probability of a correct response between nonmasters and masters

The strength of associations between measured attributes and each item were examined by using item characteristics bar charts (ICBCs). The ICBCs for each DP knowledge attributes are displayed in Figure 5.24 and Figure 5.25. As shown in these figures, there were significant differences in the probability of getting a correct response between nonmasters and masters for all 7 DP items.

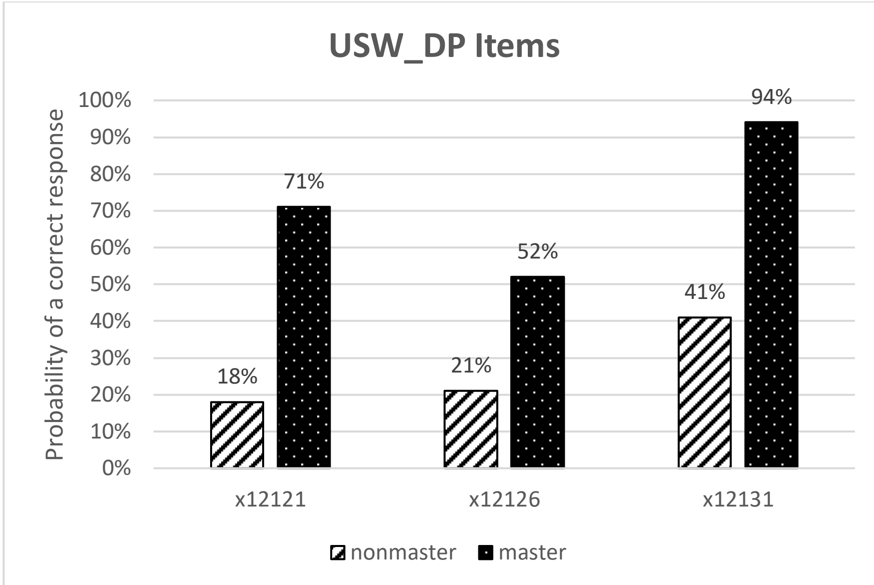


Figure 5.24. Item characteristic bar chart for USW_DP items

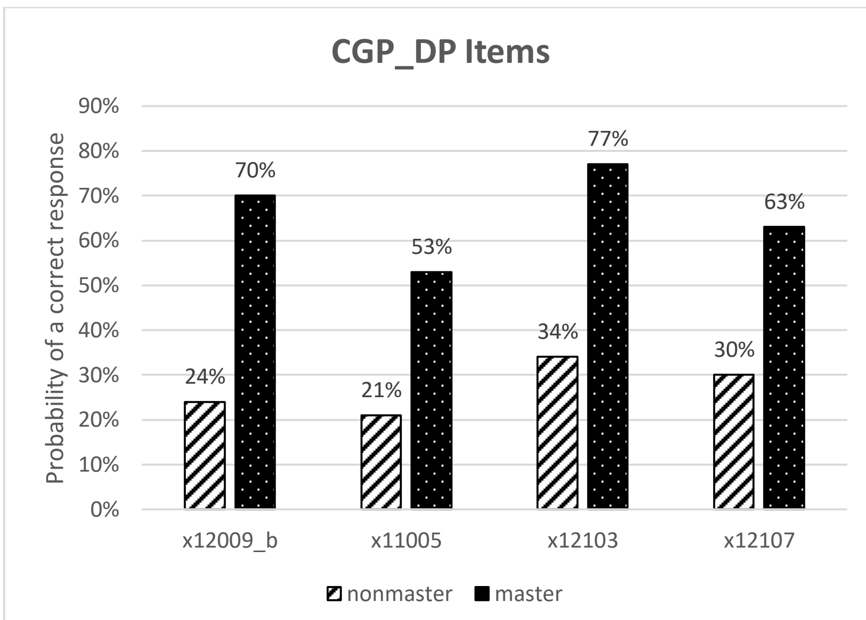


Figure 5.25. Item characteristic bar chart for CGP_DP items

5.2.2.2.4 Model comparison

Similar to the test conducted under MIRT framework, the 2-attribute LCDM model was compared to the 1-attribute LCDM model (where the attributes USW_DP and CGP_DP combine together). The result was consistent with that of IFA in that the 2-attribute model was significantly better than the 1-attribute model (($\chi^2(2) = 11.734$, $p=0.003$). Other fit statistics, such as AIC, also suggested better fit of the 2-attribute model (3818.606) than the 1-attribute model (3826.341).

In summary, tested measurement models supported my hypothesis that multidimensional models fit the item responses better than a unidimensional model within the instructional situation of geometric calculation (DP). The results supporting the hypothesis were also consistent across three different approaches: SEM, MIRT, and DCM. The items within the same situation of DP could be distinguished by different tasks of teaching – understanding students' work (USW) and choosing the givens for a problem (CGP). However, there may be a need for more items to more reliably measure two knowledge dimensions in a wider range of thetas (discussed in Chapter 6).

5.2.3 Organization of the items within the instructional situation EF

The items that involved the respondents in the instructional situation exploring a Figure in geometry (EF) were classified into two separate tasks of teaching, as shown in Table 5.24. One category is the set of items measuring teachers' mathematical knowledge in doing the task of understanding students' work (USW_EF) and another is the set of items measuring teachers' mathematical knowledge in doing the task of choosing appropriate givens for a problem (CGP_EF).

Table 5.24. Items in the task EF

Task of teaching	Instructional situation	Knowledge dimension/attribute	# items
Exploring a Figure in geometry (EF)	Understanding Students' Work (USW)	USW_EF	3
	Choosing appropriate Givens for a Problem (CGP)	CGP_EF	4

With these two sets of items, this section examines whether the items measuring teachers' knowledge used in the situation of EF can be distinguished by two different tasks of teaching. The analyses were conducted using IFA and DCM models, and the IFA was implemented under SEM-based and IRT-based framework.

5.2.3.1 Dimensionality of the items within EF using IFAs

Confirmatory item factor analyses were conducted under SEM and IRT based framework to examine whether the items commonly measuring teachers' mathematical knowledge used in the instructional situation EF can be distinguished in terms of the task of teaching. The result suggested that the factors USW_EF and CGP_EF were highly correlated as 0.928 (Figure 5.26). This high correlation between USW_EF and CGP_EF led me to compare the 2-factor model and 1-factor model where USW_EF and CGP_EF are combined (Figure 5.27).

The comparison between the 2-factor model (Figure 5.26) and the unidimensional model (Figure 5.27) using DIFFTEST and likelihood ratio test revealed that the two-factor model is not significantly better than the one-factor model. (DIFFTEST: χ^2 =0.635, df =1, p =0.4254; LR test: χ^2 =0.44, df =1, p =0.5071).

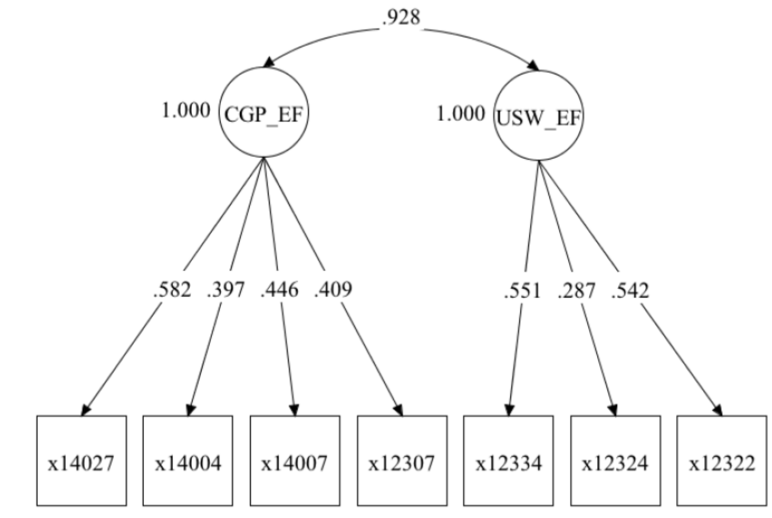


Figure 5.26. Two-dimensional model within EF

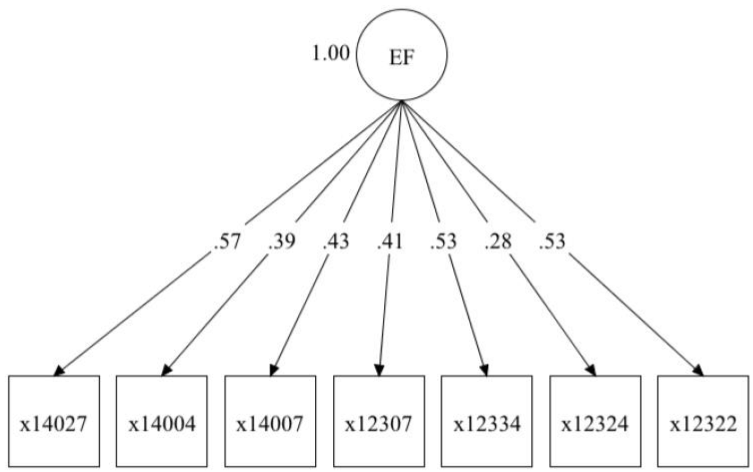


Figure 5.27. Unidimensional model within EF

Table 5.25 presents model fit statistics for each factor model as well as the result of comparison test (2-factor vs. 1-factor).

Table 5.25. Comparison of fit among IFAs within EF

	SEM-based			MIRT-based					
Model	RMSEA	CFI	TLI	Deviance (-2log likelihood) statistics	Number of free paramet-ers	Deviance-change	df change	p	AIC
2-Dimension	0.035	0.961	0.936	5433.87	18				5469.87
1-Dimension	0.033	0.963	0.944	5434.31	17	0.44	1	0.507	5468.31

Table 5.26 presents model fit statistics for each factor model as well as the result of comparison test (2-factor vs. 1-factor).

Table 5.26. Estimated standardized factor loadings within EF

		SEM-based		IRT-based	
Item	Dimension	1	2	1	2
X12322	USW_EF	0.56 (0.07)	0.58 (0.07)	0.53 (0.07)	0.54 (0.08)
X12324		0.31 (0.07)	0.31 (0.07)	0.28 (0.06)	0.29 (0.07)
X12334		0.56 (0.06)	0.58 (0.06)	0.53 (0.06)	0.55 (0.07)
X12307	CGP_EF	0.44 (0.07)	0.44 (0.08)	0.41 (0.08)	0.41 (0.08)
X14007		0.48 (0.07)	0.49 (0.07)	0.43 (0.07)	0.45 (0.08)
X14004		0.43 (0.07)	0.44 (0.08)	0.39 (0.07)	0.40 (0.08)
X14027		0.59 (0.07)	0.61 (0.07)	0.57 (0.07)	0.58 (0.08)

All item loadings are significant at the $p < 0.05$ level
Standardized loading (standard error)

In summary, the result was different from my initial hypothesis that the items within the same instructional situation of Exploring a Figure (DP) could be distinguished by different tasks of teaching – USW and CGP. My conjecture about this result is described in Appendix. In the next section, the same hypothesis on the structure of EF items was investigated under an assumption that teachers' knowledge level is discrete (DCM modeling).

5.2.3.2 Dimensionality of the items within EF using DCM

A LCDM was retrofitted to the items measuring teachers' knowledge used in the instructional situation EF (USW_EF, CGP_EF). Following the same 2-factor structure tested in IFA, the LCDM model was set to estimate the main effects of items on each of the two knowledge attributes distinguished by tasks of teaching within the instructional situation EF. The specified Q-matrix (Table 5.27) and the equation demonstrating the model parameters (Equation 5) are as follows.

Table 5.27. Q-matric for the items within EF

Item	Attribute 1: USW_EF	Attribute 2: CGP_EF
X12322, X12324_b X12334_b	1	0
X12307, X14004, X14007, X14027	0	1

Items with _b are dichotomized (originally ordinal items) responses

The equation estimating parameters of the EF items can be represented as follows:

$$\ln\left(\frac{P(X_i = 1|\alpha_{ra})}{P(X_i = 0|\alpha_{ra})}\right) = \lambda_{i,0} + \lambda_{i,1,(a)}\alpha_{ra}$$, where *a*=USW_EF or CGP_EF (5)

depending on the attribute item *i* measures

For each item, the equation (5) has one intercept and one main effect. Given that there are two hypothesized dimensions within EF, a respondent's profile is expressed as a vector with two elements α_r=[$\alpha_{rUSW-EF}, \alpha_{rCGP-EF}$], each of which has 1 when a respondent has mastered the knowledge attribute. Thus, the main effect parameter of an attribute can be included in the calculation of the logit, only if a respondent's mastery profile has "1" for that attribute.

The parameter $\lambda_{i,0}$ indicates the logit for non-masters of the attribute ($\alpha_{rUSW-EF} = 0 \; or \; \alpha_{rCGP-EF} = 0$). The parameter $\lambda_{i,1,(a)}$ is the main effect (indicated by the second subscript) for USW_EF (or CGP_EF). It represents the increase in the logit of a correct response for mastering USW_EF (or CGP_EF) for a respondent (a teacher in this study) who has not mastered the attribute.

5.2.3.2.1 Model fit

The fit of the LCDM model with EF items was evaluated by calculating a bivariate goodness of fit statistic with a $\chi^2(1)$ distribution for each pair of seven EF items (Rupp et al., 2010). The result suggested that all the item pairs except one (between x12322 and x14027) showed a good model fit (i.e., chi-square value is insignificant) when chi-square values are evaluated at a 0.05 significance value. Thus, the LCDM model estimating teachers' knowledge in the situation of EF across two knowledge attributes was considered as acceptable and the item parameters estimated from the model were further interpreted.

5.2.3.2.2 Item parameter estimates

The item parameters estimated from the LCDM for EF items are listed in Table 5.28. As shown in the table, the average intercept across items is -0.73, indicating the average predicted logit of a correct response for teachers who had not mastered any of the attributes. This means that approximately 0.32% (= $\frac{e^{-0.73}}{(1+e^{-0.73})}$) of teachers who had not mastered any of the two EF knowledge attributes answered items correctly. The average main effect is 2.06 and 1.76 for USW_EF and CGP_EF, respectively. This means that the increase in the logit of a correct response by mastering USW_EF is 2.96, which corresponds to odds ratio of 7.84, and the increase in the logit of a correct response by mastering CGP_EF is 1.76, which corresponds to odds ratio of 5.81. The main effect size of each item on each knowledge attribute was evaluated using the criterion for odds ratio suggested by Bradshaw et al. (2014, p. 6), with small effect sizes for odd ratios between 1.44 and 2.47, medium effect sizes for odd ratios between 2.47 and 4.25, and large effect sizes for odd ratios larger than 4.25). According to these suggested thresholds, 2 items had medium effects and the remaining 5 items had large effect sizes. This result suggests that all the items measured the intended attributes with significant effect sizes.

Table 5.28. Estimated item parameters (logit) within EF

Item	Intercept $\lambda_{i,0}$	USW_EF $\lambda_{i,1,(1)}$	CGP_EF $\lambda_{i,1,(2)}$
USW_EF			
X12322	-0.69 (0.25)	1.79 (0.37)	
X12324_b	-1.12 (0.25)	1.26 (0.32)	
X12334_b	-0.62 (0.41)	3.13 (0.72)	
CGP_EF			
X12307	0.21 (0.18)		1.60 (0.39)
X14004	-1.41 (0.27)		1.27 (0.34)
X14007	-0.59 (0.21)		1.77 (0.35)
X14027	-0.91 (0.24)		2.40 (0.44)
Average	**-0.73 (0.26)**	**2.06 (0.47)**	**1.76 (0.38)**

Standard errors in parentheses

5.2.3.2.3 Probability of a correct response between nonmasters and masters

The strength of associations between measured attributes and each item were examined by using item characteristics bar charts (ICBCs). The ICBCs for each EF knowledge attributes are displayed in Figure 5.28 and Figure 5.29. As shown in these figures, there were significant differences in the probability of getting a correct response between nonmasters and masters for all seven EF items.

5.2.3.2.4 Model comparison

Similar to the test conducted under MIRT framework, the 2-attribute LCDM model was compared to the 1-attribute LCDM model (where the attributes USW_EF and CGP_EF combine together). The result was consistent with that of IFA in that the 2-attribute model was not significantly better than the 1- attribute model (($\chi^2(2) = 2.842$, p=0.24). The other fit statistics such as AIC also suggested better fit of the 1- attribute model (3629.922) than the 2- attribute model (3631.08).

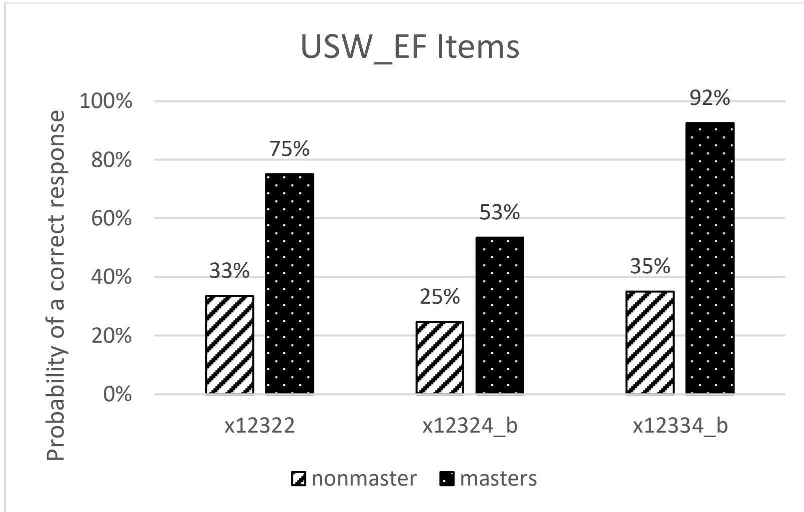


Figure 5.28. Item characteristic bar chart for USW_EF items

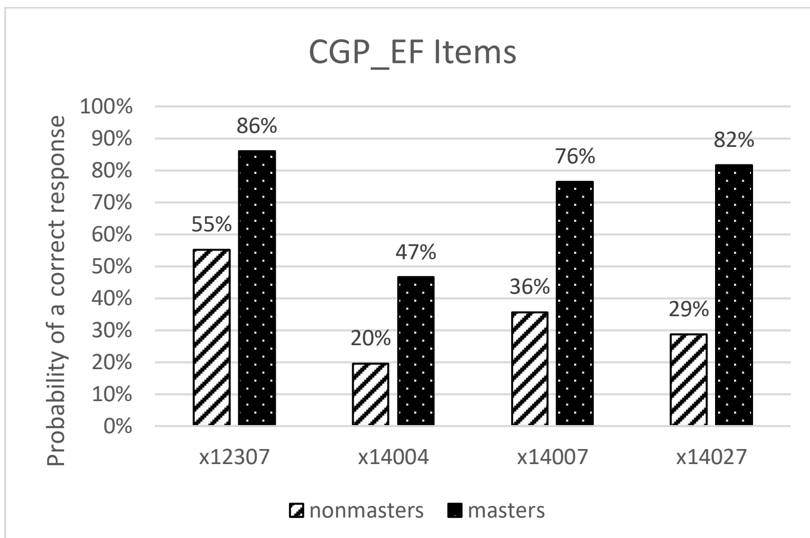


Figure 5.29. Item characteristic bar chart for CGP_EF items

In summary, for the instructional situation of exploring a figure (EF), tested measurement models did not support my hypothesis that the items representing the situation EF can be distinguished by task of teaching – USW_EF and CGP_EF. The results were also consistent across three different approaches: SEM, MIRT, and DCM. The conjectured reason for the non-distinction between USW and CGP items within EF is discussed in Chapter 6.

5.3 Dimensionality within the same course of studies

In the following sections, I report the results of the two dimensionality analyses, each of which was conducted within the same course of study. One is the model in which each dimension is characterized by one task of teaching and one instructional situation within the U.S. high school geometry course. Another is the model in which each dimension is characterized by one task of teaching and one instructional situation within the U.S. algebra 1 course.

5.3.1 Organization of the items within geometry

The items commonly measuring teacher' mathematical knowledge used in teaching U.S. high school geometry were sorted into two separate instructional situations, either CG or DP, and each set was further divided into two tasks of teaching, USW or CGP, as shown in Table 5.29. In particular, this section focuses on 1) whether the dimensions which were not distinguishable by one of the organizers (e.g., USW_CG and USW_DP) are distinguishable if they have different traits in both organizers (e.g., USW_CG and CGP_DP) and 2) whether the dimensions different in both organizers are less correlated each other than the dimensions different in one of the organizers. Note, considering the result shown in the previous section, the set of items characterized by the situation EF was not included in this analysis because its item set was not statistically distinguishable by different tasks of teaching (5.2.3). This may indicate that the items of EF used in this study do not capture a difference in teachers' knowledge between the task of USW and CGP in the situation EF.

In sum, this section examined the dimensionality among the items reflecting instructional situations (CG and DP), where problems are taught within customary ways of doing academic work in U. S. high school geometry[24] across two different tasks of teaching, USW and CGP.

Table 5.29. Items within the course of high school geometry

Instructional situation	Task of teaching	Knowledge dimension/attributes	# items
Calculation in Geometry (CG)	Understanding Students' work (USW)	USW_CG	6
	Choosing appropriate Givens for a Problem (CGP)	CGP_CG	5
Doing Proofs in geometry (DP)	Understanding Students' work (USW)	USW_DP	3
	Choosing appropriate Givens for a Problem (CGP)	CGP_DP	4

All of the four item sets were analyzed in one of the previous analyses conducted within one task of teaching or instructional situation, but they were not analyzed both across tasks of teaching and instructional situations simultaneously. This section reports the results analyzing the dimensionality across two different tasks of teaching (USW and CGP) and two instructional situations (CG and DP) using IFA and DCM models, and the IFAs were implemented under SEM-based and IRT-based modeling.

5.3.1.1 Dimensionality of the items within geometry using IFAs

To examine whether the two hypothesized dimensions characterized by different tasks of teaching and instructional situations can be distinguished within the same course of study (U.S. high school geometry), the 4-factor confirmatory item factor analyses were conducted under SEM and IRT based framework (Figure 5.30).

[24] American high school students often take a year-long course in geometry in their first or second year of high school. The topics covered, mostly within plane synthetic, transformational, and coordinate geometry and some solid geometry, can be found in commercial textbooks (Ko & Herbst, under review).

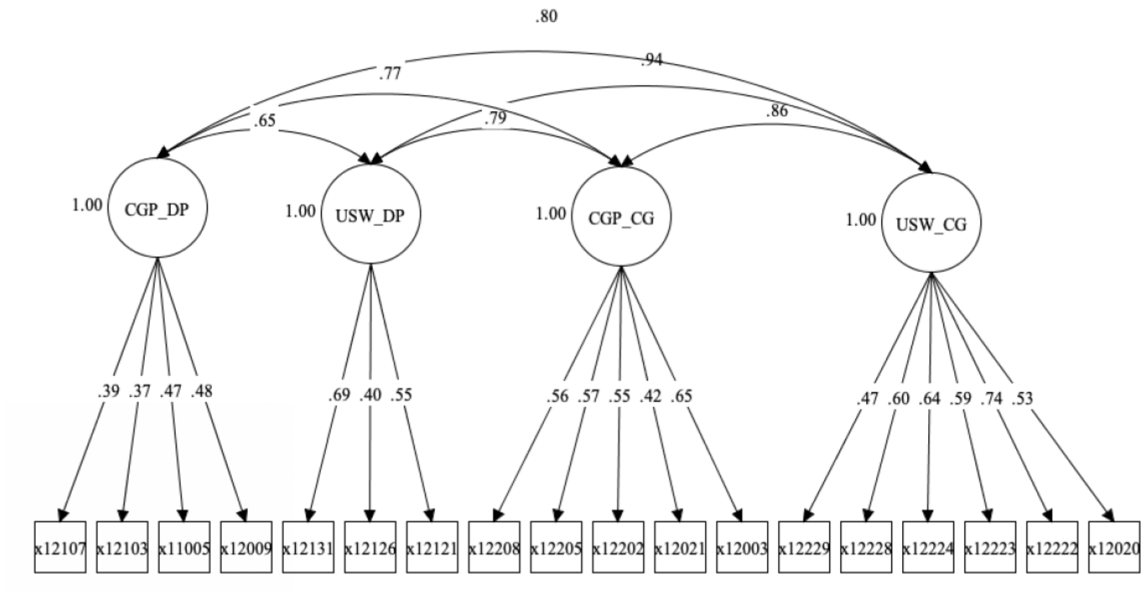


Figure 5.30. Four-dimensional model within U.S. high school geometry

Table 5.30 presents all the estimated pairwise correlations among the hypothesized knowledge dimensions in geometry. As shown in Table 5.30, the correlations between two factors which have different tasks of teaching and different instructional situations were significantly less than 1. For example, the factor of CGP_DP and USW_CG had a correlation 0.80 and the factor of USW_DP and CGP_CG had a correlation 0.79. This may indicate that items are distinguishable if they are measuring different tasks of teaching and different instructional situations, even though they are not distinguishable when having the same category regarding one of the organizers.

Table 5.30. Estimated correlations among dimensions within geometry

	Dimension			
Dimension	CGP_DP	USW_DP	CGP_CG	USW_CG
CGP_DP				
USW_DP	0.65			
CGP_CG	0.77	**0.79**		
USW_CG	**0.80**	0.94	0.86	

The factor loadings estimated by the 4-factor model under SEM- and MIRT-based models are presented in Table 5.31. All the factor loadings were significant at 0.01 level in both SEM and MIRT modeling. SEM-based model also yielded good fit statistics (RMSEA=0.017; CFI=0.990; TLI=0.988).

Table 5.31. Estimated standardized factor loadings within geometry

	SEM-based 4-factor	IRT-based 4-factor
USW_CG		
X12020	0.58 (0.05)	0.53 (0.05)
X12222	0.74 (0.04)	0.74 (0.05)
X12223	0.64 (0.05)	0.59 (0.05)
X12224	0.66 (0.04)	0.64 (0.04)
X12228	0.62 (0.05)	0.60 (0.05)
X12229	0.52 (0.06)	0.48 (0.06)
CGP_CG		
X12003	0.65 (0.04)	0.65 (0.05)
X12021	0.48 (0.06)	0.42 (0.06)
X12202	0.57 (0.05)	0.55 (0.05)
X12205	0.61 (0.06)	0.58 (0.06)
X12208	0.59 (0.06)	0.56 (0.06)
USW_DP		
X12121	0.59 (0.05)	0.55 (0.06)
X12126	0.46 (0.06)	0.40 (0.06)
X12131	0.72 (0.06)	0.69 (0.07)
CGP_DP		
X12009	0.48 (0.05)	0.48 (0.06)
X11005	0.55 (0.07)	0.47 (0.07)
X12103	0.37 (0.07)	0.37 (0.07)
X12107	0.44 (0.07)	0.39 (0.07)

All item loadings are significant at the $p < 0.01$ level
Standardized loading (standard error)

Model comparison tests were further conducted to examine whether the distinction is statistically significant for each pair of dimensions different in both task of teaching and instructional situation. Table 5.32 presents model fit statistics for each factor model as well as the result of comparison test (4-factor vs. 3-factor). As shown in the table, the 4-dimensional model better fits the data than the 3-dimensional models, where CGP_DP and USW_CG or USW_DP and CGP_CG are combined.

Table 5.32. Comparison of tit among multidimensional models within geometry

Model	SEM-based			MIRT-based					
	RMSEA	CFI	TLI	Deviance (-2log likelihood) statistics	Number of free parameters	Deviance change	df change	p	AIC
4-Dimension (All separate)	0.017	0.990	0.988	12525.26	54				12633.26
3-Dimension (CGP_DP+USW_CG)	0.019	0.987	0.985	12535.22	51	9.96	3	**0.019**	12637.22
3-Dimension (USW_DP+CGP_CG)	0.020	0.985	0.982	12538.14	51	12.88	3	**0.005**	12640.14

In summary, tested measurement models supported my hypothesis that the items within the same course of study (geometry) could be distinguished if they are different in both instructional situations and tasks of teaching. The results were consistent under both SEM and IRT based framework.

However, different from what I hypothesized, the pairs of dimensions were not necessarily less correlated when they have different tasks of teaching and different instructional situations simultaneously than when they have either the same task of teaching or the same instructional situation. This implies that the degree of distinction is dependent on the characteristics of each task of teaching or instructional situation (e.g., how similar the different tasks or situations are to each other). This point is discussed in more detail in Chapter 6. In the next section, the same hypothesis on the structure of geometry items was investigated under a DCM modeling

5.3.1.2 Dimensionality of the items within geometry using DCM

Following the same procedure, a LCDM model was applied to the items measuring teachers' mathematical knowledge for teaching high school geometry. The model was set to estimate the main effect of items on each of the four knowledge attributes characterized by one tasks of teaching and one instructional situation (USW_CG, CGP_CG, USW_DP, CGP_DP). The specified Q-matrix (Table 5.33) and the equation demonstrating the model parameters (Equation 6) are as follows.

Table 5.33. Q-matrix for the items within geometry

Items	Attribute 1: USW_CG	Attribute 2: CGP_CG	Attribute 3: USW_DP	Attribute 4: CGP_DP
X12020 X12222 X12223 X12224_b X12228_b X12229	1	0	0	0
X12003_b X12021 X12202_b X12205 X12208	0	1	0	0
X12121 X12126 X12131	0	0	1	0
X12009_b X11005 X12103 X12107	0	0	0	1

Items with _b are dichotomized (originally ordinal items) responses

Following the convention of notations, the equation estimating parameters of the geometry items can be represented as follows:

$$\ln\left(\frac{P(X_i = 1|\alpha_{ra})}{P(X_i = 0|\alpha_{ra})}\right) = \lambda_{i,0} + \lambda_{i,1,(a)}\alpha_{ra} \quad \text{, where } a\text{=USW\_CG, CGP\_CG,} \tag{6}$$

USW_DP, or CGP_DP, depending on the attribute item *i* measures

For each item, the equation (6) has one intercept and one main effect. Given that there are four hypothesized dimensions within geometry, a respondent's profile is expressed as a vector with four elements α_r=[$\alpha_{rUSW-GC}, \alpha_{rCGP-DP,}\alpha_{rUSW-DP,}\alpha_{rCGP-DP}$], each of which has 1 when a respondent has mastered the knowledge attribute. Thus, the main effect parameter of an attribute can be included in the calculation of the logit, only if a respondent's mastery profile has "1" for that attribute. The parameter $\lambda_{i,0}$ is the intercept, indicating the logit for non-masters of the attribute (USW_CG, CGP_CG, USW_DP, or CGP_DP) ($\alpha_{rUSW-GC} = 0, \alpha_{rCGP-GC} = 0, \alpha_{rUSW-DP} = 0, or\ \alpha_{rCGP-DP} = 0$). The parameter $\lambda_{i,1,(a)}$ is the main effect (indicated by the second subscript), indicating the increase in the logit for mastering the attribute for a respondent (a teacher in this study) who has not mastered the attribute.

5.3.1.2.1 Model fit

The fit of the LCDM model with CG items was evaluated by calculating a bivariate goodness of fit statistic with a $\chi^2(1)$ distribution for each pair of 18 geometry items (Rupp et al., 2010). The result suggested that 93% (142 pairs) of the 153 (=$\frac{18*17}{2}$) item pairs showed good model fit (i.e., chi-square value is insignificant) when chi-square values are evaluated at a 0.05 significance value and 2 item pairs (between x12107 and x12223, and between x12208 and x12223) showed misfit at a 0.01 significance level. Overall, my LCDM model estimating teachers' knowledge for teaching high school geometry across four knowledge attributes provided an acceptable model fit for the data used in this study. Therefore, the item parameters estimated from the model were further interpreted as follows.

5.3.1.2.2 Item parameter estimates

The item parameters estimated from the LCDM are listed in Table 5.34. As shown in the table, the average intercept across items is -0.85, indicating the average predicted logit of a correct response for teachers who had not mastered any of the attributes. This means that approximately 0.30% (= $\frac{e^{-0.85}}{(1+e^{-0.85})}$) of teachers who had not mastered any of the four geometry knowledge attributes answered the items correctly. The average main effect ranged from 1.60 to 2.40. This means that the increase in the logit of a correct response by mastering one of the attributes ranged from 1.60 to 2.40, which correspond to 4.95 to 11.02 in terms of odds ratio. The main effect size of each item on each knowledge attribute was evaluated using the criterion for odds ratio suggested by Bradshaw et al. (2014, p. 6), with small effect sizes for odd ratios between 1.44 and 2.47, medium effect sizes for odd ratios between 2.47 and 4.25, and large effect sizes for odd ratios larger than 4.25). According to this suggested thresholds, three items had medium effect sizes (x12126, x12103, and x12107) and the remaining 15 items had large effect sizes. This result suggests that all the items measured the intended attributes with significant effect sizes.

Table 5.34. Estimated item parameters within geometry

Item	Intercept $\lambda_{i,0}$	USW_CG $\lambda_{i,1,(1)}$	CGP_CG $\lambda_{i,1,(2)}$	USW_DP $\lambda_{i,1,(3)}$	CGP_DP $\lambda_{i,1,(4)}$
x12020	-0.91 (0.23)	1.94 (0.28)			
x12222	0.00 (0.20)	3.52 (0.72)			
x12223	-0.50 (0.20)	2.17 (0.30)			
x12229	-0.02 (0.18)	1.66 (0.28)			
x12224_b	-0.51 (0.22)	2.56 (0.33)			
x12228_b	-1.40 (0.29)	2.51 (0.32)			
x12021	-0.39 (0.20)		1.63 (0.27)		
x12205	-0.24 (0.21)		2.24 (0.36)		
x12208	0.44 (0.23)		2.34 (0.42)		
x12003_b	-2.86 (0.61)		3.01 (0.57)		
x12202_b	-1.38 (0.25)		1.95 (0.31)		
x12121	-1.64 (0.53)			2.37 (0.46)	
x12126	-1.38 (0.28)			1.36 (0.31)	
x12131	-0.53 (0.38)			2.86 (0.46)	
x11005	-1.59 (0.27)				1.91 (0.34)
x12103	-0.57 (0.18)				1.44 (0.35)
x12107	-0.87 (0.21)				1.41 (0.30)
x12009_b	-1.03 (0.17)				1.65 (0.41)
Average	**-0.85 (0.27)**	**2.40 (0.37)**	**2.23 (0.39)**	**2.20 (0.41)**	**1.60 (0.35)**

Standard errors in parentheses

5.3.1.2.3 Probability of a correct response between non masters and masters

Following the same approach as Bradshaw et al. (2014), the strength of associations between measured attributes and each item were examined by using item characteristics bar charts (ICBCs). The ICBCs for each knowledge attributes are displayed in Figure 5.31~Figure 5.34. By examining the difference in the probability of a correct response between nonmasters and masters, the ability of each item in distinguishing masters from nonmasters was evaluated. Overall, significant differences could be identified in the probability of getting a correct response between nonmasters and masters across items, with respect to the four geometry knowledge attributes, Again, each knowledge attribute was characterized by one task of teaching and one instructional situation.

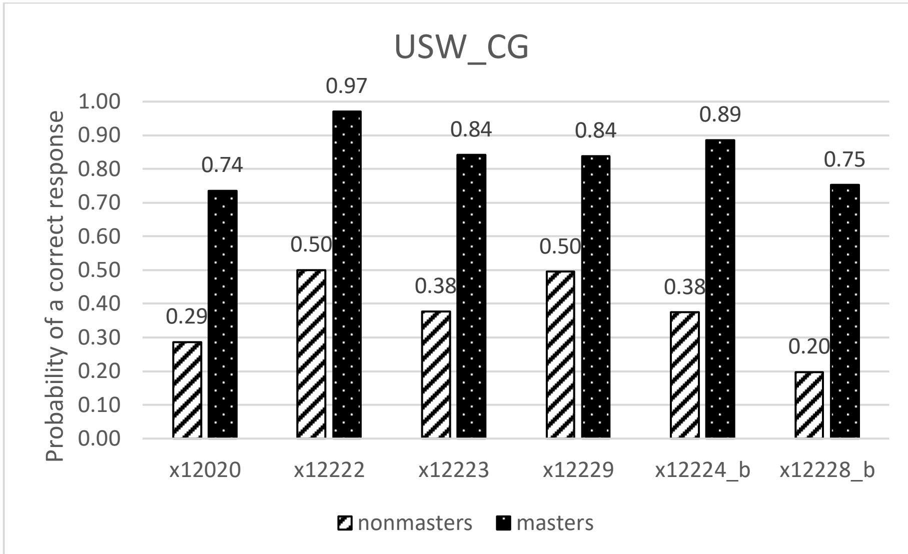


Figure 5.31. Item characteristic bar chart for USW_CG items

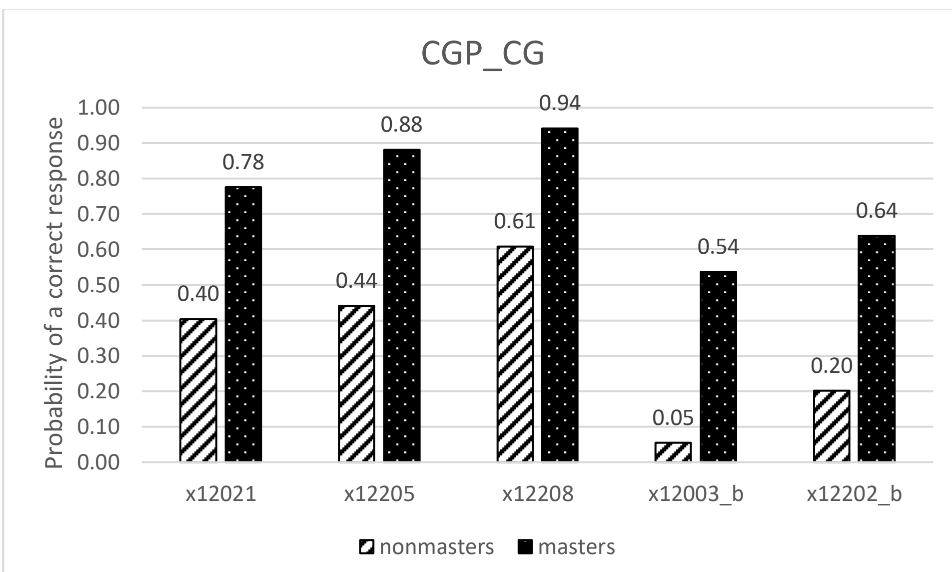


Figure 5.32. Item characteristic bar chart for CGP_CG items

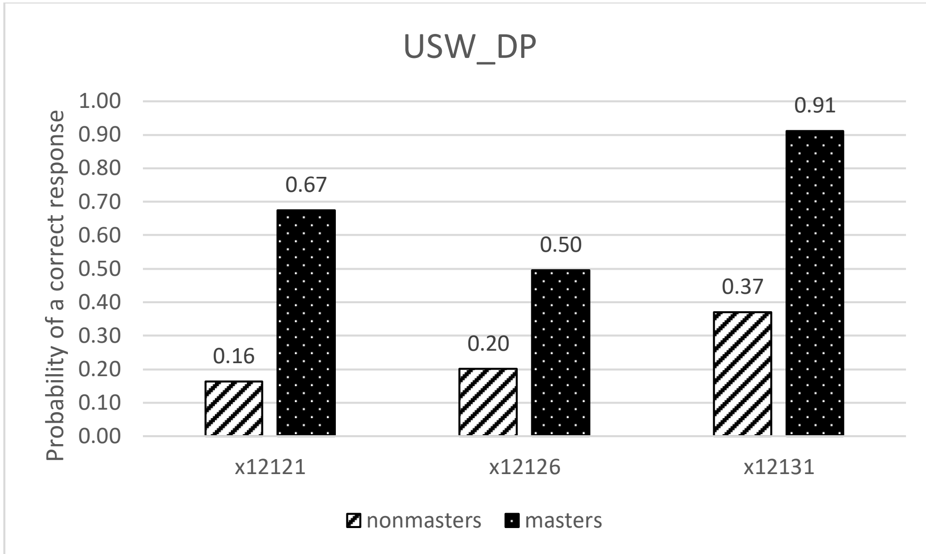


Figure 5.33. Item characteristic bar chart for USW_DP items

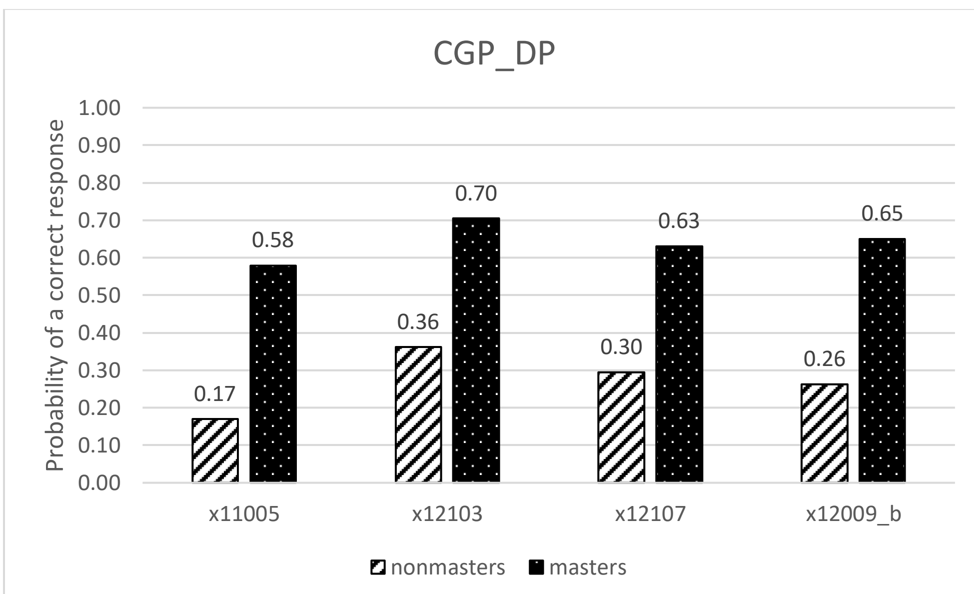


Figure 5.34. Item characteristic bar chart for CGP_DP items

Overall, ICBCs showed noticeable differences in the probability of a correct response between nonmasters and masters for all the CGP items.

5.3.1.2.4 Model comparison

Similar to the test conducted under MIRT framework, the 4-attribute LCDM model was compared to the 3-attribute LCDM models, where CGP_DP and USW_CG or USW_DP and CGP_CG are combined. The result of the distinction between USW_DP and CGP_CG was consistent with that of IFA in that the 4-attribute model was significantly better than the 3-attribute models (USW_DP and CGP_CG combined, $\chi^2(8) = 18.598$, $p=0.017$). However, the 3-factor LCDM model, where CGP_DP and USW_CG were combined, was not significantly worse than the 4-factor model ($\chi^2(8) = 14.566$, $p=0.068$). Nonetheless, considering the distinction is marginally significant and IFA models suggest separating CGP_DP and USW_CG, the distinction between CGP_DP and USW_CG may require further investigation with more items.

5.3.2 Organization of the items within algebra

The 11 items measuring teacher' mathematical knowledge commonly used in teaching algebra 1 were classified into three separate categories: items measuring 1) teachers' knowledge needed for understanding students' work in the situation of simplifying rational expressions (USW_SR); 2) teachers' knowledge needed for understanding students' work in the situation of solving a linear or quadratic equation (USW_SE); 3) teachers' knowledge needed for choosing the givens for a problem in a situation of calculating with numbers (CGP_CN) (Table 5.35).

Table 5.35. Items within algebra

Task of teaching	Instructional situation	Knowledge dimension/ attribute	# items
Understanding Students' Work (USW)	Simplifying Rational expressions in algebra (SR)	USW_SR	3
Understanding Students' Work (USW)	Solving a linear or quadratic Equation in algebra (SE)	USW_SE	5
Choosing appropriate Givens for a Problem (CGP)	Calculation with Numbers in algebra (CN)	CGP_CN	3

5.3.2.1 Dimensionality of the items within algebra using IFAs

To examine whether the hypothesized dimensions characterized by different tasks of teaching and different instructional situations can be distinguished within the same course of study (U.S. algebra 1), three-factor confirmatory item factor analyses were conducted under SEM and IRT based framework (Figure 5.35).

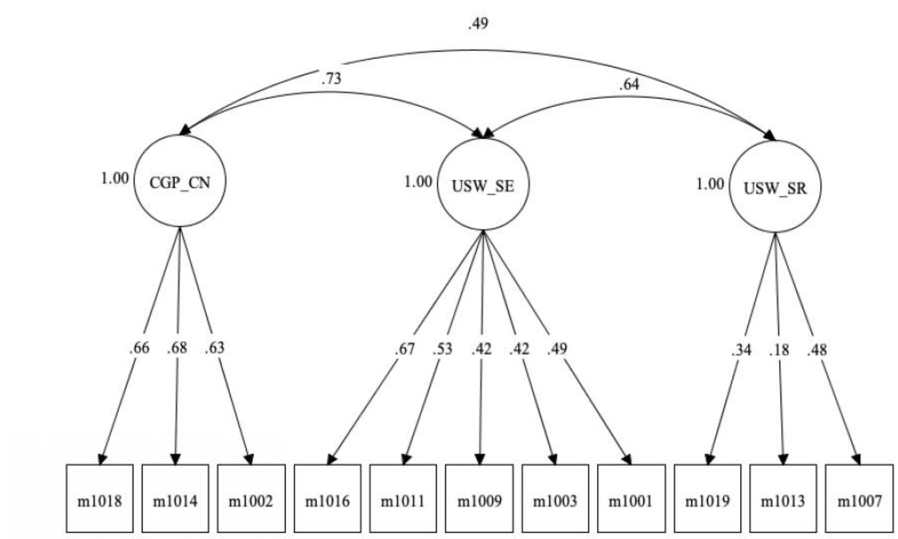


Figure 5.35. Three-dimensional model within algebra

As shown in Table 5.36, the correlations among three factors, which have different tasks of teaching and different instructional situations were significantly less than 1. For example, the factor of CGP_CN and USW_SE had a correlation 0.73 and the factor of CGP_CN and USW_SR had a correlation 0.49, indicating that these factors are statistically distinguishable from each other.

Table 5.36. Estimated correlations among dimensions within algebra

	Dimension		
Dimension	CGP_CN	USW_SE	USW_SR
CGP_CN			
USW_SE	0.73		
USW_SR	0.49	0.64	

The factor loadings estimated by the 3-factor model under SEM- and MIRT-based models are presented in Table 5.37. All the factor loadings were significant at 0.05 level in both SEM and MIRT modeling, and the SEM-based model yielded acceptable global fit statistics (RMSEA=0.024, CFI=0.968, TLI=0.957).

Table 5.37. Estimated standardized factor loadings within algebra

	SEM-based 3-factor	IRT-based 3-factor
USW_SR		
M1007	0.55 (0.14)	0.49 (0.12)
M1013	0.19 (0.10)	0.18 (0.09)
M1019	0.36 (0.10)	0.34 (0.10)
USW_SE		
M1001	0.51 (0.08)	0.49 (0.08)
M1003	0.46 (0.10)	0.42 (0.11)
M1009	0.45 (0.07)	0.42 (0.07)
M1011	0.55 (0.07)	0.53 (0.08)
M1016	0.69 (0.07)	0.67 (0.07)
CGP_CN		
M1002	0.70 (0.07)	0.63 (0.07)
M1014	0.68 (0.07)	0.68 (0.07)
M1018	0.67 (0.07)	0.66 (0.07)

All item loadings are significant at the $p < 0.01$ level
Standardized loading (standard error)

Model comparison tests were further conducted to examine whether the distinction is statistically significant for each pair of dimensions different in categories of task of teaching and instructional situation. Table 5.38 presents model fit statistics for each factor model as well as the result of comparison test (3-factor vs. 2-factor). As shown in the table, the 3-dimensional model better fits the data than the 2-dimensional models where USW_SE and CGP_CN or USW_SR and CGP_CN are combined.

Table 5.38. Comparison of tit among multidimensional models within algebra

	SEM-based			MIRT-based					
Model	RMSEA	CFI	TLI	Deviance (-2log likelihood) statistics	Number of free parameters	Deviancechange	df change	p	AIC
3-Dimension (All separate)	0.024	0.968	0.957	6224.768	28				
2-Dimension (USW_SE+CGP_CN)	0.030	0.946	0.930	6237.214	26	12.446	2	0.002	6284.21
2-Dimension (USW_SR+CGP_CN)	0.027	0.956	0.944	6232.528	26	7.76	2	0.021	6284.53

In summary, tested measurement models supported my hypothesis that the items within the same course of study (algebra 1) could be distinguished if they are different in both the categories of instructional situation and task of teaching. However, the pairs of dimensions, which involve not only having different tasks of teaching but also different instructional situations, were not necessarily less correlated than them, which involve having the same task of teaching or instructional situation. For example, the factors CGP_CN and USW_SE showed a higher correlation (0.73) than the factors USW_SE and USW_SR (0.64) (Figure 5.35). In addition, one of the USW_SR item M1013 yielded a low factor loading, indicating that the response pattern of the item is not highly correlated with that of other two items. Moreover, all the three USW_SR items showed relatively lower factor loadings (0.11 ~ 0.31) than other items when all the algebra items were loaded on one common factor of knowledge for teaching algebra 1.

This may imply that the items are not coherently measuring the knowledge of understanding students' work in simplifying rational expressions. This point is discussed in Chapter 6 in more detail.

5.3.2.2 Dimensionality of the items within algebra using DCM

Following the same procedure conducted for geometry items, a LCDM model was applied to the items measuring teachers' mathematical knowledge for teaching algebra 1. The model was set to estimate the main effect of items on each of the three knowledge attributes characterized by one tasks of teaching and one instructional situation (USW_SR, USW_SE, CGP_CN). The specified Q-matrix (Table 5.39) and the equation demonstrating the model parameters (Equation 7) are as follows.

Table 5.39. Q-matrix for the items within algebra

Items	Attribute 1: USW_SR	Attribute 2: USW_SE	Attribute 3: CGP_CN
M1007 M1013 M1019	1	0	0
M1001 M1003_b M1009 M1011_b M1016	0	1	0
M1002 M1014 M1018	0	0	1

Following the same notations used in the previous section, the equation estimating parameters of the algebra items can be represented as follows:

$$\ln\left(\frac{P(X_i = 1|\alpha_{ra})}{P(X_i = 0|\alpha_{ra})}\right) = \lambda_{i,0} + \lambda_{i,1,(a)}\alpha_{ra} \text{ , where a= USW\_SR, USW\_SE, or CGP\_CN, depending on the attribute item } i \text{ measures} \quad (7)$$

For each item, the equation (7) has one intercept and one main effect. Given that there are three hypothesized dimensions within algebra items, a respondent's profile is expressed as a vector with three elements α_r=[$\alpha_{rUSW\_SF}, \alpha_{rUSW-SE,}\alpha_{rCGP-SC}$], each of which has 1 when a respondent has mastered the knowledge attribute. Thus, the main effect parameter of an attribute can be included in the calculation of the logit, only if a respondent's mastery profile has "1" for that attribute. For example, the parameter $\lambda_{i,0}$ is the intercept, indicating the logit for non-masters of the targeted attribute ($\alpha_{rUSW\_SF}, = 0, \alpha_{rUSW-SE} = 0, or\ \alpha_{rCGP-SC} = 0$). The parameter $\lambda_{i,1,(a)}$ is the main effect (indicated by the second subscript) of the attribute, indicating the logit for mastering the attribute for a respondent (a teacher in this study) who has not mastered the attribute.

5.3.2.2.1 Model fit

The fit of the LCDM model within algebra 1 items was evaluated by calculating a bivariate goodness of fit statistic with a $\chi^2(1)$ distribution for each pair of 11 algebra 1 items (Rupp et al., 2010).

The result suggested that 95% (52 pairs) of the 55 (=$\frac{11*10}{2}$) item pairs showed good model fit (i.e., chi-square value is insignificant) when chi-square values are evaluated at a 0.05 significance value and 1 item pairs (between M1001 and M1011) showed misfit at a 0.01 significance level. Overall, my LCDM model estimating teachers' knowledge for teaching algebra 1 across three knowledge attributes provided an acceptable model fit for the data used in this study. Therefore, the item parameters estimated from the model were further interpreted as follows.

5.3.2.2.2 Item parameter estimates

The item parameters estimated from the LCDM are listed in Table 5.40. As shown in the table, the average intercept across items is -0.51, indicating the average predicted logit of a correct response for teachers who had not mastered any of the attributes. This means that approximately 0.38% (= $\frac{e^{-0.51}}{(1+e^{-0.51})}$) of teachers who had not mastered an attribute measured by an item answered the items correctly. The average main effect ranged from 1.84 to 2.68. This means that the increase in the logit of a correct response by mastering one of the attributes ranged from 1.84 to 2.68, which correspond to 6.28 to 14.53 in terms of odds ratio. The main effect size of each item on each knowledge attribute was evaluated using the criterion for odds ratio suggested by Bradshaw et al. (2014, p. 6), with small effect sizes for odd ratios between 1.44 and 2.47, medium effect sizes for odd ratios between 2.47 and 4.25, and large effect sizes for odd ratios larger than 4.25). According to this suggested thresholds, one item had small effect size (M1013) and the remaining 10 items had large effect sizes. This result suggests that all the items measured the intended attributes with significant effect sizes.

Table 5.40. Estimated item parameters (logit) within algebra

Item	Intercept $\lambda_{i,0}$	USW_SR $\lambda_{i,1,(1)}$	USW_SE $\lambda_{i,1,(2)}$	CGP_CN $\lambda_{i,1,(3)}$
M1007	-1.27 (0.67)	2.21 (0.69)		
M1013	-1.19 (0.38)	0.90 (0.44)		
M1019	-2.73 (1.21)	2.41 (1.23)		
M1001	0.56 (0.21)		1.56 (0.34)	
M1009	-0.25 (0.23)		1.65 (0.32)	
M1016	-2.01 (0.41)		2.88 (0.45)	
M1003_b	1.75 (0.25)		2.29 (0.82)	
M1011_b	0.47 (0.22)		1.59 (0.34)	
M1002	-0.25 (0.21)			2.81 (0.63)
M1014	-0.69 (0.25)			2.53 (0.42)
M1018	0.06 (0.21)			2.69 (0.57)
Average	**-0.51 (0.38)**	**1.84 (0.79)**	**1.99 (0.45)**	**2.68 (0.54)**

Standard errors in parentheses

5.3.2.2.3 Probability of a correct response between non masters and masters

Following the same approach of Bradshaw et al. (2014), the strength of associations between measured attributes and each item were examined by using item characteristics bar charts (ICBCs). The ICBCs for each knowledge attributes are displayed in Figure 5.36~ Figure 5.38. By examining the difference in the probability of a correct response between nonmasters and masters, the ability of each item in distinguishing masters from nonmasters was evaluated. Overall, significant differences could be identified in the probability of getting a correct response between nonmasters and masters for most items, with respect to the three algebra 1 knowledge attributes, Again, each knowledge attribute was characterized by one task of teaching and one instructional situation.

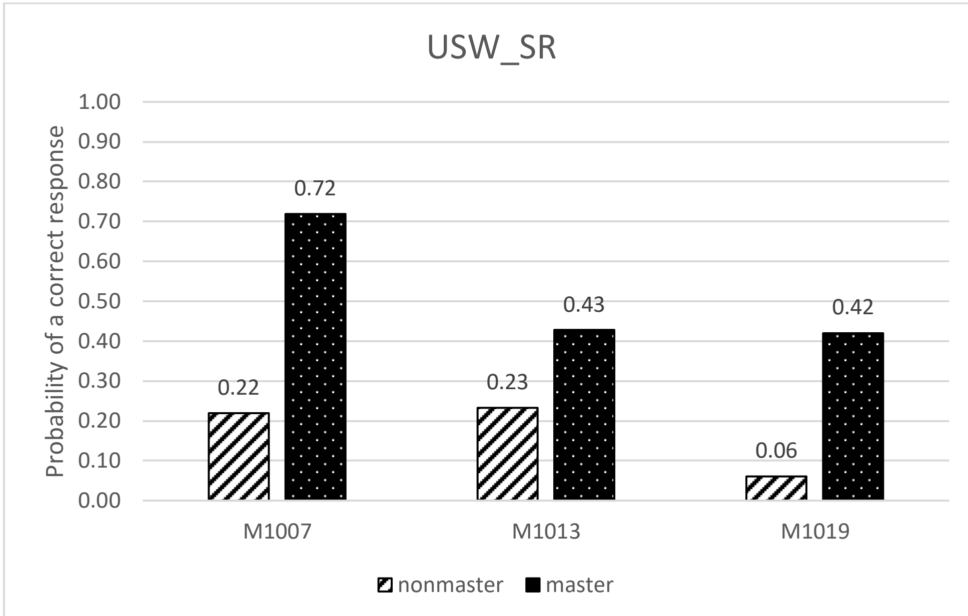


Figure 5.36. Item characteristic bar chart for USW_SR items

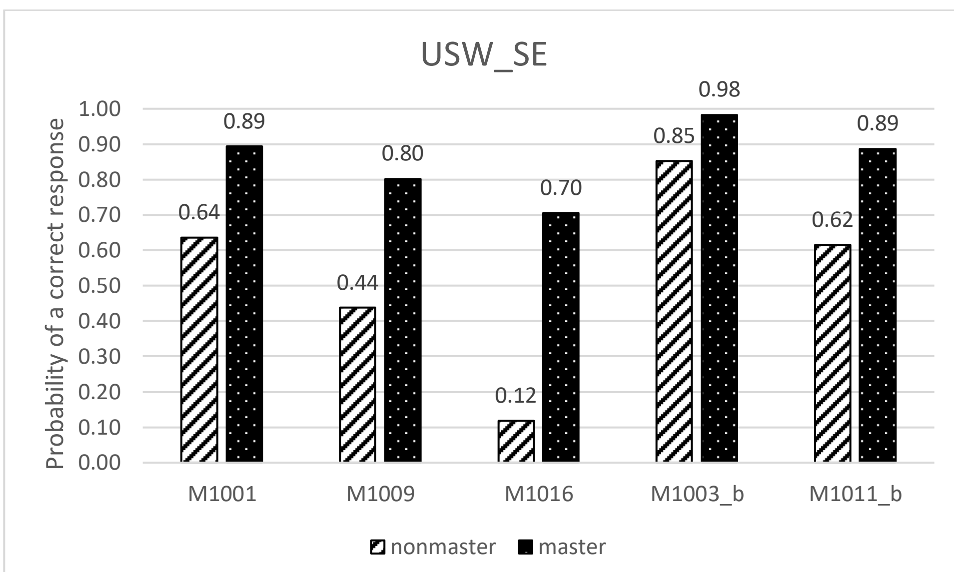


Figure 5.37. Item characteristic bar chart for USW_SE items

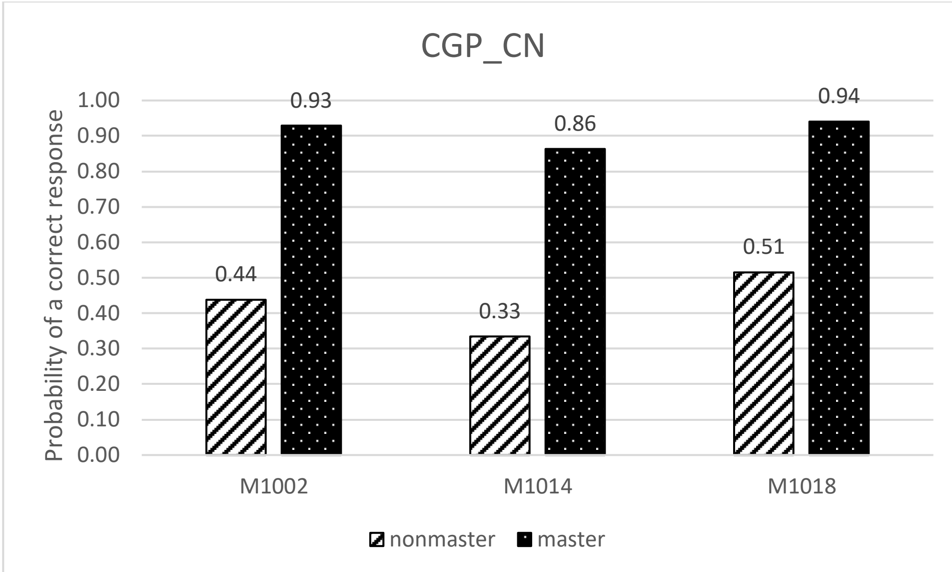


Figure 5.38. Item characteristic bar chart for CGP_CN items

One thing to note is that teachers could answer three of the five USW_SE items correctly with more than 0.5 probability, even though they did not have mastered level of knowledge. This may indicate that USW_SE items were too easy to distinguish teachers' mastery level. On the other hand, teachers could not correctly answer two of the three USW_SR items with more than 0.5 probability correctly, even though they are classified as masters of USW_SR. This may indicate these USW_SR items were too difficult to distinguish teachers. CGP_CN items were relatively good at distinguishing teachers compared to other algebra 1 items.

5.3.2.2.4 Model comparison

Similar to the test conducted under MIRT framework, the 3-attribute LCDM model was compared to the 2-attribute LCDM models (where the attributes USW_SE and CGP_CN or USW_SR and CGP_CN combine together). The result was consistent with that of IFA in that the 3-attribute model was significantly better than the 2-attribute model ($\chi^2(4)$ =14.55, p=0.006 for USW_SE and CGP_CN combined model; $\chi^2(4)$ =12.72, p=0.013 for USW_SR and CGP_CN combined model). The other fit statistics such as AIC also suggested better fit of the 3-attribute model (5182.703) than the 2- attribute models for USW_SE and CGP_CN combined model (5189.248) for USW_SR and CGP_CN combined model (5187.541).

5.4 Differences in geometry knowledge profile proportions

As shown in the previous sections, except for three pairs (USW_CG and USW_DP; CGP_EF and CGP_DP; USW_EF and CGP_EF), all hypothesized dimensions were distinguishable by the different tasks of teaching and instructional situations within and across the courses of study, geometry and algebra 1. This encouraging result answers the first research question – Can multiple dimensions be identified by either or both of the organizers within or across the course of studies? While the previous sections answer this question by using model fit statistics, this section examines differences among hypothesized dimensions in terms of the proportions of knowledge profiles. This section applies a DCM model to compare the proportions of knowledge profiles across the hypothesized dimensions.

5.4.1 Knowledge profile proportions within the same task of teaching

Considering that the EF items involve mathematical work that was more novel than the CG or DP items for students to do in the geometry course, I hypothesized that the proportion of teachers who mastered the knowledge attribute EF would be smaller than the proportion of teachers who mastered the knowledge attribute CG or DP. This hypothesis on the difference between EF and CG or DP with respect to the difficulty in managing the task USW was examined using teachers' knowledge profiles estimated from the DCM model described in 5.1.1.2. The proportion of teachers who are classified in each of the mastery profiles is presented in Figure 5.39. As the number of attributes compared is three, each attribute profile for a teacher is indicated by a vector with three elements $[\alpha_{rUSW\_CG}, \alpha_{rUSW\_DP}, \alpha_{rUSW\_EF}]$. For example, [0,0,0] indicates a profile where none of the three attributes are mastered, and [0,0,1] indicates a profile where only USW_EF attribute is mastered.

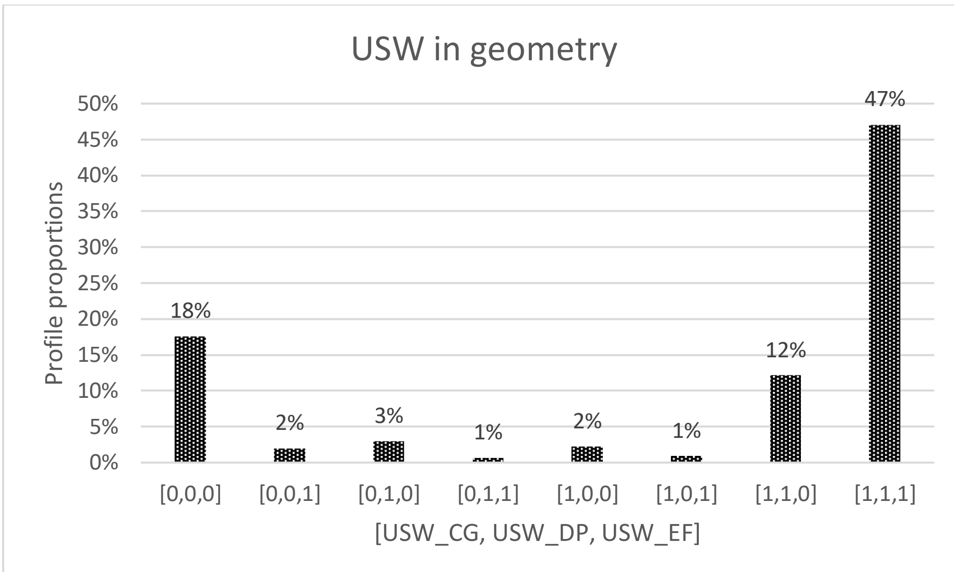


Figure 5.39. Profile proportions within the task USW

As shown in Figure 5.39[25], the most likely attribute profile is [1,1,1] (47% of teachers), where all three attributes were mastered, and the second most likely attribute profile is [0,0,0], where none of the three attributes were mastered (18% of teachers). As expected from the high correlation between USW_CG and USW_DP, very low proportions of teachers (1% ~ 3%) had profiles where USW_CG and USW_DP have

[25] As only the estimates for the geometry-related attributes are plotted, the sum of percentage is not 100% (attribute of USW_SR and USW_SE in algebra 1 are not included)

different values ([0,1,0], [0,1,1], [1,0,0], [1,0,1]). Regarding the proportions for USW_EF, only 2% of teachers had mastered USW_EF without mastering USW_CG and USW_DP (profile [0,0,1]), whereas 12% of teachers had mastered USW_CG and USW_DP, but not USW_EF (profile [1,1,0]). This difference in the proportion indicates that mastering the knowledge attribute for the instructional situations CG and DP seems to be relatively easy compared to mastering the knowledge attribute for the instructional situation EF. This result seems to support my hypothesis that USW_EF is more difficult to achieve a level of mastery than USW_CG or USW_DP.

A similar analysis was conducted to compare mastery proportions within the CGP task. In contrast to USW where the most likely attribute profile is [1,1,1], in CGP, the profile [0,0,0] (43% of teachers), where none of the three attributes were mastered, was the most likely attribute profile (Figure 5.40)[26]. As expected from the high correlations among the dimensions, each of the profiles other than [1,1,1] or [0,0,0] showed low proportions (1% ~ 7%). However, the sum of the proportions of teachers who mastered or did not master one of the knowledge attributes was a significant as 21%, which might not be identified under the 1-factor DCM model. This result is consistent with the result suggesting the 3-attribute DCM model better fit the data than the model less than the 3-attribute model (5.1.2.2.4, Model Comparison).

[26] As only the estimates for the geometry-related attributes are plotted, the sum of percentages is not 100% (attribute of CGP_CN in algebra 1 is not included).

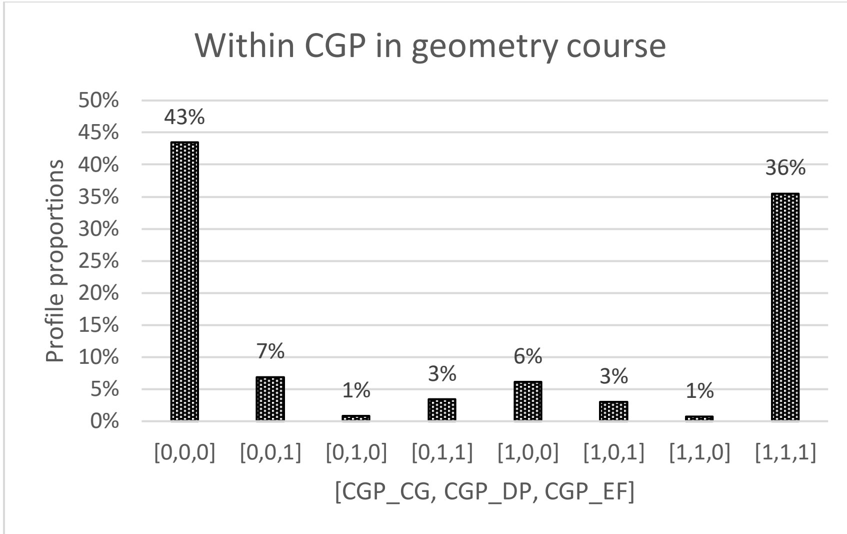


Figure 5.40. Profile proportions within the task CGP

5.4.2 Knowledge profile proportions within the same instructional situation

Section 5.2 reported the results showing whether the items are distinguishable according to the different tasks of teaching within each of the instructional situations CG, DP, or EF. The results showed that in the instructional situations common in the geometry course, such as CG and DP, the dimensions reflecting the tasks USW and CGP were statistically distinguishable, meaning that the models distinguishing the items reflecting USW and CGP fit the data significantly better than the models combining the USW and CGP items together. On the other hand, in classroom encounters where students engage in mathematical tasks that are novel or not customary, such as EF, the dimensions reflecting the tasks USW and CGP were not statistically distinguishable.

Having demonstrated in the previous section that most dimensions are distinguishable by different tasks of teaching in terms of model fit, I turn to a discussion of dimensionality in terms of the proportions of knowledge profiles. In other words, I examined the proportions of teachers who were knowledgeable only in doing USW or

CGP within a situation. The existence of those teachers would support the need of having more than one dimension.

The proportions of teachers for each of the mastery profiles were derived from the DCM models and the proportions of profiles within each of the instructional situations are plotted in Figure 5.41~ Figure 5.43. In those figures, each attribute profile is indicated by a vector with two elements $[\alpha_{rUSW\_GC}, \alpha_{rCGP\_GC}]$, $[\alpha_{rUSW\_DP}, \alpha_{rCGP\_DP}]$, and $[\alpha_{rUSW\_EF}, \alpha_{rCGP\_EF}]$ within CG, DP, and EF, respectively. For example, [0,1] in CG indicates a profile where only CGP_CG attribute is mastered and [1,0] indicates a profile where only USW_CG is mastered. Similarly, within DP, [0,1] indicates a profile where only CGP_DP attribute is mastered and [1,0] indicates a profile where only USW_DP is mastered.

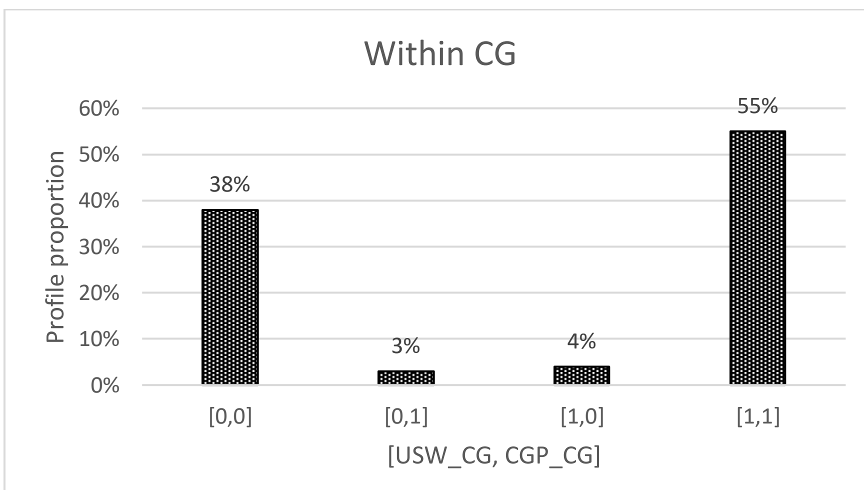


Figure 5.41. Profile proportions within CG within geometry

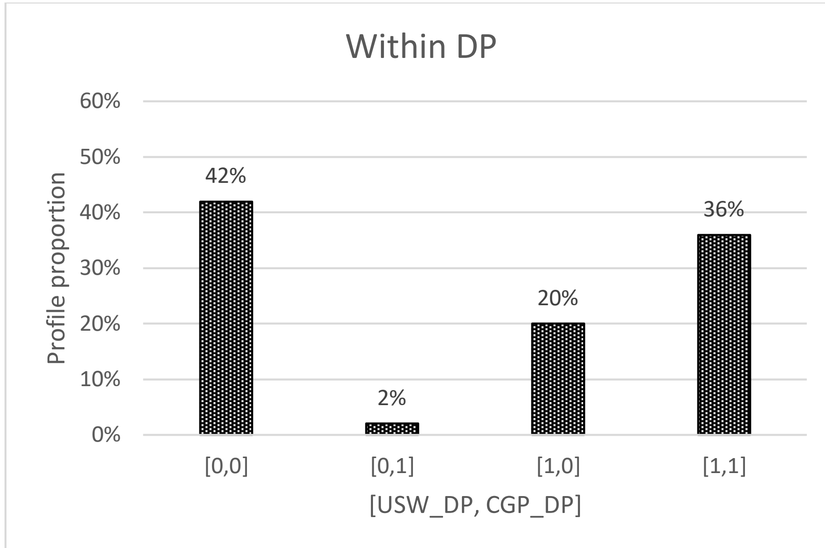


Figure 5.42. Profile proportions within DP within geometry

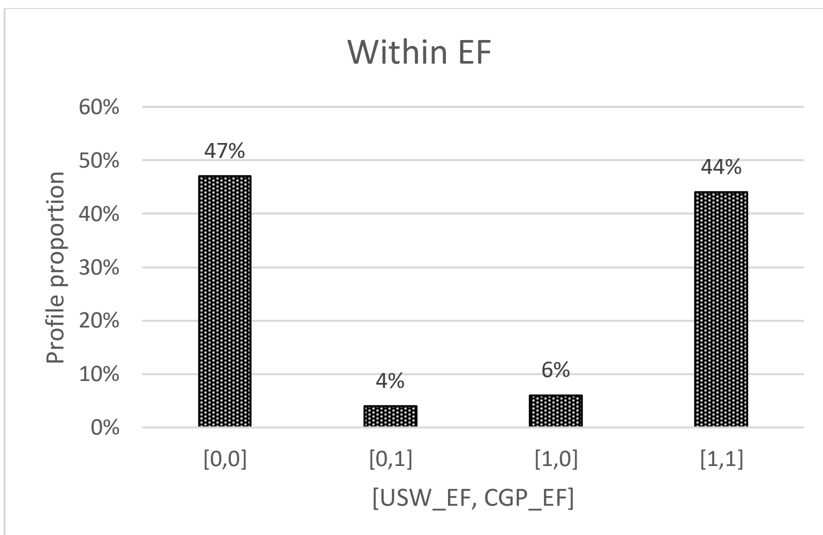


Figure 5.43. Profile proportions within EF within geometry

As shown in Figure 5.41and Figure 5.43, the number of teachers who mastered only either USW or CGP within CG or within EF was as small as 7% and 10%, respectively. This result may imply that distinguishing the two dimensions (USW_CG and CGP_CG or USW_EF and CGP_EF) within CG may not provide more information regarding knowledge profiles than a unidimensional model. In contrast, a notable number of teachers had mastered only either USW or CGP within DP (22% of teachers) (Figure 5.42). In particular, 20% of teachers mastered only USW_DP, whereas only 2% of teachers mastered only CGP_DP, implying that teachers are more highly knowledgeable in USW_DP than CGP_DP.

There is also a difference across different instructional situations in the proportion of mastery, called a base rate. The base rate is calculated by taking the sum of the probability distribution for the profiles that include mastery of each attribute in teachers' knowledge. For example, the base rate of USW_DP is 56% (=20%+36%), whereas that of CGP_DP is 38% (2%+36%) (Figure 5.42). This implies that CGP_DP is more difficult to master than USW_DP.

In sum, in spite of the small proportion of teachers whose profiles are mix of 0 and 1, i.e., masters or non-masters only in some of the dimensions, the existence of those teachers (7% ~ 22%) supports the previous results suggesting that multidimensional models organized by task of teaching and instructional situation provide valuable information that is not available in the unidimensional model. In addition, the pattern of proportions for some dimensions suggests some hierarchical relationships among the dimensions (e.g., USW_CG, USW_DP, and USW_EF) (discussed in Chapter 6).

5.5 Differences in algebra knowledge profile proportions

The differences among knowledge dimensions within a course of study were further examined in terms of the difficulty level of mastering each knowledge attribute. To compare the difficulty among the dimensions, the proportions of marginal attribute mastery (under a DCM) were plotted for algebra 1 (Figure 5.44).

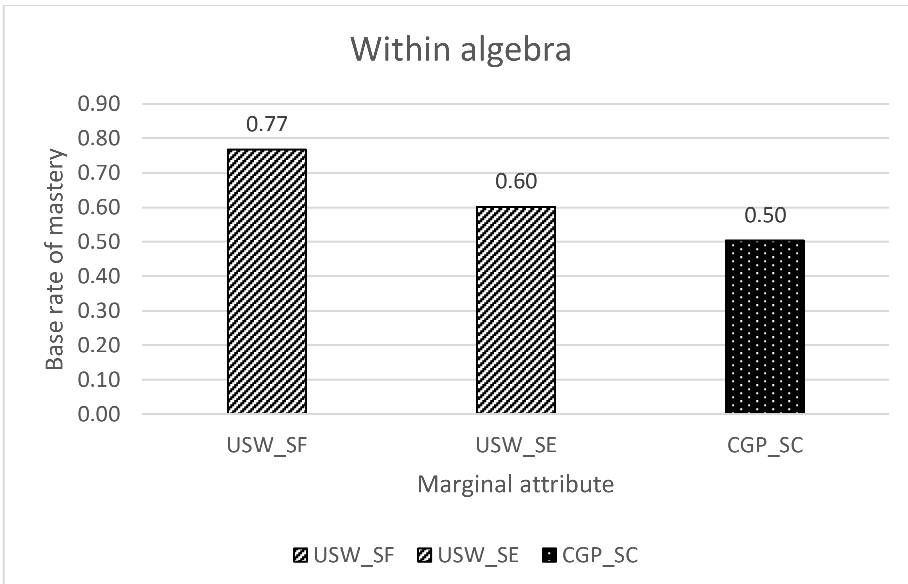


Figure 5.44. Marginal attribute mastery within algebra

As shown in Figure 5.44, within algebra, the proportion of mastery ranges from 50% for the CGP_CN attribute to 77% for the USW_SR attribute. This proportion is called the base rate, which can be calculated by taking the sum of the probability distribution for the profiles that include mastery of each attribute in teachers' knowledge. This plot indicates that USW_SE is more difficult to master than USW_SR, and CGP_CN is more difficult to master than USW_SE.

5.6 Differences in a relationship with educational and teaching experience

The relationships between each of the hypothesized knowledge dimensions and teachers' educational and teaching experience were examined to see whether the distinctions identified in the dimensionality analyses are also noticeable in terms of the different relationships with background information. The identification of different relationships among dimensions would provide additional evidence supporting the distinctions among the hypothesized dimensions. In the analysis, both measurement and structural relations were tested simultaneously in a single model so that the relational parameters could be more accurately estimated by taking into account measurement error and relations among latent constructs (Kline, 2010). The results are presented in Table 5.41 ~ 5.43.

Table 5.41. Effects of teachers' background on knowledge in doing USW

Dimensions of HS geometry	Standardized estimate	*SE*
Predictors for USW-CG		
Years of teaching Geometry	**0.381********	0.056
CollegeMathCourses	0.109	0.057
Years of teaching non-geometry	-0.040	0.058
Current Teaching	0.114	0.062
Predictors for USW-DP		
Years of teaching Geometry	**0.487*****	0.070
CollegeMathCourses	0.100	0.064
Years of teaching non-geometry	0.020	0.067
Current Teaching	0.063	0.072
Predictors for USW-EF		
Years of teaching Geometry	**0.269*****	0.075
CollegeMathCourses	0.072	0.077
Years of teaching non-geometry	-0.100	0.077
Current Teaching	-0.044	0.082

(*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$)

Table 5.42. Effects of teachers' background on knowledge in doing CGP

Dimensions of HS geometry	Standardized estimate	*SE*
Predictors for CGP-CG		
Years of teaching Geometry	**0.355*** **	0.057
CollegeMathCourses	-0.015	0.058
Years of teaching non-geometry	**0.165** **	0.058
Current Teaching	0.077	0.064
Predictors for CGP-DP		
Years of teaching Geometry	**0.189** **	0.073
CollegeMathCourses	**0.258*** **	0.070
Years of teaching non-geometry	-0.005	0.079
Current Teaching	-0.048	0.077
Predictors for CGP_EF		
Years of teaching Geometry	**0.290*** **	0.074
CollegeMathCourses	0.091	0.068
Years of teaching non-geometry	-0.055	0.067
Current Teaching	-0.065	0.080

(* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

Table 5.43. Effects of teachers' background on knowledge in algebra 1

Dimensions of algebra 1	Standardized estimate	*SE*
Predictors for USW-SR		
Years of teaching Alg1	-0.032	0.095
CollegeMathCourses	-0.002	0.095
Years of teaching non-alg1	-0.112	0.093
Current Teaching alg1	-0.092	0.099
Predictors for USW-SE		
Years of teaching Alg1	**0.168***	0.068
CollegeMathCourses	0.055	0.068
Years of teaching non-alg1	**0.215** **	0.070
Current Teaching alg1	0.009	0.074
Predictors for CGP-CN		
Years of teaching Alg1	**0.136***	0.065
CollegeMathCourses	**0.151***	0.062
Years of teaching non-alg1	**0.168***	0.072
Current Teaching alg1	0.066	0.068

(* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

As shown in Table 5.41 ~ 5.43, teaching experience specific to the course of study (years of teaching geometry or years of teaching algebra 1) consistently showed significant effects on all hypothesized geometry dimensions or algebra 1 dimensions, except for USW_SR, when controlling for other predictors. The size of the effect is, however, different across dimensions. For example, the effect of experience teaching geometry on teachers' knowledge for understanding students' work in the doing proofs situation (USW_DP) was significantly greater than the effect on teachers' knowledge for choosing the givens for a problem in the doing proofs situation (CGP_DP). For example, the effect size for USW_DP is as large as 0.49, whereas that for CGP_DP is as small as 0.19 (Figure 5.45).

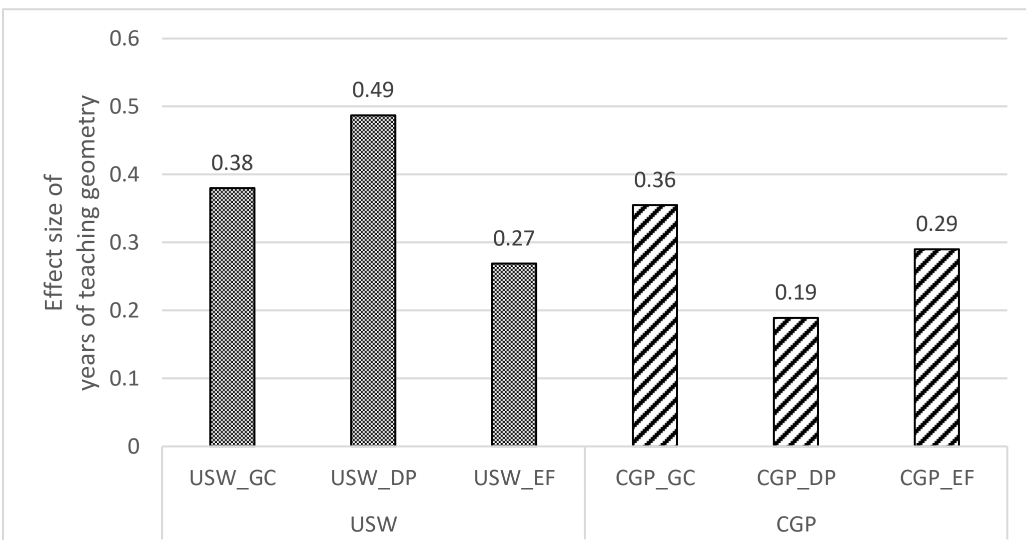


Figure 5.45. Effect sizes of years of teaching geometry within geometry attributes

On average, the effect size of experience teaching geometry on USW dimensions is 0.38, whereas that on CGP dimensions is 0.29 (the difference is mostly from DP). The difference between USW and CGP is consistent with my hypothesis that the dimensions reflecting the USW task would be more strongly associated with years of teaching experience than with the CGP task given that teachers tend to engage in USW more frequently than in CGP in their work of teaching (Ko & Herbst, under review).

Another difference in the relationship between USW and CGP is that, for CGP, not only years of teaching experience in geometry, but also years of teaching experience in non-geometry and the number of college math courses taken showed significant effects on teachers' knowledge (0.165 for CGP_CG and 0.258 for CGP_DP). Specifically, the significant effect of non-geometry teaching experience on CGP_CG implies that some aspects of mathematical knowledge required for teachers to design problems for CG could improve with more experience in teaching a mathematics course besides geometry. This result is understandable in that teachers need to consider not only geometric properties, but also the algebraic components of the problem (e.g., whether the equation is solvable) when they create problems that involve geometric calculation (Boileau & Herbst, 2015; Hsu & Silver, 2014).

The difference between geometry dimensions and algebra 1 dimensions regarding the effects of teachers' background is that the effects on the geometry dimensions were significant and moderate, but the effects on the algebra 1 dimensions were not all significant and the effect sizes were small. This might imply that teachers' mathematical knowledge used for teaching geometry is more specific to the course than the knowledge used for teaching algebra 1. This result is also consistent with Herbst and Kosko (2014)'s study showing that teachers' MKT-G was significantly related with teachers' teaching experience specific to geometry courses, but not with experience teaching mathematics courses in general.

In summary, the results provided additional findings to the previous study: course-specific teaching experience has also significant effects on subdimensions of knowledge when controlling for the experience in teaching other courses; the effect size is different depending on the involved task of teaching and instructional situations where the tasks are enacted; and the course-specific effect is more salient for the knowledge in geometry than algebra 1.

5.7 Chapter summary

This chapter presented analysis results showing that items could be distinguished by hypothesized organizers: tasks of teaching and instructional situations. In other words, most multidimensional models, which distinguish items according to their hypothesized categories (by tasks or situations), better fit the item responses than the models which combine dimensions together.

Within the same tasks of teaching, 14 among 16 ($\frac{5*4}{2}$ = 10 pairs for USW and $\frac{4*3}{2}$ =6 pairs for CGP) hypothesized pairs of dimensions were statistically distinguishable. The pairs that were not distinguishable were between USW_CG and USW_DP and between CGP_DP and CGP_EP. Within the same instructional situation, two among three hypothesized pairs of dimensions were statistically distinguishable. The pair that was not distinguishable were USW_EF and CGP_EF. The results imply that the dimensions reflecting the different tasks USW and CGP in classroom encounters where students engage in mathematical tasks that are novel were not statistically distinguishable each other, whereas they were distinguishable in the instructional situations common in high school geometry such as CG and DP. My conjectures on the results on the pairs of non-distinguishable dimensions are discussed in Chapter 6.

The distinctions among the hypothesized dimensions were also examined using proportions of mastery profiles estimated from DCM models. As expected from high correlations among dimensions, most teachers were assigned to the profiles representing masters or non-masters on all (or none) of the dimensions (within the same instructional situations or within the same tasks of teaching). In other words, if a teacher was a master or a non-master in one of the hypothesized knowledge dimensions, the teacher was very likely to be a master or non-master in other hypothesized knowledge dimensions.

As such, the proportions of teachers who are masters or non-masters only in some of the knowledge dimensions were small. However, it is important to note that there are noticeable proportions of teachers who are masters or non-masters for only some of the knowledge attributes. For example, 20% of teachers attained a mastery level of knowledge for USW_DP, but not for CGP_DP. The existence of teachers whose mastery level is different across dimensions supports the rationale for the need to measure teachers' knowledge under the multidimensional conceptualization of the knowledge.

The differences among the hypothesized knowledge dimensions were also examined in terms of the relationship with teachers' educational background and teaching experience. The SEM results showed that years of geometry teaching experience had significant effects on all of the geometry knowledge dimensions, but years of non-geometry teaching experience was not. Moreover, the differentiated effects of teachers' educational and teaching background on MKT-G and MKT-A depending on the involved task of teaching and instructional situations provided additional evidence to support the rationale for designing a teacher knowledge assessment using the categories of task of teaching and instructional situation.

Chapter 6

Discussion and Conclusion

This dissertation centered on examining the theoretical and methodological plausibility of a hypothesized knowledge framework to organize the dimensionality of teachers' mathematical knowledge for teaching high school geometry and algebra 1. This framework is based on two organizers of dimensions: tasks of teaching and instructional situations. The question of the plausibility of the framework considers whether assessment items categorized according to the framework can be used to 1) provide a fine-grained description of teachers' mathematical knowledge used in the work of teaching and 2) develop multiple distinguishable measures of teachers' mathematical knowledge for teaching. The first question asks about the conceptual plausibility and the second question asks about the methodological feasibility of the proposed framework.

To evaluate the methodological feasibility of the framework, I applied three different multidimensional measurement models (SEM, IRT, and DCM) to a sample of 602 U.S in-service high school teachers and their responses on items measuring teachers' mathematical knowledge for teaching high school geometry and algebra 1. I used the measurement models to evaluate the quality of the items and to yield scales or profiles reflecting the teachers' competencies in hypothesized multiple dimensions. Ultimately, the results of this study stand to contribute to our understanding of the organization of mathematical knowledge used in the work of teaching as well as ways to operationalize it.

In this chapter, I summarize and interpret the main findings. This interpretation is followed by theoretical and methodological implications and limitations of the study. I conclude this dissertation with recommendations for future study.

6.1 Conclusion and summary of findings

Overall, the tested measurement models support my hypothesis that items organized around an integrated knowledge trait, which can be characterized in terms of the types of teaching tasks (e.g., understanding student mathematical work) and instructional situations (e.g., doing a proof in a geometry class), can measure distinguishable dimensions of teachers' mathematical knowledge for teaching high school geometry and algebra 1. The results supporting the hypothesis are also consistent across three different measurement models: SEM, IRT (MIRT), and DCM. The main findings of this study are discussed as answers to my initially posed research questions as follows.

6.1.1 The role of the two organizers in measuring dimensions of the knowledge

Question 1 asked: Conceptually, how well do the organizers – task of teaching and instructional situation – capture multiple aspects of teachers' mathematical knowledge used in the work of teaching?

In regards to the descriptive purpose of this study (1.1.1), the evaluation of the conceptual plausibility of the item blueprint examined whether the two organizers allow for describing different aspects of teachers' mathematical knowledge for teaching. By using tasks of teaching and instructional situations as the organizers for the item blueprint, the blueprint provides a way to deconstruct teachers' mathematical knowledge situated in the work of teaching.

Furthermore, the item blueprint provides a way to capture both generic aspects of teachers' knowledge across mathematics courses of study (e.g., algebra and geometry) and the knowledge aspects specific to a subject matter (e.g., algebra or geometry). Specifically, in the framework, "task of teaching" measures generic aspects of teachers' mathematical knowledge, whereas "instructional situation" captures the subject-matter specific knowledge. Also, the hypothesized role of these organizers in capturing differences in mathematical knowledge among the teachers (3.3) is consistent with previous empirical studies. For example, Herbst and Kosko (2014) showed that whether the items are contextualized in an instructional situation or not is related to the difference in the knowledge between novice teachers and experienced teachers.

As mentioned in 1.2, this study does not claim that my suggested item blueprint is necessarily better than others (e.g., the MKT framework by Ball et al., 2008). Instead, I argue the appropriateness of the item blueprint with respect to the intended use and interpretation of the results (Kane, 2006). The intended use is to develop multiple distinguishable measures that allow us to examine multiple dimensions of teachers' knowledge used in the work of teaching. The distinguishable measures could be established by multiple sets of items that capture distinguishable traits of the work of teaching. Within this purpose, I argue that the proposed item blueprint might be more suitable than others that characterize the items by knowledge type, because it allows to define the boundaries and relationships among the dimensions with more clarity and defines the knowledge dimension in terms of the characteristics of the components of the work.

As described in 1.2.3, the nature of teachers' knowledge is complex, so it is challenging to determine whether a certain characteristic of the knowledge is solely mathematical or combined with pedagogical content knowledge. Due to this challenge, I could not clearly determine the knowledge type (e.g., CCK, SCK) associated with the items that I used in this study. Similar to the item described in 1.2.4, some items seemed to be associated with more than one knowledge type that the item was initially designed to measure. Thus, I decided not to categorize the existing items according to the knowledge type. In contrast to the criterion of knowledge type, I could determine the type of task of teaching and instructional situation for most of the items that had been developed according to a knowledge type item blueprint. This feasibility of the proposed item blueprint supports my argument that the proposed item blueprint is conceptually plausible and compatible with the purpose of this study.

Question 2 asked: Methodologically, are the knowledge scales estimated by items measuring either or both different tasks of teaching and different instructional situations statistically distinguishable?

The majority of the hypothesized dimensions could be identified by either or both of the organizers (tasks of teaching and instructional situations) within and across both geometry and algebra 1 courses of study. In other words, multidimensional models, which distinguish items according to the categories of task of teaching and instructional situation, better explained the variance among the items than the models that combined the hypothesized dimensions together. Some pairs of dimensions were not distinguishable by either task of teaching or instructional situation. However, those results are understandable considering the similarities in a teacher's mathematical work between two tasks of teaching or between two instructional situations. The need for more than one organizer to describe one dimension of teachers' mathematical knowledge was also supported by the results showing associations between task of teaching and instructional situation. For example, the dimensions reflecting different instructional situations (e.g., DP and CG) were distinguishable in one task of teaching (CGP), but they were not distinguishable in another task (USW).

Some other studies have suggested a knowledge framework structured by more than one organizer. Among them, Etkina et al. (2018) conceptualized physics teachers' content knowledge for teaching energy (CKT-E) as a multidimensional construct that is organized by the intersection of tasks of teaching (ToT) and student energy targets. Compared to my study, their first organizer – task of teaching (ToT) – is similar to one of the organizers used in my study (task of teaching), and their second organizer – student energy targets – is somewhat related to the other organizer used in my study (instructional situation) given that both consider the subject matter work of students. The concept of instructional situation, however, includes more than their "student energy targets" in that it also includes teachers' knowledge of how labor is divided in the mathematical tasks that are used to claim those targets. Also, their first dividing principle that undergirds divisions of the construct is still knowledge domains. In other words, their framework organized by ToT and student energy targets is applicable to CKT-E, but may not include PCK-E. In this study, however, knowledge-domain is not the primary dividing principle of the construct. Rather, the combination of ToT and instructional situation is the primary principle that distinguishing the dimensions, which could be a mix of different knowledge domains.

Apart from the difference in primary organizers (ToT and instructional situation instead of knowledge domain such as CK or PCK), the result of this study is unique in that teachers' mathematical knowledge was operationalized in terms of multiple scales instead of an overall unidimensional scale. Also, the multiple scales were developed using different assumptions on the characteristics of the construct being modeled by each measure representing the amount of the knowledge (continuous or discrete). In other words, even though several other studies have suggested a multidimensional knowledge framework and this framework has been used in the phase of item development, only single scales have been developed to represent an overall knowledge construct. For example, Etkina et al. (2018) conceptualized teachers' CKT-E as multidimensional, but they assigned a single score reflecting a level of CKT-E to individual teachers rather than assigning multiple scales to follow their framework. Similarly, Phelps et al. (2014) developed a unidimensional scale as a measure of teachers' mathematical knowledge for teaching algebra 1, even though they conceptualized the knowledge construct as multidimensional. Regarding assumptions on the knowledge being modeled, researchers have developed knowledge scales assuming either a continuous construct (Hill et al., 2012) or discrete (Bradshaw et al, 2014) attributes. This study, on the other hand, developed knowledge measures under both continuous and discrete assumptions and showed that how the different assumptions can provide complementary information on the characteristics of dimensions of the knowledge.

Overall, as described in Chapter 2, I do not argue that my proposed knowledge framework should be construed as a framework that accurately represents the true structure of teachers' mathematical knowledge or as a framework necessarily better than other previous frameworks. Rather, I argue that my proposed knowledge framework provides a basis for *developing a test blueprint* that allows for multiple scales reflecting distinguishable dimensions of teachers' mathematical knowledge for teaching. Also, my proposed framework accounts for the instructional specificity of the knowledge needed to do the work of teaching mathematics. By instructional specificity, here, I allude to how the transactions of content between teacher and student make those knowledge demands specific (along the lines argued by Herbst & Kosko, 2014). In doing so, it provides a way to describe teachers' mathematical knowledge. In the same vein, I did not compare my proposed factor model with other models (e.g., the MKT framework from Ball et al., 2008). The comparison between the item blueprint in terms of model fits was not even possible as I could not clearly determine the knowledge type (e.g., whether an item is measuring CCK or SCK) associated with the existing items (discussed in 1.2.4).

Following the definition of validity described in Kane (2013), according to which validity depends on arguments and uses, I evaluated my proposed framework based on an argument about the use of the test – measuring distinguishable dimensions and describing the knowledge reflected in the work of teaching. Given this use of the item blueprint, I was not interested in creating a map representing the full range of teachers' mathematical knowledge organized by all possible components of the work of teaching. Rather, I was interested in providing a way to explore the multiple dimensions of teachers' mathematical knowledge that could be anticipated by the use of the two organizers. And, I wanted to determine whether the measures entailed by using a combination of the two organizers led to feasible measures. The use of particular examples of different tasks of teaching and particular instructional situations can also be justified in that they enabled the development of distinguishable measures, thus illustrating that feasibility.

While the proposed item blueprint was validated as being methodologically feasible with particular cases, the feasibility cannot be generalized to an argument that all possible *different* kinds of tasks of teaching or instructional situations *generate distinguishable dimensions* of teachers' mathematical knowledge or that all possible *similar* kinds of tasks of teaching or instructional situations *are unable to generate distinguishable dimensions* of teachers' mathematical knowledge. Indeed, the results showed that some pairs of dimensions could not be differentiated by the two organizers, even though they were hypothesized to be different.

In spite of this methodological limitation in validating all possible tasks of teaching and instructional situations, it is reasonable to conclude that this study supports the validity of my suggested framework for organizing teachers' mathematical knowledge for teaching by task of teaching and instructional situation. The main reason for this conclusion is that the distinctions between the knowledge dimensions could be operationalized by multiple measures. The knowledge dimensions that were distinguishable were not merely convenient samples of possible knowledge dimensions, but they were knowledge dimensions hypothesized to be different by meaningful variation in different components of the work of teaching.

To further support my conclusion, next, I present my conjecture on the results that were different from my initial hypothesis. Those results refer to the lack of distinction in three pairs of dimensions: USW_CG (Understanding Students' Work in Geometric Calculation) and USW_DP (Understanding Students' Work in Doing Proof), CGP_EF (Choosing the Givens for a Problem in Exploring a Figure) and CGP_DP (Choosing Givens the for a Problem in Doing Proof), and USW_EF (Understanding Students' work in Exploring a Figure) and CGP_EF (Choosing the Givens for a Problem in Exploring a Figure).

6.1.2 Interpretations on the indistinguishable dimensions

This section describes my conjectures on the pairs of dimensions that were not shown to be distinguishable from each other. Even though some of this lack of distinction was interpretable in consideration of similarities in teachers' knowledge used in two tasks of teaching or two instructional situations, it is important to note that the interpretations described in this section are still conjectures. This means that the interpretations are plausible only as alternatives to the initial hypothesis on the distinctions. The lack of distinction might also suggest an issue with the instrumentation, such as construct underrepresentation or an issue arising from retrofitting the existing items. More discussion about this issue is discussed in the section of limitations.

6.1.2.1 Lack of distinction between USW_CG and USW_DP

The dimensionality analysis within the task of USW (Understanding Students' Work) showed that USW_EF (Understanding Students' Work in Exploring a Figure) was distinguishable from both USW_CG (Understanding Students' Work in Geometry Calculation) and USW_DP (Understanding Students' Work in Doing Proof), but USW_CG and USW_DP were not distinguishable from each other. To understand these results, I referred to the work expected from a teacher and students in each of the situations. As described in 4.3.4, the task of teaching USW includes a teacher's reading and making sense of a student's work and deciding whether the student's answer, inferred processes, and reasoning are mathematically correct. Thus, to understand the relationships among USW_EF, USW_CG, and USW_DP, we first need to compare the characteristics of the students' work expected in each situation.

In both CG and DP, student work is presented by sequential steps that connect the givens (information given in the diagram or by statements) to a solution ('prove' statement in DP and unknown '*x*' in CG). The sequential steps in both situations are derived by the student's knowledge about particular properties of the figure or related theorems. This similarity between DP and GCN (one type of CG) regarding the students' problem-solving process and students' performance is also described in Hsu's (2010) dissertation study. Specifically, the processes of using geometric properties and the sequence of them required to obtain solutions were the same for student tasks in both GCN and DP, when the given diagram and geometric properties required to find solutions were controlled (Hsu, 2010, p.149). Also, there was no significant difference in 9th grade students' performance when solving either GCN or DP problems. As Hsu described in both situations, "students need to visualize the geometric diagrams and identify the needed geometric properties in order to set up calculating sentences [GCN] or form logical proving statements [DP]" (2010, p. 5).

Based on Hsu's (2010) findings and the characteristics of students' work, my finding that USW_CG and USW_DP cannot be distinguished seems understandable. The task of USW in both situations requires a teacher to determine whether students correctly visualize the geometric diagrams, recall and apply known theorems, definition, and properties in problem solving. The teacher also would need to know whether students correctly translate the properties into symbolic notations (an equation for CG, a statement for DP). In short, both situations require a teacher to evaluate students' ability for deductive reasoning. The similarities between CG and DP in the work and the involved knowledge expected from a teacher when doing USW may pose a challenge for distinguishing USW_DP and USW_CG. However, it does not necessarily mean that the knowledge USW_DP and USW_CG are indeed not distinguishable. In spite of the similarities between CG and DP, there are differences, such as the use of symbols or the need of students' justification statements. For example, the algebraic symbols used in CG are more standard and readable than the symbolic statements used in DP. Also, DP expects students to write a justification for each statement when doing a proof problem, but CG does not (Boileau & Herbst, 2015). Whether the knowledge dimensions USW_DP and USW_CG are indeed distinguishable or not may need further investigation with more items that better operationalize these differences between the two situations.

In contrast to CG or DP where students work to find a logical path from the given information to the given conclusion, in EF, students are responsible for generating a statement of a conclusion or conjecture by freely choosing information implicitly given in a diagram. In other words, students in EF need to find conclusions that are worth making rather than validating the conclusions given by teachers (Aaron & Herbst, 2015, p.2). Given the nature of this effort, the correctness of student work in EF is likely to depend on what conjecture students generate and the resources students use. Because of this possible variability in the mathematical work, doing the task of USW in EF might require teachers to use a kind of mathematical knowledge that differs from the knowledge the teachers normally use in DP or CG.

Apart from the need to work with unpredictable conjectures and students' resource use, EF might be more challenging for teachers to manage than the situation of CG or DP, as the product of students' work cannot necessarily be anticipated due to its potential variability. For example, students could generate very divergent conjectures depending on the information individual students decide to use in making their own conjectures. In this context, a teacher needs to understand and evaluate students' different conjectures and reasoning. My conjecture on a difference in difficulty of mastering USW_EF and USW_DP or USW_CG was supported by the result of the DCM analysis, which showed a higher proportion of teachers mastering only USW_CG or USW_DP (12%) than the proportion of teachers mastering only USW_EF (2%).

6.1.2.2 Lack distinction between CGP_EF and CGP_DP

The dimensionality analysis within the task of CGP showed that CGP_DP was distinguishable from CGP_CG, but it was not distinguishable from CGP_EF. On the other hand, the items reflecting the situation CG and DP were distinguishable for CGP. Similar to the previous section, I interpret these results based on the work called forth from a teacher in doing CGP in each of the situations EF, CG, and DP.

The CGP_EF items reflect a teachers' knowledge in creating classroom tasks and test items or choosing information to be given for the tasks that enable students to explore a figure. The task – exploring a figure – is expected to provide students an opportunity to examine (look at, measure, mark, or draw) a diagram, conjecture common or general properties of the geometric objects, and state the properties in conceptual language

(Herbst, 2010). In other words, the task chosen/created by a teacher needs to give students the best opportunity to show their capacity to infer general properties from a particular diagram using their knowledge of the definitions and properties of geometric figures, skills manipulating instruments, and their ability for inductive or abductive reasoning. Given this role of student tasks in EF, the CGP_EF items ask teachers to choose the best diagrams or written tasks that can best accomplish the goal of EF.

In this study, CGP_EF was initially hypothesized to be distinguishable from CGP_DP given differences in the cognitive process students are expected to take in the two situations. Specifically, I hypothesized that there would be a difference in teachers' work between two situations in that a teacher needs to provide a context for students to demonstrate inductive reasoning in EF, whereas a teacher needs to provide a context for students to demonstrate deductive reasoning in DP. However, the dimensionality analysis showed that the knowledge dimensions CGP_EF and CGP_DP were not statistically distinguishable. This result may imply that the mathematical work teachers do (i.e., choosing the givens for a problem) or the kind of knowledge used in EF is not different enough to be distinguishable from the mathematical work done and the knowledge used in DP, even though the work students normally do in the two situations is different.

From a teacher's perspective, when doing CGP in both EF and DP, teachers may start from a general property of a theorem that their students are expected to learn, and then, consider a specific example problem that can help their students understand the property or the theorem. Teachers also need to make sure that the targeted conjecture (in EF) or the prove statement (in DP) are true. This means that the cognitive process required for teachers to take in choosing appropriate resources for a task in EF is still deductive reasoning, even though the logical process their students are expected to take in the situation is inductive or abductive reasoning. Thus, when a teacher solves a CGP_EF item, the teacher needs to choose an option representing a diagram or a written activity related to the targeted geometric property or theorem. In short, the knowledge required to answer a CGP_EF item would require teachers to use deductive reasoning instead of inductive reasoning that their students are expected to use in EF. CGP_EF, thus, might not be distinguishable from CGP_DP.

6.1.2.3 Distinction between CGP_CG and CGP_DP

In contrast to the result that the instructional situations of CG and DP were not distinguishable within the task of USW (5.1.1), CG and DP were distinguishable within the task of CGP (5.1.2). The distinction between CGP_CG and CGP_DP is understandable in light of work by Herbst and Kosko (2014), which states that "tasks of teaching could call for different kinds of mathematical work depending on specifics of the work of teaching geometry" (Herbst & Kosko, 2014, p. 7). In their study, conjectures on the specifics were described within the context of designing a problem, which is similar to the category of CGP in my study. Specifically, they explained that

> the task designing a problem would involve a teacher in different mathematical work if the designed problem was a proof problem versus a geometric calculation. While the former might involve the teacher in figuring out what the givens should be to make sure the desired proof could be done, the latter might involve the teacher in posing and solving equations and checking that the solutions of those equations represented well the figures at hand (p. 7)

Considering that the definition of CGP in my study includes the task of designing a problem, my result showing distinction between CGP_CG and CGP_DP is in line with Herbst and Kosko's (2014) conjecture on the distinction between the mathematical work teachers do in designing a problem for proof (DP) and for geometric calculation (CG).

6.1.2.4 Lack of distinction between USW_EF and CGP_EF

In contrast to my initial hypothesis, USW_EF and CGP_EF were not statistically distinguishable from each other (5.2.3). As mentioned, EF can be considered as an instructional situation where students engage in mathematical tasks that are not customary in that they are expected to generate conclusions which are customarily given by a teacher in most geometry classrooms. As EF is an instructional situation that is uncommon in high school classrooms, teachers would rarely have an opportunity to do the work of EF, regardless of whether the involved task is USW or CGP.

The lack of distinction between USW_EF and CGP_EF in the results was interpreted in consideration of the characteristics of the work teachers are expected to do for USW and CGP in EF. USW_EF requires a teacher to determine whether students' conjectures on a given figure are correct and whether students correctly chose and used the given information to make their conjectures. CGP_EF requires a teacher to determine whether the information given in a problem (e.g., a geometric object or a tool) enables students to generate conjectures stating correct properties of the geometric objects at hand. In EF, the work of evaluating the correctness of a conjecture or the appropriateness of a conclusion of the problem should be done not only for CGP but also for USW. This is because for EF a teacher in EF does not constrain a specific object or property so that students can make diverse conjectures and conclusions, such as a conjecture about a perimeter of a triangle or a conjecture about a triangle congruence. In other words, students have the freedom to choose information among the givens as well as the content of their conjectures. Therefore, a teacher's work – evaluating a conclusion – may need to be done in both USW and CGP. These similarities in teachers' work when doing USW_EF and CGP_EF may lead to the lack of distinction between the two dimensions in my results. Yet, this is a conjecture and whether USW_EF and CGP_EF are indeed distinguishable or not needs further investigation with more diverse items.

6.1.3 Different relationships with teachers' background among dimensions

Question 3 asked: Are there differences among the hypothesized dimensions of teachers' mathematical knowledge for teaching geometry and algebra in terms of their relationships with teachers' self-reported background (number of mathematics courses taken in college, number of geometry teaching years, number of non-geometry teaching years)?

Overall, most knowledge dimensions were significantly associated with subject specific teaching experience. Specifically, a 10-year increase in teaching geometry experience was significantly associated with approximately 0.31 ~ 0.97 increase in IRT scores for geometry dimensions. This result of effect size is somewhat consistent with Hill's study (2007) showing that additional 10 years of teaching experience was associated with approximately 0.5 increase in the middle school teachers' overall MKT IRT score. Furthermore, the result that only years of teaching geometry had significant relationships with three USW dimensions is consistent with Herbst and Kosko's (2014) study which showed a significant relationship between teachers' MKT-G scores and their teaching experience specific to geometry courses, but not with teaching experience in general.

Compared to these previous studies that examined the effect of teachers' teaching experience on the overall MKT (or MKT-G) scores derived from a unidimensional model, my study examined the effect of teachers' teaching experience on each of the subdimensions derived from multidimensional models. The application of the multidimensional models provided several new findings.

First, the magnitude of the effect from the subject specific teaching experience was different across knowledge dimensions. In particular, the effect was stronger for the dimensions involving a frequent task (e.g., USW) or a customary instructional situation (e.g., CG or DP) than for the dimensions involving a task that teachers less frequently engage in (e.g., CGP) or a less-customary instructional situation (e.g., EF). This result implies that the familiarity with doing a task of teaching or managing an instructional situation gained from teaching experience enables teachers to develop their mathematical knowledge required in doing the task or managing the situation. This positive effect of teaching experience on teachers' knowledge was also reported in a study that measured early career teachers' general pedagogical knowledge (König, Blömeke, & Kaiser, 2015). In particular, König et al. (2015) found that the effect was especially crucial when the experience is related to "teachers' deliberate efforts to reflect on and improve their teaching" (what they called "deliberate practice") (p.335).

Thus, experience in teaching a course of studies may increase teachers' mathematical knowledge if that experience provides teachers opportunities to engage in the task or the situation that is called forth by the work of teaching the course. For example, if teachers are less knowledgeable in the task CGP than the task USW because they did not have enough experience in engaging CGP, providing teachers with professional learning experience involving task design (Zaslavsky & Sullivan, 2011) could significantly increase CGP related MKT-G scores.

Second, the answer to the question – whether subject matter preparation or experience in teaching mathematics courses in general influence teachers' mathematical knowledge or not – depends on which dimension the teachers' mathematical knowledge refers to. The finding of a previous study, which shows no significant effect of non-geometry teaching experience on teachers' knowledge (Herbst & Kosko, 2014), was consistent with the result for the USW-related dimensions, but not for the CGP-related dimensions. In other words, not only years of teaching geometry but also the predictors representing years of teaching courses other-than-geometry or the number of college mathematics courses taken yielded significant effects on teachers' mathematical knowledge if the measured knowledge is used in the task CGP. The effect of years of teaching mathematics courses in general or the number of college mathematics courses taken on teachers' MKT-G in doing CGP would not have been recognized if the MKT-G is operationalized only by a single score. In other research, König et al (2015), which examined early career mathematics teachers' general pedagogical knowledge, also showed differences in the association with teachers' educational and teaching background between different knowledge domains. Specifically, in their study, teachers' declarative knowledge was predictable by teacher education grades, whereas teachers' skill to interpret classroom situations was predictable by the amount of time spent on teaching (König et al., 2015, p. 335). These findings, from prior studies and the present study, demonstrate how multidimensional measures of teachers' knowledge allow detecting differences different knowledge dimensions, in terms of the relationship with characteristics of teachers' subject matter preparation and teaching experience.

6.2 Implications

The findings of this study and the methods used have implications for at least four areas: 1) theory of mathematics teaching, 2) the practice of teacher education, 3) theory of measurement, and 4) item development. In the following sections, I describe implications of this study along with the relevant findings.

6.2.1 Implications for the development of theory of mathematics teaching

First, this study provides a way of allowing for the fine-grained description of teachers' mathematical knowledge by suggesting two organizers that distinguish multiple aspects of teachers' mathematical knowledge. By using two organizers, the suggested framework allows measuring not only generic aspects of teachers' knowledge across mathematics (by tasks of teaching) but also the knowledge aspects specific to the mathematics at stake in a course of studies (by instructional situations). Moreover, by examining teachers' knowledge in combinations of the mathematical work that teachers do and the mathematical work that students do in a particular course of study, the description of the teachers' knowledge takes into account the association between the task of teaching and the work students do.

Another contribution of this study is that the identified similarities and differences among knowledge dimensions could provide a reference back to the similarities and differences among different tasks of teaching or different instructional situations. The suggested knowledge framework provides a way to understand and describe the complexity of teaching practice in terms of what teachers do and what mathematical tasks students are expected to do. For example, by not being able to confirm a distinction between teachers' knowledge in USW_CG and USW_DP and look for an alternative explanation that would recognize similarities between the instructional situations CG and DP when teachers do the task USW, I have been able to contribute a conjecture that might support further examination of how tasks of teaching and instructional situations interact. As such, this study could contribute to the integration of a theory of tasks of teaching and a theory of instructional situations in light of the similarities and differences in mathematical work called forth from teachers and students.

This study also addresses the issue of elusive boundaries among traditionally defined knowledge domains (e.g., CK and PCK) by employing the alternative organizers, task of teaching and instructional situation. These organizers enabled developing multiple scales (continuous or discrete) representing multiple distinguishable dimensions of knowledge, while staying connected to the knowledge used in teaching practice.

6.2.2 Implications for the practice of teacher education

The benefits to teacher education of a multidimensional understanding of teachers' mathematical knowledge were described in 1.2.2. These potential benefits include tools for the diagnosis of teacher knowledge and resources for the design of teacher training and teacher certification or badging. In this section, the implications are described around these benefits that can be expected from the results of this study.

The results derived from multidimensional assessments of teachers' mathematical knowledge could inform the status of individual teachers' knowledge and this diagnosis could, if this is desired, contribute to the development of personalized teacher training programs. As the measures provide diagnostic information of the knowledge teachers have across different tasks of teaching and instructional situations, teachers (or teacher educators) can recognize their (viz., their students) specific weaknesses or strengths. The weaknesses and strengths assessed would be closely related to the actual work of teaching, so the diagnostic information could provide teachers and teacher educators with practical support in teaching through targeted professional development.

Specifically, this study's findings showed that the application of multidimensional models can provide detailed information about teachers' mathematical knowledge, such as knowledge profiles, which indicate which knowledge attributes teachers have mastered or have not mastered. For example, the result of applying DCM models allowed identifying teachers who are in need of gaining knowledge for doing a certain task in a given instructional situation. For example, the result showed that there is a considerable number of teachers (20%) who have attained a mastery level for the task understanding students' work, but not for the task choosing the givens for a problem in doing proofs. This result may suggest the need of a curriculum designed to spend more time on improving the knowledge required to design appropriate proof problems for teachers who lack this knowledge.

To have a curriculum map specified according to diverse tasks of teaching and instructional situations may not be necessary in general, given that the majority of the teachers had attained mastery level for all the dimensions or none of the dimensions. However, I would argue that there is benefit in having a specified curriculum for a particular group of teachers who need specialized training. For example, some teachers who had taught middle school mathematics but are going to teach high school geometry may need to attain knowledge required for doing tasks specific to the instructional situations in high school geometry. Beginning teachers may also need a specialized professional development for gaining mathematical knowledge that can only be attainable through years of teaching experience. In this regard, my proposed item blueprint could provide diagnostic information on this need and the way to design a curriculum with respect to the task of teaching and instructional situation specific to a particular need.

The specified knowledge dimensions also could be applied to a badging system. If mastery of a specific knowledge is recognized, a teacher could earn a badge. This system would enable teachers and teacher educators to identify new competency areas and signal to them a professional life of active learning and ongoing development.

The identified knowledge structure could offer a curriculum map laying out a sequence of multiple knowledge dimensions that may need to be taught in teacher education or professional development programs. As shown in the results of the DCM model, there is an implied hierarchical structure among knowledge attributes (e.g., very few teachers mastered USW_EF without mastering USW_CG or USW_DP). This finding could inform how to sequence a curriculum or training for teachers. For example, curriculum could be designed to follow the hierarchical sequence of knowledge attributes that would most quickly lead teachers to mastery level.

6.2.3 Implications for the theory of measurement

Methodologically, this study showed how different measurement models (SEM-based, IRT-based, and DCM-based models) could be used to provide complementary information on the dimensionality of a construct (i.e., knowledge). Most importantly, three different models yielded the same results regarding the factor-model comparisons.

How to model measures with categorical variables has been challenging for researchers as there is no single best method that can resolve a number of issues (e.g., data type, sample sizes, model parameterizations; see Wirth & Edwards, 2007). To offer a review on the methods, some studies compared the performance of SEM-based (WLSMV estimation) versus IRT-based models (FIML estimation). For example, Wang, Su, and Weiss (2018) compared SEM-based and IRT-based models (MIRT) when the normality assumption of the latent responses is violated, and they found that "FIML (MIRT models) consistently outperformed WLS (SEM models) when there were one or multiple skewed latent trait distributions" (p.403). Similarly, Wirth and Edwards (2007) compared several IFA models and estimation methods within SEM and IRT frameworks and presented the potential advantages and disadvantages of the methods (Wirth & Edwards, 2007).

Even though these studies suggested some recommendations to consider when applying measurement models with categorical data, the results and suggestions were based on simulated data rather than on real data collected from human participants. In addition, to determine which specific assumptions and conditions that the data used in a study satisfy or violate would be challenging for researchers. In this regard, I decided to apply all different measurement models that I considered for the data rather than gauging the appropriateness of each method.

The different models yielded consistent results. This consistency assured the reliability of the item parameters and of the results of the dimensionality estimated in this study. Inconsistencies in the results would have suggested that at least some of the results were not accurate. The consistency between IFA models (SEM and MIRT) and DCM models with real data further solidify the results of the dimensionality analysis. One of the issues in using IFA models with categorical variables is the violation of the multivariate normality assumption of the latent response variables (Wirth & Edwards, 2007; Wang et al., 2018). Considering that DCM models do not make a normality assumption for latent variables, the consistency of the results among the models implies that the violation of the normality assumption is not a concern in this study.

Moreover, the models provided different supplemental results supporting the claim of multidimensionality. The continuous scales of knowledge estimated from MIRT models provided results informing the effects teachers' educational and teaching experience had on their knowledge, whereas the binary scales of knowledge estimated from DCM models allowed criterion-referenced interpretations of teachers' knowledge across multiple knowledge attributes (e.g., how many more teachers have mastered USW_DP than other dimensions).

Lastly, this study showed the benefit of DCM models with respect to the reliability of the measures. In other words, diagnostic classification measurement models (DCM) yielded higher reliabilities for the estimates of the teachers' mastery status than the estimates of the amounts of teachers' mathematical knowledge derived from IRT-based models. This result informs researchers that a DCM model could be used as an alternative to an IRT model, as suggested by Bradshaw et al. (2014), if a construct can be presented on a binary scale and the construct is measured with a small number of items.

6.2.4 Implications for item development

The results of this study provide some insights into the development of items measuring distinguishable dimensions of teachers' mathematical knowledge for teaching. First, by showing that the suggested item blueprint allows for multiple distinguishable scales, this study provides guidance for future item development around this new organizational scheme, that is, creating items to focus on tasks of teaching and instructional situations.

Second, the results provide guidance for determining the grain size of each knowledge dimension. As discussed in Chapter 3, different researchers conceptualize tasks of teaching at different levels of specification. For example, Haertel (1991) described tasks using general categories, such as creating a structure for learning, which is not specific to the subject content. On the contrary, Ball et al. (2008) used finer distinctions in describing tasks of teaching by focusing on a mathematical difference among different tasks of teaching rather than a general purpose of tasks such as classroom management. Similarly, instructional situations could be described at various levels of detail (Chapter 4). For example, one could conceptualize a situation of exploration with two sub-situations according to different tools available to students in the situation (protractor and ruler vs. dynamic geometry software). In this regard, the results of distinction (e.g., USW_EF and USW_CG) or non-distinction (USW_CG and USW_DP) could help determine whether to operationalize a knowledge dimension as one scale or multiple scales. If two sets of items are valid and reliably measure two hypothesized constructs, but these constructs are not distinguishable from each other, researchers need to examine whether the lack of distinction is due to some common characteristics between two constructs or due to an unanticipated factor (e.g., item format, wording). If an unanticipated factor is identified, researchers may need to revise the items or develop new items and examine whether the hypothesized dimensions can be distinguishable with those items. On the other hand, if two sets of items are distinguishable from each other, researchers could examine if further distinctions in the form of attributes are feasible to make.

The benefits of this organizing scheme (combination of tasks of teaching and instructional situations) could extend beyond the field of mathematics education. The scheme could be considered for the items intended to measure distinguishable aspects of any professionals' knowledge. For example, assessment items measuring surgeons' medical knowledge could be structured in a way that combines surgeon's tasks and types of surgery. The example of types of surgery could be cardiothoracic surgery, plastic surgery, etc. The tasks could be categorized according to different time frames, such as preoperative care (e.g., diagnosis of the patient, surgical plan), intraoperative care (performing operation, treatment of unanticipated findings), and postoperative care (treatment of complications) (Griffen et al., 2007).

6.3 Limitations and future research

While the results of this study support the item-blueprint organized by tasks of teaching and instructional situations, there are several limitations pointing to the need for further research. First, the small number of items, ranging from three to six for one dimension, is a clear limitation, which threatens content validity, in particular, "construct underrepresentation" (Furr & Bacharach, 2014). In other words, three items may not cover the full range of content that is relevant to a targeted knowledge construct.

The limitation of small number of items in this study was also revealed in the low reliabilities in the IRT scales. For example, the dimensions formed by three items (e.g., USW_SR) yielded low reliabilities for the entire range of the construct. The results also showed that most IRT scales provided relatively precise measurements for teachers with weak knowledge, but they were less precise for teachers with strong knowledge. While identifying teachers who are in need of supports to gain knowledge for teaching geometry or algebra could have more practical importance than identifying teachers who are strong in knowledge for teaching geometry, additional items are necessary to measure all levels of knowledge with high reliability.

To mitigate the issue of low reliability, this study applied DCM models as a way to improve reliability of knowledge estimates. The DCM models indeed provided better reliability for knowledge estimates (mastery or lack of mastery) even with three items. However, again, to develop reliable scales representing teachers' knowledge on a continuous scale, items for each hypothesized dimension need to be developed. Otherwise, the definition of a hypothesized dimension or conclusions of this study may need to be narrowed.

In this study, the small number of items for each dimension was, however, inevitable in that some of the items were not initially developed by the suggested item blueprint. To evaluate the suggested framework, the items were retrofitted to the models, which classified items by tasks of teaching and instructional situations. In the process of retrofitting, several items were excluded when they involved a unique task or instructional situation and did not fit into a cluster with other items. The exclusion of those items might have limited the possibility of a holistic understanding of the characteristics of teachers' mathematical knowledge. In the same line of thought, two tasks of teaching and five instructional situations do not fully represent all the variation that could be found among the universe of tasks of teaching or instructional situations. To generalize my argument on the role of tasks of teaching and instructional situations as related to the dimensionality of teachers' mathematical knowledge for teaching geometry, more diverse items that represent various tasks of teaching and instructional situations need to be developed.

Apart from the issue of the small number of items, not all the items used in this study were initially developed by the suggested framework and this leads to the possibility that factors other than the organizers could have influenced the degree to which the item clusters were associated with each other. For example, the reason for the lack of distinction between USW_CG and USW_DP items could be due to construct underrepresentation. Even though lack of distinction was interpretable in consideration of students' problem-solving process in two instructional situations, the interpretations are still conjectural. This means that the interpretations cannot fully support the argument that the two knowledge constructs are not distinguishable because of the similarities in teachers' mathematical knowledge. To address this issue, further investigation is needed to determine whether the two dimensions are indeed distinguishable or not with more items that are different only in the involved instructional situations, but the same regarding other factors, such as mathematical topic, item format, or wording style. Systematically developed test items would allow for the generalization of my argument on the feasibility of the suggested framework to measure distinguishable dimensions of teachers' mathematical knowledge for teaching. To do this, a study which incorporates various tasks of teaching and various instructional situations, supported by a theory of tasks of teaching and a theory of instructional situations, may need to be further investigated.

Another limitation could be a possibility of biased measurement model parameters due to the existence of "person heterogeneity", which "occurs when a trait structure is qualitatively different for distinct examinee samples" (Reise & Gomel, 1995, p.342). Given the significant relationship between years of teaching geometry and teachers' MKT-G (Chapter 5 in this study), there could be a possibility that the same measurement model does not fit the data from novice teachers and experienced teachers equally. The hypothesis regarding the difference in the dimensionality of novice teachers and experienced teachers is also reasonable given the dependency of depth and structure of knowledge (Jong & Ferguson-Hessler, 1996) and the differences among novice teachers and expert teachers (Berliner, 2001). In other words, there is a high probability that experienced (or expert) teachers' knowledge dimensions are more organically linked to each other than are those of novice teachers. Krauss et al. (2008) also showed that content knowledge and pedagogical content knowledge were not distinguishable for teachers with strong content knowledge, whereas the two knowledge domains were distinguishable for less knowledgeable teachers. Thus, more dimensions may be needed to measure novice teachers' mathematical knowledge, while fewer may be needed to measure expert teachers' mathematical knowledge. If there is a strong evidence that the population of teachers is not homogeneous, a mixed-measurement model that accounts for heterogeneity in measurement may need to be applied (Cohen & Bolt, 2005) to the data. Specifically, in a future study, a DIF (differential item functioning) method can be applied to the data with a variable indicating a novice or experienced teacher to identify DIF items. If DIF is identified, a two-group mixture IRT model could be applied to allow different item parameter values to different groups of teachers. Any difference in the dimensionality of knowledge can then be investigated between groups.

Further study can also examine which kinds of experience teachers have in the total years of experience so that the extent to which different kinds of experience teaching geometry (or algebra 1) can increase teachers' mathematical knowledge for teaching geometry (or algebra 1) can be examined. If a future study identifies experience-specific relationships across different knowledge dimensions, then professional development could be designed around the experiences that can accelerate improvement of teachers' mathematical knowledge for teaching geometry (or algebra 1).

Lastly, the current study is limited in its ability to capture the specific aspect of teachers' mathematical knowledge required in the actual practice of teaching. Even though all the items are situated in a specific context of teaching, the context was presented by verbal-vignettes instead of visual-vignettes (e.g., animation, video) or in real settings. The context described in writing requires teachers to hypothesize the situation rather than experience the situation while solving the items. Thus, teachers' knowledge required to appropriately interpret and recognize a context could not be captured. This limitation may have caused some hypothesized dimensions to be indistinguishable.

The use of an instrument that requires respondents' interpretations of events has relevance for construct measurement, which is context-specific (Ambrose, Clement, Philipp, & Chauvot, 2004). This study hypothesized differences between knowledge dimensions, which reflect different instructional situations that imply different expectations for the teachers' and students' work. What is therefore needed is an instrument that can provide teachers with scenarios in which they are called on to recognize the situation. With this consideration, future work can use animations or storyboards that depict classroom interaction to present the different contexts of teaching in measuring teachers' mathematical knowledge. The viability of scenario-based instruments has been reported in other studies (Herbst & Chazan, 2015; Herbst, Chazan, Kosko, Dimmel, & Erickson, 2016). For example, animations of cartoon characters were used to investigate the norms that govern classroom interactions, which are expected in the instructional situations "installing a theorem in geometry" and "doing word problems in school algebra" (Herbst, Nachlieli, & Chazan, 2011; Chazan, Sela, & Herbst, 2012). In particular, among the studies measuring teachers' mathematical knowledge, the TEDS-M group developed and used a video-vignette assessment to measure teachers' general pedagogical knowledge, which is context-dependent, in their follow-up study TEDS-FU (Blömeke, Kaiser, & König, 2009; Klein, Suhl, Busse, & Kaiser König, Blömeke, Klein, Suhl, Busse, & Kaiser, 2014).

Similar to these studies, future research could develop items that present an intended instructional situation using animation or storyboard to measure teachers' mathematical knowledge used in the situations. Such approaches would remove the reliance on written word and better capture the aspect of situational teacher knowledge.

6.4 Concluding remarks

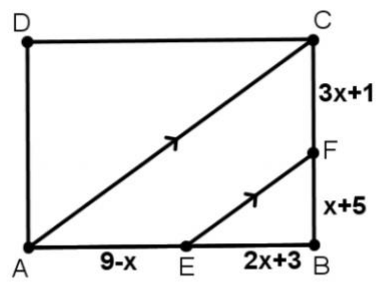
"Methodological sophistication cannot substitute for theoretical cogency" (Garvin & Kirkland, 1977, p.24), whereas theoretical cogency cannot guarantee sound operation. The challenge of attaining both methodologically and theoretically sound knowledge framework has motivated this study. Despite some limitations that need further research, this study has shown the possibility of measuring multiple dimensions of teachers' mathematical knowledge in terms of multiple scales, which has not yet been documented in the mathematics education literature. Having described the suggested framework and the results supporting it, my work shows a promise of an item blueprint that is not only theoretically warranted but also methodologically feasible.

Appendix

Sample Items

USW_CG (Understanding Students' Work in Calculation in Geometry)

Mr. Collins assigned his geometry class a problem using the diagram below. He asked his students to find the value of x if $\overline{AC} \parallel \overline{EF}$.

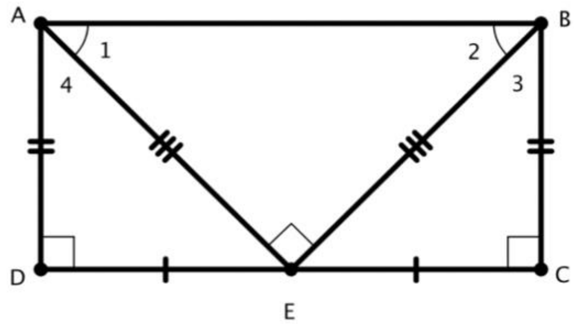


The following responses are among the ones his students produced. While all students arrived at the correct answer, only one appears to have used a valid method. Which one used a valid method?

$(2x+3) = (x+5)$ $2x - x = 5 - 3$ $x = 2$
$\frac{2x+3}{9-x} = \frac{x+5}{3x+1}$ $(2x+3)(3x+1) = (x+5)(9-x)$ $6x^2 + 2x + 9x + 3 = 9x - x^2 + 45 - 5x$ $7x^2 + 7x - 42 = 0$ $x^2 + x - 6 = 0$ $(x-2)(x+3) = 0$ $x = 2$ or $x = -3$ If $x = -3, BE < 0$. So $x = 2$ is the only answer.
$(9-x) = (2x+3)$ $9 - 3 = 2x + x$ $6 = 3x$ $2 = x$
$(9-x) + (2x+3) = (3x+1) + (x+5)$ $(12+x) = 4x + 6$ $12 - 6 = 4x - x$ $6 = 3x$ $2 = x$

USW_DP (Understanding Students' Work in Doing Proof)

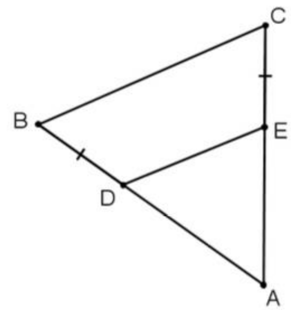
While proving a claim on the board about the figure below, Joe wrote "1 + 2 = 90." Ms. Staples ponders how to correct that statement. Of the following, what is the best alternative?



A	Do nothing. The statement is correct as is.
B	Add the degree symbol so that the statement reads 1 + 2 = 90°.
C	Replace what Joe wrote; write instead that "m∠A + m∠B = 90°".
D	Replace what Joe wrote; write instead that "m∠EAB + m∠EBA = 90°".

USW_EF (Understanding Students' Work in Exploring a Figure)

Mr. King assigned his geometry students an open-ended problem. He drew a diagram on the board (shown below) in which $\overline{BD} \cong \overline{CE}$. He asked students to "Pick an additional given and say what conjecture that additional given would allow you to make about the figure. Explain your reasoning."



The work of four students is shown below.

Alex	If I could assume angle C is greater than angle B, I think $DA > AE$. Because the side opposite to the greater angle is longer than the side opposite to the smaller angle. Then $BA > CA$. So, $DA > AE$, by subtraction of equals.
Bert	I think if $AD = AE$ is given, then $\overline{DE} \parallel \overline{BC}$. Because if a line divides any two sides of a triangle in the same ratio, the line is parallel to the third side.
Gabby	I think if $\overline{DE} \parallel \overline{BC}$, then $\frac{AD}{DB} = \frac{AE}{EC} = \frac{DE}{BC}$. Because ADE and ABC are two similar triangles, their corresponding sides are proportional.
Delia	I think if D is the midpoint of $\overline{AB}$, I can prove that ΔADE and ΔABC are similar. Because $\angle A$ is shared by both triangles and the ratios of AD to AB and AE to AC are equal.

Which of Mr. King's students are correct?

A. Only Alex, Bert and Gabby are correct.

B. Only Alex and Bert are correct.

C. Only Bert and Gabby are correct.

D. Only Gabby and Delia are correct.

USW_SR (Understanding Students' Work in Simplifying Rational expressions)

$$\frac{2(a+1)}{3a} + 3 - \frac{2}{3a} - \frac{6a-2}{6}$$
$$= \frac{2a+2}{3a} + 3 - \frac{2}{3a} - \frac{6a}{6} + \frac{2}{6}$$
$$= \frac{2a}{3a} + 2\frac{6}{6} + \frac{2}{6} - a$$
$$= \frac{2}{3} + 2\left(\frac{6}{6} + \frac{1}{6}\right) - a$$
$$= \frac{2}{3} + 2\frac{7}{6} - a$$
$$= \frac{4}{6} + \frac{14}{6} - a$$
$$= \frac{18}{6} - a$$
$$= 3 - a$$

Mr. Anderson asked his students to simplify the following algebraic expression.

$$\frac{2(a+1)}{3a} + 3 - \frac{2}{3a} - \frac{6a-2}{6}$$

One of his students gave the incorrect solution show to the left.

Of the following descriptions, which best characterizes what is wrong with this student's work?

A	This student used the distributive property incorrectly.
B	This student confounded mixed fractions with factors
C	This student forgot to cancel common factors in several places.
D	This student needs to apply a more formal procedure by finding the common denominator and then adding all terms.

USW_SE (Understanding Students' Work in Solving Equations)

Having taught her students to factor quadratics with integer coefficients, integer roots, and a leading coefficient of 1, Ms. Quezada explained that she was going to give them a harder problem. She then asked them to solve the following.

$$3x^2 - 3x - 6 = 0$$

After a few minutes of work, the class discussed their solutions. Letitia said that x was –1 or 2 and explained, "I added 3x to both sides and divided by 3."

$$3x^2 - 6 = 3x$$

$$x^2 - 2 = x$$

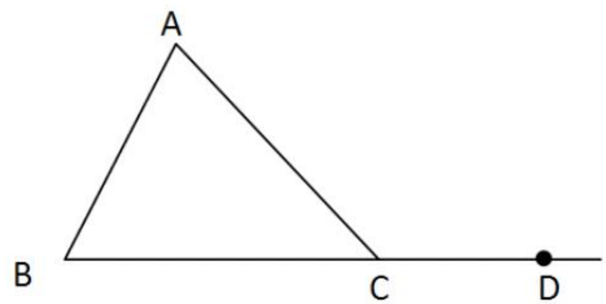
She then continued, "The parabola's just down a little and the line's at 45 degrees, so it's just below zero and about 2 to the right. x can be –1 and 2, and those are the only possible ones."

Of the following, which best characterizes Letitia's approach to this problem?

A	Letitia's <u>method</u> is wrong because she should have first divided by 3 and then factored the left side of the equation.
B	Letitia's <u>method</u> is wrong because this is a parabola and you could graph it, but you would have to graph the original equation and look for the roots.
C	Letitia's <u>reasoning</u> is correct, but her method often leads to points of intersection that might be hard to determine visually.
D	Letitia's <u>reasoning</u> is correct, but her method requires knowledge of calculus.

CGP_CG (Choosing appropriate Givens for a Problem in Calculation in Geometry)

Mrs. Marton's geometry class has just studied the angle sum and exterior angle theorems for triangles. She is now looking for a problem that would allow her students to apply the exterior angle theorem. She is writing a problem for a quiz in which students will be given algebraic expressions that stand for the angle measures in the triangle below and will ask them to "Find the value of x." Of the following sets of algebraic expressions, indicate which one is the best for her to use.



A. $m\angle A = x + 18, m\angle B = x,\ m\angle ACD = 3x - 22$

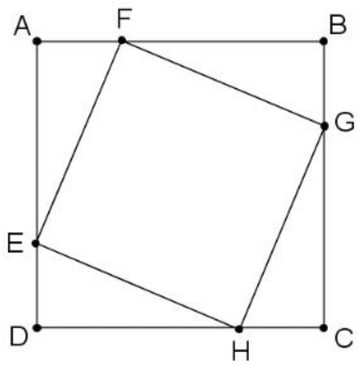
B. $m\angle A = x + 3, m\angle\ B\ = x, m\angle\ ACD = 2x - 10$

C. $m\angle A = 4x - 45, m\angle B\ = x, m\angle ACD = 3x + 45$

D. Option A and C are equally good

CGP_DP (Choosing appropriate Givens for a Problem in Doing Proofs)

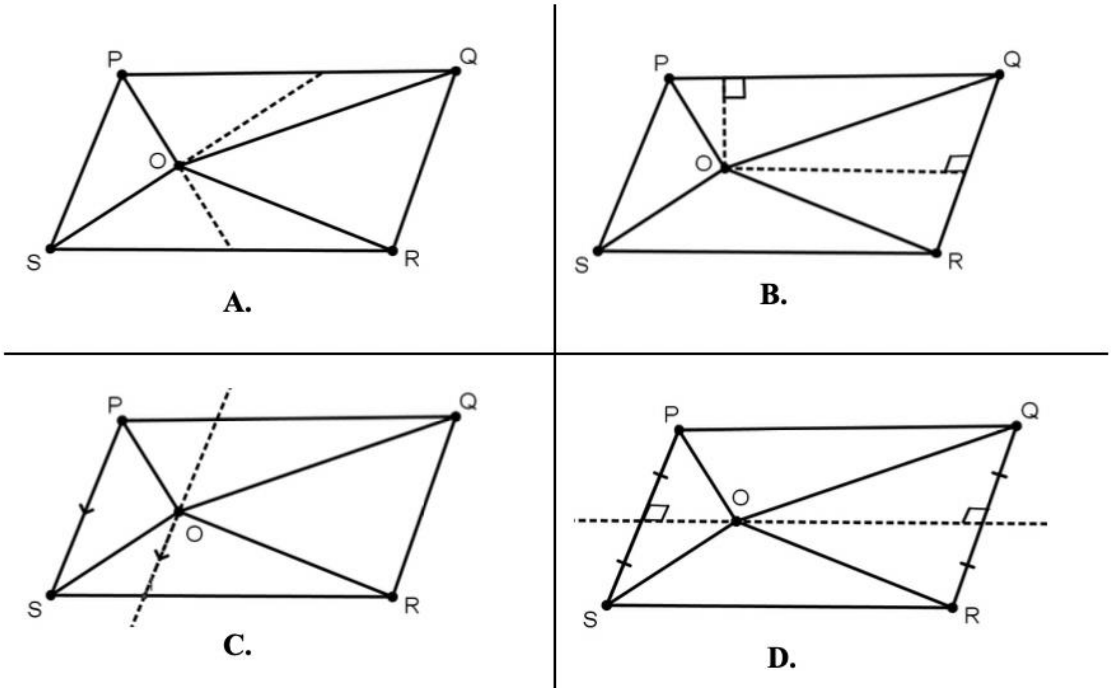
Mr. Taylor wants to give his geometry class an assignment and he is considering various possible proof problems about squares. For each of the proof problems below, indicate whether or not the students have enough information to do the proof.



	Would students have what they need to complete the proof?	Yes	No
i.	Given: *ABCD* is a square. $\overline{AF} \cong \overline{BG} \cong \overline{CH} \cong \overline{DE}$ Prove: *EFGH* is a square		
ii.	Given: $\Delta GBF \cong \Delta FAE \cong \Delta EDH \cong \Delta HCG$ Prove: *EFGH* is a square		
iii.	Given: *ABCD* is a square. $\angle BFG \cong \angle AEF \cong \angle DHE \cong \angle CGH$ Prove: *EFGH* is a square		
iv.	Given: *ABCD* is a square. $\overline{AF} \cong \overline{HC}, \overline{BG} \cong \overline{DE}$ Prove: *EFGH* is a square		

CGP_EF (Choosing appropriate Givens for a Problem in Exploring a Figure)

After finishing area of plane figures, Mr. Reed draws a diagram on the board, in which O is a point inside a parallelogram and will ask his students to make conjectures. To help students make valid conjectures on the areas of the triangles, Mr. Reed will draw auxiliary lines on the diagram. Of the auxiliary lines shown in dotted pattern below, identify the one that would best help students make mathematically correct conjectures.



CGP_CN (Choosing appropriate Givens for a Problem in Calculating with Numbers)

A lesson in Ms. Taylor's textbook states the associative and commutative properties of addition. To motivate the students to learn the properties, she tells her students that the properties can often be used to simplify the evaluation of expressions.

She wants to give her students an example that will focus their attention on how these properties can be useful in evaluating expressions. Of the following expressions, which would best serve her purpose?

A	(455 + 456) + (457 + 458)
B	(647 + 373) + (227 + 456)
C	(551 + 775) + (49 + 225)
D	Each of these expressions would serve her purpose equally well.

Bibliography

Aaron, W. R., & Herbst, P. G. (2015). Teachers' perceptions of students' mathematical work while making conjectures: an examination of teacher discussions of an animated geometry classroom scenario. *International Journal of STEM Education*, *2*(1), 1-13.

Ambrose, R., Clement, L., Philipp, R., & Chauvot, J. (2004). Assessing prospective elementary school teachers' beliefs about mathematics and mathematics learning: Rationale and development of a constructed-response-format beliefs survey. *School Science and Mathematics*, *104*(2), 56-69.

Association of Mathematics Teacher Educators. (2017). *Standards for Preparing Teachers of Mathematics.* Available online at amte.net/standards.

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

Ball, D. L., & Bass, H. (2003b). Toward a practice-based theory of mathematical knowledge for teaching. In B. Davis & E. Simmt (Eds.), *Proceedings of the 2002 annual meeting of the Canadian Mathematics Association Study Group* (pp. 3-14). Edmonton, AB: CMESG/GDEDM.

Ball, D. L., & Forzani, F. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*(5), 497-511.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching what makes it special? *Journal of teacher education*, 59(5), 389-407.

Ball, D. L., Sleep, L., Boerst, T. A., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. *The Elementary School Journal, 109*(5), 458-474.

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. *Handbook of research on teaching*, 4, 433-456.

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. New York: Guilford Publications.

Baumert, J., Kunter, M., Werner, B., Brunner, M.,Voss, T., Jordan, A., . . . Tsai, Y. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133-180.

Begle, E. (1979). *Critical variables in mathematics education: Findings from a survey of empirical research.* Mathematical Association of America/National Council of Teachers of Mathematics, Washington, DC.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International journal of educational research*, *35*(5), 463-482.

Blömeke, S., Houang, R. T., & Suhl, U. (2014). Diagnosing Teacher Knowledge by Applying Multidimensional Item Response Theory and Multiple-Group Models. In *International Perspectives on Teacher Knowledge, Beliefs and Opportunities to Learn* (pp. 483-501). Springer, Netherlands.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education*, *56,* 35-46.

Boileau, N. & Herbst, P. (2015, November). *Teachers' Expectation About Geometric Calculations in High School Geometry.* Paper to be presented at PME-NA, East Lansing, MI.

Bradshaw, L., Izsák, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic classification framework. *Educational measurement: Issues and practice*, *33*(1), 2-14.

Buchbinder, O., Chazan, D. I., & Capozzoli, M. (2019). Solving Equations: Exploring Instructional Exchanges as Lenses to Understand Teaching and Its Resistance to Reform. *Journal for Research in Mathematics Education*, *50*(1), 51-83.

Conference Board of the Mathematical Sciences (CBMS). (2012). *The mathematical education of teachers II*. Providence, RI, and Washington, DC: American Mathematical Society and Mathematical Association of America. Retrieved from https://www.cbmsweb.org/archive/MET2/met2.pdf

Chazan, D., & Lueke, H. M. (2009). Exploring relationships between disciplinary knowledge and school mathematics: Implications for understanding the place of reasoning and proof in school mathematics. In D. Stylianou, E. Knuth, & M. Blanton (Eds.), *Teaching and learning mathematics proof Across the grades* (pp. 21-39). Erlbaum: Hillsdale, NJ.

Chazan, D., Sela, H., & Herbst, P. (2012). Is the role of equations in the doing of word problems in school algebra changing? Initial indications from teacher study groups. *Cognition and Instruction*, *30*(1), 1-38.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41*(4), 778-820.

Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, *42*(2), 133-148.

Cooney, T. J., Davis, E. J., & Henderson, K. B. (1975). *Dynamics of teaching secondary school mathematics*. Boston, MA: Houghton Mifflin Company.

Cunningham, R. F., & Yacone, E. (2013). Algebra Textbooks' and Teachers' Methods: Simplifying rational expressions. *International Journal of Research in Educational Methodology*, *4*(1), 457-468.

De Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational psychologist*, *31*(2), 105-113.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods*, *4*(2), 144-192.

ElHitti, S., Bonanome, M., Carley, H., Tradler, T., Zhou, L., (2017). *Arithmetic | Algebra*, *CUNY Academic Works*. Retrieved from: http://academicworks.cuny.edu/ny_oers/5

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling*, *8*(3), 430-457.

Etkina, E., Gitomer, D., Iaconangelo, C., Phelps, G., Seeley, L., & Vokos, S. (2018). Design of an assessment to probe teachers' content knowledge for teaching: An example from energy in high school physics. *Physical Review Physics Education Research*, *14*(1), 010127.

Educational Testing Service. (2018). *The Praxis ® Study companion: Mathematics: Content knowledge*. Retrieved from: http://www.ets.org/s/praxis/pdf/5161. pdf.

Fernández, S., Figueiras, L., Deulofeu, J., & Martínez, M. (2011). Re-defining HCK to approach transition. In *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education* (pp. 2640-2649).

Ferrini-Mundy, J., & Findell, B. (2001). The mathematical education of prospective teachers of secondary school mathematics: Old assumptions, new challenges. In *CUPM discussion papers about mathematics and the mathematical sciences in 2010: What should students know?* (pp. 31–41). Washington, DC: Mathematical Association of America

Franklin, C., Bargagliotti, A., Case, C., Kader, G., Schaeffer, R., & Spangler, D. (2015). *Statistical education of teachers*. Arlington, VA: American Statistical Association. Retrieved from https://www.amstat.org/asa/files/pdfs/EDU-SET.pdf

Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology.* SAGE Publications Ltd.

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. (2nd Ed.). SAGE Publication.

Gitomer, D., Phelps, G., Weren, B., Howell, H., & Croft, A. (2014). Evidence on the validity of content knowledge for teaching assessments. In T. Kane, K. Kerr, & R. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 493–528). San Francisco: Jossey-Bass.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual review of psychology*, *40*(1), 631-666.

Good, T. L., & Grouws, D. A. (1979). The Missouri Mathematics Effectiveness Project: An experimental study in fourth-grade classrooms. *Journal of Educational Psychology, 71*(3), 355-362.

Grossman, P. L., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45*(1), 184–205.

Haertel, E. H. (1991). New forms of teacher assessment. *Review of Educational Research, 17*, 3-29.

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of public economics*, *95*(7-8), 798-812.

Heid, M. K., Blume, G. W., Zbiek, R. M., & Edwards, B. S. (1998). Factors that influence teachers learning to do interviews to understand students' mathematical understandings. *Educational Studies in Mathematics, 37*(3), 223-249.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191-210.

Herbst, P. and Brach, C. (2006). Proving and 'doing proofs' in high school geometry classes: What is 'it' that is going on for students and how do they make sense of it? *Cognition and Instruction, 24*(1), 73-122.

Herbst, P. & Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM The International Journal of Mathematics Education,* 44(5), 601-612.

Herbst, P. and Chazan, D. (2015). Using multimedia scenarios delivered online to study professional knowledge use in practice. *International Journal of Research and Method in Education*, *38*(3), 272-287

Herbst, P., Chazan, D., Kosko, K. W., Dimmel, J., & Erickson, A. (2016). Using multimedia questionnaires to study influences on the decisions mathematics teachers make in instructional situations. *ZDM*, *48*(1-2), 167-183.

Herbst, P., & Kosko, K. W. (2014). Mathematical knowledge for teaching and its specificity to high school geometry instruction. In J. Lo, K. R. Leatham, & L. R. Van Zoest (Eds.), *Research trends in mathematics teacher education* (pp. 23-46). New York: Springer.

Herbst, P. (2004). Interactions with Diagrams and the Making of Reasoned Conjectures in Geometry. ZDM, *Zentralblatt für Didaktikk der Mathematikk, 36*(5), 129-139

Herbst, P. Chen, C., Weiss, M., and González, G., with Nachlieli, T., Hamlin, M., and Brach, C. (2009). "Doing proofs" in geometry classrooms. In M. Blanton, D. Stylianou, and E. Knuth (Eds.), *Teaching and learning of proof across the grades: A K-16 perspective* (pp. 250- 268). New York: Routledge.

Herbst, P. with González, G., Hsu, H. Y., Chen, C., Weiss, M., and Hamlin, M. (2010). Instructional situations and students' opportunities to reason in the high school geometry class. Manuscript. Deep Blue at the University of Michigan. http://hdl.handle.net/2027.42/78372

Herbst, P. (2002b) 'Establishing a custom of proving in American school geometry: evolution of the two-column proof in the early twentieth century'. *Educational Studies in Mathematics 49*(3), 283-312.

Herbst, P. (2006). Teaching geometry with problems: Negotiating instructional situations and mathematical tasks. *Journal for Research in Mathematics Education, 37*(4), 313-347.

Herbst, P., Aaron, W., Dimmel, J., and Erickson, A. (2013, April). Expanding students' involvement in proof problems: Are geometry teachers willing to depart from the norm? Paper presented at the 2013 meeting of the American Educational Research Association. Deep Blue at the University of Michigan. http://hdl.handle.net/2027.42/97425

Herbst, P., Nachlieli, T., and Chazan, D. (2011). Studying the practical rationality of mathematics teaching: What goes into “installing” a theorem in geometry? *Cognition and Instruction, 29*(2), 218-255.

Hill, H. C. (2007). Mathematical knowledge of middle school teachers: Implications for the No Child Left Behind policy initiative. *Educational Evaluation and Policy Analysis, 29*(2), 95-114.

Hill, H. C. (2010). The nature and predictors of elementary teachers' mathematical knowledge for teaching. *Journal for Research in Mathematics Education, 41*(5), 513-545.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for research in mathematics education, 39*(4), 372-400.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers’ mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30.

Hill, H. C., (2016, July 27) Measuring Secondary Teachers’ Knowledge of Teaching Mathematics: Developing a Field, Hamburg.

Hoover, M. Mosvold, R., & Fauskanger, J. (2014). Common tasks of teaching as a resource for measuring professional content knowledge internationally. *Nordic Studies in Mathematics Education, 19* (3-4), 7-20.

Thames, M. H. (2009). *Coordinating mathematical and pedagogical perspectives in practice-based and discipline-grounded approaches to studying mathematical knowledge for teaching (K-8)*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.

Hsu, H-Y, & Silver, E. A. (2014). Cognitive complexity of mathematics instructional tasks in a Taiwanese classroom: An examination of task sources. *Journal for Research in Mathematics Education, 45*(4), 460- 496.

Hsu, H. (2010). *The study of Taiwanese students’ experiences with geometric calculation with number (GCN) and their performance on GCN and geometric proof (GP)*. Unpublished doctoral dissertation, The University of Michigan, Ann Arbor, MI.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Jonassen, D.H., Tessmer, M., & Hannum, W.H. (1999). *Task analysis methods for instructional design*. Mahwah, NJ: L. Erlbaum Associates

Jurich, D. P., & Bradshaw, L. P. (2014). An illustration of diagnostic classification modeling in student learning outcomes assessment. *International Journal of Testing*, *14*(1), 49-72.

Kane, M. T. (2006). Validation. *Educational measurement*, *4*(2), 17-64.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1-73.

König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., & Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education*, *38*, 76-88.

Kline, R. B. (2010). *Principles and practice of structural equation modeling*. New York: Guilford publications.

Ko, I., & Herbst, P. (2017). Investigating the Dimensionality of the Instrument Measuring Teachers' Mathematical Knowledge for Teaching Secondary Geometry. Paper presented at AERA Annual Meeting, San Antonio, TX. April 2017.

Ko, I., & Herbst, P. (2017, under review). Subject Matter Knowledge of Geometry Needed in Tasks of Teaching: Relationship to Prior Geometry Teaching Experience. Manuscript submitted for publication.

Ko, I., & Herbst, P. (2019). Using Assessment Items to Identify Distinguishable Dimensions of Teachers' Mathematical Knowledge for Teaching Secondary Geometry. Poster will be presented at the Annual Meeting of the American Educational Research Association. Toronto, Canada.

König, J., Blömeke, S., & Kaiser, G. (2015). Early career mathematics teachers' general pedagogical knowledge and skills: Do teacher education, teaching experience, and working conditions make a difference? *International Journal of Science and Mathematics Education*, *13*(2), 331-350.

Koponen, M., Asikainen, M. A., Viholainen, A., & Hirvonen, P. E. (2016). Teachers and their Educators-Views on Contents and their Development Needs in Mathematics Teacher Education. *The Mathematics Enthusiast*, *13*(1), 149-170.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, *100*(3), 716–725.

Krauss, S., Baumert, J., & Blum, W. (2008). Secondary mathematics teachers’ pedagogical content knowledge and content knowledge: validation of the COACTIV constructs. *ZDM*, *40*(5), 873-892.

Lee, D. (1986). Academic achievement, task characteristics, and first job performance of young engineers. *IEEE Transactions on Engineering Management*, EM-33(3), 127–133.

Lehrer, R., & Franke, M. L. (1992). Applying personal construct psychology to the study of teachers' knowledge of fractions. *Journal for Research in Mathematics Education, 23*(3), 223-241.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural equation modeling, 9*(2), 151-173.

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). *Descriptive Analysis in Education: A Guide for Researchers*. NCEE 2017-4023. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Manouchehri, A., & Goodman, T. (2000). Implementing mathematics reform: The challenge within. *Educational Studies in Mathematics, 42*(1), 1-34.

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D., & Senk, S. L. (2012). Knowledge of algebra for teaching. *Journal for Research in Mathematics Education*, *43*(5), 584-615.

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

Mohr-Schroeder, M., Ronau, R. N., Peters, S., Lee, C. W., & Bush, W. S. (2017). Predicting student achievement using measures of teachers' knowledge for teaching geometry. *Journal for Research in Mathematics Education*, *48*(5), 520-566.

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

Muthén, L.K. and Muthén, B.O. (1998-2015). Mplus User’s Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén

National Council of Teachers of Mathematics and the Council for the Accreditation of Educator Preparation (NCTM & CAEP). (2012b). *NCTM CAEP Standards— Secondary (Initial Preparation)*. Retrieved from http://www.nctm.org/uploadedFiles/Standards_and_Positions/CAEP_Standards/NCTM%20CAEP%20Standards%202012%20-%20Secondary.pdf

New Zealand Ministry of Health. (2014). *Knowledge and Skills Framework for Cancer Nursing (KSFCN).* Wellington. Ministry of Health.

NCTM (1991). *Professional standards for teaching mathematics*. Reston, VA, National Council of Teachers of Mathematics.

Pepin, B., & Haggarty, L. (2001). Mathematics textbooks and their use in English, French and German classrooms. *ZDM Mathematics Education*, *33*(5), 158-175.

Petrou, M., & Goulding, M. (2011). Conceptualising teachers' mathematical knowledge in teaching. In *Mathematical knowledge in teaching* (pp. 9-25). Springer Netherlands.

Phelps, G., & Schilling, S. (2004). Developing measures of content knowledge for teaching reading. *The Elementary School Journal*, *105*(1), 31-48.

Phelps, G., Weren, B., Croft, A., & Gitomer, D. (2014). Developing Content Knowledge for Teaching Assessments for the Measures of Effective Teaching Study. *ETS Research Report Series, 2014*(2), 1–92.

Phelps, G., Kelcey, B., Jones, N., & Liu, S. (2016). Informing estimates of program effects for studies of mathematics professional development using teacher content knowledge outcomes. *Evaluation review*, *40*(5), 383-409.

Reckase, M. (2009). *Multidimensional item response theory*. New York: Springer.

McCrory, R., Floden, R., Ferrini-Mundy, J., Reckase, M. D., & Senk, S. L. (2012). Knowledge of algebra for teaching: A framework of knowledge and practices. *Journal for Research in Mathematics Education*, *43*(5), 584-615.

Reise, S. P., & Gomel, J. N. (1995). Modeling qualitative variation within latent trait dimensions: Application of mixed-measurement to personality assessment. *Multivariate behavioral research, 30*(3), 341-358.

Reynolds, A. (1992). What is competent beginning teaching? A review of the literature. *Review of educational research*, *62*(1), 1-35.

Reynolds, A. (1995). The knowledge base for beginning teachers: Education professionals' expectations versus research findings on learning to teach. *The Elementary School Journal, 95*(3), 199–221.

Robová, J. (2013). The impact of web sites on teaching and learning mathematics. *Procedia-Social and Behavioral Sciences, 93*, 631-635.

Rowan, B., Chiang, F. S., & Miller, R. J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education, 70*(4), 256-284.

Rowland, T., & Ruthven, K. (2011). Introduction: Mathematical knowledge in teaching. In *Mathematical knowledge in teaching* (pp. 1-5). Springer Netherlands.

Royal College of Nursing (Great Britain). (2005). *NHS knowledge and skills framework outlines for nursing posts: RCN guidance for nurses and managers in creating KSF outlines in the NHS*. London: RCN.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic measurement: Theory, methods, and applications. New York: Guilford Press.

Samejima, F. (1997). Graded response model. In W.J. Van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory.* New York: Springer. Chap. 5.

Schilling, S., & Bock, R. D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70*(3), 533-555.

Schilling, S. G. (2007). The role of psychometric modeling in test validation: An application of multidimensional item response theory. *Measurement, 5*(2-3), 93-106.

Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education*. Boston: Kluwer Academic.

Senk, S. L., Tatto, M. T., Reckase, M., Rowley, G., Peck, R., & Bankov, K. (2012). Knowledge of future primary teachers for teaching mathematics: an international comparative study. *ZDM Mathematics Education, 44*(3), 307-324.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton, Mifflin and Company.

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

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

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1-23.

Silverman, J., & Thompson, P. W. (2008). Toward a framework for the development of mathematical knowledge for teaching. *Journal of mathematics teacher education*, *11*(6), 499-511.

Speer, N. M., King, K. D., & Howell, H. (2015). Definitions of mathematical knowledge for teaching: using these constructs in research on secondary and college mathematics teachers. *Journal of Mathematics Teacher Education*, *18*(2), 105-122.

Steffe, L. P. (2004). On the construction of learning trajectories of children: The case of commensurate fractions. *Mathematical thinking and learning*, *6*(2), 129-162.

Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). Teacher Education and Development Study in Mathematics (TEDS-M): *Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.

Tatto, M. (2014). Teacher Education Development Study-Mathematics (TEDS-M). In *Encyclopedia of Mathematics Education* (pp. 586–592). Dordrecht: Springer Netherlands

Tchoshanov, M. A. (2011). Relationship between teacher knowledge of concepts and connections, teaching practice, and student achievement in middle grades mathematics. *Educational Studies in Mathematics*, *76*(2), 141-164.

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, *30*(2), 251-275.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, *11*(3), 287.

Thompson, P. W. (2015). Researching mathematical meanings for teaching. In English, L., & Kirshner, D. (Eds.), *Third Handbook of International Research in Mathematics Education* (pp. 435-461). London: Taylor and Francis.

Tirosh, D., & Even, R. (2007). Teachers' knowledge of students' mathematical learning: An examination of commonly held assumptions. In *Mathematics knowledge in teaching seminar series: Conceptualising and theorizing mathematical knowledge for teaching (Seminar1)*. Cambridge, MA: University of Cambridge.

Vanthienen, J., & De Witte, K. (Eds.). (2017). *Data Analytics Applications in Education*. Boca Raton: CRC Press.

Wang, C., Su, S., & Weiss, D. J. (2018). Robustness of parameter estimation to assumptions of normality in the multidimensional graded response model. *Multivariate behavioral research*, *53*(3), 403-418.

Wilson, P., & Heid, M. K. (2015). From situations to framework. In Heid, M. K., Wilson, P. S., & Blume, G. W. (Eds.), Mathematical Understanding for Secondary Teaching: A Framework and Classroom Based Situations (pp.41-56). Charlotte, NC: Information Age.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah (New Jersey): Lawrence Erlbaum Associates.

Zaslavsky, O., & Sullivan, P. (2011). Setting the Stage: A Conceptual Framework for Examining and Developing Tasks for Mathematics Teacher Education. In O. Zaslavsky & P. Sullivan (Eds.), *Constructing Knowledge for Teaching Secondary Mathematics* (pp. 1–19). Boston, MA: Springer US.

Zazkis, R., & Mamolo, A. (2011). Reconceptualizing knowledge at the mathematical horizon. *For the Learning of Mathematics*, *31*(2), 8-13.