

Determining and Testing an Alternative Model for the Evaluation of Responses to  
a Standardized Test of Mathematics

Fraser Bocell

A thesis

submitted in partial fulfillment of the  
requirements for the degree of

Master in Education

University of Washington

2012

Committee:

Catherine S. Taylor

Elizabeth Sanders

Program Authorized to Offer Degree:

College of Education

University of Washington

**Abstract**

Determining a Mathematical Thinking Skills Model for the Evaluation of  
Responses to a Standardized Test of Mathematics

Fraser Bocell

Chair of the Supervisory Committee:  
Professor Catherine S. Taylor  
Educational Psychology

State standardized tests of mathematics are generally organized around a set of learning objectives set forth by the state or some other body. Items are sampled from these learning objectives and student reports are often partially based on the number of items within each objective(s) answered correctly. This paper sets forth an alternative method for organizing items on a standardized test of mathematics based on the thinking skills necessary to answer each item. This model is then compared to the traditional learning objective model using confirmatory factor analysis procedures along with a brief analysis of the information provided by each model. While both models provide a reasonable fit to the data, the thinking skills model should provide more useful information for teachers.

*Keywords:* Mathematics, Middle School, Thinking Skills, Learning Objectives

## Table of Contents

|  | Pages |
|--|-------|
| LIST OF FIGURES.....                                   | ii    |
| LIST OF TABLES.....                                    | iii   |
| Chapter I: Introduction and Statement of Problem ..... | 1     |
| Chapter II: Literature Review.....                     | 3     |
| Number Sense .....                                     | 7     |
| Part-Whole Relationship.....                           | 8     |
| Sequencing .....                                       | 8     |
| Algebraic Reasoning .....                              | 9     |
| Visuospatial Reasoning .....                           | 10    |
| Thinking Skills Models.....                            | 11    |
| Chapter III: Methods .....                             | 11    |
| Instrument .....                                       | 12    |
| Test Content.....                                      | 12    |
| Data Analyses .....                                    | 14    |
| Coding.....  | 14    |
| Quantitative analyses.....                             | 16    |
| Chapter IV: Results.....                               | 18    |
| Chapter V: Discussion.....                             | 19    |
| Limitations.....                                       | 21    |
| Recommendations for Future Research .....              | 22    |
| Appendix A .....                                       | 33    |

## LIST OF FIGURES

| Figure Number  | Page |
|--|------|
| 1. Example WASL Math Item .....                        | 29   |
| 2. Path Diagram for the Learning Objective Model ..... | 31   |
| 3. Path Diagram for Thinking Skills Model .....        | 31   |
| 4. Path Diagram for the Single Factor Model .....      | 32   |

## LIST OF TABLES

| Table Number  | Page |
|---|------|
| 1. Number of Items by Type and Total Items for Grade 7 WASL Mathematics Tests .....       | 27   |
| 2. Grade 7 Mathematics Test Map .....   | 27   |
| 3. Chi-Square Test for Difference Testing Between Proposed Model and 1-Factor Model ..... | 27   |
| 4. Model Fit Information .....  | 28   |

## **Chapter I: Introduction and Statement of Problem**

Testing based on learning objectives provides minimal data on student knowledge of material covered in those objectives. The items on the test are only a sample of the universe of items possible for the specified objectives. Additionally, it is not possible to cover all of the learning objectives and provide enough items on each to adequately cover all of the material in an objective taught during the period before the test. Consequently, the material covered on the test is only a sample of the material a student has potentially learned. Thus, information gleaned from the test, especially formative information, is of minimal use.

Although the goal of the test is to determine the student's proficiency in mastering the learning objectives set forth for that year, the test itself contains a potential wealth of information that is not captured in traditional scoring models. The learning objective model used to develop most standardized tests does not provide a tool for making meaningful inferences about students from their responses. Instead, I propose that by applying a cognitive-based thinking skills model to the responses to test items, patterns of student responses could potentially provide more meaningful information about the student. Information about a student's mastery of the thinking skills necessary to succeed at math could be helpful for making instructional decisions about the student. Considering the time and expense of developing standardized tests, information relevant to decisions made about future coursework or the necessity of remediation for students should be welcome. Especially since the proposed method of analysis would not add significant time or cost to test development.

Core mathematical learning can be described using cognitive processes. The thinking skills based on these cognitive processes are critical to the development of mathematical

competency. Should an analysis of student responses to standardized test items based on mathematical thinking skills provide a useful model, the data on student mathematical competence could prove valuable. Information on student responses to items involving particular thinking skills could potentially provide a basis for targeted remediation of mathematical skills.

First, it is important to show that a model based mathematical thinking skills can sufficiently describe student responses on standardized tests of mathematics. Particularly, it is hoped that the thinking skills-based model will better describe the data than the traditional objective-based model. The test to be used to investigate this question is the *Washington Assessment of Student Learning* (WASL). It is important to note that the math portion of the WASL was designed to measure mathematical applications, problem-solving, and reasoning rather than the ability to use mathematical algorithms. Consequently, a model of scoring based on mathematical thinking skills should be relevant to the test. Based on this, the following research questions are explored:

1. Does the multi-factor Essential Academic Learning Requirements (EALRs) focused model fit the test item data better than a 1-factor model for each of the two years of data collected?
2. Does the multi-factor learning-focused model fit the test item data better than a 1-factor model for each of the two years of data collected?
3. Based upon the results of the first two questions, what can we speculate about which model should be selected?

## Chapter II: Literature Review

The literature on the cognitive basis of mathematics suggests that mathematics is composed of a distinct cognitive skill set (Anderson, 1982; Schoenfeld, 1987). Essentially, the cognitive processes involved in mathematical thinking are separate from the processes involved in other subjects. Numerical and non-numerical domains in the brain appear to be separate. Cappelletti and Giardino (2007) cited evidence suggesting "...that numerical knowledge is a functionally distinct semantic category within the semantic system, meaning that to some extent numerical processing is independent from other cognitive processes" (p.78). Consequently, it is reasonable to theorize that these processes play a role in a student's mathematical competence. Thus, evaluating student responses using a model based on these cognitive processes could be useful in identifying areas where a student may benefit from remediation.

Cognitive science has also led to the further definition of the cognitive basis for mathematics. Some of the bases for the cognitive processes involved in math appear to be biological. "Based on studies with animals and babies, there appears to be some innate mathematical sense built into our brains. Basically, babies recognize the values of numerals. They have 'some numerical competence'" (Cappelletti & Giardino 2007, p.77). Simple recognition of mathematical representations appears to be innately ingrained into our cognitive processes. The recognition of values constitutes the very basis for our mathematical thinking.

In addition to numerical value, we also appear to have developed separate processes for dealing with the different uses of numbers.

For example, one patient showed a selective (although transient) incapacity to perform number comparison tasks (e.g. to decide which is the larger of '4' and '7') despite being able to count and to tell what number comes before or after another (Delazer and Butterworth, 1997). His pattern of performance therefore suggests a clear distinction between some cardinal and ordinal uses of numbers (Cappelletti & Giardino, 2007).

It appears that a number is not simply a number in all contexts. Numbers contain different meanings depending on the situation. Numbers can be represented in many ways, as ordered rankings, quantities, or as names. Research has shown that these separate uses for numbers are governed by separate cognitive processes (Greeno, 1987). Consequently, it is important to be aware of the different roles that numbers play in mathematics, and also consider that separate cognitive processes likely govern the use of numbers in these roles.

In attempting to build a cognitive model for mathematical thinking skills, it is important to define what makes a person a competent mathematician. How do we define what is necessary for mathematical competence? In other words, what does it take to do well in mathematics? Schoenfeld (1992) argues that there are two key components to learning to think mathematically. He writes, "learning to think mathematically means (a) developing a mathematical point of view—valuing the processes of mathematization and abstraction and having the predilection to apply them, and (b) developing competence with the tools of the trade, and using those tools in the service of the goal of understanding structure—mathematical sense-making" (p. 335). The "mathematical point of view" Schoenfeld writes about is akin to thinking and reasoning about mathematics, while "competence with the tools of the trade" can be likened to the knowledge of mathematical facts and procedures, and

reasoning in mathematics. That is, thinking inside of a mathematical framework, using the knowledge of mathematical procedures and mathematical reasoning to solve problems, rather than just applying algorithms.

Thinking in mathematics is essential for taking the step from applying algorithms to becoming a competent mathematician. Schoenfeld argues further that students must be able to think metacognitively about mathematics; however, this must come after students have become proficient at mathematical thinking skills, or what I am terming: thinking in mathematics. Thus, I seek to create a scoring model for a standardized test of mathematics based on the mathematics thinking skills that appear in the literature about the cognitive basis of mathematics and mathematical problem solving.

As Schoenfeld (1992) says, mathematics is more than just applying formulas or algorithms. Certain types of mathematical thinking play a key role in doing well in mathematics. Resnick et al. (1989) provide support for the importance of reasoning in mathematics. They note, “a close consideration of recent research on mathematical cognition suggests that in mathematics, as in reading, successful learners understand the task to be one of *constructing meaning*, of doing interpretive work rather than routine manipulations” (p. 13). The idea of constructing meaning, doing interpretive work, thinking and reasoning in mathematics, all relate to the cognitive based processes, or thinking skills, necessary for success in mathematics.

Unfortunately, there appears to be a disconnect between instructional practice and the development of mathematical thinking skills. Young children enter school with the ability to solve simple mathematical problems. Children who have not received formal mathematic instruction show the ability to solve simple arithmetic word problems. Semantic understanding

appears to be a significant factor in problem solving (Silver, 1987). After receiving mathematics instruction, many children abandon the use of semantics and “prefer instead to choose an arithmetical operation on the basis of the problems surface features” (Silver, p. 46).

Mathematics instruction may rob children of the innate strategies they use to solve problems. Instead, they begin to refer to a tool bag of algorithms taught in school.

Other researchers have reached similar conclusions regarding the interplay of a child’s natural mathematical ability and school. Similar to children’s abilities to solve simple arithmetic problems, children also appear to have a natural mathematical intuition before they receive instruction (Cappelletti & Giardino, 2007). Unfortunately, this natural intuition is not fostered and built upon in mathematics classes. “There is substantial evidence that children’s difficulty in learning school mathematics derives in large part from their failure to recognize and apply the relations between formal rules taught in school and their own independently developed mathematical intuitions” (Resnick, et al., 1989, p. 13). Many of the connections between what is taught in school and what a student naturally understands about mathematics are not explicitly integrated during instruction. Students with an inclination towards mathematical thinking, through whatever avenue, recognize these implicit connections and develop the mathematical thinking skills on their own. However, many students will not make these connections and develop the necessary thinking skills, leaving them to struggle with various learning objectives, depending on the thinking skills they lack.

According to the literature (Anderson, 1982; Cappelletti & Giardino, 2007; Schoenfeld, 1987), there are a number of thinking skills, related to cognitive processes, important to mathematical problem solving. These thinking skills form the basis for success in mathematics.

However, they are not explicitly taught in school. Consequently, students are left to develop the thinking skills on their own, or struggle through the numerous concepts affected by these skills. In what follows is a review of the thinking skills that appear in the literature and connect to the important objectives for 7<sup>th</sup> graders in the state of Washington. It is important to note that these thinking skills are important throughout a multitude of mathematics courses, but have been chosen as the most important to middle school mathematics, especially 7<sup>th</sup> grade.

### **Number Sense**

As referenced earlier, a number does not always act the same in all situations. Numbers can function in different ways and have different properties depending on their use. Greeno makes note of this when he writes, "Identifying a quantity involves recognizing a component of information that specifies a number of some kind of object, and amount of some kind of substance, a rate or some other quantitative element" (1987, p. 64). Using numbers to represent quantities can take many forms. The different forms of a number, as quantities, are implied in the examples and problems in classrooms. Additionally, numbers can be used nominally, to name items. They can be used on an ordinal scale, and on an interval scale. All of these uses require a different approach. Thus, making sense of number attributes constitutes a thinking skill involving the ability to distinguish the properties of a number in a problem situation and act accordingly. However, students are not always explicitly taught how to identify the use of a number and how to approach it in a problem. This thinking skill, involving making sense of number attributes, will be coded as number sense in this study.

## **Part-Whole Relationship**

The relationship between numbers used as a part versus numbers used as a whole is another thinking skill that is identified through examples (Bransford et al., 2000). The thinking skill involving part-whole relationships involves the ability to thoughtfully distinguish between a part and a whole in terms of representation, understanding, and use. The difference in the operations done on parts (fractions, decimals, ratios, time, etc.) versus those done on whole numbers are connected to the understanding of the differences between a part and a whole. Bransford et al writes, “Rational numbers (fractions) do not behave like whole numbers, and attempting to treat them as such leads to serious problems” (p. 91). While students try to learn the different algorithms for dealing with rational numbers, they do not always fully develop an understanding of differences between rational numbers and whole numbers. This lack of understanding deprives students of an thinking skill essential to successfully navigating not only rational numbers, but other aspects of mathematics, including time, ratios, and others. Consequently, the part-whole relationship is another critical thinking skill.

## **Sequencing**

Additionally, the problems mathematicians and students are faced with are complex and multifaceted. The number of processes involved in solving a single problem can be astounding once the problem is broken into all of its component parts. Silver wrote, “One common difficulty that students have when solving complex or novel mathematics problems is the apparently overwhelming number of processes to control” (1987, p. 40). Solving a problem is not always a clear, straightforward task. Students must manage both discrete and inter-related information while working a problem. In his work with computer models of

mathematical problem solving, Anderson noted, "Because the system can have only one goal at any moment in time, productions from only one of these subroutines can apply at any one time. This forces a considerable seriality into the behavior of the system" (1982, p. 372).

Essentially, because only one piece of a problem can be managed at a time, the processes necessary to solve the problem are forced to be serial in nature. Thus, students must be able to process the problem in a correct order, or potentially create errors in their solution.

Consequently, an important thinking skill appears to be the concept of sequencing. Students must manage a large number of processes that are forced to be serial in nature, in order to solve a problem. Sequencing refers to the ability of students to break down problems, solve the parts of a problem, and control the processes necessary throughout.

### **Algebraic Reasoning**

In addition to sequencing, algebraic reasoning is a major thinking skill in mathematics. Algebraic reasoning (AR) is pervasive throughout all levels of mathematics. Additionally, AR can take many forms within mathematics. For instance, Greeno writes, "Recognizing patterns that involve quantities and relations among quantities is a major requirement for successful performance in arithmetic word problems" (1987, p. 69). Matching patterns within and among problems leads to stronger problem solving abilities. Recognizing that patterns can be generalized and equated to many problems aids the student in formulating strategies to solve problems that may appear fundamentally different on the surface, but essentially equivalent in underlying processes. Silver argues this point when he writes,

Students who build similar representations for mathematically related problems are far more likely to notice their similarity and to use the relationship about one problem in

solving the other. Conversely, the similarity between mathematics problems with isomorphic representations can go unnoticed if the solver does not represent the two problems in similar ways (1987, p. 45).

Students should view equality as a larger concept encompassing many mathematical ideas. Things do not have to be exactly the same to be equal. Sometimes you are matching problems and ideas based on underlying concepts or processes. Forming the relationships between problems that are represented differently but are fundamentally the same is just as important as understanding the relationship between the different representations for a ratio. In the end, students must understand how to reason algebraically within mathematics, allowing them to make connections between discrete concepts and pieces of information.

### **Visuospatial Reasoning**

The final thinking skill identified in the literature involves reasoning with visual images and within three-dimensional space. This type of reasoning can be termed visuospatial reasoning and is associated with visuospatial sketchpad, or visuospatial working memory (Baddeley, 1992; Delgado & Prieto, 2004; Kyttala & Lehto, 2008). According to Baddeley, "There are separable spatial and visual components, with different tasks differentially recruiting the two" (p. 558). Within the visuospatial sketchpad, different types of tasks appear to be processed by different components. One component handles visual information, such as shapes and pictures, while the other component is responsible for orienting things spatially. However, this distinction is not necessary to explore item types related to visuospatial reasoning. It is necessary to note that visuospatial reasoning is related to items involving both visual and spatial material. Students may be required to either spatially manipulate objects or make

judgments based on visual information in order to solve items. It is likely that a problem requiring visuospatial reasoning will require geometry. Geometry is defined as “a branch of mathematics that deals with the measurement, properties and relationships of points, lines, angles, surfaces and solids” (*Merriam-Webster's Collegiate Dictionary Tenth Edition*, 1994, p. 488). It would stand to reason that the properties of geometric figures may be analyzed and deciphered using visuospatial reasoning.

### **Thinking Skills Models**

Although reorganizing mathematics items according to the type of thinking skills necessary to solve the problem is potentially fertile ground, little has been done to explore this line of research. In comparing methods of item classification, Taylor (2000) used  $Q_3$  statistics (Yen, 1984) and confirmatory factor analysis to compare objective-based item groupings with item clusters based on residuals. Aside from Taylor's paper, little work has been done concerning item classification and clustering around students' thinking. However, grouping items, and eventually content, according to how students perceive and think about them is an important avenue for research.

### **Chapter III: Methods**

The essence of this research project is a comparison of two models for the evaluation of responses to items on a standardized test of mathematics. The model used for comparison is based on the learning objectives with which the items were aligned. These objectives come from the Washington State Essential Academic Learning Requirements (EALRs). The theoretical model under investigation is a model based on a review of mathematical thinking skills present in the literature surrounding mathematics and cognition.

For the experiment, two years worth of test data is analyzed, 2003 and 2004.<sup>1</sup> For the purposes of this study, students identified as English language learners or receiving special education services are not included in analysis. These students' answers may be affected by factors other than their knowledge of mathematics and mathematical thinking skills. Students receiving a score of zero on the test will also be removed from analysis, as the test is presumably not an appropriate measure of their mathematical ability. There were approximately 85,000 students per grade level for each test year. For each year of test data, a sample of ten percent of the total eligible seventh grade students was randomly chosen using SPSS 17 statistical package.

### **Instrument**

The tests are criterion-referenced mathematics tests composed of multiple-choice items (scored 0, 1), short-answer items (scored 0, 1, 2), and extended-response items (scored 0, 1, 2, 3, 4). Test scores were equated using year-to-year common item equating procedures. About 30 to 35% of the items were common across contiguous years.

### **Test Content**

Table 1 presents the number of multiple-choice and constructed-response items on each test as well as the total number of items on each test. Table 2 presents the test maps for the seventh grade mathematics tests.

For the mathematics tests, the concepts and procedures for the test items were specific to grade level expectations; however, the same categories of concepts and procedures were assessed at all grade levels. The mathematics content strands included conceptual

---

<sup>1</sup> The age of the data is unavoidable, as it is the most recent set of complete data available to the researcher.

understanding and procedural skills in number, measurement, geometry, algebra, and statistics/probability. The process strands were problem-solving, reasoning, communication/representations, and connections (i.e., equivalent representations or application of two content strands to solve a problem). The item shown in Figure 1 is an example of a mathematical connections item. Note that the rubric requires students to demonstrate algebraic and geometric conceptual understanding as well as correct algebra procedures.

Multiple-choice items tested conceptual understanding, simple applications of procedures, identification of features of a problem, accurate representations of data, and accurate translation of information from word problems into equations and other symbolic forms. Constructed-response items assessed multi-step applications of concepts and procedures, estimation, mathematical problem-solving and reasoning, mathematical communication and representations, and connections across mathematical strands.

What follows is an example of a constructed response item, previously included in the test, which has since been released to the general public. This item was originally coded according to the Washington State EALRs as representing geometric sense. According to the thinking skills model developed here, the item is similarly coded as Visuospatial Reasoning. Students are required to spatially manipulate the triangle, both reflecting and translating the triangle away from its original position. This differentiates the item from other geometry problems which ask students to name the properties of various two and three dimensional figures or locate points on a coordinate grid, as this item requires students to actively manipulate the figure.

## Data Analyses

### Coding.

The first step in conducting a comparison of different models and their fit to the data is to identify clusters of items. For this study, all items were already classified according to their objectives. The initial stage of the data analysis involved the coding of the items for the first year of test data, 2003. As suggested by Miles and Huberman, a 'start list' of codes based on the conceptual framework and theory are the starting point for the coding process (1984, p. 57). The objective based list of codes was based on the Washington State EALRs. Essentially each item was coded using the appropriate EALR. Next, a set of codes further reflecting the type of problem within the content was created. Both of these sets of codes are what Miles and Huberman term "descriptive codes". Wherever possible, EALRs for a item provided in released items and published testing materials was used for coding the items.

The next set of codes was interpretative codes, based on the researcher's evaluation of the data. In this case, the mathematical thinking skills previously mentioned were used to code the data, Number Sense, Part-Whole Reasoning, Sequencing, Algebraic Reasoning, and Visuospatial Reasoning. While coding the data, the possibility that one or more items did not fit the coding scheme existed. If the item appeared to fit with more than one code, all of the possible codes were then written down.

Once all of the codes had been determined for all of the items, a decision on items with multiple codes was made. A determination for the final coding of the items with multiple codes was made after considering several factors. First, patterns in the data were taken into account. If, for instance, one code did not account for more than a few items, then the alternate codes

for those items were used. Second, similarities between items with multiple codes and other items were researched. The item under consideration then received the code of the item it was most similar to, as its final code. In the case of this study, the tetrachoric and polychoric correlations between items were also used to inform decisions made on items with multiple codes. Correlations among the items were calculated using the statistical software Mplus. The items were then organized by their possible thinking skills code, and correlations among items within codes were compared. Correlations with items in other codes were also taken into account when deciding between codes for a particular item.

At times, none of the codes currently under use fit an item. In such a case, the interpretive codes were adjusted based on experience with the data and previous experience with middle school math students. A new code was then created to account for the new items type. At the same time, the test was reviewed again using the new code to determine whether it was appropriate for other items. Then, decisions based on patterns and similarities were made to determine the final coding. It should be noted, further reading in the literature was necessary to ground these changes to the coding within the current theoretical framework.

For the second year of test data, 2004, the coding process described for the 2003 items was adjusted to better account for the need for efficiency in large scale test development. It is necessary for the coding scheme to apply quickly and easily to a new test, limiting the expense necessary to interpret the test data. Thus, the final coding scheme developed for the 2003 test was used for the coding and analysis of the 2004 test. Each item was only coded once before analysis, using the coding scheme derived for the 2003 items. If the results of the analysis of the

2004 test mimic the results of the 2003 test, then evidence was provided for the utility of coding and interpretive scheme.

### **Quantitative analyses.**

The basis of this comparison was a confirmatory factor analysis of the objectives-based (see Figure 2) and the thinking skills models (see Figure 3). The test was composed of two types of items, multiple choice and constructed response (CR). The constructed response items were scored using partial credit up to a total of two points for short answer items or four points for extended response items. Treating the responses to the constructed response items as falling on an interval scale was not tenable, as the cognitive distance between a score of zero and score of one was not likely to be the same as the cognitive distance between a score of one and score of two. Given the items and the rubrics used to score student responses, it was not likely that the knowledge needed move from one score level to the next is equal among different score levels in the same CR item. Equal distance between levels of the variable is required to treat to assume an interval scale underlying the variable. Consequently, both the multiple choice and CR items are treated as ordinal variables in the analysis.

The data were analyzed using the statistical software Mplus version 6. The default estimator for analyzing categorical variables is a robust weighted least squares estimator, using means and variances to adjust the chi-square test statistic (WLSMV). These are a “weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistic that use a full weight matrix” (Muthén & Muthén, 1998-2007, p. 484). Weighted Least Squares (WLS) “applies the fitting function

$$F_{WLS} = [s - \sigma(\theta)]'W^{-1}[s - \sigma(\theta)] \quad (1)$$

where  $\mathbf{s}$  is a vector of sample statistics (i.e., polychoric correlations),  $\boldsymbol{\sigma}(\boldsymbol{\theta})$  is the model-implied vector of population elements in  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ , and  $\mathbf{W}$  is a positive-definite weight matrix” (Flora & Curran, 2004, p. 469). The WLSMV estimator is a different, though similar, approach to obtain parameter estimates.

With this method, parameter estimates are obtained by substituting a diagonal matrix,  $\mathbf{V}$ , for  $\mathbf{W}$  in Equation 1, the elements of which are the asymptotic variances of the thresholds and polychoric correlation estimates (i.e., the diagonal elements of the original weight matrix). Once a vector of parameter estimates is obtained, a robust asymptotic covariance matrix is used to obtain parameter standard errors. Calculation of this matrix involves the full weight matrix  $\mathbf{W}$ ; however it need not be inverted (Flora & Curran, p. 470).

When using this estimator, the chi-square values of the two models are not directly comparable because the difference in the chi-square is not distributed as chi-square distribution. Consequently, the two models for student responses, thinking skills and objective based, are not directly comparable using this estimator.

In order to compare the results of the confirmatory factor analysis for each model, both models were compared to a one factor model (see Figure 4) using the DIFFTEST option in Mplus. The DIFFTEST option in Mplus allows the user to test nested models with a mean and variance adjusted chi-square statistic (See Asparouhov & Muthén, 2006). Next, the change in the chi-square value for each model was compared to evaluate the difference in the model fit when comparing the thinking skills model to the objective-based model.

## Chapter IV: Results

The chi-square comparison of the more restrictive single factor model to the less restrictive thinking skills and learning objective models was used to assess the difference in explanatory power of the models.

The single factor model was termed the more restrictive model; it can be thought of as the same model as the thinking skills or learning objective model, only having the correlations between each of the factors set to 1 (Kline, 2005, pp. 182-183). The easing of restrictions placed on the data by the thinking skills model or the learning objective model may or may not increase the explanatory power of the model over the more restrictive single factor model. If the fit of the less restrictive model is similar to or better than as the fit of the more restrictive model, the relaxing of restrictions on the model is acceptable and may help to better explain patterns in the data.

In the case of the current analysis, both of the models provide a significantly better fit to the data than the single factor model, as shown in Table 3 for each year of testing. The difference between the chi-square values, as calculated by the DIFFTEST procedure in Mplus (Muthén & Muthén, 1998-2010) demonstrates that the difference between the single factor model and the less restrictive model under comparison is significant ( $p < 0.05$ ), providing evidence that both the thinking skills and the learning objectives models provided a statistically significant improvement over the single factor model. The evidence points to the interpretation that both of the models under consideration provided more explanatory power, in terms of patterns of responses within the mathematics portion of the WASL, than the model containing only a single factor. This supports the interpretation that the mathematics portion of the WASL

does not simply assess a single “math” factor, but rather, is constructed from items that cluster together in separate categories contained within mathematics as a whole.

However the interpretation of the improvement of both models over the single factor model should be met with some caution. As seen in Table 4, the fit statistics for all of the models run independently were very similar, and, aside from the chi-square statistics, appear to show a reasonably good fit of the models to the data. A root mean square error of approximation (RMSEA) less than 0.05 indicates a close approximate fit of the model to the data (Browne & Cudeck, 1993). At the same time, a comparative fit index greater than approximately 0.90 may indicate a reasonably good fit of the model (Hu & Bentler, 1999). Using these criteria, all of the models tested appeared to fit the data reasonably well. Though at the same time, the chi-square statistics were all significant, indicating that they are not a good approximate fit to the data. However, chi-square statistics can be sensitive to large sample sizes, making even small deviations in the model fit appear large, and inflating the chi-square statistic (Brown, 2006, p. 81).

## **Chapter V: Discussion**

As shown in Table 3, both models provide a statistically significant improvement over the single factor model for both years of data. Unfortunately, there is no method available to compare the thinking skills and learning objectives models directly. In order to make a direct comparison of the models using a chi-square fit statistic the use of a maximum likelihood estimator is necessary. According to Muthén and Muthén this becomes “increasingly more computationally demanding as the number of factors and the sample size increase” (1998-2010, p. 58) Consequently, it was impossible to run the analysis using a maximum likelihood

estimator using the hardware available to the researcher. However, even without a direct statistical comparison of the two models, they can be logically compared based on the utility of the information they provide.

As stated in the introduction, the information provided by the learning objectives model is of limited use. The objectives are usually broad, covering multiple smaller topics within an objective. Additionally, these topics are generally grouped because of their historic relatedness rather than the relationship between the thought processes needed to reason and solve problems within the objective. Plus, the nature of a standardized test means that only a small sample of items from all of the topics within each objective will be covered. Consequently, the information available based on student responses is piecemeal and incomplete at best.

In contrast, the thinking skills factors are set up to account for the different ways that students reason about mathematics, and more specifically, the types of mathematics necessary to successfully navigate the concepts covered in middle school mathematics. These factors can provide a useful insight into the types of thinking in which students may need additional support. If a teacher focuses on re-teaching an objective based on students incorrectly answering a single problem sampled from that objective, the teacher may spend valuable time covering large parts of the objective when students only have need of covering a small portion. If data based on thinking skills were used in place of the learning objectives, a teacher could instead focus on the types of thinking that students struggle with, spending the same amount of time targeting a variety of problems connected to the thinking skill with which students are struggling across different content stands (objective groupings).

The nature of these factors means that the thinking skills are present across and within various objectives within the middle school curriculum. If a student struggles with a certain thinking skill, it could affect him/her throughout the curriculum, rather than in one section, designated by a set of objectives. The characteristics of the thinking skills model lend themselves to diagnosing deeper structures within a student's thinking. This enables teachers to use this information in better supporting his or her students throughout their mathematics career. When a teacher targets a thinking skill for remediation, she/he is not focusing on a single concept or type of problem. Rather, the teacher is working on the foundation that allows the student to successfully navigate a variety of problems.

### **Limitations**

As discussed earlier, one of the primary limitations of this research is the ability to directly compare the fit statistics of the competing models. Given the lack of a direct statistical comparison, we are left to rely on the separate estimates of the model fit along with a conceptual evaluation of the perceived utility of the competing models. While this still provides evidence for the utility of the thinking skills model, in the end indirect comparison leaves the study lacking.

At the same time, items were constrained to a single factor and were not allowed to load on multiple factors. Essentially, each item could only represent one learning objective or thinking skill. However, items are not necessarily represented well by a single factor. One item may require students to use more than one thinking skill or may cover more than one objective. Consequently, the fit of the model is diminished by constraining the item to one factor.

Additionally, the research surrounding the underlying cognitive processes involved in mathematics is currently divided and incomplete. As the research moves forward, the cognitive processes underlying mathematical thinking will continue to improve. At the same time, research on how students approach and think about various problems is still either general in its approach (Anderson, 1982; Greeno, 1987; Schoenfeld, 1987) or is focused on particular mathematical concepts and therefore incomplete in its coverage (Cappelletti & Giardino, 2007; Stavridou & Kakana, 2008). Consequently the thinking skills developed in this research, though based on solid research, is still somewhat speculative. The lack of strong model fit for the thinking skills model may be a direct result of a lack of well identified thinking skills. The possibility certainly exists that more research into the ways that students perceive and cognitively process mathematical problems is necessary before clear and distinct thinking skills can be defined. Thus, more research and refinement is necessary in order to provide a truly useful model of student responses for teachers.

### **Recommendations for Future Research**

As mentioned in the limitations, the lack of strong model fit for the thinking skills model may stem from thinking skills that were not well defined. Future research must be undertaken concerning the cognitive processes involved in middle school mathematics problem solving, as well as the thinking strategies and skills used by students to solve different types of mathematics problems. This will lead us down a path of better understanding the types of connections that exist between mathematical concepts that have been previously taught in isolation. The goal being to find ways to organize concepts in a manner that is beneficial to the ways different students think about them.

Additionally, it is necessary to evaluate and refine the methods of research and analysis used to determine thinking skills and evaluate patterns of student responses. The process of assessing relationships among items through use of correlations could be refined, as well as expanded to look at variance, covariance matrices, as well as correlations among residuals. Data describing the relationships among items could be better used to inform decisions about item placement within thinking skill factors, as well as developing well defined thinking skills. Having determined a set of thinking skills based on theory and correlations among items on one assessment, the technique's usefulness and model's generalizability could be tested on multiple assessments.

Finally, the method used to evaluate the model fit of the thinking skills model must be reevaluated and refined. There may be a benefit to allowing items to load on more than one factor, meaning allowing an item to be represented by more than one thinking skill. At the same time, different methods of analysis may provide further insight into the descriptive ability of the thinking skills model. Other methods of analysis, using item response theory for instance, may bring forth relationships not found in the previous analysis.

The most important result of this line of research is determining how traditionally disparate mathematical concepts are cognitively related. Knowing these relationships will allow teachers to better use test data and organize lessons. As research moves forward, it is important to remember utility to teachers and students is the ultimate goal.

## References

- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.
- Asparouhov, T., & Muthén, B. (2006). Robust Chi Square Difference Testing With Mean and Variance Adjusted Test Statistics (May 26, 2006 ed.).
- Baddeley, A. (1992). Working Memory. *Science*, 255(5044), 556-559.
- Bransford, J., National Research Council . Committee on Developments in the Science of, L., National Research Council . Committee on Learning, R., & Educational, P. (2000). *How people learn : brain, mind, experience, and school*. Washington, D.C.: National Academy Press.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park: Sage Publications.
- Cappelletti, M., & Giardino, V. (2007). The Cognitive Basis of Mathematical Knowledge. In M. Leng, A. PASEAU & M. D. POTTER (Eds.), *Mathematical knowledge*. Oxford; New York: Oxford University Press.
- Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, 32(1), 25-32.
- Flora, D. B., & Curran, P. J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods*, 9(4), 466-491.
- Greeno, J. G. (1987). Instructional Representations Based on Research About Understanding. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus new alternatives. *Structural Equation Modeling*, 6(1).
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Press.

- Kyttala, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education, 23*(1), 77-94.
- Merriam-Webster's Collegiate Dictionary Tenth Edition.* (1994). Springfield, MA: Merriam-Webster, Inc.
- Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis : a sourcebook of new methods.* Beverly Hills: Sage Publications.
- Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus User's Guide. Fifth Edition.* Los Angeles, Ca: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus User's Guide. Sixth Edition.* Los Angeles, Ca: Muthén & Muthén.
- Resnick, L. B. E., Klopfer, L. E. E., Association for, S., & Curriculum Development, A. V. A. (1989). *Toward the Thinking Curriculum: Current Cognitive Research. 1989 ASCD Yearbook:* Association for Supervision and Curriculum Development, 1250 N. Pitt St., Alexandria, VA 22314-1403 (\$15.95).
- Schoenfeld, A. H. (1987). Cognitive Science and mathematics Education: An Overview. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education.* Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Schoenfeld, A. H. (1992). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning.* New York; Toronto; New York: Macmillan ; Maxwell Macmillan Canada ; Maxwell Macmillan International.
- Silver, E. A. (1987). Foundations of Cognitive Theory and Research for Mathematics Problem-Solving. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Stavridou, F., & Kakana, D. (2008). Graphic Abilities in Relation to Mathematical and Scientific Ability in Adolescents. *Educational Research, 50*(1), 75-93.

Taylor, C. (2000). *A Comparison of Methods for Item Classification as Evidence for the Validity of Objective-level Scores*. Paper presented at the Annual Meeting of the National Council on Measurement in Education.

Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement Applied Psychological Measurement, 8*(2), 125-145.

Table 1

Number of Items by Type and Total Items for the Grade 7 WASL Mathematics Tests

| Multiple Choice | Short Answer | Extended Response | Total |
|-----------------|--------------|-------------------|-------|
| 26              | 11           | 4                 | 42    |

Table 2

Grade 7 Mathematics Test Map

| Strands                    | Multiple Choice | Short Answer | Extended Response |
|----------------------------|-----------------|--------------|-------------------|
| Number Sense               | 3-5             | 1-2          | 0                 |
| Measurement Concepts       | 3-5             | 1-2          | 0                 |
| Geometric Sense            | 3-5             | 1-2          | 0                 |
| Probability and Statistics | 3-5             | 1-2          | 0                 |
| Algebraic Sense            | 3-5             | 1-2          | 0                 |
| Solves Problems            | 0-2             | 2-4          | 1-2               |
| Reasons Logically          | 0-2             | 1-2          | 1-2               |
| Communicates Understanding | 0               | 1-2          | 1-2               |
| Making Connections         | 0-2             | 1-2          | 0                 |
| Total Number of Items      | 27              | 11           | 4                 |

Table 3

Chi-Square Test for Difference Testing Between Proposed Model and Single Factor Model

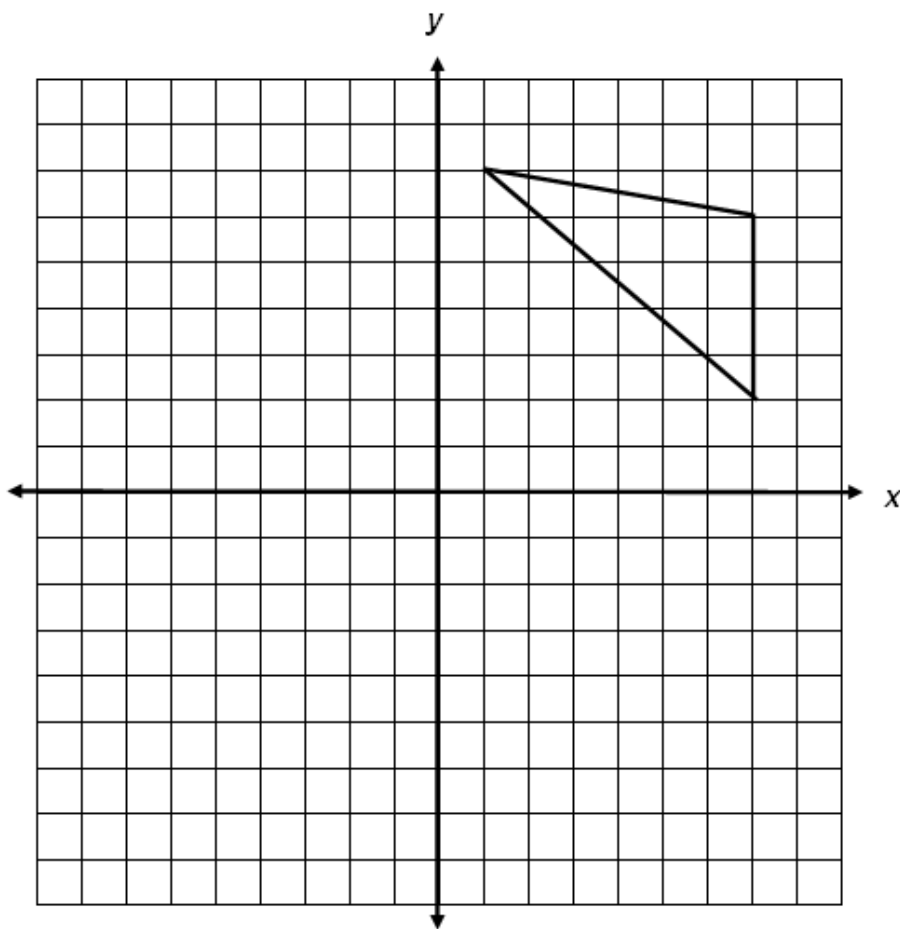
|                              | 2003                 |                     | 2004                 |                     |
|------------------------------|----------------------|---------------------|----------------------|---------------------|
|                              | Thinking Skill Model | Learning Obj. Model | Thinking Skill Model | Learning Obj. Model |
| Difference in $\chi^2$ Value | 91.1                 | 118.2               | 80.07                | 213.327             |
| Degrees of Freedom           | 10                   | 28                  | 10                   | 28                  |
| P - Value                    | < 0.001              | < 0.001             | < 0.001              | < 0.001             |

Table 4.

| Model Fit Information |                |               |               |                |               |               |
|-----------------------|----------------|---------------|---------------|----------------|---------------|---------------|
|                       | 2003           |               |               | 2004           |               |               |
|                       | Thinking Skill | Learning Obj. | Single Factor | Thinking Skill | Learning Obj. | Single Factor |
| $\chi^2$ Value        | 1776.69        | 1745.15       | 1855.35       | 1883.09        | 1761.25       | 1954.97       |
| Degrees of Freedom    | 809            | 791           | 819           | 809            | 791           | 819           |
| P - Value             | < 0.001        | < 0.001       | < 0.001       | < 0.001        | < 0.001       | < 0.001       |
| CFI                   | 0.98           | 0.99          | 0.98          | 0.99           | 0.99          | 0.99          |
| RMSEA                 | 0.01           | 0.01          | 0.01          | 0.02           | 0.01          | 0.02          |

**12** Reflect (flip) the triangle over the x-axis and then translate (slide) the triangle up two units.

- Mark the three points for the angles of the reflected triangle.
- Use a straight edge (ruler) to draw the rectangle in its final position.



*Figure 1.* Example WASL Math Item. This figure illustrates an example of a constructed response item showing Geometric Sense and Visuospatial Reasoning.

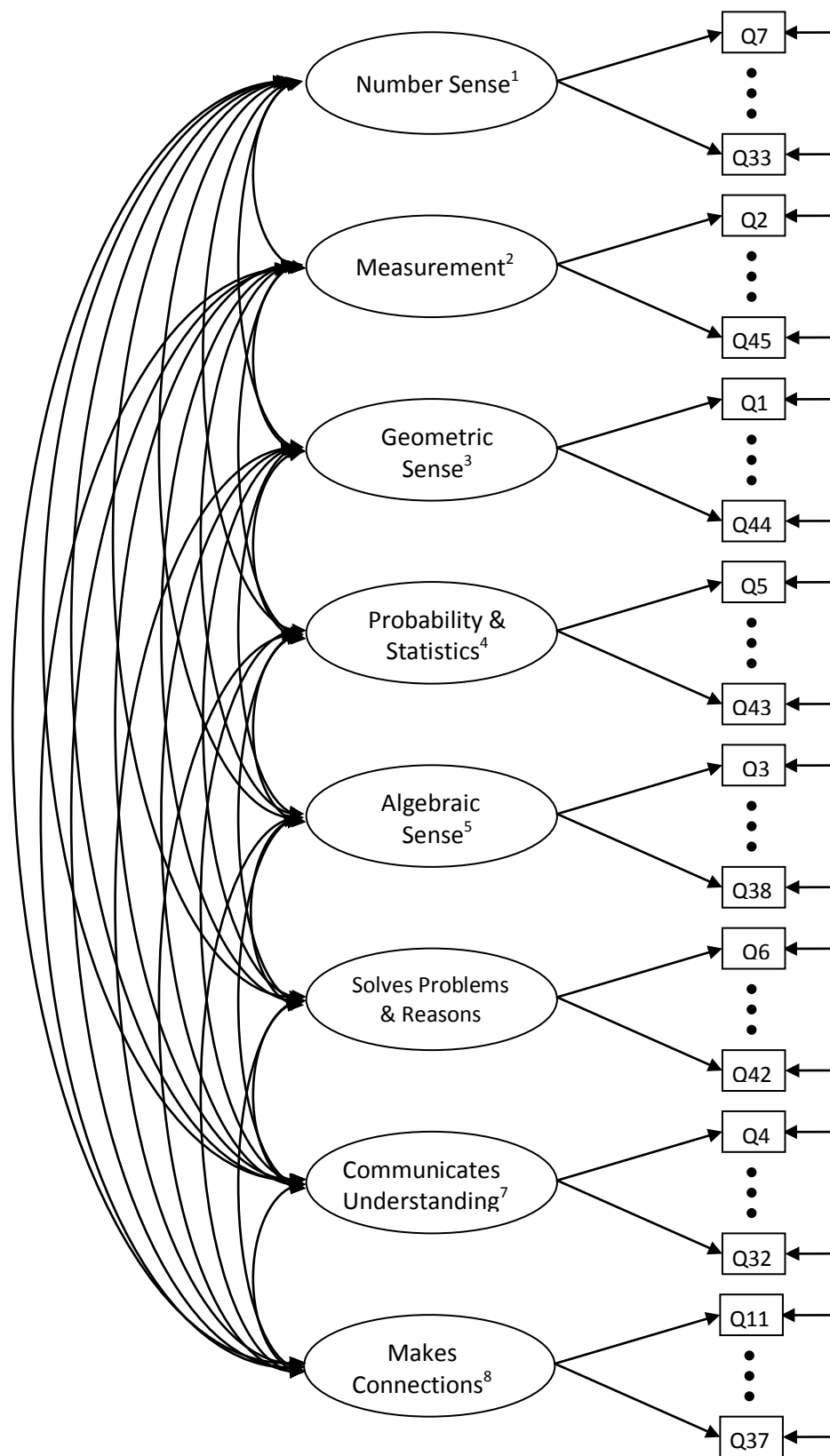


Figure 2. Path Diagram for the Learning Objective Model. Test consists of Questions (Q) 1 to 21 and 24 to 45. Dots indicated items not pictured in the diagram.<sup>1</sup> – Questions 7, 13, 26, 27, 30, 33. <sup>2</sup> – Questions 2, 12, 14, 25, 31, 45. <sup>3</sup> – Questions 1, 9, 16, 41, 44. <sup>4</sup> – Questions 5, 19, 21, 34, 39, 43. <sup>5</sup> – Questions 3, 8, 10, 15, 29, 36, 38. <sup>6</sup> – Questions 6, 17, 28, 40, 42. <sup>7</sup> – Questions 4, 18, 32. <sup>8</sup> – Questions 11, 20, 35, 37.

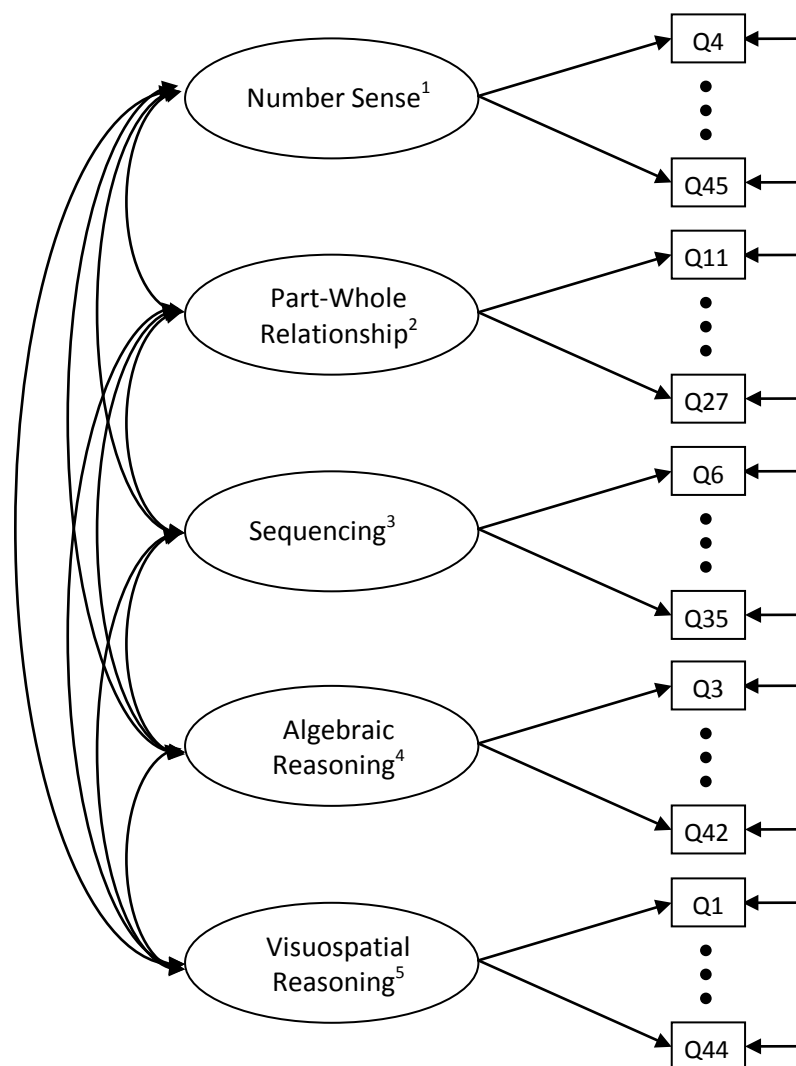
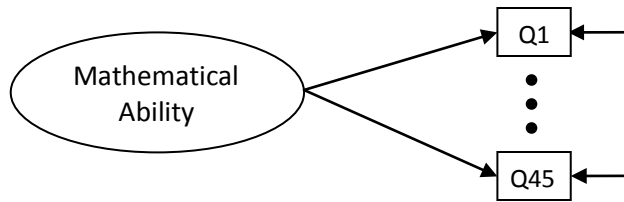


Figure 3. Path Diagram for Thinking Skills Model. Test consists of Questions (Q) 1 to 21 and 24 to 45. Dots indicated items not pictured in the diagram. 1 – Questions 4, 5, 7, 18, 26, 30, 33, 34, 36, 37, 39, 45. 2 – Questions 11, 12, 19, 20, 25, 27. 3 – Questions 6, 13, 17, 21, 29, 31, 35. 4 – Questions 3, 8, 10, 15, 28, 32, 38, 40, 42. 5 – Questions 1, 2, 9, 14, 16, 41, 43, 44.



*Figure 4.* Path Diagram for the Single Factor Model. Test consists of Questions (Q) 1 to 21 and 24 to 45. Dots indicated items not pictured in the diagram.

## Appendix A

**Mplus Syntax for the Data Analysis****Syntax for Learning Objective model.**

TITLE: MATH 7 2003 Learning objectives Model All Categorical  
 DATA: FILE IS H:\Thesis\WASLG7M03EDIT.dat;  
 VARIABLE: NAMES ARE q1-q21 q25-q45;  
           CATEGORICAL ARE q1-q21 q25-q45;  
           MISSING ARE q1-q21 q25-q45 (9);  
 MODEL: f1 BY q7 q13 q26 q27 q30 q33;  
        f2 BY q2 q12 q14 q25 q31 q45;  
        f3 BY q1 q9 q16 q41 q44;  
        f4 BY q5 q19 q21 q34 q39 q43;  
        f5 BY q3 q8 q10 q15 q29 q36 q38;  
        f6 BY q6 q17 q28 q40 q42;  
        f7 BY q4 q18 q32;  
        f8 BY q11 q20 q35 q37;  
 SAVEDATA: DIFFTEST IS derive03LO.dat;  
 OUTPUT: STANDARDIZED;  
        TECH4;

**Syntax for the Single Factor Comparison Model.**

TITLE: MATH 7 2003 SINGLE FACTOR COMPARISON All Categorical  
 DATA: FILE IS H:\Thesis\WASLG7M03EDIT.dat;  
 VARIABLE: NAMES ARE q1-q21 q25-q45;  
           CATEGORICAL ARE q1-q21 q25-q45;  
           MISSING ARE q1-q21 q25-q45 (9);  
 ANALYSIS: DIFFTEST IS derive03LO.dat;  
 MODEL: f1 by q1-q21 q25-q45;  
 OUTPUT: STANDARDIZED;  
        TECH4;

**Syntax for the Thinking Skills Model.**

TITLE: MATH 7 2003 THINKING SKILLS MODEL All Categorical  
 DATA: FILE IS H:\Thesis\WASLG7M03EDIT.dat;  
 VARIABLE: NAMES ARE q1-q21 q25-q45;  
           CATEGORICAL ARE q1-q21 q25-q45;  
           MISSING ARE q1-q21 q25-q45 (9);  
 MODEL: f1 BY q4 q5 q7 q18 q26 q30 q33 q34 q36 q37 q39 q45;  
        f2 BY q11 q12 q19 q20 q25 q27;

f3 BY q6 q13 q17 q21 q29 q31 q35;  
f4 BY q3 q8 q10 q15 q28 q32 q38 q40 q42;  
f5 BY q1 q2 q9 q14 q16 q41 q43 q44;  
SAVEDATA: DIFFTEST is derive03TS.dat;  
OUTPUT: STANDARDIZED;  
TECH4;