Title
Profiling Individuals that Vary in Mathematic Abilities: A Latent Class Analysis Approach

Author
Olide, Andres Fernando

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Profiling Individuals that Vary in Mathematic Abilities: A Latent Class Analysis Approach

A Dissertation in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Andres Fernando Olide

December 2015

Dissertation Committee:
Dr. H. Lee Swanson, Chairperson
Dr. Rollanda O'Connor
Dr. Austin Johnson

The Dissertation of Andres Fernando Olide is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Profiling Individuals that Vary in Mathematic Abilities: A Latent Class Analysis Approach

by

Andres Fernando Olide

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, December 2015
Dr. H. Lee Swanson, Chairperson

The importance of mathematics at an early age stems from the finding that early competence is likely to lead to future mathematics achievement (Duncan, Dowsett, Claessens, Magnuson, Huston, Klebanov, Pagani, Feinstein, Engel, Brooks-Gunn, Sexton, Duckworth, & Japel, 2007). Moreover, individuals that do well in mathematics are likely to have increased overall income earnings (Paglin & Rufolo, 1990; Rivera-Batiz, 1992), while individuals that do not perform as well in mathematics are essentially limiting their potential for career options and labor market earnings (e.g., Paglin & Rufolo, 1990). Studies attempting to better understand the processes related to mathematic underachievement have proposed different ability groups to explain deficits in mathematics. The purpose of this study is to determine whether discrete latent classes can be observed in children within the continuum of mathematic achievement ability. The findings provided partial evidence for the hypothesized ability subgroups along the math achievement ability distribution. The ability groups observed were further tested for external validity. Evidence for the external validity of measures suggests that the domain specific indicators played a critical role in distinguishing between ability groups. The findings gathered from this research support the notion of mathematics learning ability subgroups and further improves the identification of children at risk for math difficulties.

Table of Contents

List of Tables

Chapter 1: Introduction

Mathematic achievement has become an important consideration in contemporary education theory and its worth appears to have more and more significance in terms of academic and economic success. In early childhood, children experience a window that prepares them to develop the skills and knowledge required for mathematic achievement. Mathematic achievement at an early age is likely to result in increased achievement in both the short and long term (Duncan et al., 2007; Ritchie & Bates, 2013), and potentially lead to careers in STEM fields (Ing & Nylund-Gibson, 2013). The study of mathematic achievement has also received attention in an international context of the continued intellectual prowess of nations (Hyde & Mertz, 2009), whereas in the U.S. mathematics achievement is valued at all education levels (National Mathematics Advisory Panel, 2008).

As the importance of math achievement in both short and long-term outcomes has increased, so has the need to better understand children who may be at risk for math difficulties (MD) and low achievers (LA). There appears to be a subgroup of children who perform poorly in math, which is distinct from another group of children who display symptoms of difficulty in learning mathematics though the underperformance is less severe (Geary, 2013). The estimates of individuals with math difficulties differ according to state, as does the definition, or percentile that a subgroup belongs to (Cortiella & Horowitz, 2014; U.S. Department of Education, 2013). Issues with identifying individuals at risk for math disabilities likely stem from the assumption that the group is homogenous; that is, that all individuals with math learning difficulty will show the same symptoms and have the same deficits in the respective domains (Senf, 1986; Geary, 1993). It is more likely, however, that the subgroups with learning difficulties in mathematics have deficits in a combination of achievement domains and, that due to their complexity, by their very nature cannot truly be assessed with a single test of mathematic achievement.

Statement of Problem

School age children displaying characteristics of math difficulties (MD) have been estimated to be about 6.4% of the public school population (Badian, 1983; Geary, 1993). The estimates of prevalence, however, vary (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2005; Desoete, Roeyers, & DeClercq, 2004; Lewis, Hitch & Walker, 1994; Geary, Hoard, Nugent, & Bailey, 2012; Shalev, Manor, & Gross-Tsur, 2005) and in some estimates comprise up to 10% of the public school population (Geary, 2013). Further, theoretical work has suggested that various subgroups of underperforming individuals make-up the classification of MD (e.g., Geary, 1993; Murphy, Mazzocco, Hanich, & Early, 2007). An example is the subgroup of children who display similar deficits manifested as those with MD but to a lesser degree, children who are reported to represent an additional approximately 10-15 percent of the public school population (Berch & Mazzocco, 2007; Dowker, 2005; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Murphy et al., 2007). This group has been termed “low achievers” (LA) (e.g., Geary, 2013). For comparison purposes other subgroups include, depending on the researcher and research focus, “typical achievers” and “high achievers.” This theoretical perspective proposing achievement subgroups argues for the existence of at least three to four subgroups, depending on the categorization, that lie on the continuum of math achievement (Geary, 2013). An alternative view, however, could be that no subgroups exist that fit into the presently used mathematic achievement continuum and previous evidence of subgroups has been an artifact of the data, thus, putting into question the notion of subgroups in the context of the mathematic achievement continuum. If this is true, the data on mathematic achievement ability are best conceptualized as a continuum, leaving researchers to come up with an alternative explanation for the deficits in the math achievement domains. Because this alternative possibility is usually ignored in MD research, it is explored in this investigation. In addition to determining the purported existence of these subgroups, it is argued that this investigation will improve the classification and allow for a better understanding of the potential deficits of school-aged children with math difficulty and low achievers relative to typical achieving and high achieving groups.

One approach used in previous investigations that identified children at risk for MD includes a discrepancy between IQ and math achievement scores (for a review see, Murphy, et al., 2007). Studies that have considered the IQ-achievement discrepancy model, or achievement as a proxy for IQ, have found that students’ achievement scores are inconsistent with what their IQ score would predict (Shalev, 2007, p.50; Swanson, 2011). Extreme cases, where MD cases are found, can potentially be overlooked when using this approach. Shalev (2007, p. 50) points out that children with high intelligence but poor study habits are also likely to be misdiagnosed with this approach. Similarly, children who are average achievers yet perform poorly on intelligence tests do not fit this criteria yet would clearly benefit from academic intervention. Researchers have criticized the IQ-achievement discrepancy approach by noting that it is susceptible to error and likely to yield unreliable results (Murphy et al., 2007; Shalev, Auerbach, Manor, & Gross-Tsur, 2000; Swanson, 2011). Finally, the DSM-V no longer uses the previous operationalization—the IQ-achievement discrepancy model—to identify individuals with math difficulty (Shalev, 2007, p.50). As the evidence suggests, alternative approaches to the IQ-achievement discrepancy model should be considered in the process of identification of mathematic achieving ability groups in order to reduce misidentification.

Another approach to identify MD or LA is the use of single domain specific indicators. One example is intelligence, where general intelligence is reportedly the best predictor of mathematics performance for the general population (Deary, Strand, Smith, & Fernandes, 2007). Researchers, however, question the use of this indicator as the only single item used for identification of children at risk for math difficulties (MD, Geary et al., 2007; Murphy et al., 2007). As a result, researchers have warned about the increased possibility of misidentification associated with using single indicators such as IQ or performance scores in mathematics to diagnose individuals and subsequently place them in corresponding math achievement ability groups (Francis, Fletcher, Stuebing, Lyon, Shaywitz, & Shaywitz, 2005). The potential for error when using single indicators increases, particularly since different measures used to assess mathematic achievement ability yield different results because of the domain(s) assessed by the test or depending on the cutoff score used by the researcher to determine the severity of deficit for MD subgroup identification (Murphy et al., 2007). Using multiple domains as indicators of mathematic ability (e.g. reading comprehension, word problem solving accuracy, fluid intelligence, and mathematics calculation) is an approach that will allow for better identification of the complexity of the underperforming groups; therefore, using multiple indicators of mathematic ability to reduce misidentification is highly recommended. Consequently, new identification strategies that account for previous methodological flaws (e.g. unreliable, no longer used operationalizations of MD, and use of single indicators) should be implemented in determining the existence of discrete math achievement groups.

Geary (2013) points out that the current consensus in the field includes two years of continued underperformance for MD diagnosis (below 10th percentile) and low achievers (LA) diagnosis (between 11th and 25th percentile). While this identification approach may be conservative and helpful to children who should not be labeled incorrectly, it is impractical, particularly since a two-year delay in diagnosis may be detrimental given the critical stage of development and learning of the children. That is, the delay would impact the critical period when children are most cognitively open to developing the skills and knowledge required for mathematic achievement. The best approach in identifying these subsamples along the mathematic achievement continuum is arguably the cutoff scores approach (Murphy et al., 2007; Francis, Fletcher, Stuebing, Lyon, Shaywitz, & Shaywitz 2005). Concerns with cutoff scores may result from the notion that they seem arbitrary, particularly since error in classification is likely to occur around the cutoff score and because it is difficult to determine disability from difficulty (Vukovic & Siegel, 2010; Francis et al., 2005). As a result, the term “difficulty” is used in this research as a label for individuals with math difficulty and “low achievers” is used for individuals with a less severe deficit. Moreover, cutoff scores can be critical for exploring the sample heterogeneity (Collins & Lanza, 2010), used here to explore the mathematic achievement ability sample characteristics. This investigation informs the current debate on strategies used to explore and identify ability groups with a focus on individuals with math learning difficulty.

Purpose of Investigation

The present study explores whether latent classes of children with MD (mathematic difficulty) and LA (low achievers) can be identified. The aim here is therefore to determine whether subgroups can be observed based on mathematic achievement ability into discrete subgroups. Because the 25th percentile is commonly used to identify children at risk for math achievement difficulty (e.g., MD and LA) (Fletcher et al., 1998; Fuchs et al., 2005; Fuchs, Fuchs, & Prentice, 2004; Geary, 2013; Geary, Bailey, Littlefield, Wood, Hoard, & Nugent, 2009; Lyon et al., 2001; Siegel & Ryan, 1989; Swanson & Beebe-Frankenberger, 2004; Vukovic & Siegel, 2010), this cutoff point will be used to explore the complexity of the data to determine the existence of discrete subgroups. Previous investigations have provided the justification for using cut off scores.

An approach used to assess the existence of discrete groups that are dichotomized, using cutoff scores in this case, is the use of latent class analysis (LCA). Appendix A, is a figure of a typical example used to describe the latent class analysis. The figure shows an overall distribution. It is important to point out that the overall distribution can also be represented by two separate distributions that are in turn represented by separate means and standard deviations (Collins & Lanza, 2010; Marcoulides, Gottfried, Gottfried, Oliver, 2008; Masyn, 2013). As the example suggests, the advantage of this approach is the ability of latent class analysis to identify the likelihood of heterogeneity in a distribution (Collins & Lanza, 2010; Muthén, 2002). In other words, this approach can be used to determine and potentially identify if discrete groups hypothesized in the literature can be observed. Heterogeneity or discreteness of groups might be readily observed (as in the case of gender or grade characteristics) or unobserved (as in performance groups) (Marcoulides, et al., 2008). Similar to other investigations (see Marcoulides et al., 2008), it is acknowledged that thinking of mathematics achievement ability as a continuum might be preferred by some researchers over the use of cut off scores; however, it is argued that the benefit of determining the existence and learning about the potential heterogeneity of the distribution of mathematic ability to better understand groups with mathematic learning difficulties gives justification to an exploratory approach such as the latent class analysis (Collins & Lanza, 2010; Marcoulides et al., 2008). Moreover, the literature has presented substantial evidence for the importance of cutoff scores and similar to other investigations the results will be assessed based on the criteria of how the data will be represented in terms of model fit (Marcoulides et al., 2008).

In addition to exploring the existence of subgroups this investigation aims to relate the findings to the present literature by searching for supporting evidence for the subgroups that exist in the mathematic achievement continuum. The results will be further tested to assess whether other domains related to mathematic achievement ability can predict the different group profiles.

Research questions

1. Are latent subgroups of individuals observed using the 25th percentile cutoff score on norm referenced achievement measures?

2. Do the latent subgroups that are observed fit the profiles commonly listed in the literature (e.g. high achievers, typical achievers, low achievers, and math difficulty)?

3. Do latent classes differ on external measures? These external measures include working memory (executive function, phonological loop, visual-spatial sketchpad), inhibition, numeracy, estimation, naming speed, and problem solving components.

The subsequent sections of this investigation will describe the relevant research as it relates to the identification of individuals with math difficulty. Specifically, the focus will be on the current classification procedures, the drawbacks of these approaches, and ways this investigation will improve on previous procedures. This will be followed by a description of the cognitive mechanisms that have been used to explain deficits in mathematics. Following will be a description of the classification measures and their importance in identifying individuals with MD. Chapter three will give a detailed description of the sample in question and the procedures used to collect the data. A detailed description of the measures used in the analysis will follow. Measures used in the investigation will be separated by its use in this investigation; where there will be a section for measures used to classify individuals and another section that will be used to describe the external validity measures. Chapter four describes the analyses and discusses the results as they relate to the research questions. Subsequent sections will include a discussion of the findings, potential limitations, conclusions, and final recommendations.

Chapter 2: Literature Review

In reviewing the literature a heavy emphasis will be placed on Geary (2013), whose summary of the literature and theoretical work on better understanding children with MD has heavily influenced this area of study. This chapter will introduce the factors that will be used in identifying the profiles on the mathematic achievement continuum. Following will be a detailed description of the current classification procedures and the importance of using cut off scores, particularly when the complexity of the data demands an exploratory statistical approach (e.g. latent class analysis). In order to place the complexity of the classification of individuals in context, a review of the cognitive mechanisms used to explain mathematic difficulty with a focus on the domains will be reviewed here. Finally, a description with an emphasis on the importance of the classification measures, used here to explore the different profiles, will be followed by a summary and will conclude the chapter.

In a synthesis of the literature, Geary (2013) summarized the most recent findings in the domains of mathematics achievement and cognitive ability as they relate to the mathematic achievement continuum. As the name suggests children diagnosed with math difficulty tend to underperform in mathematics—this is also the case, though less severe, for individuals labeled low achievers. In this investigation mathematic calculation will be one of the components used to identify individuals into achievement subgroups. Similar to other investigations, it is argued here that computation skill is not the only indicator of the complexity of mathematic ability (Swanson & Beebe-

Frankenberger, 2004); as a result, word problem solving accuracy will be an additional domain used to explore the mathematic achievement subgroups. Fluid intelligence, as well, has been implicated with mathematic underperformance (Geary, 2013). Moreover, a high percentage of children with MD are also diagnosed with reading difficulty and underperform in the reading domain (Ashkenazi et. al., 2013; Barbaresi et al., 2005; Vukovic, 2012), where the same is true for a subgroup of low achievers (Geary, 2013). These four domains (e.g. mathematic calculation, word problem solving components, fluid intelligence, and readings skills), have been widely used, represent the complexity of mathematic achievement ability, and will be used as key indicators in the exploratory latent class analysis approach. The added benefit of using these measures in the classification analysis part of the investigation is that these are norm-referenced measures. That is, participants will not be evaluated on the immediate tasks or their cohort group but a "general reference group" (Thorndike & Thorndike-Christ, 2010, p. 69). In effect, there is an advantage of potentially increasing the generalizability that these measures contribute to the investigation and will therefore be used as part of the classification process as standard scale scores.

Current Classification Procedures

This investigation will consider the latent subgroups that will emerge in a sample of elementary school children that may vary on a continuum of math achievement ability. Although, previous investigations have attempted to establish subgroups related to math achievement ability, subgroup categorizations are not always in agreement (Geary, Bailey, Littlefield, Wood, Hoard, & Nugent, 2009; Murphy, et al., 2007), nor is the approach used to identify subgroups of children with math learning difficulty (Geary, 2013; Swanson & Alloway, 2012). Because criteria used to classify individuals into ability groups is still being established (Geary, 2013), and previous operationalizations such as the DSM-IV, two-year achievement deficit, and the IQ-achievement discrepancy model are prone to error or no longer applicable according to current standards (Murphy et al., 2007; Shalev, 2007) the best current classification approach is arguably the performance percentile cut-off score (Murphy et al., 2007; Geary, Hoard, Nugent, & Bailey, 2012).

In the present investigation the percentile cutoff score approach will be used to facilitate classification. Participant's mathematic achievement ability scores will be initially dichotomized into high and low math performers based on the common cut-off score for determining risk status (i.e., 25th percentile). This cut-off score has been reported to be conservative relative to other interpretations of math difficulty, yet inclusive of individuals that may have symptoms of underperforming subgroups (e.g. MD and LA groups) (Mazzocco, 2007). Murphy et al., (2007) noted that higher percentile scores have been documented in the literature though this approach is generally used to increase the low performing sample size or because eligibility for school services dictates a higher cutoff score (Jordan et al., 2003; Geary, 1990; Geary, Bow-Thomas, Yao, 1992). It is important to note that studies using higher percentile scores may not capture the unique differences between the theorized math achievement ability groups. The cut-off score criterion is important because it is the most commonly used and accepted in the literature (Fletcher et al., 1998; Geary, 2013; Geary, Bailey, Littlefield, Wood, Hoard, & Nugent, 2009; Siegel & Ryan, 1989); however, the 25th percentile score also represents two subgroups of individuals who are low achieving and those with math difficulty (Murphy et al., 2007). Therefore using a percentile cutoff score should, at a minimum, differentiate between the low achieving groups and typical achieving groups.

In effect, this process of dichotomizing the data hypothesizes that the complexity of the data will be reduced and therefore produce more homogenous classes (Lanza, Dziak, Huang, Wagner, & Collins, 2013) according to those predicted by the literature (e.g. math difficulty, low achievers, typical achievers, and/or higher achievers). Because it is believed that the data are best treated or operationalized as categorical when considering the subgroups along the math achievement distribution the data will be treated as such. Moreover, by using the 25th percentile the importance of the cut off criterion is implicitly being tested. Consequently the cut off score will aid in understanding the potential profiles of individuals and should yield a better understanding of the heterogeneity of the subgroups on the math achievement ability continuum. As a result, the aim here is to use this exploratory investigation to determine if latent classes of math achievement ability emerge in the data set or whether mathematic achievement ability is best treated as a continuum.

An exploratory approach that can be used to identify the existence of discrete groups, or heterogeneity is mixed modeling. Specifically, a statistical tool used to classify unobserved (e.g. mathematic achievement ability subgroups) or latent classes that are not readily observed from the data, as in this case, is the latent class analysis (LCA) approach (Lanza, Collins, Lemmon, & Schafer, 2007; Masyn, 2013; Nylund, Asparouhov, & Muthén, 2007). This statistical approach uses the individual's categorical response pattern, or performance pattern in this instance, as it relates across mathematic achievement ability domains to determine the existence of given subgroups or classes. This statistical approach has been previously used in the area of education and has been successfully applied in identifying children's attitudes toward the areas of STEM and persistence in pursuit of STEM careers (Aschbacher, Ing, & Tsai, 2014; Denson & Ing, 2014; Ing, & Nylund-Gibson, 2013). The advantage of the latent class model with categorical indicators is that there are no assumptions about the distributions; rather it is assumed that the factors for each latent class are independent of each other (i.e. local independence) (Lanza et. al., 2007). An additional benefit is that one can test the importance of a critical cut off score to dichotomize a sample presented in the literature. While it may be argued that math performance should be strictly treated on a continuum or homogenous and not discretely heterogeneous, it is argued here that substantial research has provided evidence for the heterogeneity between subgroups in the area of math learning difficulty on the math achievement continuum (Fletcher et al., 1998; Geary, 2013; Geary, et al., 2009; Siegel & Ryan, 1989). In other words, research has found evidence for the math achievement performance ability group profiles that explains the deficits of individuals with MD and LA, and the strengths of typical and high achievers (e.g. Geary, 2013). Consequently, work done on MD and LA populations will be used to make predictions on the profiles and model fit criteria will be used to determine whether the profiles can be observed and in turn elucidate whether math achievement ability should be treated on a continuum or discrete.

Cognitive Mechanisms Associated With Mathematic Achievement Ability Deficits

Given the complexity of the area researchers have sought to explain the mechanisms associated with deficits such as a MD diagnosis. Researchers have offered several explanations that describe the deficits and allow for better understanding of the processes of how these deficits operate (Ashkenazi, et al., 2013; Geary, 2013; Swanson, 2011). The first of these mechanisms involves the employment of slower problem solving strategies for numerical calculation related problems. The second mechanism is a deficit in the area responsible for remembering relevant information resulting in an inhibition problem of irrelevant information. Thirdly, this is a mechanism that explains the deficit in the area that influences performance via mental mapping of symbols and the numbers these symbols may represent. A hybrid approach has been proposed as the fourth mechanism (Ashkenazi et al., 2013), and involves increased mistakes in recalling of relevant information and errors in mapping of symbols. Next a discussion will follow with an emphasis on the implications of these mechanisms as they relate to the mathematic achievement continuum.

The first mechanism used to explain math achievement underperformance is the use of counting to solve arithmetic problems due to the use of slower problem solving strategies (e.g. counting fingers). This deficit results in not counting properly or making mistakes in the process of counting in a sequence (Swanson, 2011). Geary (2013) notes that a deficit in this mechanism will likely impact the phonetic representational systems therefore influencing counting in a way that increases the likelihood of errors. If the phonetic representational systems are indeed being affected by this mechanism then group differences based on performance ability are expected for the phonological loop domain since individuals with MD would be more likely to have these systems impacted (Geary, 2013). Naming speed is a good indicator of the math fluency of mathematic achievement ability and has been found to be a distinct construct from untimed measures (Petrill et al., 2012). This is important for children with MD and LA as the mechanism would likely predict underperformance in the domain of naming speed. Similarly, another domain where MD children are expected to underperform due to a deficit would be numeracy (e.g. quantity estimation) (Mazzocco, Feigenson, & Halberda, 2011; Ashkenazi, et al., 2013). A deficit in this domain specific mechanism would predict underperformance in arithmetic or calculation where counting and relevant domains are concerned.

The second mechanism allows for recalling information and remembering facts. A deficit in this mechanism results in increased mistakes in recalling information and remembering facts (Barrouillet, Fayol, & Lathuliere, 1997), also known as a working memory—specifically an executive function deficit (Ashkenazi, et al., 2013). This deficit allows for irrelevant information to enter into the thought process that results in increased solution times and more mistakes. According to Swanson (2011), this shortfall in the working memory's executive function is likely a defining feature of MD (e.g. Geary, 1993). This domain general deficit would predict underperformance in the number sense domain (Geary, 2004; Ashkenazi, et al., 2013). Based on this mechanism, it is expected that children with MD deficits would underperform in the naming speed task since there will be a delay in the retrieval of information and facts. It is also likely that the tasks related to inhibition will result in underperformance for children with MD since errors are likely to influence performance on the relevant tasks (Geary, 2013). Moreover, because inhibition is related to the executive function, there is an increased likelihood of a deficit in that respective domain.

The third mechanism involves mental mapping of symbols and the numbers these symbols may represent (Butterworth, 2005). In other words, children are likely to have a deficit in visual spatial information (Swanson, 2011). Children with deficits in the areas associated with this mechanism are likely to underperform in comparing numbers. This mechanism is likely to be best reflected by the domain of the visual-spatial sketchpad, where children with MD are expected to underperform. Another domain where children with MD are predicted to underperform is the area of numeracy where children are estimating numbers on a number line. Similarly, it is expected that children with MD symptoms are likely to underperform when they are expected to compare a series of numbers, also the numeracy domain. An alternative view of this deficit (Geary, Hoard, Byrd-Craven & DeSoto, 2004), however, suggests that rather than a deficit in visual spatial information it is the inability to follow the sequence of steps of a problem, problem solving operation or algorithm (Swanson, 2011). Based on this alternative approach, it would be expected for underperformers to have a deficit in the domain of word problem solving components, instead of the visual-spatial sketchpad and numeracy domains. This prediction will be indirectly tested as part of this investigation.

The fourth, and hybrid mechanism, argues for a combination of deficits between mechanisms representing executive function (mechanism two) and visual-spatial information (mechanism three) (Ashkenazi, et al., 2013). In all likelihood a MD diagnosis or deficit results from a combination of some of these mechanisms. Determining the importance of these mechanisms is beyond the scope of this investigation. The purpose of describing the mechanisms here is to guide the predictions in the external validity part of the investigation.

The mechanisms described (e.g. deficit in counting, errors in working memory—fact retrieval and error rates, and errors processing visual spatial information) are expected to yield differences between the theorized groups (i.e. math difficulty, low achievers, typical achievers, and high achievers). It is expected that low achievers will outperform individuals with math difficulty in all domains except for the mechanism that influences domains related to errors in working memory processing (Geary, 2013). For comparison purposes, typical achievers are expected to outperform those with math difficulty. Lastly, and similarly, high achievers are expected to outperform individuals in the group labeled MD. These predictions will be implicitly tested in this investigation. Appendix B shows a table summary of the latent constructs separated by measures used to compose manifest measures to determine latent classes and external measures used for the external validity of classes. One can observe that most deficits occur for individuals with math difficulty (MD). Low achievers (LA) are the next group to have evidence of deficits. Note that the main differences between the groups MD and LA is a matter of severity with LA characterized as having less severe deficit. Typical achievers (TA) and high achievers (HA) are not expected to show deficits. In the case of high achievers, individuals in this group are expected to excel on these indicators.

Influence of External Measures on Mathematic Achievement Ability

For the purpose of this investigation classification measures will be used to categorize or profile individuals into a group on the math achievement continuum. Measures critical to the identification of subgroups will be the following norm referenced indicators of math calculation, word problem solving, reading ability, and fluid intelligence. All constructs have multiple manifest variables that will be used in classifying individuals. Additional measures that have been found to be relevant to mathematic achievement will be used in assessing the external validity of the latent classes observed.

Math ability is also associated with the “flexible” use of mathematical knowledge (McCloskey, 2007; Siegler & Booth, 2004) and the ability to understand, compare and contrast, a given quantity relative to another and thought to be a basic skill (McCloskey, 2007)—also known as numeracy. The term numeracy is interpreted here to be synonymous with quantitative literacy. In the context of math ability, those with MD would be less likely to successfully estimate and compare and contrast given quantity(ies); individuals who are low achievers are also likely to have a deficit in this domain though the severity will be less. A deficit in numeracy influences representation and processing of numbers and hinders subsequent higher order processes related to calculation and arithmetic (Butterworth, Varma, & Laurillard, 2011; Geary, 2013; Geary, Hoard, Nugent, & Byrd-Craven, 2008). In a task related to numeracy, mapping or estimation of a given representation of a number can also influence the mistakes a child makes when solving math problems (Geary, Bailey, & Hoard, 2009; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007). Children with MD would underperform in the above domains as well as the domain of visual-spatial (sketchpad) information.

Previous work has shown that working memory has been found to be a reliable predictor of math difficulty (DeStefano & LeFevre, 2004; Geary et al., 2007; McLean & Hitch, 1999; Swanson & Sachse-Lee, 2001).

One major conclusion is that individuals with math difficulty have been found to underperform on indicators of working memory (Bull, Johnson, & Roy, 1999; McLean & Hitch, 1999; Swanson, 1993; Swanson, Jerman, & Zheng, 2009; Swanson & Sachse-Lee, 2001). Geary (2013) reports that children identified as LA perform average in the domains of phonological loop and visual-spatial sketchpad, noting that other studies reported conflicting evidence for indicators of the visual-spatial sketchpad (McLean & Hitch, 1999). Those diagnosed with MD, however, were reported to have difficulties in both domains (e.g. phonological loop and visual-spatial sketchpad) (Geary, 2013).

Specifically, compared to the phonological loop the visual-spatial sketchpad has been found to be the best predictor of measures assessing math ability (Swanson & Alloway, 2012). Moreover, the executive function, or the attention system, is responsible for moderating the storage subsystems (e.g. visual-spatial sketchpad and the phonological loop) of working memory (Baddeley, 2000). The executive function therefore is responsible for processes such as inhibition and random number generation that identify an individual's ability to ignore or suppress irrelevant information in order to focus on what is pertinent. Children with the ability to suppress unrelated information can increase the fluency (e.g. naming speed) of processing relevant information resulting in faster task completion and reduce error in problem solving (Bull & Scerif, 2001; Passolunghi & Siegel, 2004). Children identified with MD and LA that have an inhibition deficit will likely result in instances of retrieval errors that will likely lead to underperformance in mathematics (Bull & Scerif, 2001; Murphy et al., 2007; Passolunghi, Cornoldi, & DeLiberto, 1999; Passolunghi & Siegel, 2004). Thus, lower mathematic ability is in all likelihood due to a deficit in the process of suppressing irrelevant information (Swanson, 2011). Therefore, underperforming in the cognitive domains is likely to also result in errors in the execution of math problem solving (word problem solving components).

A critical factor in identifying math difficulty is performance in working memory related tasks. The literature has reported a positive correlation between the ability to retain and process information, via the central executive function, and higher academic performance; where individuals manifesting MD symptoms would likely perform worse on retention and processing of information (Bull, Espy, & Wiebe, 2008; Geary, 2013; Mazzocco & Kover, 2007). While most of the work has clearly identified the executive function as critical in identifying MD all components of working memory (e.g. phonological loop and visual-spatial sketchpad) play a key role in different contexts (Geary, 2013). Nevertheless, the components of working memory are critical factors in understanding MD in children (Swanson & Alloway, 2012). Therefore, exploring the existence of subgroups of ability groups on fluency, inhibition, working memory (visual- spatial sketchpad, executive function, phonological loop), word problem solving components, numeracy and estimation, should yield an understanding of the profiles of individuals on the cognitive measures and other domains that are related to the math achievement continuum.

Classification Measures

The assessments used to classify individuals into latent classes were chosen because they are core indicators that are norm-referenced measures. Another way to understand the use of these core measures of math difficulty is that their use controls for more severe problems related to overall achievement—where MD is a specific learning difficulty. Multiple measures were used to compose the domains of mathematic calculation, word problem solving accuracy, reading skills, and fluid intelligence. Because MD is a complex phenomenon domain specific and domain general indicators can potentially explain the occurrence and deficits when it comes to achievement. Domain specific and domain general factors will be carefully considered to create a profile of the mathematic achievement ability continuum (Geary, 2013; Swanson & Alloway, 2012). A domain general indicator, readings skills, have been found to be central in mathematic achievement ability because of the likelihood of MD and reading difficulty to co-occur (Swanson & Alloway, 2012; Vukovic & Siegel, 2010). Fluid intelligence has been considered synonymous with academic achievement and its importance and contribution to mathematic achievement ability will be carefully considered (Geary, 2013; Swanson & Alloway, 2012). Word problem solving components is a candidate indicator for exploring the profiles because it is an assessment of a skill that involves application of the child's mathematic ability (Swanson & Beebe-Frankenberger (2004). The final indicator considered is mathematic calculation because it is the most widely used indicator when investigating MD (Geary, 2013). Following will be a discussion of the four indicators that will be considered for exploring the profile of mathematic achievement ability.

Reading Skills. Reading skills are a critical component to mathematic achievement ability particularly. Children who are found to have a math difficulty are also likely to have a deficit in reading skills (Swanson & Alloway, 2012; Vukovic & Siegel, 2010). The occurrence of one phenomenon (e.g. reading difficulty or math difficulty) is therefore less likely. Rather, MD in combination with reading difficulty is more probable and even expected (Ashkenazi et al., 2013; Badian, 1983; Dirks, Spyer, van Lieshout, & de Sonneville, 2008; Rubinsten, 2009; Siegel & Ryan, 1989; Swanson & Alloway, 2012; Vukovic, Lesaux, Siegel, 2010). Including reading skills as a potential core deficit for identifying math difficulty is warranted. A reliable indicator of an individuals' reading skill is the phonological loop component of working memory (Swanson & Alloway, 2012). The phonological loop is the mechanism that maintains short-term information related to language (e.g. speech based). This mechanism is critical for children's reading ability but also as it relates to word problem solving ability (Swanson & Beebe-Frankenberger, 2004; Vukovic & Siegel, 2010; Vukovic, 2012). Recent research has provided support for the importance of reading as it relates to a core component of math difficulty. In a longitudinal study that investigated the characteristics of underperformance in mathematics Vukovic and Siegel (2010) followed a sample of elementary school children for a period of four years (e.g. first to fourth grade). Among other characteristics they found that reading ability was an important factor for children with the most severe deficit of math ability. Additionally, using a cross-sectional sample to assess the likelihood for math difficulty researchers also found that children at risk for MD have a higher chance of experiencing reading difficulty (Swanson & Beebe-Frankenberger, 2004). As an indicator, reading skills will therefore inform the profile of children with MD since these individuals will likely underperform in the working memory component of phonological loop, and word problems. To summarize, reading skills is an important domain to consider when investigating the characteristics used to better understand math difficulty and other mathematic achievement ability groups.

Fluid Intelligence. In contrast to crystallized intelligence (knowledge of a specific area), fluid intelligence pertains to the ability to reason abstractly in novel situations or conditions (e.g. Cattell, 1971; Horn, 1980; Horn & Noll, 1997, Swanson, & Alloway, 2012**)**. One important point to make is that while fluid intelligence as a domain is not a good indicator of learning difficulties (LD, e.g. MD and RD) when used in the context of the achievement discrepancy model it is however a reliable predictor of achievement (Swanson, 2011). Some controversy surrounds this domain and its relation to working memory, specifically the executive function (Geary, 2013; Swanson & Alloway, 2012). In a review of the literature, Geary (2013) noted a high association between the ability for abstract reasoning, the executive function component of working memory (WM), and the time for task completion. Research has provided evidence supporting the notion that WM is best represented by intelligence (Colom, Rebollo, Palacios, Juan-Espinosa & Kyllonen, 2004; Kyllonen & Christal, 1990). At a lower level of analysis, however, WM might be more telling of learning and academic achievement (Gathercole, Alloway, Willis, & Adams, 2006); that is, that separate components that make up working memory may help identify and explain the mechanisms associated with differences in mathematic achievement ability groups (Bull & Johnston, 1997; Deary, Strand, Smith & Fernandes, 2007; Geary, 2011, 2013). In their review Swanson and Alloway (2012), they pointed out that the findings that support working memory and fluid intelligence being essentially the same is likely due to age. In an effort to explain the directional relationship evidence has been reported showing that working memory, specifically the executive function, contributes to the intelligence of children (e.g. Case, 1992; Schweizer & Koch, 2001). In another investigation these findings were further supported, suggesting that the executive function component of working memory indeed accounted for the relationship between age and fluid intelligence, and age and math, respectively (Swanson, 2008; Swanson & Alloway, 2012). Given the reported role that WM plays in fluid intelligence and mathematics, using this domain (e.g. fluid intelligence) to explore the profiles of individuals mathematic achievement ability is warranted requires careful consideration.

Word Problem Solving Components. Another core construct of mathematic achievement ability is the domain of word problem solving components. The word problem solving components domain is the ability to solve everyday problems as they pertain to words related to mathematics, word problems, and story problems (Swanson & Beebe-Frankenberger, 2004). For example, children successful at word problem solving components are proficient in word phrases, sentences, and proposition comprehension. This proficiency allows for correct understanding interpretation of word problems (Swanson & Alloway, 2012). Involved in this domain is an individual's ability to successfully account for information that is presented and may need a given solution (e.g. Andersen, Reder, & Lebiere, 1996; Kail & Hall, 1999; Mayer & Hegarty, 1996; Swanson & Beebe-Frankenberger, 2004; Swanson & Sachse-Lee, 2001). While the process is not clear the mechanisms associated with working memory are thought to play a key role in successful completion of mathematical word problems (LeBlanc & Weber-Russell, 1996; Swanson & Alloway, 2012; Swanson & Sachse-Lee, 2001). In an effort to better understand word problem solving components Swanson & Beebe-Frankenberger (2004) followed a sample of elementary school children (first, second and third grade). The researchers found that working memory, in particular, the executive function, was an overall better predictor of mathematical problem solving than other indicators (Swanson & Alloway, 2012). That is, children who might be at risk for math difficulty were more likely to underperform in working memory and mathematical problem solving. As a component word problem solving components address the complexity of mathematic achievement ability. Accounting for these components in the characteristics that represent the mathematics achievement ability continuum is an approach that will allow for a better profile identification of children who may have deficits related to math difficulty.

Math Calculation. The main focus of researchers who study MD populations is the domain of mathematic calculation. In his review, Geary (2013) notes that common approaches to solving arithmetic, for instance, involves direct retrieval and decomposition of a given problem. Using the direct retrieval approach the child finds the answer from long term memory storage that results from a given problem (e.g. 2+3+5). That is, the child does not have to resort to less sophisticated approaches such as counting fingers or another less fluent strategy. The second approach entails making sense of a problem from its parts and coming up with the correct solution (e.g. 5+7 = 5+5+2). In this example, Geary notes that the main areas of deficit for MD or LA, though less severe, children are procedural errors in solving simple arithmetic and word problems. Similarly, fact retrieval, or the ability to retrieve basic facts from long-term memory is inconsistent and results in mistakes or errors in solutions.

Domains related to these deficits include the domain of inhibition; where retrieving key information may be hindered, as in retrieving an answer from a simple arithmetic problem. The domains of inhibition and working memory such as the executive function and the phonological loop play a role in attention, inhibition, processing of short term and long-term information. The domain of inhibition is likely to influence performance by the reduction in ability to ignore irrelevant information (Bull & Scerif, 2001; Passolunghi & Siegel, 2004). Similarly, a deficit in fact retrieval can be attributed to the executive function (Geary, 2013). According to Geary (2013) the working memory subsystem, the phonological loop, facilitates naming of numbers when counting (Krajewski & Schneider, 2009), solving word problems (Swanson & Sachse-Lee, 2001), and evidence has been found that this domain facilitates the retrieval of factual information when computing arithmetic (Fuchs et al., 2006; Geary, 1993, 2013). Because these functions are critical, a deficit in any of these domains would indicate a deficit likely associated with MD or LA. Since arithmetic calculation is one of the key factors in understanding potential deficits related to MD and LA symptoms, its inclusion as a classification measure would complete a profile of mathematic achievement ability.

To summarize the indicators used to explore the profile of mathematic achievement ability here will be reading ability, fluid intelligence, word problem solving components, and mathematic calculation. Furthermore, partial evidence for the domains considered to explore the profiles here has been supported by other investigations (Vukovic & Siegel, 2010; Geary, et al., 2009). These indicators arguably make up the most parsimonious number of indicators used to make the achievement profile. Moreover, using these indicators will allow for the control of more severe conditions related to overall achievement or aptitude. Using these indicators will therefore exclude the alternative explanation that the findings can potentially be a result of low overall achievement or aptitude and not the condition of MD. Another critical reason for selecting these items is because these are norm referenced assessment tools that will allow for a certain degree of generalizability. Performance between groups is expected to occur on a gradient. Individuals with MD are likely to have a profile that reflects difficulties in, fluid intelligence, word problem solving components, reading, and math calculation. Low achievers have difficulty in the domain of mathematic calculation, and if there is a deficit on other domains it will be less severe than that compared to children in the MD group. Typical achievers do not show any deficits in the previously discussed domains and will have the largest sample of participants in that group (Masyn, 2013). The high achieving group is expected to excel in these domains relative to other groups.

Statement of Purpose and Questions

The purpose of this exploratory investigation was to determine if latent classes of math ability emerge in the data set or whether math ability is best treated as a continuum. In other words, are there unobserved subsamples (latent classes) in the domains related to math performance based on the 25th percentile cut off? Do model fit criteria suggest that the predicted profiles are indeed present? And, are the subgroups of math problem solving in agreement with previous investigations? (Geary, 2013; Geary, Bailey, Littlefield, Wood, Hoard, & Nugent, 2009; Murphy et al., 2007). Since the cut off score (i.e. 25th percentile) reflects two distinct groups, evidence for more than two existing groups will be evidence for theorized math achievement groups reported in the literature and for the existence of subgroups that reflect the discrete nature of the math achievement ability groups. Lastly, to assess the validity of the results external measures will be used to predict the latent classes.

Research Questions

1. Are subgroups (latent classes) of individuals observed using the 25th percentile cutoff score?
2. Do observed groups have sufficiently different profiles (e.g. high achievers, typical achievers, low achievers, and math learning difficulty)?
3. Do latent classes differ on external measures of working memory (executive function, phonological loop, and visual-spatial sketchpad), numeracy, estimation, computational fluency, speed, and problem solving components?

Hypotheses

1. Using the domain specific measures of math performance, it is expected that subgroups of individuals, or latent classes, will be observed using the 25th percentile cutoff.
2. It is expected that the observed latent classes will reflect distinct profiles and those latent classes will reflect those discussed in the literature (e.g. high achievers, typical achievers, low achievers, and math difficulty).
3. The latent classes observed will reflect differently on external measures such as working memory (executive function, phonological loop, and visual-spatial sketchpad), numeracy, estimation, computational fluency, speed, and problem solving components.

Chapter 3: Methodology

Participants

A sample (*N= 385*) from four public schools in the Southwest region of the United States participated in this study. The gender composition for the sample was 201 female and 184 male participants. Participants represented a diverse group of elementary school students in the Southwestern U.S. at the time (2009-2013) the data were collected (*Age = 6-10 years;* $Mean_{age}$ *= 7.97, SD=0.57*). The elementary school sample of children primarily consisted of third graders *(n= 276)* and was European American (*n=219*), followed by Hispanic (*n=87*), African American (*n=24*), Asian American (*n=21*), Mixed race (*n=31*), and other (*n=2*), respectively. Students were selected from typical performing classes to represent the diversity a given researcher may find in a mainstream elementary school class.

The research for this study was carried in accordance of the Human Subjects committee and written informed consent at the University of California-Riverside protocol number (HS-O6-099) and Federal grant number USDE R324A090002 Institute of Education Sciences. Written informed consent was received from parents and/or guardians prior to testing. These data was gathered as part of a larger research project that occurred from 2009 to 2013.

Measures and Procedures

Once parental consent was obtained the students were then administered a battery of tests originally intended as pretests of a larger investigation (Swanson, Orosco, & Lussier, 2013). The assessment period consisted of 45-minute sessions over a 2-day period. Twenty-nine assessments representing domain general and domain specific cognitive ability, reading, math ability, and IQ were administered. A description of the assessments follows.

Manifest Measures Used to Determine Latent Classes

Math Calculation. Two measures were used to assess calculation ability and computational fluency. The arithmetic computation subtests for the Wide Range Achievement Test (WRAT-3; Wilkinson, 1993) and the numerical operations subtest for the Wechsler Individual Achievement Test (WIAT; Psychological Corporation, 1992) were administered to measure basic calculation ability. Both subtests required the participant to perform written computation on number problems that increased in difficulty. Cronbach's alpha coefficient was acceptable for both tests (.92 WRAT-III and.93, WIAT).

Word Problem Solving Accuracy. Story problem subtests from two separate measures were used to measure problem solving accuracy.

Test of Mathematical Abilities (TOMA). The story problem subtest (TOMA-2; Brown, Cronin, & McIntire, 1994) was administered to assess word problem solving ability. Specifically, the task measures math performance on the two traditional major skill areas in math (i.e., story problems and computation) as well as attitude, vocabulary, and general application of mathematical concepts in real life. Raw scores for the subtest ranged from 0 to 25. Reliability coefficients for the subtests were above .80. The TOMA-2 Story Problems subtest involved children silently reading a short story problem that ended with a computational question about the story and then working out the answer in the space provided on their own.

Comprehensive Mathematical Abilities Test (CMAT; Hresko, Schlieve, Herron, Swain, & Sherbenou, 2003). The CMAT is a normed assessment of children's word problem solving accuracy and is composed of story problems that increase in difficulty. Participants were given 1 point for every correct answer for a total of 24 points (minimum score is 0). Composite standardized score for the test is a mean of 100 and standard deviation of 15. Reliability for the subtests was good (> .86).

Reading Skills. Two tasks were used to assess reading skills.

Test of Reading Comprehension-Third Edition (TORC-3; Brown, Hammill, & Weiderholt, 1995). The TORC-3 was used to assess passage comprehension. The task consisted of participants reading a set of questions followed by a short, one paragraph story. Participants were then asked to answer five comprehensive multiple-choice questions about the text. Performance is assessed on the number of questions answered correctly with the test scores ranging from 0-60. Cronbach's coefficient alpha for the TORC-3 subtests were good at .90 and above.

Wide Range Achievement Test-Third Edition (WRAT-3; Wilkinson, 1993). The WRAT-3 measured the ability to recognize words. Children were asked to identify and correctly pronounce a list of words. Words increased in difficulty and are not discernable from the context presented. The task ends when ten errors are made. Number of words correctly identified ranged between 15- 57. Cronbach's coefficient alpha was acceptable at .87 to .99 with a median of .92.

Fluid Intelligence. One indicator, the overall scores were used to assess fluid intelligence.

Raven Colored Progressive Matrices (RCMT; Raven, 1976). Fluid intelligence is the ability to answer novel and relatively abstract problems while adapting to new contexts. Participants solved patterns with missing pieces that increased in difficulty. The task required participants to circle the correct pattern. Raw score ranged between 0-36. This normative measure yields a standard score of 100 and standard deviation of 15. Cronbach's coefficient alpha was acceptable .88.

Measures Used to Assess External Validity of Latent Classes

In contrast to the normative scales used to establish latent classes, the measures used to compare latent classes on external measures were factor scores. Factor scores were created by output from a confirmatory factor analysis. The tasks representing the constructs are reviewed below.

Inhibition. Two measures were used to create a latent variable for inhibition.

Random Number and Random Letter Generation Task. An adapted version (e.g., Baddeley, 1996; Towse, 1998) of the measure assesses participant’s inhibition responses and suppression of alternate number (or letter) responses that may be associated with well-learned sequences. Participants were asked to write a sequence of numbers (from 0 to 9) as quickly as possible for thirty seconds. The second time participants were asked to write the number (or letter) sequence as fast possible, but this time out of order for 30 seconds. An index for inhibition was calculated as the number (or letter) of sequential numbers minus the number of correctly unordered numbers (or letters) divided by the number of sequential numbers plus the number of unordered numbers. Cronbach's alpha has been previously reported in similar samples at .89, and Random Letter Generation was .91.

Naming Speed. An indicator of speed was measured using two separate constructs intended to measure the child's ability to rapidly name letters and numbers.

Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 2000). As a measure of speed the CTOPP subtests of the rapid digit naming and rapid letter naming were used. Specifically, rapid digit naming measures the speed that individuals can recall and name numbers, respectively. The participants are given a sheet containing single digit numbers that had been randomly selected and placed in a page containing four rows and nine columns. Participants are then asked to name two sheets with the same quantities as quickly as possible totaling seventy-two numbers. Speed is measured using a stopwatch. Performance is the total time it takes the participant to name both arrays of numbers. The child's standard score is the total number of seconds taken to name all the numbers. Rapid letter naming was assessed in the same fashion using randomly selected letters instead of numbers. Coefficient alphas ranged from .87 to .96 with an average of .92 indicating overall high reliability.

Working Memory. Three latent variables that represented working memory were used to assess the ability to process the accumulation of information that increases in difficulty while simultaneously problem solving. Consistent with the Baddeley and Logie (1999) multicomponent model, measures of this construct were the visual spatial - sketchpad, and short-term memory/ phonological loop and executive function.

Visual-Spatial Sketchpad (WM). Two measures were used to create a latent variable for the visual- spatial sketchpad. These measures were used to assess visual stimuli in relation to spatial working memory.

Visual Matrix Task. The Visual Matrix task assessed the working memory ability to remember visual sequences within a matrix (Swanson, 1992, 1995). Participants were given five seconds to review the pattern presented of a series of dots in a matrix (a grid made of squares). After removing the matrix participants were asked if they observed any dots in the first column. Next, students were asked to draw the dots they remembered seeing in the corresponding boxes of a blank matrix in their response booklets. The task difficulty ranged from a matrix of four squares with two dots to a matrix of 45 squares with 12 dots. Performance was measured on the highest set recalled correctly (range of 0 to 11). Cronbach's alpha was acceptable at .86.

Mapping and Directions Span Test (S-CPT). The S-CPT measured participant's ability to recall visual-spatial sequence of directions on a map (Swanson, 1992, 1995). Prior to map stimulus presentation, the child was shown a card depicting four strategies for encoding visual-spatial information for later recall, such as: elemental, global, sectional, or backward processing of patterns. Each strategy was described before administration of targeted items. The stimulus, a street map was presented for ten seconds with lines connected to a number of dots, illustrating the direction a bike would go to get out of the city, with difficulty ranging from 4 to 19 dots. Buildings were squares, dots were stoplights, and lines and arrows were directions to travel. After the removal of the map, the child was asked to recall if there were any stoplights in the first column and then to select a strategy, from an array of four strategies, that best represented how the information was recalled. The child was then instructed to draw lines, stoplights, and arrows on a blank map. Task difficulty ranged from a map with two arrows and two stoplights to a map with two arrows and twelve stoplights. The number of questions answered correctly determined performance. Cronbach's alpha was .94.

Central Executive Function (WM). Three measures were used to assess the working memory's central executive function. These tasks entailed receiving and retaining information that increases in complexity and at the same time answering questions about the tasks. If a question was not answered correctly or if items could not be recalled then the task was discontinued. In effect, WM measures were intended to be a balance between the storage of information and the correct answer to questions with an emphasis on the discrimination of information (e.g. old versus new).

Listening Sentence Span. This task assesses the child's ability to remember numerical information embedded in a short sentence (Daneman & Carpenter, 1980; Just & Carpenter, 1992; Swanson, 1992, 1995, 2011). Participants listened as the tester read a series of sentences, asked a question about the topic of one of the sentences, and then asked the child to remember and repeat back the last word in each sentence in order. Listening sentence sequences were presented in increasing set sizes. Children completed three practice trials with a set size of two sentences. Then participants were presented with four possible listening sentence sequences in groups of two through five. Participants received points towards their span score for correctly answering the topic question, for the number of correctly recalled words in order, and for the number of correctly recalled words out of order. Additionally, a count of the total number of wrong words inserted into the sequence was accounted for in the final score.

Word Problem Span. To assess the successful shifting of tasks or mental sets the Word Problem Span task (Swanson, 1999) was modified for children. The experimenter read a series of sentences, asked a question about the topic of one of the sentences, and then asked the child to remember and repeat back the last word in each sentence in order. However, this measure differs from the Listening Sentence Span Test in that the last word of a sentence included both numbers and words naming objects or things while Listening Sentence Span Test only included words naming objects and things as the item to remember at the end of each sentence. Children received points towards their span score for correctly answering the process topic question, for the correct number of recalled words in order, and for the number of correctly recalled words out of order. Additionally, total number of wrong words inserted into the sequence was also noted.

Updating. To measure the ability to renew and scrutinize new or updating of information so that information that is no longer relevant may be substituted (e.g., Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; St Clair-Thompson, & Gathercole, 2006) participants were asked to take the experimental updating task adapted from Swanson, Sáez, Gerber, and Leafstedt (2004). For the present task participants were given single digit sequences of numbers of different lengths. Participants were informed that the sequence of numbers could vary from three, five, seven, or nine digits and were asked to recall the latter three digits in their respective order. Thus, the task requires renewal and scrutiny of information as the new sequence of numbers was presented that would replace the previously presented information. Performance was assessed on the number of sequences correctly answered (range: 0-16).

Short Term Memory Measures (STM)/ Phonological loop (WM). Three scales were used in the creation of the latent variable short-term memory measure.

Digit Span. The digit forward portion of the Digit Span subtest for the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) was administered to measure short-term memory (STM). The Digit Forward subtest involves a series of orally presented numbers which the child repeats back verbatim. There are eight number sets with two trials per set, with the number of digits to repeat back increasing in range of difficulty starting at two digits and going up to nine digits. The test was scored by getting the highest set of items recalled in order with a range of 0 to 16 for items and 0 to 8 for sets. Internal consistency was reported for all subtests as moderate to excellent at .61 to .92.

Word Span (Swanson & Ashbaker, 2000) and **Phonetic Memory Span** (Swanson & Berninger, 1995). Word Span and the Phonetic Memory Span/ Pseudoword Span tasks were presented in the same manner as the Digit Span measure. The word stimuli used in the Word Span task were one or two syllable high-frequency words (e.g., apple, zoo). Participants were read a list of common but unrelated nouns that increased in difficulty and then they were asked to recall the words.

Phonetic Memory Span. The Phonetic Memory Span task used strings of nonsense words (one syllable long), which were presented one at a time in sets of two to seven non-words (e.g., des, seeg, seg, geez, deez, dez). Performance was assessed based on the highest set of items retrieved in the correct serial order (range of 0 to 7 for Word Span and 0 to 6 for Phonetic Memory Span). Cronbach's alpha was previously reported as .62 for the Word Span task and .82 for the Phonetic Memory Span task (Swanson & Beebe-Frankenberger, 2004).

Word-Problem Solving Component. Six measures assessed the processes assumed to be involved in solving word problems based on the work of Mayer and Heagarty (1996).

Word-Problem-Solving Processes (WPS). The Word Problem Solving Memory Booklet (WPS) was an experimental instrument used to assess the child's ability to identify and recover processing components of word problems—in effect assessing the understanding of the subcomponents that make up a given problem (Swanson & Beebe-Frankenberger, 2004; Swanson & Sachse-Lee, 2001). Tasks were adapted for each grade from each of the word problem solving grade booklets (Swanson, Cooney, & Brock, 1993). The experimenter read aloud each problem and all multiple-choice response options as the students followed along. Performance was assessed on ability to [1]identify the question proposition, [2]numbers in the proposition, and [3]goals in the assignment proposition of each story problem, [4]the operations, and [5]algorithm, and [6]extraneous propositions (Swanson, 2006).

Numeracy. Two tasks were combined to create a latent variable on numeracy or quantitative literacy.

Numerical Estimation (Magnitude Subtest). A component of spatial thinking number line performance was used as a measure of numerical estimation (adapted from Siegler & Opfer, 2003; Seigler & Booth, 2004). This task is critical to numeracy as it is a test of understanding the quantity of the number and its given representation relative to other numbers (McCloskey, 2007). To assess number sense this numeracy task required participants to identify the smallest and largest number, out of three total numbers. Participants were then given 30 seconds to identify the number. This process of correctly identifying numbers occurred for two consecutive times, or two consecutive pages. The total score was obtained by combining the number of correctly identified numbers.

Line Estimation (Estimation Judgment). Estimation task used equivalent forms A and B. The individual was asked to estimate the location of a number using an "x" on a line between 0 and 100. Raw scores were the count of the number of units the estimate "x" was from the correct answer. Lines were standardized at every quarter inch for the standardized test, where the non standardized lines were arithmetically equivalent distances between the estimated "x" and the correct answer on the number line.

Statistical Analysis

Assessing the Factor Structure. Several tasks were administered to determine those constructs external to the classification measures that predicted latent classes. Thus, prior to the latent class analysis, it was important to ascertain whether the factor structure for the proposed constructs provided a good fit to the data. As a first step in the analysis, a confirmatory factor analysis (CFA) was computed to test the a priori specified factor structure for the external validation measures.

Exploring the Existence of Math Achievement Ability Groups. Once the factor structure for the external measures was determined to be satisfactory the next step in this investigation was to explore the existence of latent classes for the math achievement ability groups. As will be described in detail in the results section below, this analysis followed a series of steps that required careful weighing of several sources of information to determine the number of classes. Because LCA is an exploratory approach by nature and it is unlikely that the sources of information would clearly be in favor or against the hypotheses (Masyn, 2013), extra caution was used in interpreting the number of classes produced in the analysis. A series of models were tested for a predicted number of classes that were then compared and contrasted to assess the number of classes that were determined to have best represented the data.

Exploring the classes, labeling the classes, and determining whether enough evidence existed for within class homogeneity and heterogeneity were the steps taken of selecting the number of classes. Several statistical indicators were used in interpreting the latent class analysis results. Evidence was weighed in favor and against the predicted four-class model (e.g. class homogeneity, class separation, and the proportion of endorsement across the domains) the next step was to apply the LCA results by testing the external validity of the results.

Assessing External Validity of the Latent Classes. The results of the latent class analysis were applied by using the posterior probabilities produced by the statistical software to conduct further analyses related to the external validity of the results. The posterior probabilities produced for each class represented the likelihood (e.g. likelihood of class membership) of a given individual belonging to that particular class. A regression approach, for example, was applied to make predictions on performance between the classes (e.g. Clark & Muthén 2009). A multinomial logistic regression was used to determine the influence of variables external to the classification measures that were used to predict the latent classes. For this investigation, the external measures represented domain specific and domain general assessments of cognitive processing.

Chapter 4: Results

This chapter will report the analysis of the data in three parts. First, a preliminary analysis will include a confirmatory factor analysis that will describe the results of the test of the factor structure for the classification and external validation measures. This was done to determine if the factor structure of the variables was consistent with the a priori classification and to produce factors scores for the external validation measures. Second, a latent class analysis was computed to determine if the data set yielded distinct subgroups based on math achievement ability. Finally, a multinomial logistic regression was computed to determine if the latent classes yielded distinct performance differences on constructs independent of the classification measures.

Preliminary Analysis

Table 1 reports the mean standard scores for the normed referenced measures used in the determination of latent classes. All measures are standard scores (*M=100, SD=15*) based on the norm-referenced measures. The use of standard scores therefore allows the participants of this investigation to be compared relative to a larger reference group of elementary school children. Also shown in Table 1 are the skewness and kurtosis for the classification measures that are in acceptable range and that reliability was adequate for this investigation. Of interest here is to determine whether four unique factors or latent variables for reading skills (e.g. TORC and WRAT-reading), calculation (e.g. WIAT, WRAT), word problem solving and accuracy (e.g. TOMA and CMAT), and fluid intelligence (e.g. raven) provided an adequate fit to the data. A confirmatory factor analysis was computed, via a SAS PROC CALIS program (SAS Institute, 9.2). In order to ensure a more parsimonious model fit the variances of the model were set to 1(Kline, 2011). The chi-square test of model fit did not show evidence of acceptable model fit, $\chi^2(8)$ 27.23, $p < 0.001$. However, other fit criteria confirmed that the tested model was also an acceptable model and confirmed the predicted factor structure. Indicators critical to assessing model fit were as follows: the root mean square standard error of approximation estimate was near the suggested threshold of 0.05 (*RMSEA*=. *08*; *CI* = *.05* - *.12*) (Widaman & Thompson, 2003). The comparative fit index (CFI) estimate was well above the value of .90 that is considered good fit (*CFI* =0.97) (Bentler, 1990), and the non-normed index (*NNI =0.93*) (Bentler & Bonnett, 1980) that follows similar criteria of above .90 and is considered good model fit. The results of the confirmatory factor analysis suggested that the tested factor structure for the classification measures was confirmed. The loadings for each task related to each factor are reported in Table 2.

The descriptive measures used to establish external validity of the latent classes are also reported in Table 1. To determine if the organization of the external validity measures fit the hypothesized organization, a confirmatory factor analysis was computed, via a SAS PROC CALIS program (SAS Institute, 9.2). The confirmatory factor analysis tested an a priori task classification of 22 measures into an eight-factor model for the external measures, respectively. For variables considered for the external validation part of the investigation, the latent variables tested were, working memory (executive function, phonological loop, and visual-spatial sketchpad), numeracy, estimation, speed, and problem solving components. To obtain optimal model fit the variances of the model were fixed at 1 (Kline, 2011). The chi-square statistic test did not provide evidence for acceptable model fit, $\chi^2(432)$ 589.86, $p < 0.001$. However, other fit criteria did provide evidence for good model fit and confirmed the predicted factor structure. The root mean square standard error of approximation estimate was below the suggested threshold of 0.05 (*RMSEA=.03*; *CI = .03 - .04*) (Widaman & Thompson, 2003). The comparative fit index (CFI) estimate was well above the value of .90 that is considered good fit (*CFI* =0.96) (Bentler, 1990), and non-normed index (*NNI* =0.95) (Bentler & Bonnett, 1980) that follows similar criteria of above .90 and is considered acceptable model fit. The loadings for each task related to each factor are reported in appendix C.

For the correlation matrix of all the measures used in the classification analysis and measures used for the external validation part of the analysis see appendix D.

A representation of the confirmatory factor analysis formula is as follows:

$$\Sigma = \Lambda\Psi\Lambda' + \Theta$$

This formula represents the matrix expanded form, where Σ signifies the covariance matrix of the indicators, the Λ_y represents the factor loadings, Ψ is the matrix for the factor covariances for the indices, and the final term Θ represents the unique variances of the indicators (Widaman, 2012, p. 363).

In summary, the preliminary analysis suggested that the a priori classification of the variables provided a good fit to the data.

Latent Class Analysis

SAS PROC LCA software (Lanza, Dziak, Huang, Wagner, & Collins, 2013) was used to determine whether latent classes could be identified in the data set of math calculation, math problem solving components, reading, and fluid intelligence. The purpose of the LCA was to find the number of best possible classifications or profiles in the data, estimate the likelihood of being part of a given class, and classify individuals into a given class (Lanza, Dziak, Huang, Wagner, & Collins, 2013). The formula below describes the latent class analysis as it applies to the predicted four-class model. This example was taken from Masyn (2013, p. 559) and applied to the present investigation.

$$\Pr(u1i, u2i, u3i, u4i, u5i, u6i, u7i) = \sum_{k-1}^{4} \left[\pi_k \left(\prod_{m-1}^{7} \omega_{m|k} \right) \right]$$

To the left of the equal sign is the probability response pattern for all seven-core variables of interest used to estimate the latent class analysis. To the right of the equal sign one can observe that a latent variable is estimated for a given number of classes. In this example the model is estimated for the predicted four-classes or groups (e.g. k=4). This model is estimated using the seven dichotomized core indicators for mathematic achievement ability (m=7 (e.g. <.25)). Whereas $\omega_{m|k}$ represents the probability of a child of a given class (e.g. class k, 1-4) endorsing or performing below or above the 25[th] percentile cutoff on a given indicator (e.g. Pr (u_{mi} = 1|ci = k) = $\omega_{m|k}$). In other words, what is being assessed is the profile of children that perform at or below the 25th percentile versus those that perform above the 25th percentile for the four hypothesized classes.

Indicators used to determine potential group status were the aforementioned categorical variables. To ensure that a proper number of classes were assessed the LCA model was computed for up to seven latent groups (i.e., 1-7 classes). Because of the nature of the data there are no assumptions made about the distribution of the factors entered into the model, however, it was assumed that variables within the classes are observed and the errors are independent (Lanza, Collins, Lemmon, & Schafer, 2007). Additionally, in order to compute the latent class analysis using the SAS program the data are required to be dichotomized (Lanza, Dziak, Huang, Wagner, & Collins, 2013). The manifest variables were dichotomized (above and below the 25th percentile) based on the norms for these standard measures. That is, based on each of these normed referenced measures, performance was dichotomized as reflecting children at or below the 25th percentile and children above the 25th percentile. The 25th percentile however was based on the standard score normative data making the cutoff score below 91. As previously stated the 25th percentile cutoff score is a well-established score in the literature used to identify at risk groups (Fletcher et al., 1998; Geary, 2013; Geary, Bailey, Littlefield, Wood, Hoard, & Nugent, 2009; Siegel & Ryan, 1989).

Two major steps were used here in interpreting a latent class analysis (Masyn, 2013). First, the number of classes was determined based on key indicators (e.g. latent class selection). The second step of the latent class analysis included interpretation or labeling of the classes once the best model representing the data had been established. For the second step of the LCA, as the name suggests, the results were interpreted and subsequently the classes were labeled based on the criteria of class homogeneity, class interpretation, and item proportion endorsement. The sections below report the steps taken in the latent class analysis in the context of the proposed hypotheses.

Were Latent Classes of Math Achievement Ability Observed?

Latent Class Model Selection. Once an LCA was computed for a number of classes (e.g. 1-7 classes), then it was necessary to determine the best fit according to the indices. Indicators used for determining the number of classes include the Bayesian Information Criterion (BIC) (Schwartz, 1978), and the adjusted Bayesian Information Criterion (adjusted BIC) (Sclove, 1987), where smaller values for these indicators suggest a better statistical model (Nylund, Asparouhov, & Muthén, 2007). When statistical indicators conflict, the BIC and adjusted BIC are generally the best indicators used for latent class analysis, while the BIC has been found to be the best overall indicator of model fit when the sample is not large (i.e. n=200), there are unequal class sizes, and the data are categorical (e.g., Nylund, Asparouhov, & Muthén, 2007).

Another indicator included was the measure of entropy, with values ranging between 0 and 1, where a value closer to 1 is evidence for differences between profiles of individuals or latent classes (e.g. class differentiation) (Ramaswamy, Desarbo, Reibstein, & Robinson, 1993). Suggested guidelines for the use of the entropy indicator are, a value of .4 and below, values of around .6, and values greater than .8 are evidence for low, medium, and high-class differentiation, respectively (Clark & Muthén, 2009). Lastly, other indicators used to aid in the interpretation of the information criteria were the Bayes Factor (BF) pairwise comparison that was used to evaluate the relative fit of two latent class models; where Jeffrey's Scale of Evidence (Wasserman, 2000) recommends that a BF score between 1 and 3 is weak, BF between 3 and 10 is moderate, and BF greater than 10 is strong evidence for heterogeneity (Masyn, 2013).

In addition to the above indicators the Lo-Mendell-Rubin (LMR) likelihood test of model fit (Lo, Mendell, & Rubin, 2001) were used to compare a given class model with the previous class tested. A p-value that is due to chance ($p > .05$) indicates that the previously tested model (e.g. k-1, one less class than the one estimated) was the better model for the data. Another test of model rejection commonly used was the Bootstrap Likelihood Ratio Test (BLRT) (Nylund et al., 2007). For this indicator, bootstrap draws or datasets were created from the original dataset, the comparison test was computed and then compared between the tested model and the previously tested (e.g. smaller or k-1 classes) model. Similar to the LMR, a significant p-value for this test indicated that the larger model (e.g. k) was an improvement or preferred model to the k-1 classes (e.g. previously tested model). The LMR and BLRT were used as critical indicators to determine the profile (e.g. number of classes) that was best represented by the data. Present consensus suggests that the BLRT is an overall better indicator for selecting classes (Nylund et al., 2007), and the results for this investigation depended heavily on this indicator. If the LMR and BLRT conflict the literature suggests that preference be given to the BLRT (Nylund et al., 2007).

Table 3 shows the fit indices for all the latent class models computed. For ease of interpretation the bolded values indicate the critical values for interpretation of the models. The table shows that class two was the indicator with the lowest BIC. Subsequent models for classes three, and four had scores that were near that of class two and were also acceptable models. Another indicator, the population adjusted BIC indicator also supported the conclusion with models two, three, four, five, and six having the lowest values, respectively. The LMR results suggested that no improvement was observed after the fifth latent class analysis. Similarly, the other indicator of model rejection, the BLRT showed that no improvement was observed for the same, the fifth class model, and subsequent models.

Taken together the results show that most indicators aligned in favor of model four. The BIC and population adjusted BIC agreed that model four was the most parsimonious, as did the LMR and BLRT indicators. Minimal change was observed for the entropy index, which ranged between .61 and .79. The entropy index guildelines suggested moderate certainty of the selection of the four-class model indicator. Another indicator, the Bayes Factor pairwise comparison test (BF) provided evidence that the competing model was not the preferred model and therefore provided evidence that the four-class model was homogenous and ideal, in this case.

Another indicator of class selection is the participant distribution for a given number of classes. To ensure that the data were not over-distributed (e.g. spread too thin) amongst the classes it was important to observe that the class proportions for the selected number of classes were not approaching zero and confirmed that classes can be distinguished (e.g. separated) (Masyn, 2013). For example, a class that was observed as spread too thin will likely result from misidentification and wrong class selection in the previous step. Table 4 shows the proportions and class counts for the models tested. The proportion of participants for the latent classes ranged from zero to one hundred percent. Note that classes five, six, and seven were spread too thin or over-distributed (e.g. class count less than 3 percent) and this made these classes potentially more susceptible for misidentification. This piece of evidence further confirmed the conclusion that the four-class model was the most parsimonious. In the four-class model forty percent of individuals were in the first group, twenty-nine percent in the second group, sixteen percent in class three, and fourteen percent in class four. The following section describes the remaining results, including labeling, and the rest of the groups in detail.

To summarize, a latent class analysis was computed for up to seven classes. The evidence supported the four-class model as the best representation of the data. Because four classes emerged the results were interpreted to suggest that the sample was best judged as reflecting a mixture or discrete distribution rather than a continuous distribution.

Do the Latent Classes Observed Reflect Distinct Profiles of Mathematic Achievement Ability?

Class Interpretation. The next step in the analysis required interpretation of the classes based on the criteria described above. This part of the analysis allowed the identification of the different profiles produced in the analyses based on the mathematic achievement ability indicators. Two criteria were used to determine evidence for class endorsement: class homogeneity and class separation (Collins & Lanza, 2010; Masyn, 2013). Class separation determined the degree to which classes are different from each other (Masyn, 2013). This part of the interpretation process ensured that the profiles were unique to warrant its own profile. Masyn (2013) explained that one approach to calculate the degree to which classes were different from each other was to use the odds ratio test of endorsement. As the researcher explained, the formula was used to compute the "differentness" between classes' j and k for a particular core item used in the LCA. The formula produced an odds ratio that represents the class separation via the endorsement of a given core variable.

$$\hat{OR}_{mj/k} = (\hat{\omega}_{m|j/k} - \hat{\omega}_{m|j})/(\hat{\omega}_{m|k/1} - \hat{\omega}_{m/k})$$

The left side of the equals sign corresponds to the resulting odds ratio of participants (e.g. j) performing below the 25^{th} percentile for a given core item (e.g. m), for a given class (e.g. k). An odds ratio greater than 5 represented a high degree of separation and an odds ratio less than .2 also indicated a high degree of separation (Masyn, 2013).

Class homogeneity was used here as the overall endorsement of a given class for a specific core factor. Class homogeneity is reported as a probability score that varies from 0 to 1.0. Here, although not a widely used standard, probabilities of .50 and greater on a given core item were considered meaningful for determining risk status (e.g. high endorsement). Another criteria used to assign endorsement were the class proportions, where higher proportion endorsement or low proportion endorsement determined the likelihood of a given deficit for that class. That is, the more defined the deficit or strength of a given item was observed would facilitate the interpretation for the class.

Table 5 shows the item response probabilities to assess homogeneity within the classes. This table informs the uniqueness of the classes as well as the labeling of the classes. To facilitate in the labeling of the latent classes, Table 5 shows the item response probabilities that represented the manifest measures in the four-class model. A high probability for this table represented an increased likelihood that these individuals will perform at or below the 25^{th} percentile. Class 1 yielded probability scores in the low range (.01-.31) suggesting that this group was a relatively high achieving group that performed well below the .50 probability threshold. Class 2 also resulted in low probability of risk except for one of the core domains that was high. For class 2, the core domain that indicated risk was the area of word problem solving accuracy that was above the .50 threshold. Class 3 was the lowest performing group with most of the indicators, with the exception of the WRAT-read and Raven, probabilities greater than the predetermined cutoff. Lastly, class 4 was at risk for underperformance on core items that represented calculation (probabilities greater than .50 for WIAT and WRAT), and a component of word problem solving accuracy (probability greater than .50 for TOMA). These probabilities therefore resulted in profiles that yielded unique deficits that resulted from the interaction of the at risk deficit domains. It is argued that Table 5 provided evidence for the homogeneity between classes for the four-class model, which shows that the different profiles produced in the latent class analysis are likely to perform differently as a result of the deficits observed. A low probability of performing (e.g. < .5) below the 25th percentile across all the measures was indicative of low risk for a given indicator.

In table 6, odds ratios were computed as an indicator of between class performances. For the odds ratio comparisons, class one scores appeared to be most pronounced in that evidence for degree separation was found when compared to other classes. That is, most of the odds ratios were below the <. 20 cut-off. The wide range achievement test was close but did not achieve the criteria for separation. When compared to class three, class 1 was highly separated with most indicators below the <. 20 cutoff. Similar results were observed when comparing class one versus class four; however, the TOMA and the Raven were not observed to be below the separation cut off scores. There was also evidence of separation for class two that was observed when compared to class three. The Raven was an ambiguous indicator when comparing class two versus class three, however. Class two and four were highly separated classes by the calculation items and the CMAT representing word problem solving accuracy. Lastly, class three was separated from class four on the core items representing reading, word problem solving accuracy, and fluid intelligence. The items representing calculation did not differentiate between classes three and four. To summarize, evidence for separation and homogeneity between classes was mostly supported across the core items representing each of the four latent constructs. It can be confidently stated that the profiles observed are unique representations of individuals that make up the mathematic achievement ability groups.

Appendix E shows the performance of the groups across the core domains (e.g. manifest variables) used in the LCA. Note that class three was the lowest performing group with some overlap on the calculation domain (e.g. WIAT and WRAT-Math) with class four. Class three performed below the 25^{th} percentile (the lowest level) on items related to reading, math calculation, and word problem solving accuracy indicating that this class represented the low achieving group. Class one, performed the best of all groups with the lowest probability of performing below the 25^{th} percentile on all the domains considered. Class two showed that this group performed averaged however with difficulties that increased the risk in the core area of word problems. Class four was at-risk as a result of underperformance in the areas of calculations and word problem solving accuracy.

In the previous paragraph class performance was discussed; following will be how class homogeneity and overall endorsement were considered in the naming of the classes that were found to best represent the data. When considering within class performance it was clear that class one was found to have the lowest probability of performing below the 25^{th} percentile on all of the core items. Therefore, class one performed the lowest or reflected no risk of a deficit on any of the core domains. Since one of the seven core items was not clearly a low score (e.g. TOMA at .31) this group is best termed typical achievers and will now be referred to with that label. Latent class two had high item probabilities for the domain representing word problem solving accuracy (e.g. TOMA and CMAT). The subtest required children to read a word problem and solve the problem on a separate sheet of paper. Class two was therefore labeled at risk for word problem solving accuracy. Class three individuals were found to have a clear risk (e.g. high probability, >.50) of performing below the 25^{th} percentile on items related to word problem solving accuracy as reflected by the high probability on the CMAT and the TOMA, calculation due to underperformance on WIAT and WRAT-3, and partial evidence for the TORC representing the reading domain. The third class was labeled low achievers with the highest risk for word problem solving, calculation, and some evidence for risk in reading skills. For class three, the second item representing reading skills (WRAT-3 reading), calculation (CMAT and TOMA), and the Raven representing fluid intelligence were domains that the lowest performing group was not at risk of underperforming. Due to the severity of the underperformance on nearly all domains used for classification class 3 will be labeled the low achieving group and will now be referred to by that name. Individuals representing class four performed average, however, where likely to be at risk for calculation and word problem solving accuracy with the probability of performing below the 25^{th} percentile being above the >.50 threshold. Class four will be labeled at risk for MD because it was found to underperform in the critical core domains that characterize MD. Class four will from here on be referred to as the MD group.

In summary, the results show that the sample is best represented by four subgroups or profiles and that the class descriptions or profiles provide evidence for the classes described in the present literature on math learning difficulty. Consistent with our hypothesis, the groups observed represented characteristics of the math achievement ability groups observed in the literature.

In the next section, this investigation continues to answer questions about the profiles observed. Specifically, the aim of the following section is to assess the importance of the profiles to external measures relevant to math achievement ability. This follow-up question is a test of validity of the observed profiles on measures that impact the ability groups differently.

Multinomial Logistic Regression

Do Observed Latent Classes Reflect Differently on External Validation Measures?

One of the benefits of conducting an LCA is that one can obtain posterior probabilities (e.g. likelihood of participants being part of a given class). Appendix F shows a figure that describes the purpose of multinomial logistic regression in the context of the latent class analysis. For this investigation, once evidence for the four classes (e.g. u_1, u_2, u_3, u_4) was found then the posterior probabilities (e.g. the likelihood of an individual being part of a given group) was obtained using the software used to compute the LCA. These data were useful when determining the importance of variables, or auxiliary (e.g. X in the figure) variables that are important to the construct (e.g. math achievement abilities). It is important to note that the variables for external validity were not used in the latent class analysis, and therefore had no influence in producing the profiles or posterior probabilities. Therefore, using the posterior probabilities in further analyses would determine the predictive validity of the classification groups or profiles (Clark & Muthén, 2009). A multinomial logistic regression allowed an number estimation of the importance (e.g. predictive validity) of the cognitive measures in predicting the latent classes (e.g. math ability group). This statistical approach aided in determining the contribution of the executive function, phonological loop (short term memory), visual-spatial sketchpad, inhibition, naming speed, numeracy, estimation, and problem solving components to the different profiles of math ability.

The formula below describes how the multiple logistic regression was estimated.

LN (P(HA) P(TA) P(LA)P(MD)) = b_{10} + b_{11} (executive function) + b_{12} (phonological loop) + b_{13} (visual-spatial sketchpad)+ b_{14} (inhibition)+ b_{15} (naming speed)+ b_{16} (number estimation) + b_{17} (magnitude estimation) + b_{18} (problem solving components).

The above formula describes the multiple logistic regression model as a whole. However, the model here was estimated using a given profile or class as a reference (e.g. Low Achievers) and then compared to other groups or classes. In other words, multiple profiles are compared on given predictors. In doing so, the impact of a predictor on a given mathematic achievement ability group was observed. The resulting outcome should be consistent with the current existing literature.

Table 7 shows the Wald chi-square estimates, odds-ratios, and an effect size indicator based on the logistic distribution d_{cox} (Sanchez-Meca, Chacon Moscoso, & Marin-Martinez, 2003) for the multinomial logistic regression. All models tested were contrasted with the lowest performing group or class three, low achievers (LA). The first model contrasted the LA children profile to children in the typical achievers (TA) profile. The estimates were significantly in favor of the typical achievers on measures related to naming speed, correctly identifying problem solving components, and number estimation. No other estimates were significant (ps > .05) between the highest and lowest achieving groups. The second model compared typical achieving individuals at risk for word problems to those children with LA characteristics. Similar to the previous model the estimates for correctly identifying problem solving components and number estimation were in favor of the typical achieving group. No other estimates approached significance for the second model. The subsequent model compared those at risk for poor calculation to low achievers. Also favoring the higher typical achieving group were significant estimates for the latent variables for naming speed and number estimation. No other effects between performance groups were observed for the third model. When comparing MD to low achievers, an effect was observed for the indicators of problem solving components and the number estimation indicator. The last model for table 7 used MD as the reference group. Reported here is the comparison between the highest achieving (TA) group and MD. Consistent with previous models the problem solving components indicator was in favor of the high achieving group beyond chance—no other effects were observed. The effect size indicator, d_{cox}, shows that on two occasions (e.g. model one and three) and for two separate groups the number estimation indicator effect was low (e.g. <. 20). All other effect sizes were considered medium (e.g. = .5) or large (e.g. =.8).

To summarize, this section considered the question of whether the latent classes observed reflected differently on measures external to the core classification measures yet related to math achievement ability. Results showed that indeed the latent classes differed on some of the measures of external validation as evidenced by the Wald chi-square statistics test and the d_{cox} estimate of effect size. In the next chapter the results will be interpreted in detail and discussed in the context of the present literature.

Chapter 5: Discussion

This chapter provides a summary and a discussion as it relates to this investigation's research questions, related literature, and theoretical work on the study of mathematic achievement ability. First, the results will be interpreted in light of each research question and the interpretations to produce a summary of the results and findings. The first question posed above inquired about the existence of different performance groups. Here, the results overwhelmingly support the existence of latent classes on mathematic achievement ability. Second, it was asked whether the observed profiles were distinct from other performance groups. It was observed that children reflected different and unique profiles of mathematic achievement ability. The evidence for this conclusion comes from the latent class analysis where children were categorized into four different profile groups that were represented differently on the core indicators of mathematic achievement ability. Emphasis will be placed on the limitations and considerations of this investigation. The chapter will conclude with a discussion on the potential applications to educational practice, as the hope is that this investigation might be able to contribute to this area.

Are subgroups (latent classes) of individuals observed using the 25th percentile cutoff score?

The results of this investigation give support to the notion that subgroups exist on the mathematic achievement continuum. Of the seven models tested the combined evidence supports the existence of four classes as the most parsimonious model and the best representation of the data. The sample used for this investigation is therefore best characterized by a mixed distribution where different groups or profiles are represented in the data. Groups of individuals represented in the results did not however reflect all profiles described in the literature. For example, the results did not produce clear evidence of high achievers. Evidence indicated the existence of the math difficulty group, where individuals were expected to perform poorly on the key indicators. However, the evidence that was found, mainly for typical achievers, at-risk for word problem solving, low achievers, and math difficulty did reflect differences in the student's performance that would likely have tangible consequences on academic performance and academic endeavors.

Do observed groups have sufficiently different profiles (e.g., typical achievers—no risk, at risk for word problems, low achievers, and math difficulty)?

Some ambiguity was observed for the different classes in that the thresholds (e.g. >.50) for risk status were not quite reached but children generally reflected clear deficits on core items representing ability groups. Therefore, children fit the criteria that were used to label the respective performance groups. The typical achieving group reflected the most class homogeneity, followed by the at-risk groups that represented typical achieving, low achieving group, and MD group. Furthermore, when comparing the different classes, evidence for separation was observed amongst all the classes. In particular, however, the most evidence appears to be between the low achievers and the typical achieving groups. What define the classes are the at-risk domains that make up the various groups. Deficits faced by the MD group may not be as severe to impact their overall performance but would likely have an impact on performance of key items related to math achievement ability—as will be discussed in the following section. With the manifest measure endorsement for the performance below the 25th percentile the MD group showed evidence for deficit on the domain of calculation and on word problem solving accuracy. Evidence reported to compare and contrast the different profiles is used here to support the prediction that the characteristics between the different ability groups do indeed represent individuals that represent the math achievement ability groups, where different profiles have been observed.

Do latent classes differ on external measures of working memory (executive function, phonological loop, and visual-spatial sketchpad), numeracy, estimation, computational fluency, speed, and problem solving components?

This investigation sought to better understand children in elementary school who might have deficits that impede typical learning of mathematics. Children considered here to be at the lower end of the mathematic achievement distribution have a disadvantage in learning mathematic content that is critical for early and later life academic achievements and goals. Therefore, the purpose of this investigation was to determine the existence of profiles associated with at-risk for group underperformance. The second purpose was to explore the profiles associated with the different performance groups and relate them to what has been found in the literature. Thirdly, the findings were tested for accuracy on external measures related to the performance groups, in particular the lowest performing group.

Evidence for the Existence of Mathematic Achievement Ability Groups

Overall, the findings were consistent in their support for the existence of mathematic achievement ability groups. Key indicators aligned in support of the existence of unique profiles for three (e.g. models two, three, and four) of the seven proposed LCA models, where the four-class model was selected as the most parsimonious model. While the notion of discrete ability groups has been previously assumed in investigations, exploring the existence of the discrete distribution for mathematic achievement ability groups allowed for the identification of the profiles to better understand these ability groups. Subsequent analyses were therefore aimed at exploring the profiles observed in the discrete distribution. Observing distinct profiles is important because it provides evidence for the existence of discrete groups of children with severe deficits. Finding evidence for these subgroups has been previously corroborated in some form or another in the literature (Geary, 2013; Geary et al., 2009). To substantiate the existence of the subgroups observed, further comparisons were made to assess performance differences on external measures. This type of testing allowed for evidence of external validation of the observed profiles.

The third question proposed in this investigation is important because it would determine whether the latent classes could be estimated by measures externally related to the classification profiles. Latent variables that were successful in distinguishing performance differences between the latent classes were naming speed, problem solving components, and number estimation. As discussed in the introduction, research has provided support for the respective use of the latent variables in differentiating between math achievement ability groups: naming speed, and problem solving components (Swanson & Beebe-Frankenberger, 2004), and number estimation (Geary, Bailey, & Hoard, 2009; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007). The results produced evidence for differentiation in performance between classes when using external latent measures that are domain specific and domain general, as in the case of naming speed.

Most of the evidence found to have successfully differentiated between classes in this investigation is best explained by the first mechanism described in chapter 2. The first described mechanism argues that the deficit children with MD and LA have is mainly due to a deficit in fluency that is generally compensated with the use of slower and “immature” counting practices (Swanson, 2011). This explanation accounts for the differences in-group performance observed in the naming speed domain. For example, one explanation makes the case that naming speed can expedite access to numbers as well as reduce the risk for computation underperformance (Swanson & Beebe-Frankenberger, 2004). This lack of fluency can potentially manifest itself in other important domains such as word problem solving accuracy where children are expected to follow a sequence of steps to solve a problem, as in the case of a problem solving operation or algorithm (Geary, Hoard, Byrd-Craven & DeSoto, 2004; Swanson, 2011). This second finding can potentially support the third mechanism since it is critical and would explain the deficit children may have in the domain of word problem solving accuracy. This domain has been found to be critical to working memory (e.g. executive function) in that it is important to attend to information related to word problem solving accuracy (Ashkenazi, et al., 2013; Geary, 2004). This investigation found evidence that higher order processes (e.g. cognition) played a role in predicting differences between groups. Moreover, this investigation found evidence, direct and indirect, supporting the three mechanisms discussed in chapter 2. The findings highlights the complexity of the domains associated with mathematic achievement ability and emphasizes the importance of higher order process on the lower order processes.

Lastly, evidence supported the notion that subgroups exist within the math achievement ability distribution and that this phenomenon can be thought of as discrete rather than on a continuum. This comparison was tested and provided support for this notion. Students who take standardized tests are consistently placed in honor classes if they perform well or remediation courses if the opposite is true. Classifying individuals can assist in finding ways to help students' overcome or compensate for their deficits. Such an approach will provide more definitive evidence on the conclusion of the present results. Additionally, the use of demographic covariates (i.e. ethnicity, gender, socio-economic status, motivation, bilinguals) in the model should be tested to assess whether key variables influence the categories produced when using the LCA approach.

Limitations and Considerations

Several limitations are evident in this preliminary work. One major limitation was the lack of longitudinal data. To be able to determine if the same latent classes can be observed over time; that is, consistency within the latent classes, would support the literature suggesting that the participants' latent class assignment is fixed and not fluid. Also, as it is usually the case, a larger sample would increase the power to detect the hypothesized profiles. Future investigations should run the latent class models using dichotomous and continuous data and testing for similarities between the two approaches (see Ruscio & Walters, 2009). Another limitation to consider is that the present results would not likely generalize to special populations such as English language learners. Finally, designing an investigation to potentially disentangle the hierarchy of the mechanisms may further explain how to best address the deficits associated with math achievement ability.

Potential Application to Educational Practice

The aim of this investigation was to explore the observed latent classes in order to better identify individuals and understand the unique profiles of those who may be at risk for math learning difficulty. Using a more conservative cutoff score and finding unique profiles of individuals should yield critical information on the differences between the identified groups. Also, a better understanding of the math achievement ability groups should aid in the problem of misclassification of children into special education. In combination with other literature, the results of this investigation shed light into the domains where participants with math learning difficulty and low achievers have the most deficits. This investigation should be used to highlight the major deficits so that researchers working on interventions apply these results in a way that might reduce underperformance in math achievement. One of the ways the results of this investigation can be used to target the profiles observed here in an intervention is to identify the lowest performing profiles of individuals and target the deficits. For example, the findings revealed that all children who are at risk in this sample would benefit from targeted intervention addressing deficits in the word problem solving accuracy domain. On the other hand, an intervention in the domain of calculation for the low achievers and MD group would be most beneficial for them. It is clear from the findings here that targeting low achievers for more general domains, such as reading, would benefit the group with a reading deficit and not the other profiles of individuals deficits such as those observed here. Thus, targeting the deficits in an intervention will ensure that the child will gain the most from the added instruction. Finally, addressing the deficits children may have in learning mathematics can potentially improve the learning process but also their experience of education.

References

Andersen, J.R., Reder, L.M., & Lebiere, C. (1996). Working memory: Activation and limitations on retrieval. *Cognitive Psychology, 30,* 221-256. doi: 10.1006/cogp.1996.0007

Aschbacher, P., Ing, M., & Tsai, S. (2014). Is science me? Exploring middle school students' STE-M career aspirations. *Journal of Science Education and Technology, 23*(6), 735-743. doi: 10.1007/s10956-014-9504-x

Ashkenazi, S., Black, J.M., Abrams, D.A., Hoeft, F., & Menon, V. (2013). Neurobiological Underpinnings of Math and Reading Learning Disabilities. *Journal of Learning Disabilities, 46*(6) 549-569. doi: 10.1177/0022219413483174

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49*(A), 5-28. doi: 10.1080/027249896392784

Barbaresi, W.J., Katusic, S.K., Colligan, R.C., Weaver, A.L., & Jacobson, S.J. (2005). Math learning disorder: Incidence in a population-based birth cohort, 1976-1982, Rochester, Minn. *Ambulatory Pediatrics, 5*, 281-289. doi: 10.1367/a04-209r.1

Barrouillet, P., Fayol, M., & Lathuliere, E. (1997). Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning disabilities. *Journal of Behavioral Development, 21,* 253-275. doi: 10.1177/1053451212472232

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246. doi: 10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588. 0033-2909/80/8803-0$88$00.75

Berch, D. B., & Mazzocco, M. M. (2007). *Why Is Math So Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities*. Brookes Publishing Company; Baltimore, MD.

Brown, V.L., Cronin, M.E., & McIntire, E. (1994). *Test of mathematical abilities (2nd ed.)*. Austin, TX: PRO-ED.

Brown, V.L., Hammill, D., & Weiderholt, L. (1995). *Test of reading comprehension.* Austin, TX: PRO-ED.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205-228. doi: 10.1080/87565640801982312

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology, 65*(1), 1-24. doi: 10.1006/jecp.1996.2358

Bull, R., Johnson, R.S. & Roy, J.A. (1999). Exploring the roles of the visual-spatial sketchpad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology, 15*, 421-442. doi: 10.1080/87565649909540759

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19*(3), 273-293. doi:10.1207/s15326942dn1903_3

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, 46*, 3-18. doi: 10.1111/j.1469-7610.2004.00374.x

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. *Science, 332*(6033), 1049-1053. doi: 10.1126/science.1201536

Case, R. (1992). *The minds staircase: Exploring the conceptual underpinnings of children's thought and knowledge.* Hillsdale, NJ: Erlbaum.

Cattell, R.B. (1971). *Abilities: Their structure, growth and action*. Boston, MA: Houghton Mifflin.

Clark, S.L. & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. Retrieved from: www.statmodel.com/download/relatinglca.pdf

Collins, L. M., & Lanza, S. T. (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences (Vol. 718). Wiley. com. doi: 10.1002/9780470567333

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32*(3), 277-296. doi: 10.1016/j.intell.2003.12.002

Cortiella, C., Horowitz, S.H. (2014). *The state of learning disabilities: Facts, trends, and emerging issues*. NY, NY: National Center for Learning Disabilities.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi: 10.1016/s0022-5371(80)90312-6

Deary, I.J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13-21. doi: 10.1016/j.intell.2006.02.001 S0160289606000171

Denson, N., & Ing, M. (2014). Latent class analysis in higher education: An illustrative example using pluralistic orientation. *Research in Higher Education, 55*(5), 508-526. doi:10.1007/s11162-013-9324-5

Desoete, A., Roeyers, H., & De Clercq, A. (2004). Children with mathematics learning disabilities in Belgium. *Journal of Learning Disabilities, 37*(1), 50-61. doi: 10.1177/00222194040370010601

DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353-386. doi: 10.1080/09541440244000328

Dirks, E., Spyer, G., van Lieshout, E.C., & de Sonneville, L. (2008). Prevalence of combined reading and arithmetic disabilities. *Journal of Learning Disabilities, 41,* 460-473.

Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 324. doi:10.1177/00222194050380040801

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428. doi: 10.1037/0012-1649.43.6.1428

Fletcher, J.M., Francis, D.J., Shaywitz, S.E., Lyon, G.R., Foorman, B.R., Stuebing, K.K., et al. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice, 13,* 186-203. doi: 10.1177/00222194050380020101

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*(2), 98-108. doi: 10.1177/00222194050380020101

Fuchs, L.S., Compton, D.L., Fuchs, D., Paulsen, K., Bryant, J.D., & Hamlett, C.L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493-513. doi: 10.1037/0022-0663.97.3.493

Fuchs, L.S., Fuchs, D., Compton, D.L., Powell, S.R., Seethaler, P.M., Caprizzi, A.M. et al., (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*,29-43. doi:10.1037/0022-0663.98.1.29

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem- solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities, 37*, 293-306. doi: 10.1177/00222194040370040201

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49*(3), 363-383. doi: 10.1016/0022-0965(90)90065-g

Geary, D.C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114,* 345-362. doi: 10.1037//0033-2909.114.2.345

Geary, D.C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4-15. doi: 10.1177/00222194040370010201

Geary, D.C. (2011). Cognitive predictors of individual differences in achievement growth in mathematics: A five-year longitudinal study. *Developmental Psychology, 47,* 1539-1552.

Geary, D.C. (2013). Learning Disabilities in Mathematics: Recent Advances. In H. L. Swanson, K. R. Harris, and S. Graham (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 593-606). New York, NY: The Guilford Press. doi: 10.1016/b978-0-12-374748-8.00002-0

Geary, D.C., Bow-Thomas, C., Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology, 54*, 372-391. doi: 10.1016/0022-0965(92)90026-3

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., & Nugent, L.

(2009). First-grade predictors of mathematical learning disability: A latent class trajectory analysis. *Cognitive Development, 24*(4), 411-429. doi: 10.1016/j.cogdev.2009.10.001

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool the number sets test. *Journal of Psychoeducational Assessment, 27*(3), 265-279. doi: 10.1177/0734282908330592

Geary, D.C., Hoard, M.K., Byrd-Craven, J., & DeSoto, C.M. (2004). Strategy choices is simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 74,* 213-239. doi: 10.1016/j.jecp.2004.03.002

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*(1), 206-223. doi: 10.1037/a0025398

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*(3), 277-299. doi: 10.1080/87565640801982361

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78*, 1343-1359. doi: 10.1111/j.1467-8624.2007.01069.x

Horn, H.L. (1980). Concepts of intellect in relation to learning and adult development. Intelligence, 4, 285-317. doi: 10.1016/0160-2896(80)90025-2

Horn, H.L. & Noll, J. (1997). Human cognitive abilities: Gf-Gc theory. In Flanagan, D.P., Genshaft, J.L., & Harrison, P.L. (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York, NY: Guilford Press.

Hresko, W., Schlieve, P.L., Herron, S. R., Swain, C., & Sherbenou, R. (2003). *Comprehensive Math Abilities Test*. Austin, TX: PRO-ED.

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, 106*(22), 8801-8807. doi: 10.1073/pnas.0901265106

Ing, M., & Nylund-Gibson, K. (2013). Linking early science and mathematics attitudes to long term science, technology, engineering and mathematics career

attainment: Latent class analysis with proximal and distal outcomes. *Educational Research and Evaluation, 19*(6), 510-524. doi: 10.1080/13803611.2013.806218

Jordan, N.C., Hanich, L.B., Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development, 74,* 834-850. doi: 10.1111/1467-8624.00571

Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149. doi; 10.1037//0033-295x.99.122

Kail, R., & Hall, L.K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition, 29,* 1-9. doi: 10.3758/bf03195735

Kline, R.B. (2011). *Principles and practice of Structural Equation Modeling*. New York, NY: Guilford Press.

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual–spatial working memory, and preschool quantity–number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology, 103*(4), 516-531. doi: 10.1016/ j.jecp.2009.03.009

Kyllonen, P.C. & Christal, R.E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14,* 389-433. doi: 10.1016/S0160-2896(05)80012-1

Lanza, S. T., Collins, L. M., Lemmon, D. R., & Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling, 14*(4), 671-694. doi: 10.1080/10705510701575602

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A., & Collins, L. M. (2013). *PROC LCA & PROC LTA users' guide* (Version 1.3.0). University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu

LeBlanc, M. D., & Weber- Russell, S. (1996). Text integration and mathematical connections: a computer model of arithmetic word problem solving. *Cognitive Science*, *20*(3), 357-407. doi: 10.1207/s15516709cog2003_2

Lo, Y., Mendell, N., & Rubin, D.B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767-778. doi: 10.1093/biomet/88.3.767

Lyon, G.R., Fletcher, J.M., Shaywitz, S.E., Shaywitz, B.A., Torgesen, J.K., Wood, F.B., et al., (2001). Rethinking learning disabilities. In C.E. Finn, Jr., A.J. Rotherham, & C.R. Hokanson, Jr. (Eds.), *Rethinking special education for a new century* (pp. 259-287). Washington, D.C.: Fordham Foundation.

Marcoulides, G. A., Gottfried, A. E., Gottfried, A. W., & Oliver, P. H. (2008). A latent transition analysis of academic intrinsic motivation from childhood through adolescence. Educational Research and Evaluation, 14(5), 411-427. doi: 10.1080/13803610802337665

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 375-393). Oxford, UK: Oxford University Press.

Mayer, R.E., & Hegarty, M. (1996). The process of understanding mathematical problem solving. In R.J. Sternberg & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 29-54). Mahwah, NJ: Erlbaum.

Mazzocco, M.M.M. (2007). Defining and differentiating mathematical learning disabilities and difficulties. In Berch, D.B., & Mazzocco, M.M.M. (Ed.), *Why is math so hard for some children?: The nature and origins of mathematical learning difficulties and disabilities* (pp. 29-48). Baltimore, Maryland: Paul H. Brookes Publishing Co., Inc.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child development, 82*(4), 1224-1237. doi: 10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M., & Kover, S. T. (2007). A longitudinal assessment of executive function skills and their association with math performance. *Child Neuropsychology, 13*(1), 18-45. doi: 10.1080/09297040600611346

McCloskey, M. (2007). Quantitative literacy and developmental dyscalculias. In Berch, Daniel B. (Ed); Mazzocco, Michèle M. M. (Ed), (2007). *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (p.415-429). Baltimore, MD: Paul H Brookes Publishing.

McLean, J.F., & Hitch, G.J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74*, 240-260. doi: 10.1006/jecp.1999.2516

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology*, *41*(1), 49-100. doi: 10.1006/cogp.1999.0734

Murphy, M. M., Mazzocco, M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*(5), 458-478. doi: 10.1177/00222194070400050901

Muthén, B.O. (2002). Latent variable mixture modeling. In G.A. Marcoulides, & R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp.1-33). Mahwah, N.J.: Lawrence Erlbaum Associates.

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the National Mathematics Advisory Panel*. Washington, D.C.: United States Department of Education. Retrieved from: http://www.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569. doi: 10.1080/10705510701575396

Paglin, M., & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics, 8*(1), 123-144. doi: 10.1086/298239

Passolunghi, M.C., Cornoldi, C., & DeLiberto, S. (1999). Working memory intrusions of irrelevant informationin a group of specific poor problem solvers. *Memory and Cognition, 27*, 779-790. doi: 10.3758/bf03198531

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88*(4), 348-367. doi: 10.1016/j.jecp.2004.04.002

Petrill, S., Logan, J., Hart, S., Vincent, P., Thompson, L., Kovas, Y., & Plomin, R. (2012). Math Fluency Is Etiologically Distinct From Untimed Math Performance, Decoding Fluency, and Untimed Read Perform Evidence From a Twin Study. *Journal of Learning Disabilities, 45*(4), 371-381. doi:10.1177/0022219411407926

PROC LCA & PROC LTA (Version 1.3.0) [Software]. (2013). University Park: The

Methodology Center, Penn State. Retrieved from http://methodology.psu.edu

Psychological Corporation (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Harcourt Brace & Co.

Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*(1), 103-124. doi: 10.1287/mksc.12.1.10

Raven, J.C. (1976). *Colored progressive matrices*. London, U.K.: H.K. Lewis.

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science, 24*, 1301-1308. doi: 10.1177/0956797612466268

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources, 27*(2), 313-328. doi: 10.2307/145737

Ruscio, J., & Walters, G. D. (2009). Using comparison data to differentiate categorical and dimensional data by examining factor score distributions: Resolving the mode problem. *Psychological Assessment, 21*(4), 578-594. doi:10.1037/a0016558

Sanchez-Meca, J., Marin-Martinez, F., Chacon-Moscoso, S. (2003). Effect size indices for dichotomized outcomes in meta analysis. *Psychological Methods, 8,* 448-467.

SAS Institute Inc. 2008. Base SAS® 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464. doi: 10.1214/aos/1176344136

Schweizer, K. & Koch, W. (2001). A revision of Cattell's invesment theory: Cognitive properties influence learning. *Learning and Individual Differences, 13,* 57-82. doi: 10.1016/S1041-6080(02) 00062-6

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333-343. doi: 10.1007/bf02294360

Senf, G.M. (1986). LD research in sociological and scientific perspective. In Torgesen, J.K., & Wong, B.Y.L. (Ed.), *Psychological and educational perspectives on learning disabilities* (pp. 27-53). Orlando, FL: Academic Press.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled. *Child Development, 60*, 973-980. doi:10.1111/j.1467-8624.1989.tb03528.x

Siegler, R.S., & Opfer, J. (2003). The development of numerical estimation: Evidence formultiple representation of numerical quantity. *Psychological Science, 14*, 237-243. doi: 10.1111/1467-9280.02438

Seigler, R. S. & Booth, J. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428-444. doi: 10.1111/j.1467-8624.2004.00684.x

Shalev, R.S., (2007). Prevalence of developmental dyscalculia. In Berch, D.B. & Mazzocco, M.M.M. (Ed.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (p. 49-60). Baltimore, MD: Paul H Brookes Publishing.

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia- prevalence and prognosis. *European Child & Adolescent Psychiatry, 9*(2), S58-S64. doi: http://dx.doi.org/10.1007/s007870070009

Shalev, R.S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: a prospective 6- year follow-up of a common learning disability. *Developmental Medicine & Child Neurology, 47*, 121-125. doi: 10.1111/j.1469-8749.2005.tb01100.x http://dx.doi.org/10.1111/j.1467-8624.2004.00684.x

St. Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*(4), 745-759. doi: 10.1080/17470210500162854

Swanson, H. L. (1992). Generality and modifiability of working memory among skilled and less skilled readers. *Journal of Educational Psychology, 84*, 473-488. doi: 10.1037//0022-0663.84.4.473

Swanson, H.L. (1993). Working memory in learning disability subgroups. *Journal of Experimental Child Psychology, 56*, 87-114. doi: 10.1006/jecp.1993.1027

Swanson, H. L. (1995). *S-Cognitive Processing Test (S-CPT): A Dynamic Assessment Measure*. Austin, TX: PRO-ED.

Swanson, H. L. (1999). What develops in working memory? A life span perspective. *Developmental Psychology, 35*, 986-1000. doi: 10.1037/0012-1649.35.4.986

Swanson, H. L. (2006a). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology, 98*(2), 265-281. doi: 10.1037/0022-0663.98.2.265

Swanson, H. L. (2006b). Cognitive processes that underlie mathematical precociousness in young children. *Journal of Experimental Child Psychology, 93*(3), 239-264. doi: 10.1016/j.jecp.2005.09.006

Swanson, H. L. (2011b). Working memory, attention, and mathematical problem solving: A longitudinal study of elementary school children. *Journal of Educational Psychology, 103*(4), 821-837. doi: 10.1037/a0025114

Swanson, H.L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology, 100*, 597-602. doi: 10.1037/0022-0663.100.3.581

Swanson, H. L., & Ashbaker, M. (2000). Working memory, short-term memory, articulation speed, word recognition, and reading comprehension in learning disabled readers: Executive and/or articulatory system? *Intelligence, 28*, 1-30. doi: 10.1016/s0160-896(99)00025-2

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence, 21*(1), 83-108. doi: 10.1016/0160-2896(95)90040-3

Swanson, H. L., Cooney, J. B., & Brock, S. (1993). The Influence of Working Memory and Classification Ability on Children′ s Word Problem Solution. *Journal of Experimental Child Psychology, 55*(3), 374-395. doi:10.1006/jecp.1993.1021

Swanson, H.L. & Alloway, T.P. (2012). Working memory, learning, and academic achievement. In Harris, K.R., S. Graham, & Urdan, T. (Eds.) *APA Educational Psychology Handbook: Vol. 1. Theories, Constructs, and Critical Issues*. (pp. 327-366). Washington, D.C., U.S.: American Psychological Association. doi: 10.1037/13273-012

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not a risk for serious math difficulties. *Journal of Educational Psychology, 96*, 471-491. doi: 10.1037/0022-0663.96.3.471

Swanson, H.L., Jerman, O., & Zheng, X. (2009). Math disabilities and reading disabilities: Can they be separated? *Journal of Psychoeducational Assessment, 27*, 175-196. doi: 10.1177/0734282908330578

Swanson, H.L., Orosco, M.J., Lussier, C.M. (2013). The effects of mathematics strategy instruction for children with serious problem solving difficulties. *Exceptional Children, 80*(2), 149-168. doi: 10.1177/001440291408000202

Swanson, H.L. & Sachse-Lee, C., (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79*, 294-321. doi: 10.1006/jecp.2000.2587

Swanson, H.L., Sáez, L., Gerber, M., & Leafstedt, J. (2004). Literacy and cognitive functioning in bilingual and nonbilingual children at risk for reading disabilities. *Journal of Educational Psychology, 96*, 3-18. doi: 10.1037/0022-0663.96.1.3

Thorndike, R.M. & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education*. Massachusetts: Pearson Education Inc.

Towse, J. N. (1998). On random generation and the central executive of working memory. *British Journal of Psychology, 89*, 77-101. doi: 10.1111/j.2044-8295.1998.tb02674.x

U.S. Department of Education, National Center for Education Statistics. (2013). *Digest of Education Statistics, 2012* (NCES 2014-015). Retrieved from: http://nces.ed.gov/fastfacts/display.asp?id=64

Von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology*, 49(11), 868-873. doi: 10.1111/j.1469-8749.2007.00868.x

Vukovic, R. K. (2012). Mathematics difficulty with and without reading difficulty-Findings and implications from a four-year longitudinal study. *Exceptional children, 78*(3), 280-300. doi: 10.1177/001440291207800302

Vukovic, R. K., Lesaux, N. K., & Siegel, L. S. (2010). The mathematics skills of children with reading difficulties. *Learning and Individual Differences, 20*(6), 639-643. doi: 10.1016/j.lindif.2010.08.004

Vukovic, R. K., & Siegel, L. S. (2010). Academic and cognitive characteristics of persistent mathematics difficulty from first through fourth grade. *Learning Disabilities Research & Practice, 25*(1), 25-38. doi: 10.1111/j.1540-5826.2009.00298.x

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44*, 92-107. doi: 10.1006/jmps.1999.1278

Wagner, R., Torgesen, J., & Rashotte, C. (2000). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: Psychological Corporation.

Widaman, K.F. (2012). Exploratory factor analysis and confirmatory factor analysis. In Cooper, H. (Eds.),(2012). *APA handbook of research methods in psychology: Vol. 3 Data analysis and research publication* (pp. 361-389). Washington, D.C.: American Psychological Association. doi: 10.1037/13621-018

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8*(1), 16-37. doi: 10.1037/1082-989x.8.1.16

Wilkinson, G. S. (1993). *The Wide Range Achievement Test*. Wilmington D.E.: Wide Range, Inc

Table 1

Descriptive Statistics for Standard Scores for Classification Measures

Measure for Classification	Mean	SD	Skewness	Kurtosis	Reliability
Fluid Intelligence					
RAVEN	101.71	13.57	-0.86	1.22	
Calculation					
WIAT	99.07	12.82	0.15	-0.16	0.92
WRAT	99.17	10.39	0.06	0.42	0.90
Reading Skills					
TORC	101.71	13.57	-0.86	1.22	0.83
WRAT-Reading	104.12	12.5	0.22	1.18	0.95
Word Problem Solving Accuracy					
TOMA	89.62	10.91	0.61	-0.35	0.86
CMAT	99.29	15.06	-0.21	-0.92	0.92

Note. Reliability=Cronbach's alpha or KR_{20}. Raven = Colored Progressive Matrices Test. WIAT = Wechsler Individual Achievement Test. WRAT = Wide Range Achievement Test. TORC = Test of Reading Comprehension. WRAT-Reading = Wide Range Achievement Test Reading Task. Question, Number, Goal, Operation, Algorithm, and Irrelevant = Word Problem Solving Memory Booklet. TOMA = Test of Mathematical Abilities. CMAT = Comprehensive Mathematical Abilities.

Table 1 continued

Descriptive Statistics Measures for External Validation

Measures for External Validation	Mean	SD	Skewness	Kurtosis	Reliability
Inhibition					
Random Number Generation	5.96	3.63	0.82	1.18	0.94
Random Letter Generation	7.73	3.40	0.19	0.82	0.93
Central Executive					
Listening Sentence Span	3.18	2.18	0.57	1.53	0.84
Auditory Digit Sequence	5.41	1.82	0.83	1.79	0.75
Update	4.92	3.00	0.63	1.77	0.92
Visual-Spatial Sketchpad					
Mapping and Directions	5.89	3.03	1.28	3.97	0.91
Visual Matrix	13.49	2.43	0.49	0.40	0.86
Phonological Loop					
Digit Forward Span	6.76	1.71	0.40	0.63	0.75
Word Problem Span	8.22	3.45	0.39	-0.49	0.93
Phonetic Memory Span	3.06	2.24	1.01	1.54	0.84
Speed					
Rapid Digit Naming	70.69	9.24	-0.85	0.84	0.71
Rapid Letter Naming	67.30	10.63	-0.67	4.99	0.68
Word Problem Components					
Question	2.18	0.89	-0.83	-0.21	0.70
Number	2.36	0.81	-1.08	0.31	0.71
Goal	1.83	0.93	-0.32	-0.80	0.70
Operation	1.89	0.83	-0.36	-0.47	0.69
Algorithm	1.79	0.85	-0.34	-0.47	0.70
Irrelevant	2.39	0.82	-1.19	0.58	0.76
Estimation					
Number Estimation Subtest 1	17.38	2.09	-1.61	4.83	0.82
Number Estimation Subtest 2	18.13	1.31	-1.74	5.18	0.56
Magnitude Judgment					
Magnitude Subtest 1	18.90	6.51	-0.14	0.48	0.98
Magnitude Subtest 2	10.36	3.73	0.13	1.88	0.94

Note. Reliability=Cronbach's alpha or KR_{20}.

Table 2

Factor Loadings for Confirmatory Factor Analysis of Classification Measures

	Reading Skills	Calculation	Word Problem Solving Accuracy	Fluid Intelligence
RAVEN	0.00	0.00	0.00	**1.01**
WIAT	0.00	**0.80**	0.00	0.00
WRAT-math	0.00	**0.82**	0.00	0.00
TORC	**0.68**	0.00	0.00	0.00
WRAT-reading	**0.70**	0.00	0.00	0.00
TOMA	0.00	0.00	**0.75**	0.00
CMAT	0.00	0.00	**0.69**	0.00

Note. Factor loadings representing the factor are in boldface. Raven = Colored Progressive Matrices Test. WIAT = Wechsler Individual Achievement Test. WRAT = Wide Range Achievement Test. TORC = Test of Reading Comprehension. WRAT-Reading = Wide Range Achievement Test Reading Task. Question, Number, Goal, Operation, Algorithm, and Irrelevant = Word Problem Solving Memory Booklet. TOMA = Test of Mathematical Abilities. CMAT = Comprehensive Mathematical Abilities.

Table 3

Fit Indices for the Number of Latent Classes Based on the Manifest Variables Separated Above and Below the 25th percentile

C	LL	DF	BIC	BIC_{adj}	VLMR	BLRT	E	$BF_{A,B}$
1	-1378.78	7	2799.22	2777.01				
2	-1277.11	15	**2643.52**	**2595.92**	0.00	0.00	0.71	1.5433E-34
3	-1259.19	23	**2655.30**	**2582.32**	0.08	0.00	0.74	3.6213E+02
4	-1248.37	31	**2681.30**	**2582.94**	**0.02**	**0.04**	**0.61**	**4.4131E+05**
5	-1243.86	39	2719.89	2596.15	0.15	1.00	0.76	7.9982E-26
6	-1240.09	47	2759.97	2610.85	0.02	0.60	0.79	1.5196E+42
7	-1236.54	55	2800.51	2626.00	0.76	1.00	0.79	6.3428E+08

Note. C=Number of Classes, LL = Log Likelihood Ratio, DF = Degrees of Freedom, BIC = Bayesian Information Criterion; BIC_{adj} = Sample Adjusted Bayesian Information Criterion, Adjusted LMR-LRT p-value (H_0: K classes; H_1: K+1 classes), Bootstrapped Likelihood Ratio Test= BLRT p-value (H_0: K classes; H_1: K+1 classes), E= Entropy, Bayes Factor Pairwise Comparison Test = BF.

Table 4

Prevalence Proportions of the Sample as a Function of the Number of Latent Classes

No. Classes	1	2	3	4	5	6	7
1	1.00	0.73	0.66	**0.40**	0.01	0.02	0.26
2	--	0.27	0.11	**0.29**	0.36	0.34	0.11
3	--	--	0.23	**0.16**	0.28	0.01	0.10
4	--	--	--	**0.14**	0.18	0.28	0.31
5	--	--	--	--	0.17	0.18	0.02
6	--	--	--	--	--	0.16	0.03
7	--	--	--	--	--	--	0.17

Note. The table represents the proportion of the total sample reflected in the latent classes.

Table 5

Probabilities for the Indicators in the Four-Class Model

Class	TA (n=155)	At-risk (n=112)	LA (n=63)	MD (n=55)
Test of Reading Comprehension (TORC-3) —Reading Skills	0.01	0.21	**0.77**[a]	0.27
Wide Range Achievement Test (WRAT-3) —Reading Skills	0.01	0.03	0.40	0.04
Wechsler Individual Achievement Test (WIAT) —Calculation	0.02	0.09	**0.64**[a]	**0.61**[a]
Wide Range Achievement Test (WRAT-3) —Calculation	0.00	0.06	**0.57**[a]	**0.62**[a]
Test of Math Abilities (TOMA) —WPSA	0.31	**0.77**[a]	**1.00**[a]	**0.57**[a]
Comprehensive Math Abilities Test (CMAT) — WPSA	0.00	**0.51**[a]	**0.87**[a]	0.13
Raven Colored Progressive Matrices —Fluid Intelligence	0.10	0.18	0.35	0.07

[a] Item response probability .5 or greater suggests that a latent class is at for performing at or below the 25th percentile.

Note. WPSA = Word Problem Solving Accuracy. Class 1 = Typical Achievers (TA), Class 2= at risk- word problems (at-risk), Class 3= low achievers (LA), Class 4= math difficulty (MD at risk--calculation and word problem solving accuracy).

Table 6

Odds Ratios Assessing Degree of Separation that Latent Class Groups are Similar on Manifest Variables

Class Comparison	TA vs. Risk	TA vs. LA	TA vs. MD	Risk vs. LA	Risk vs. MD	LA vs. MD
Test of Reading Comprehension (TORC-3)	**0.03**	**0.00**	**0.02**	**0.06**	0.73	**8.96**
Wide Range Achievement Test (WRAT-3)	0.21	**0.01**	**0.15**	**0.04**	0.71	**16.22**
Wechsler Individual Achievement Test (WIAT)	**0.19**	**0.01**	**0.01**	**0.06**	**0.06**	1.16
Wide Range Achievement Test (WRAT-3)	**0.00**	**0.00**	**0.00**	**0.05**	**0.04**	0.82
Test of Math Abilities (TOMA)	**0.14**	**0.00**	0.35	**0.00**	2.50	**0.00**
Comprehensive Math Abilities Test (CMAT)	**0.00**	**0.00**	**0.00**	**0.15**	**7.06**	**46.83**
Raven Colored Progressive Matrices	0.50	0.21	1.46	0.41	2.90	**7.03**

Note: Bold is high degree of separation (Odds ratios <0.20 or > 5); Reading = TORC-3, WRAT-3; Calculation = WIAT, WRAT-3; Word Problem Solving Accuracy = TOMA, CMAT; Fluid Intelligence = Raven Colored Progressive Matrices. Class 1 = Typical Achievers (TA), Class 2= at risk- word problems, Class 3= low achievers, Class 4= MD at risk--calculation and word problem solving.

Table 7

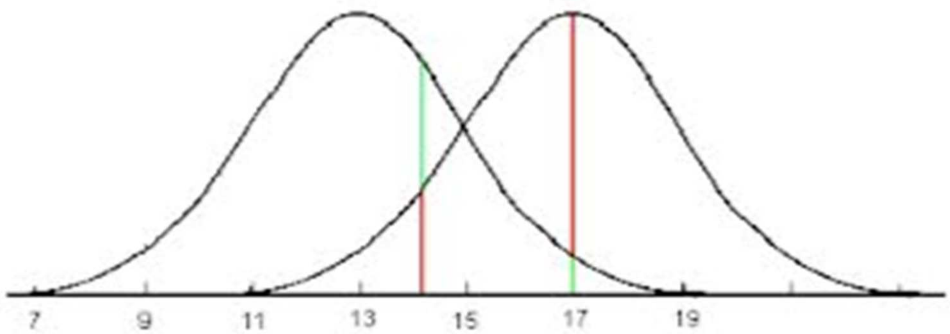
Multinomial Logistic Regression Predicting Latent Classes from Variables External to the Classification Variables

TA (no risk) vs. LA	*Wald χ^2*	*Odds Ratio*	*p-value*	d_{cox}
Intercept	34.31		<.0001	
Inhibition	1.27	0.69	0.26	0.42
Executive Function (WM)	0.06	0.73	0.80	0.44
Visual Spatial Sketchpad (WM)	0.11	0.91	0.74	0.55
Phonological Loop (WM)	0.55	2.25	0.46	**1.37**
Naming Speed	4.16	0.51	0.04	0.31
Problem Solving Components	22.39	5.43	<.0001	**3.29**
Number Estimation	38.63	0.26	<.0001	0.16
Magnitude Judgment	0.76	1.38	0.38	**0.84**
At-Risk for word problems solving vs. LA				
Intercept	28.71		<. 0001	
Inhibition	0.22	0.87	0.64	0.53
Executive Function (WM)	0.15	1.59	0.70	**0.97**
Visual Spatial Sketchpad (WM)	0.23	0.87	0.63	0.53
Phonological Loop (WM)	0.04	0.82	0.85	0.49
Naming Speed	1.36	0.70	0.24	0.42
Problem Solving Components	12.46	3.16	0.00	**1.92**
Number Estimation	9.71	0.57	0.00	0.34
Magnitude Judgment	0.35	0.82	0.55	0.49
MD vs. LA				
Intercept	7.99		0.00	
Inhibition	0.05	0.92	0.82	0.56
Executive Function (WM)	0.24	0.52	0.63	0.31
Visual Spatial Sketchpad (WM)	0.02	1.05	0.88	0.64
Phonological Loop (WM)	0.92	3.08	0.34	**1.87**
Naming Speed	0.40	0.80	0.53	0.48
Problem Solving Components	6.09	2.48	0.01	**1.50**
Number Estimation	23.72	0.32	<.0001	0.19
Magnitude Judgment	0.01	1.04	0.92	0.63
TA (no risk) vs.MD				
Intercept	18.10		<.0001	
Inhibition	1.22	0.75	0.27	0.45
Executive Function (WM)	0.15	1.41	0.70	**0.86**
Visual Spatial Sketchpad (WM)	0.50	0.86	0.48	0.52
Phonological Loop (WM)	0.16	0.73	0.69	0.44
Naming Speed	2.91	0.63	0.09	0.38
Problem Solving Components	6.76	2.19	0.01	**1.33**
Number Estimation	1.02	0.82	0.31	0.49
Magnitude Judgment	0.96	1.33	0.33	**0.80**

Note. R^2 = .35 (Cox & Snell), .37 (Nagelkerke). Model χ^2 (24) = 98.37, p < .001. *p<.05,**p<.01, ***p<0.001. TA = Typical Achiever; LA = Low Achiever; Effect size > .80 bold based on Cohen.

Appendices

Appendix A. Overall distribution can also be represented by two separate distributions



Appendix B. Profiling Individuals that Vary in Math Ability (Geary, 2013; Swanson & Alloway, 2012)

Manifest Measures Used to Determine Latent Classes	MD	LA	TA	HA
* Word Problem Solving Accuracy	x	x		
* Reading Comprehension	x			
* Fluid Intelligence	X			
* Math Calculation	X	X		
Measures Used to Assess External Validity of Classes				
Problem solving components				
Inhibition	X	X		
Naming Speed	x			
Number Estimation	X			
Magnitude Judgment	X			
-Working Memory-				
Central Executive Function (WM)	X	X		
Visual-Spatial Sketchpad (WM)	x	x		
Short Term Memory Measures (STM)/ Phonological loop (WM)	X	x		

Note. X = Clear deficit, x= Evidence suggests deficit. MD = Math difficulty, LA = Low Achievers, TA = Typical Achievers, HA = High Achievers.

Appendix C. Factor Loadings for Confirmatory Factor Analysis of External Measures

	F1	F2	F3	F4	F5	F6	F7	F8
Inhibition								
Random Number Generation	**0.60**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Random Letter Generation	**0.53**	0.00	0.00	0.00	0.00	0.00	0.00	0.00
WM-Executive								
Listening Sentence Span	0.00	**0.33**	0.00	0.00	0.00	0.00	0.00	0.00
Auditory Digit Sequence	0.00	**0.40**	0.00	0.00	0.00	0.00	0.00	0.00
Update	0.00	**0.60**	0.00	0.00	0.00	0.00	0.00	0.00
WM-Visual-Spatial								
Mapping and Directions	0.00	0.00	**0.69**	0.00	0.00	0.00	0.00	0.00
Visual Matrix	0.00	0.00	**0.69**	0.00	0.00	0.00	0.00	0.00
WM-Phonological-Loop								
Digit Forward Span	0.00	0.00	0.00	**0.66**	0.00	0.00	0.00	0.00
Word Problem Span	0.00	0.00	0.00	**0.64**	0.00	0.00	0.00	0.00
Phonetic Memory Span	0.00	0.00	0.00	**0.33**	0.00	0.00	0.00	0.00
Naming Speed								
Rapid Digit Naming	0.00	0.00	0.00	0.00	**0.88**	0.00	0.00	0.00
Rapid Letter Naming	0.00	0.00	0.00	0.00	**0.70**	0.00	0.00	0.00
Problem Solv. Components								
Question	0.00	0.00	0.00	0.00	0.00	**0.56**	0.00	0.00
Number	0.00	0.00	0.00	0.00	0.00	**0.47**	0.00	0.00
Goal	0.00	0.00	0.00	0.00	0.00	**0.66**	0.00	0.00
Operation	0.00	0.00	0.00	0.00	0.00	**0.42**	0.00	0.00
Algorithm	0.00	0.00	0.00	0.00	0.00	**0.48**	0.00	0.00
Irrelevant	0.00	0.00	0.00	0.00	0.00	**0.35**	0.00	0.00
Line Estimation								
Estimation Subtest 1	0.00	0.00	0.00	0.00	0.00	0.00	**0.78**	0.00
Estimation Subtest 2	0.00	0.00	0.00	0.00	0.00	0.00	**0.85**	0.00
Number Judgment								
Magnitude Subtest 1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.95**
Magnitude Subtest 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.88**

Note. Factor loadings representing the factor are in boldface. F1 = Inhibition, F2 = Central Executive, F3 = (WM) Visual Spatial F4 = (WM) Phonological Loop, F5 = Naming Speed, F6 = Word Problem Components, F7 = Line Estimation, F8 = Magnitude Judgment.

Appendix D. Correlation matrix across the classification and external validation measures

	1	2	3	4	5	6	7	8	9	10
1. TORC	1.00	0.47	0.39	0.36	0.39	0.33	0.20	0.22	0.14	0.11
2. WRAT-Reading	0.47	1.00	0.34	0.47	0.44	0.32	0.15	0.15	0.13	0.21
3. WIAT	0.39	0.34	1.00	0.67	0.39	0.39	0.18	0.16	0.09	0.13
4. WRAT	0.36	0.47	0.67	1.00	0.44	0.37	0.18	0.20	0.15	0.15
5. TOMA	0.39	0.44	0.39	0.44	1.00	0.51	0.20	0.16	0.08	0.11
6. CMAT	0.33	0.32	0.39	0.37	0.51	1.00	0.29	0.05	0.07	0.22
7. RAVEN	0.20	0.15	0.18	0.18	0.20	0.29	1.00	0.10	0.03	0.01
8. RNG	0.22	0.15	0.16	0.20	0.16	0.05	0.10	1.00	0.34	0.14
9. RLG	0.14	0.13	0.09	0.15	0.08	0.07	0.03	0.34	1.00	0.08
10. SENTENCE	0.11	0.21	0.13	0.15	0.11	0.22	0.01	0.14	0.08	1.00
11. AUDITORY	0.23	0.20	0.14	0.12	0.24	0.23	0.06	0.12	-0.03	0.14
12. UPDATE	0.22	0.28	0.22	0.22	0.26	0.25	0.12	0.09	0.01	0.17
13. DIRECTIONS	0.09	0.07	0.14	0.13	0.23	0.21	0.16	0.06	0.02	0.07
14. MATRIX	0.02	-0.04	0.06	0.07	0.16	0.15	0.17	-0.01	-0.03	-0.01
15. DIGIT	0.22	0.26	0.11	0.15	0.16	0.28	0.13	0.05	0.05	0.14
16. WORDSPAN	0.17	0.13	0.07	0.10	0.15	0.28	0.11	0.06	0.01	0.15
17. PHONETIC	0.07	0.16	0.11	0.05	0.14	0.12	0.08	0.10	0.06	0.11
18. RDNTIME	-0.27	-0.23	-0.24	-0.18	-0.19	-0.08	0.16	-0.24	-0.18	-0.09
19. RLNTIME	-0.27	-0.23	-0.27	-0.17	-0.18	-0.13	0.08	-0.19	-0.15	-0.08
20. QUESTION	0.32	0.23	0.25	0.23	0.27	0.29	0.11	0.09	0.04	0.11
21. NUMBER	0.23	0.14	0.12	0.14	0.21	0.26	0.06	0.06	0.05	0.00
22. GOAL	0.33	0.23	0.31	0.30	0.34	0.36	0.24	0.16	0.12	0.13
23. OPERATION	0.33	0.24	0.20	0.16	0.22	0.25	0.14	0.15	0.14	0.18
24. ALGORITHM	0.25	0.19	0.22	0.23	0.21	0.26	0.07	0.24	0.15	0.10
25. RRELEVANT	0.13	0.16	0.09	0.14	0.21	0.23	0.04	0.10	0.01	0.19
26. ESTIMATION1	-0.27	-0.26	-0.20	-0.21	-0.29	-0.33	-0.17	-0.16	-0.15	-0.10
27. ESTIMATION2	-0.31	-0.21	-0.24	-0.20	-0.35	-0.40	-0.20	-0.16	-0.17	-0.07
28. NUMERACY1	0.13	0.18	0.32	0.24	0.26	0.42	0.15	0.11	0.17	0.12
29. NUMERACY2	0.13	0.13	0.27	0.17	0.18	0.35	0.09	0.11	0.15	0.13

TORC = Test of Reading Comprehension. WRAT-Reading = Wide Range Achievement Test Reading Task. WIAT = Wechsler Individual Achievement Test. WRAT = Wide Range Achievement Test. TOMA = Test of Mathematical Abilities. CMAT = Comprehensive Mathematical Abilities Test. Raven = Colored Progressive Matrices Test. RNG = Random Number Generation. RLG = Random Letter Generation. Sentence = Listening Sentence Span. Auditory = Auditory Digit Sequence. Directions = Mapping and Directions. Matrix = Visual Matrix. Digit = Digit Forward Span. Wordspan = Word Problem Span. Phonetic = Phonetic Memory Span. RDNTIME = Rapid Digit Naming. RLNTIME = Rapid Letter Naming. Question, Number, Goal, Operation, Algorithm, and Irrelevant = Word Problem Solving Memory Booklet. Estimation1 = Estimation Subtest 1. Estimation2 = Estimation Subtest 2. NUMERACY1 = Magnitude Subtest 1. NUMERACY2 = Magnitude Subtest 2.

Appendix D continued. Correlation matrix across the classification and external validation measures

	11	12	13	14	15	16	17	18	19	20
1. TORC	0.23	0.22	0.09	0.02	0.22	0.17	0.07	-0.27	-0.27	0.32
2. WRAT-Reading	0.20	0.28	0.07	-0.04	0.26	0.13	0.16	-0.23	-0.23	0.23
3. WIAT	0.14	0.22	0.14	0.06	0.11	0.07	0.11	-0.24	-0.27	0.25
4. WRAT	0.12	0.22	0.13	0.07	0.15	0.10	0.05	-0.18	-0.17	0.23
5. TOMA	0.24	0.26	0.23	0.16	0.16	0.15	0.14	-0.19	-0.18	0.27
6. CMAT	0.23	0.25	0.21	0.15	0.28	0.28	0.12	-0.08	-0.13	0.29
7. RAVEN	0.06	0.12	0.16	0.17	0.13	0.11	0.08	0.16	0.08	0.11
8. RNG	0.12	0.09	0.06	-0.01	0.05	0.06	0.10	-0.24	-0.19	0.09
9. RLG	-0.03	0.01	0.02	-0.03	0.05	0.01	0.06	-0.18	-0.15	0.04
10. SENTENCE	0.14	0.17	0.07	-0.01	0.14	0.15	0.11	-0.09	-0.08	0.11
11. AUDITORY	1.00	0.24	0.13	0.14	0.14	0.20	0.12	-0.20	-0.21	0.10
12. UPDATE	0.24	1.00	0.07	0.09	0.36	0.31	0.22	-0.07	-0.06	0.15
13. DIRECTIONS	0.13	0.07	1.00	0.49	0.09	0.09	0.17	-0.06	-0.04	0.09
14. MATRIX	0.14	0.09	0.49	1.00	0.00	0.08	0.10	-0.10	-0.06	0.00
15. DIGIT	0.14	0.36	0.09	0.00	1.00	0.46	0.23	0.08	0.06	0.15
16. WORDSPAN	0.20	0.31	0.09	0.08	0.46	1.00	0.17	0.02	0.01	0.12
17. PHONETIC	0.12	0.22	0.17	0.10	0.23	0.17	1.00	-0.02	-0.02	0.11
18. RDNTIME	-0.20	-0.07	-0.06	-0.10	0.08	0.02	-0.02	1.00	0.77	-0.15
19. RLNTIME	-0.21	-0.06	-0.04	-0.06	0.06	0.01	-0.02	0.77	1.00	-0.14
20. QUESTION	0.10	0.15	0.09	0.00	0.15	0.12	0.11	-0.15	-0.14	1.00
21. NUMBER	0.09	0.04	0.05	-0.02	0.07	0.06	0.10	-0.09	-0.16	0.32
22. GOAL	0.21	0.16	0.13	0.09	0.10	0.16	0.10	-0.17	-0.17	0.42
23. OPERATION	0.10	0.14	0.11	0.03	0.11	0.15	0.12	-0.05	-0.03	0.28
24. ALGORITHM	0.11	0.17	0.08	0.06	0.07	0.20	0.15	-0.08	-0.09	0.23
25. RRELEVANT	0.10	0.11	0.09	0.10	0.13	0.14	0.06	0.00	-0.02	0.28
26. ESTIMATION1	-0.23	-0.24	-0.20	-0.17	-0.18	-0.16	-0.11	0.08	0.09	-0.19
27. ESTIMATION2	-0.22	-0.20	-0.22	-0.18	-0.21	-0.18	-0.17	0.05	0.09	-0.14
28. NUMERACY1	0.11	0.17	0.11	0.12	-0.06	0.06	0.08	-0.17	-0.19	0.20
29. NUMERACY2	0.14	0.15	0.08	0.09	-0.02	0.07	0.11	-0.17	-0.18	0.18

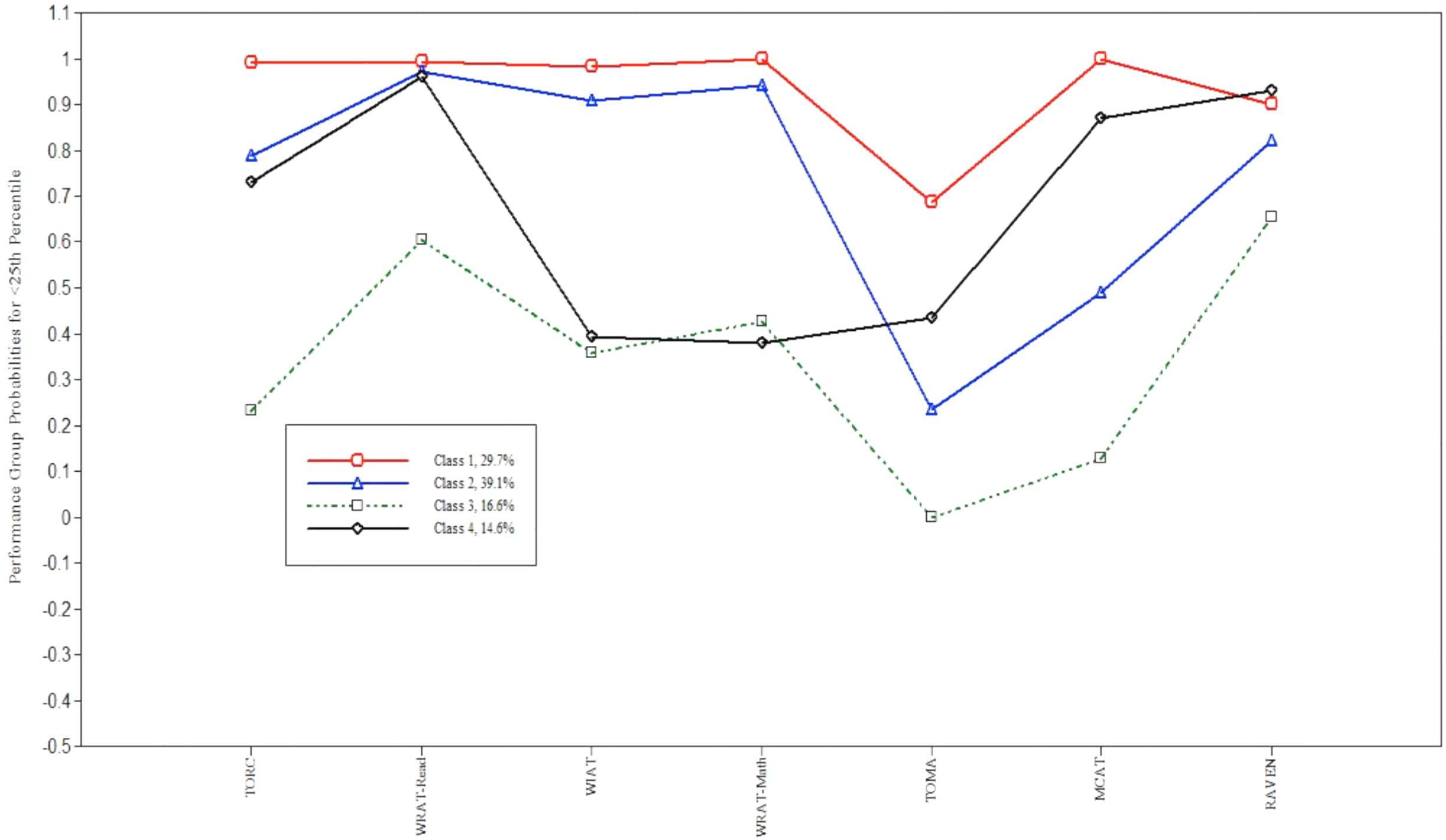
TORC = Test of Reading Comprehension. WRAT-Reading = Wide Range Achievement Test Reading Task. WIAT = Wechsler Individual Achievement Test. WRAT = Wide Range Achievement Test. TOMA = Test of Mathematical Abilities. CMAT = Comprehensive Mathematical Abilities Test. Raven = Colored Progressive Matrices Test. RNG = Random Number Generation. RLG = Random Letter Generation. Sentence = Listening Sentence Span. Auditory = Auditory Digit Sequence. Directions = Mapping and Directions. Matrix = Visual Matrix. Digit = Digit Forward Span. Wordspan = Word Problem Span. Phonetic = Phonetic Memory Span. RDNTIME = Rapid Digit Naming. RLNTIME = Rapid Letter Naming. Question, Number, Goal, Operation, Algorithm, and Irrelevant = Word Problem Solving Memory Booklet. Estimation1 = Estimation Subtest 1. Estimation2 = Estimation Subtest 2. NUMERACY1 = Magnitude Subtest 1. NUMERACY2 = Magnitude Subtest 2.

Appendix D continued. Correlation matrix across the classification and external validation measures

	21	22	23	24	25	26	27	28	29
1. TORC	0.23	0.33	0.33	0.25	0.13	-0.27	-0.31	0.13	0.13
2. WRAT-Reading	0.14	0.23	0.24	0.19	0.16	-0.26	-0.21	0.18	0.13
3. WIAT	0.12	0.31	0.20	0.22	0.09	-0.20	-0.24	0.32	0.27
4. WRAT	0.14	0.30	0.16	0.23	0.14	-0.21	-0.20	0.24	0.17
5. TOMA	0.21	0.34	0.22	0.21	0.21	-0.29	-0.35	0.26	0.18
6. CMAT	0.26	0.36	0.25	0.26	0.23	-0.33	-0.40	0.42	0.35
7. RAVEN	0.06	0.24	0.14	0.07	0.04	-0.17	-0.20	0.15	0.09
8. RNG	0.06	0.16	0.15	0.24	0.10	-0.16	-0.16	0.11	0.11
9. RLG	0.05	0.12	0.14	0.15	0.01	-0.15	-0.17	0.17	0.15
10. SENTENCE	0.00	0.13	0.18	0.10	0.19	-0.10	-0.07	0.12	0.13
11. AUDITORY	0.09	0.21	0.10	0.11	0.10	-0.23	-0.22	0.11	0.14
12. UPDATE	0.04	0.16	0.14	0.17	0.11	-0.24	-0.20	0.17	0.15
13. DIRECTIONS	0.05	0.13	0.11	0.08	0.09	-0.20	-0.22	0.11	0.08
14. MATRIX	-0.02	0.09	0.03	0.06	0.10	-0.17	-0.18	0.12	0.09
15. DIGIT	0.07	0.10	0.11	0.07	0.13	-0.18	-0.21	-0.06	-0.02
16. WORDSPAN	0.06	0.16	0.15	0.20	0.14	-0.16	-0.18	0.06	0.07
17. PHONETIC	0.10	0.10	0.12	0.15	0.06	-0.11	-0.17	0.08	0.11
18. RDNTIME	-0.09	-0.17	-0.05	-0.08	0.00	0.08	0.05	-0.17	-0.17
19. RLNTIME	-0.16	-0.17	-0.03	-0.09	-0.02	0.09	0.09	-0.19	-0.18
20. QUESTION	0.32	0.42	0.28	0.23	0.28	-0.19	-0.14	0.20	0.18
21. NUMBER	1.00	0.36	0.30	0.34	0.21	-0.08	-0.09	0.12	0.11
22. GOAL	0.36	1.00	0.26	0.33	0.26	-0.19	-0.24	0.24	0.26
23. OPERATION	0.30	0.26	1.00	0.60	0.11	-0.17	-0.16	0.18	0.16
24. ALGORITHM	0.34	0.33	0.60	1.00	0.15	-0.19	-0.18	0.21	0.21
25. RRELEVANT	0.21	0.26	0.11	0.15	1.00	-0.14	-0.09	0.13	0.12
26. ESTIMATION1	-0.08	-0.19	-0.17	-0.19	-0.14	1.00	0.71	-0.29	-0.26
27. ESTIMATION2	-0.09	-0.24	-0.16	-0.18	-0.09	0.71	1.00	-0.33	-0.31
28. NUMERACY1	0.12	0.24	0.18	0.21	0.13	-0.29	-0.33	1.00	0.91
29. NUMERACY2	0.11	0.26	0.16	0.21	0.12	-0.26	-0.31	0.91	1.00

TORC = Test of Reading Comprehension. WRAT-Reading = Wide Range Achievement Test Reading Task. WIAT = Wechsler Individual Achievement Test. WRAT = Wide Range Achievement Test. TOMA = Test of Mathematical Abilities. CMAT = Comprehensive Mathematical Abilities Test. Raven = Colored Progressive Matrices Test. RNG = Random Number Generation. RLG = Random Letter Generation. Sentence = Listening Sentence Span. Auditory = Auditory Digit Sequence. Directions = Mapping and Directions. Matrix = Visual Matrix. Digit = Digit Forward Span. Wordspan = Word Problem Span. Phonetic = Phonetic Memory Span. RDNTIME = Rapid Digit Naming. RLNTIME = Rapid Letter Naming. Question, Number, Goal, Operation, Algorithm, and Irrelevant = Word Problem Solving Memory Booklet. Estimation1 = Estimation Subtest 1. Estimation2 = Estimation Subtest 2. NUMERACY1 = Magnitude Subtest 1. NUMERACY2 = Magnitude Subtest 2.

Appendix E. The item profile plot shows the performance between classes on the seven core items that represent mathematic achievement ability. Class 1 = Typical Achievers, Class 2 = Typical Achievers--risk for word problems, Class 3 = Low Achievers, Class 4 = Math Difficulty



Appendix F. A latent class analysis model with more than three classes that is predicted by a covariate