THE IMPORTANCE OF REPRESENTATIONAL SHIFT: AN INVESTIGATION OF THE COGNITIVE MECHANISMS AND INDIVIDUAL DIFFERENCES UNDERLYING MATH PERFORMANCE

by

Allison S. Liu

B.A., University of Chicago, 2011

M.S., University of Pittsburgh, 2014

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Allison S. Liu

It was defended on

August 13, 2018

and approved by

Daniel Ansari, Professor, Department of Psychology

Julie A. Fiez, Professor, Department of Psychology

Melissa E. Libertus, Assistant Professor, Department of Psychology

Dissertation Advisor: Christian D. Schunn, Professor, Department of Psychology

THE IMPORTANCE OF REPRESENTATIONAL SHIFT: AN INVESTIGATION OF THE COGNITIVE MECHANISMS AND INDIVIDUAL DIFFERENCES UNDERLYING MATH PERFORMANCE

Allison S. Liu, Ph.D.

University of Pittsburgh, 2018

College dropout is a significant issue that particularly plagues students with poor math preparation. Many studies have attempted to understand the factors that contribute to math performance and create training interventions that can effectively improve math achievement. However, few studies have investigated these questions within populations similar to the lower achieving adults who are most at risk to drop out from college. Further, few have looked at many potential mechanisms at once to determine how each contributes to math performance. In this dissertation, we investigated the relationships between a number of cognitive and individual difference measures and different types of math performance. We also evaluated the effectiveness of an estimation-based training program that targeted one of these cognitive mechanisms to determine the factors that predict progress during the intervention and the mechanisms that explain math improvements. Importantly, we recruited participants with relatively low math skill level to better examine the underlying math foundations in adults who are most likely to struggle with math in college. Across both studies, we find evidence of a representational shift that supports higher math performance in procedural and complex math; specifically, in higher math-skilled individuals, procedural math relies more on mechanisms that involve non-symbolic number representations, while complex math draws upon mechanisms that involve primarily symbolic number representations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

As postsecondary education becomes increasingly more important for future opportunities and success, the high rate of dropout from colleges remains a significant problem. A recent report of six-year completion outcomes of college students in the United States stated that only 57% of students starting college were able to complete college within six years (Shapiro et al., 2017). The college completion rate is even lower for underrepresented minority students; Black students in particular were more likely to drop out than complete college, with a completion rate of only 39%. For comparison, White students and Asian students showed much higher completion rates of 66% and 69%, respectively.

The rate was particularly low for students who enroll in community college, with fewer than 30% of students obtaining an Associate's degree within three years. Many of the students who continued on to four-year colleges also failed to receive a Bachelor's degree (Symonds, Schwartz, & Ferguson, 2011). In an investigation of California community colleges, Moore and Shulock (2010) found that 70% of students had yet to attain a certificate or degree within six years, and rates were again worse for minority students, rising to 75% for Black students and 80% for Latinx students. Only 15% of the students who had not finished were still enrolled in college; the rest had dropped out.

Poor math preparation is a central contributor to these high college dropout rates (Balfanz & Legters, 2004; Bynner & Parsons, 1997). More broadly, colleges in the United States estimate that approximately 60% of freshmen are unprepared for college level work (Grubb et al., 2011). This lack of preparation is particularly prevalent in math (Attewell, Lavin, Domina, & Levey, 2006), despite the high importance attached to math competency within the K-12 system. Again, minority students show particularly low levels of math mastery: the National Assessment of Educational Progress (2015) reported that only 36% Black, 47% Hispanic, and 46% American Indian/Alaska Native 12th grade students reached a basic level of math mastery. Students entering college without the requisite math knowledge are set up to fail during college math courses, which is reflected in the high rate of failure in basic math courses, such as college algebra (Saxe & Braddy, 2015).

Many universities have raised the minimum math skill requirement for admission to alleviate the issue of poor math preparation (Symonds et al., 2011), which has led to a greater number of students enrolling in remedial math courses to prepare for college-level math courses. However, the percentage of students who complete these remedial courses is very low (Bailey, Jeong, & Cho, 2010). For example, the City University of New York stated that 76% of their students needed remedial math, but the pass rate for the highest level remedial course was only 38% (CUNY Office of Institutional Research, 2015a, 2015b). Many students also put off completing their remedial math courses (Bailey et al., 2010). Thus, remedial math courses, meant to help students to reach math requirements necessary to graduate, instead create another barrier that further prevents graduation for struggling students (Attewell et al., 2006).

These adults struggling in math are often equally capable of doing complex tasks as the adults who succeed in their math classes but may have poorly developed foundational resources necessary to do math well. The problem is identifying the cognitive and individual difference resources that relate to math performance. In particular, it is important to determine whether the resources used for math performance depend on one’s initial math skill level and the complexity of the math being performed, so that we know if different resources need to be taught and emphasized depending on how much math an individual knows and what kind of math they are trying to learn.

Once these resources are identified, it is also necessary to examine the effectiveness of training interventions that target these resources to improve general math abilities—in other words, the goal is to create interventions that transfer performance improvements to tasks that were not directly practiced. However, there is an increasing concern that the impressive prior demonstrations of learning transfer from training interventions are actually confounded by people’s overall skill level or are illusions stemming from problems in their experimental designs (Sala & Gobet, 2017). That is, for some studies, they may have simply documented that smarter individuals are more likely to excel at both training interventions and measured outcomes; more generally, interventions found to be successful for typically high performing students may not generalize successfully to lower performing students. In addition, many observed training effects may actually be placebo effects, as training effect sizes generally become much smaller or even null when active control conditions are included. Thus, important questions are whether progress can be made with training interventions by all learners, or whether barriers prevent some people from fully benefitting from training. We must also determine which changes in cognitive or individual difference mechanisms explain improvements in math, and whether this again depends upon the complexity of the math being done.

Prior studies have investigated the foundational resources that relate to math performance (reviewed in depth in Chapter 2), and the effects of various training interventions on math performance (reviewed in depth in Chapter 3). Importantly, these studies frequently study populations that are very different from those populations struggling in college—the very large problem described in the beginning of this chapter. The prior studies have typically involved young children, whose math skills and resources are likely unstable due to ongoing development, or undergraduate students from high-ranking colleges and universities who likely already have high math skills. Further, prior studies are limited in the number of foundational resources that they investigate simultaneously, usually focusing on one or two mechanisms at a time. While this makes it easier to examine the effects of these individual factors on math performance in depth, these studies cannot determine how these mechanisms relate to each other and which more directly contribute to math performance when the other mechanisms are taken into account. This is the next step needed to fully understand the foundations that critically support math.

This dissertation presents two (linked) studies that together aim to answer these questions and address these limitations. In Study 1 (Chapter 2), we investigated how various cognitive and individual difference measures related to basic procedural math performance and to complex applied math problems. We specifically recruited adults of varying math skill level to determine whether the relationship between these factors, as well as the unique predictors of procedural and complex math, depended upon overall math skill level. We found significant differences in the kinds of resources that mid-skill and low-skill adults utilized during math performance, particularly for more complex math.

In Study 2 (Chapter 3), we investigated changes in procedural and complex math after multiple days of training with an estimation task that targeted the refinement of number representations. We examined training effects in the same mechanisms included in Study 1, as well as the factors that predicted progress on the training task and improvements in procedural and complex math. Here, we found that individual differences may determine which learners make progress during the training task, and that different mechanisms may explain improvement in simple procedural math versus more complex applied math.

Each of the empirical chapters discusses the theoretical and practical implications of each specific study. In Chapter 4, we look across the findings of Study 1 and Study 2 and discuss what the combined findings mean for math performance in adults. We also provide recommendations for future studies based on what was found in these two studies.

2.0 STUDY 1

Mathematical competence is critical for academic and lifelong success. Poor math preparation has been linked to lower performance in later math courses (Balfanz, McPartland, & Shaw, 2002; Pelavin & Kane, 1990), greater rates of school dropout (Balfanz & Legters, 2004; Bynner & Parsons, 1997), and disadvantages with future employment (Bynner & Parsons, 1997). Given its importance, many instructional methods and training programs have aimed to improve mathematical competence with varying results. To determine effective methods to improve math skills and performance, one must understand which cognitive resources learners draw on when performing math, whether these resources vary depending on the type of math being done and learners' initial math skill, and how these resources relate to each other and various individual difference factors. Many studies have investigated potential mechanisms that underlie math performance, including representational acuity and symbolic integration, as well as individual differences such as working memory and math anxiety, and have found varying relationships between each of these factors and math achievement. Each of these mechanisms and how they relate to math performance is discussed in detail below.

2.1 POTENTIAL MECHANISMS UNDERLYING MATH PERFORMANCE

2.1.1 Representational Acuity

"Number sense," defined as an elementary intuition about quantity (Dehaene, 1997), is thought to rely on a cognitive subsystem in the brain called the Approximate Number System (ANS) that allows us to represent and process quantities without the use of symbols or language. Normally, these non-symbolic quantity representations are inexact; neurons that respond optimally to a specific quantity will also respond to a lesser degree to slightly smaller or slightly larger quantities (Ansari, 2008; Dehaene, 1997). Consequently, quantities with a high degree of representational overlap (typically quantities at close ratios, e.g., 30 and 33) are harder to discriminate than quantities with little representational overlap (typically quantities at distant ratios, e.g., 30 and 45) (Dehaene, 1992; Gallistel & Gelman, 2000). When children learn symbols that represent quantities (e.g., Arabic numerals), they also acquire a symbolic number system; these symbolic number representations are generally more precise than non-symbolic number representations, though there is still individual variation in people's symbolic and non-symbolic representational acuity. Having low **representational acuity** (i.e., a high degree of representational overlap) can make it difficult to discern quantities during math tasks, leading to poor math performance. In contrast, having more acute representations can strengthen the mapping between numerical symbols and their non-symbolic representations (Brankaer, Ghesquière, & De Smedt, 2014; Pinheiro-Chagas et al., 2014) or facilitate online error detection, allowing people to better determine whether their calculated responses are reasonable (Lourenco, Bonny, Fernandez, & Rao, 2012).

Many studies have shown a connection between representational acuity and symbolic mathematics performance in children and adults (e.g., Dehaene, 1992; Fazio, Bailey, Thompson, & Siegler, 2014; Feigenson, Dehaene, & Spelke, 2004; Gallistel & Gelman, 1992; Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Halberda, Mazzocco, & Feigenson, 2008; Holloway & Ansari, 2009; Libertus, Feigenson, & Halberda, 2013; Lourenco et al., 2012; Piazza, Izard, Pinel, Bihan, & Dehaene, 2004; Schneider et al., 2016). ANS acuity has also been shown to predict performance on mathematics achievement tests for pre-school and kindergarten students up to two years later (Gilmore, McCarthy, & Spelke, 2010; Mazzocco, Feigenson, & Halberda, 2011). Converging evidence is provided by neurological studies. The horizontal segment of the intraparietal sulcus, where quantity representations appear to be localized, is also activated during both symbolic and non-symbolic number manipulation, with greater activation when tasks require more quantity processing (Dehaene, Piazza, Pinel, & Cohen, 2003).

Some studies have found that the relationship between representational acuity and math achievement may depend upon the complexity of the math task being investigated. For example, a study involving primary school children and undergraduates found that higher symbolic acuity was associated with better performance on several tasks that required more complex mathematical reasoning, but not with tasks involving basic mathematical fluency (Pina, Castillo, Kadosh, & Fuentes, 2015). However, another study with elementary school children found the opposite, in that symbolic and non-symbolic acuity were correlated with arithmetic computation, but not mathematical reasoning (Zhang, Chen, Liu, Cui, & Zhou, 2016). Meanwhile, non-symbolic acuity has been found to correlate with both procedural math in elementary school students and adults and applied math problems in elementary school children (Lourenco & Bonny, 2017; Lourenco et al., 2012).

Additionally, several findings have suggested that a shift may occur in which non-symbolic acuity becomes less important than symbolic acuity with age and math experience. In a meta-analysis of 284 effect sizes, Schneider and colleagues (2016) found that both non-symbolic and symbolic acuity were associated with mathematical competence, but the relationship was stronger for symbolic than non-symbolic measures. Similarly, Fazio, Bailey, Thompson, and Siegler (2014) found that both types of acuity independently predicted math achievement in elementary school children, but symbolic acuity explained almost four times as much variance as non-symbolic acuity; further, the unique contributions of non-symbolic acuity weakened after six years of age. Elementary school children have also shown correlations between non-symbolic acuity and calculation math scores, while adults showed no such correlations with either procedural or complex math tasks (Inglis, Attridge, Batchelor, & Gilmore, 2011).

More direct evidence about the causal connections between representational acuity and mathematics performance comes from training studies that have successfully trained representational acuity, leading to significant math improvements, though these studies have primarily involved children or high-math-achieving college students (Fischer, Moeller, Bientzle, Cress, & Nuerk, 2011; Liu, Kallai, Schunn, & Fiez, 2015; Obersteiner, Reiss, & Ufer, 2013; Park & Brannon, 2013, 2014; Ramani & Siegler, 2011; Whyte & Bull, 2008). For example, Park and Brannon (2013) trained high-achieving adults to perform approximate addition and subtraction on dot arrays, teaching them to perform exact calculations on approximate representations. After training, participants were able to solve a greater number of symbolic addition and subtraction problems in three minutes than participants who did not complete the training, suggesting that improvements in the acuity of adults' ANS representations may transfer to mathematical performance. Thus, ANS representational acuity is a highly plausible foundation for math performance.

However, other studies have found a more tenuous link between representational acuity and mathematics achievement (see De Smedt, Noël, Gilmore, & Ansari, 2013 for a review). For example, many kindergarten students with high math achievement also showed high representational acuity, but some students with low representational acuity still showed high math achievement (Hart et al., 2016). These results may suggest that representational acuity is not always necessary for school-relevant mathematics, meaning that other mechanisms may be more important for math performance.

2.1.2 Symbolic Integration

It is commonly assumed that the numerical symbols typically seen in math (e.g., Arabic numerals) gain meaning when symbolic number representations are mapped onto their corresponding non-symbolic representations. The robustness of this mapping is known as **symbolic integration**. Separately from representational acuity, the strength of symbolic integration has been associated with math achievement in children and adults (Wilson, Dehaene, Dubois, & Fayol, 2009) and is thought to support math problem solving by providing a more intuitive understanding of magnitudes, leading to faster and more accurate processing (Geary, 2013; Holloway & Ansari, 2009). Several studies have also found that symbolic integration fully mediates the relationship between non-symbolic representational acuity and symbolic arithmetic (Jang & Cho, 2018; Wong, Ho, & Tang, 2016). This relationship may not be limited to symbolic integration strength; another study found that integration precision (i.e., the variability of people's number estimates) rather than integration accuracy significantly predicted math performance even when controlling for age, vocabulary, and non-symbolic acuity, as well as a similar fully mediated relationship between non-symbolic acuity and math performance (Libertus, Odic, Feigenson, & Halberda, 2016).

Conversely, weak symbolic integration may lead to math problem solving that is entirely dependent on verbal fact retrieval and procedural strategies, with little meaningful support from ANS representations (Lyons, Ansari, & Beilock, 2012). In a review on symbolic integration, Leibovich and Ansari (2016) concluded that current results fail to find strong levels of symbolic integration in which number symbols are grounded in non-symbolic representations in either children or adults. Reynvoet and Sasanguie's (2016) review reached a similar conclusion that evidence for a strong mapping between the ANS and symbolic representations was questionable. Most importantly, however, relatively few studies have looked in detail at symbolic integration, and even fewer have looked at integration's different aspects such as integration accuracy, precision, and strength in a single study. Instead, most studies have investigated only integration strength through number comparison tasks (in which participants are presented with sequences of symbolic and non-symbolic stimuli, and performance is compared between the two formats), or investigated integration accuracy and precision through quantity estimation tasks (in which they are shown non-symbolic stimuli and asked to estimate the quantity shown). Further, the studies utilizing number comparison tasks assume non-symbolic veridicality: that the ANS representations being accessed precisely match the non-symbolic quantities being shown without substantial bias. However, many studies have shown that people consistently underestimate larger non-symbolic quantities and overestimate smaller ones (Minturn & Reese, 1951) with estimates fitting a power function (e.g., Indow & Ida, 1977; Izard & Dehaene, 2008; Krueger, 1982, 1984), suggesting that such bias needs to be taken into account when measuring integration. Therefore, symbolic integration's connections with math problem solving is a currently under-explored possible factor for supporting math performance.

2.1.3 Working Memory Capacity

Working memory capacity differences have been broadly implicated in complex task performance (e.g., Bull, Espy, & Wiebe, 2008; McClelland, Acock, & Morrison, 2006; St Clair-Thompson & Gathercole, 2006), including mathematics (Clark, Pritchard, & Woodward, 2010). When people are trained to memorize and retrieve basic math facts to reduce the amount of explicit computation needed during math problem solving, brain regions involved during basic mathematical calculations shift away from fronto-parietal networks that are involved in attentional processing to the left angular gyrus, a region that has been associated with retrieval processes (Delazer et al., 2003; Ischebeck, Zamarian, Egger, Schocke, & Delazer, 2007). This lowered working memory load can then allow attentional resources to be allocated toward more complex non-retrieval math strategies (e.g., Ayres, 2001; Campbell & Charness, 1990; Imbo, Duverne, & Lemaire, 2007). Thus, one's working memory capacity could influence math performance, beyond the cognitive resources of representational acuity or symbolic integration, by influencing whether strategies supported by these cognitive resources can be fully utilized. Further, the previously observed correlations between math achievement and measures related to ANS precision or connectivity strength may be partially explained by working memory processes, such as working memory load; for example, the representational strength of numbers could influence functional working memory size.

Working memory is separated into verbal working memory, which stores phonological information, and visuospatial working memory, which stores visual and spatial information

(Baddeley & Hitch, 1974), and there is debate over which type of working memory is more important for math performance. Some studies have shown a stronger relationship between verbal working memory and math performance (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Friso-van den Bos, Van der Ven, Kroesbergen, & van Luit, 2013), and others have found that visuospatial working memory contributes equally or even more to math performance (Andersson & Ostergren, 2012; Cragg, Keeble, Richardson, Roome, & Gilmore, 2017; McLean & Hitch, 1999; Miller & Bichsel, 2004; Schuchardt, Maehler, & Hasselhorn, 2008; Szucs, Devine, Soltesz, Nobes, & Gabriel, 2014). The recruitment of verbal versus visuospatial working memory during math may depend upon the format of presented math problems. For example, verbally presented problems are likely processed primarily in verbal working memory, while visually presented problems may be processed in either visual or verbal working memory depending on whether people translate the presented information into a phonological code (Imbo, Vandierendonck, & Rammelaere, 2007; Logie, Gilhooly, & Wynn, 1994; Noel, Desert, Aubrun, & Seron, 2001). Further, one’s age and existing math knowledge may also affect whether one primarily relies upon verbal or visuospatial working memory. Huttenlocher, Jordan, and Levine (1994) have suggested that young children initially solve math problems using mental models and then rely more strongly on verbal codes as they develop their language and verbal working memory. Consistent with this, visuospatial working memory appears to be the best predictor of math performance in preschoolers, while verbal working memory becomes a better predictor of math performance by first grade (Rasmussen & Bisanz, 2005). Additionally, Holmes and Adams (2006) found that visuospatial working memory predicted performance on both easy and difficult math problems in younger children; meanwhile in older children, verbal working memory predicted performance on easy problems, while visuospatial working memory predicted performance on difficult problems. In other words, visuospatial working memory may be most important at younger ages and for unfamiliar math skills, while verbal working memory may be recruited for older students and when one is already proficient in that math skill.

Because the current studies involved adults who likely had at least some familiarity in the math skills required for most tasks, we chose to use a verbal working memory task. Specifically, we use the forward and backward digit span tasks, which are common measures of working memory in which participants are asked to repeat a sequence of numbers verbatim (forward task) or in reverse (backward task). The digit span tasks have been argued to be domain-specific measures of working memory (LeFevre, DeStefano, Coleman, & Shanahan, 2005) because they involve number stimuli, and that digit span tasks may overestimate the role of working memory in math performance (Raghubar, Barnes, & Hecht, 2010). However, non-numerical working memory tasks have still been found to relate to math achievement, suggesting that this relationship is domain-general and not fully explained by the numerical nature of the working memory task (Cragg & Gilmore, 2014; Hubber, Gilmore, & Cragg, 2014).

2.1.4 Math Anxiety

Affective factors have also been considered increasingly important for math performance. Math anxiety, defined as “a feeling of tension, apprehension, or fear that interferes with math performance” (Ashcraft, 2002), has been particularly well studied. Most theories posit that math anxiety is not the primary driver of math performance, but rather works by impeding other resources that are related to math. In the short-term, math anxiety can impair math achievement by disrupting working memory resources. During math tasks, highly math-anxious individuals are thought to focus on worries about their math performance, and these ruminations can co-opt working memory processes needed for higher-level math strategies (Ashcraft, 2002). In the long-term, highly math-anxious individuals tend to avoid math; this prevents them from participating in math experiences and persisting in math courses, leading to lower math achievement (e.g., Hembree, 1990; Tobias, 1978). Further, math anxiety correlates with low self-perceptions of math ability (Hendel, 1980; Meece, Wigfield, & Eccles, 1990), which has been found to predict lower graduation rates for incoming college freshmen (Larson et al., 2015). Studies have found correlations between higher math anxiety and lower performance in both procedural math fluency and more complex applied math (Braham & Libertus, 2018; Hart et al., 2016). It is still not completely clear whether math anxiety can have a unique influence on math performance beyond its impacts on other cognitive resources. In addition, there is evidence that associations between ANS acuity and complex math may depend on one's math anxiety level (Braham & Libertus, 2018). This brings up questions as to what other individual differences are important in the relationship between math anxiety and math performance.

2.2 THE CURRENT STUDY

While many prior studies have looked at a couple of the above mechanisms or individual difference measures in relation to math achievement, fewer studies have looked at all of these factors together and are therefore unable to determine which mechanisms are the primary foundations of math performance. Studies have also typically focused on younger children whose math abilities and foundational resources may still be rapidly developing, providing an unstable look at the resources involved in math performance, or adults with relatively high math achievement. Many students enter college without even basic math mastery, particularly minority students; in 2015, only 36% Black, 47% Hispanic, and 46% American Indian/Alaska Native 12th grade students reached a basic level of math mastery, compared to 73% White and 78% Asian students (U.S. Department of Education, 2015). Further, there is a very high rate of failure in basic math courses in college, with 50% of students receiving a D or F in college algebra (Saxe & Braddy, 2015). Given these issues, it is important to study what underlies success in mathematics in adults who are not already high-achieving in mathematics to aid in their remediation and ensure that they can succeed.

In the current study, we investigated how representational acuity, symbolic integration, working memory, math anxiety, and several basic demographics variables relate to basic procedural math performance and complex applied math problems. We recruited adult participants whose math scores ranged from the 80th percentile to under the 50th percentile on national assessments. Their foundational resources were likely more stable, while the varying skill levels still allowed us to still investigate what distinguishes adults who generally perform well at mathematics versus those who do not. It is possible that those at the low end of our math skill sample have had very different prior math experiences compared to those with higher math skills, and consequently rely on a different set of foundational resources. This is likely to be most evident on tests of complex math, where lower skilled participants may not even have the declarative knowledge needed to solve these problems, making foundational math resources unusable. In addition, our study examined symbolic integration in detail by including several measures of integration: integration accuracy (how correct one's quantity estimates are), integration precision (how consistent one's quantity estimates are), and integration strength (how much one's integration affects performance). Participants completed a battery of math tests, cognitive tasks, and questionnaires. We investigated the relationship between these different cognitive and individual difference factors, as well as which factors uniquely predicted procedural and complex applied math performance, accounting for participants' initial math skill level.

2.3 METHOD

2.3.1 Participants

Eighty-one adults were recruited through advertisements in local colleges and universities in the metropolitan Pittsburgh area and through the University of Pittsburgh's Pitt+Me research participant registry. Participants were required to have quantitative SAT or ACT scores below the 80th percentile rank (620 on the SAT, 25 on the ACT) and were categorized as either Mid or Low math skill based on the score: scores between the 50th and 80th percentile (520-620 on the SAT, 21-25 on the ACT) were classified as Mid skill, and scores below the 50th percentile (below 520 on the SAT, below 21 on the ACT) were classified as Low skill. Participants were also restricted to non-quantitative college majors (e.g., engineering, mathematics) to avoid substantial changes in math skill from instruction after entry exams. In total, 51 Mid skill participants (33 female, 10 male, 8 unreported; age *M* = 24.3, *SD* = 6.05, range = 18-44; verbal SAT *M* = 596, *SD* = 73, 29 unreported; 28 White, 6 Asian, 8 Black/African-American, 1 Black & Pacific Islander, 2 Hispanic/Latinx, 8 unreported; 42 with at least some post-secondary education, 8 unreported) and 30 Low skill participants (17 female, 12 male, 1 non-binary; age *M* = 25.1, *SD* = 5.14, range = 18-35; verbal SAT *M* = 578, *SD* = 81, 17 unreported; 24 White, 6 Black/African-American, 1 Hispanic/Latinx; 27 with at least some post-secondary education) completed the study. Participants were compensated US$10. Two participants were excluded from analyses because they showed low effort on several tasks (e.g., random estimates on the Dot Estimation task).

2.3.2 Materials

Table 1 summarizes each construct being investigated, their associated measure, and the means and standard deviations of the Mid- and Low-Skill groups for each measure. Variance was approximately equal for both skill groups across measures, and no measure showed ceiling effects. An independent samples t-test found that the Mid- and Low-Skill groups only significantly differed on the Applied Problems test, $t(77) = 2.72$, $p = .01$, $d = 0.64$. The measures are described in more detail below.

Table 1. The study's constructs, their associated measures, and Mid-Skill and Low-Skill groups' mean and standard deviations on each measure

Construct	**Complex math**	**Procedural fluency**	**Symbolic integration**			**Symbolic representational acuity**	**Non-symbolic representational acuity**	**Working memory**	**Math anxiety**
Measures	Applied Problems	Math Facts	Dot Estimation (accuracy)	Dot Estimation (precision)	Number Decision, Number Comparison (strength)	Number Comparison (symbolic trials)	Number Comparison (non-symbolic trials)	Digit Span	Abbreviated Mathematics Anxiety Rating Scale
Mid	50.5 (5.3)	.61 (.29)	.75 (.14)	.21 (.05)	.06 (.69)	.02 (.72)	-.06 (.69)	9.6 (1.3)	67.8 (17.5)
Low	47.0 (5.4)	.57 (.25)	.76 (.12)	.21 (.04)	-.06 (.84)	-.04 (.69)	.10 (.69)	9.5 (1.3)	69.1 (17.1)

2.3.2.1 Applied Problems test.

Participants completed the *Applied Problems* sub-test of the Woodcock-Johnson III Tests of Achievement (WJ-III ACH), which measures one's ability to analyze and solve practical quantitative problems. The test consisted of 33 word problems and showed good reliability

(Cronbach's α = .86). Each question was visible to the participant and was also read aloud by an experimenter. Questions began at a difficulty approximately equivalent to middle-school mathematics (problem #30 of the original WJ-III sub-test; e.g., "Jay's car holds 15 gallons of gas, Ana's car holds 10 gallons of gas, and Ellen's car holds 20 gallons of gas. How many more gallons does Jay's car hold than Ana's car?"). They increased in difficulty, ending at approximately college-level mathematics (question #63 of the original sub-test; e.g., "If a chord 8 inches long is 4 inches from the center of a circle, what is the radius of the circle?"). Participants were allowed to use scratch paper to solve the problems and were required to orally answer each question. A participant was tested until they reached the end of the test or incorrectly solved six consecutive items. Participants received one of two forms of the test; the two forms did not differ significantly in accuracy ($t(74) = 0.89$, $p = .38$) and were therefore combined for analyses. Accuracy was recorded, and accuracy scores more than two standard deviations from the group average were winsorized (1 participant).

2.3.2.2 Math Facts test.

To measure procedural fluency with simple addition and subtraction, participants were asked to solve addition and subtraction problems for their exact solutions as quickly and as accurately as possible. The task included four sets of problems, with 16 problems per set, for a total of 64 problems (α = .87). The first two sets of problems consisted of addition and subtraction problems, respectively, made up of one double-digit operand and one single-digit operand. The last two sets of problems consisted of addition and subtraction problems, respectively, made up of two double-digit operands. Both operands in a problem were shown simultaneously and horizontally at the center of the screen for 10 s. Participants were asked to type their responses and press the Enter key on the keyboard to register their response within this 10 s window.

Accuracy and response time was collected for each problem. Due to ceiling effects and the resulting lower reliability of scores in single-digit problems and non-carry problems, only double-digit carry problems were included in analyses. Trials with response times faster than 100 ms were also removed to exclude trials in which participants had not fully processed the stimuli. Otherwise, response time was not used as an outcome measure because the response time data failed some basic sanity checks (i.e., carry problems and double-digit problems being slower than non-carry problems and single-digit problems). The addition and subtraction accuracies were combined to produce a *Math Facts* score, as the correlations between them were relatively high ($r = 0.63$). Scores that were more than two standard deviations away from the average were winsorized (3 participants).

2.3.2.3 Dot Estimation task.

The Dot Estimation task asked participants to give a symbolic number estimate for presented dot quantities. In a direct manner, it was used to measure the accuracy and precision of participants' symbolic integration. Indirectly, it also contributed to the analysis of the Number Decision and Number Comparison tasks (see below) by measuring systematic biases in participants' estimates of non-symbolic quantities. Stimuli consisted of random letter pairs overlaid with black dot arrays (see Figure 1); the large letter pairs are included in this task to fully match the stimuli used in the Number Decision task to determine perceptual bias functions in that task, but they are irrelevant in measuring participants' accuracy and precision.

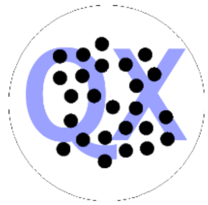

Figure 1. Example stimulus from the Dot Estimation task

The dot arrays varied in magnitude from 7 to 95 dots. To control for perceptual differences in the dot cloud stimuli, six image variations (using three different dot sizes and two different total areas occupied by the dots) were created for each dot quantity using a MATLAB script written by Dehaene, Izard, and Piazza (2005). Participants were shown the stimuli images one at a time and asked to estimate the number of dots shown by entering their estimates on a computer keyboard. The stimuli were only shown for 400 ms to prevent exact counting, but participants had an unlimited amount of time to enter their responses. Participants completed 216 trials, separated into 6 blocks of 36 problems each. Dot estimates were entered into PsiMLE (Odic, Im, Eisinger, Ly, & Halberda, 2015), a program that uses a maximum-likelihood approach to estimate individuals' psychophysical scaling, to estimate participants' individual estimation power curves. Power curves were used instead of linear or log curves based on prior findings showing that quantity estimates generally fit a power function (e.g., Indow & Ida, 1977; Izard & Dehaene, 2008; Krueger, 1982, 1984) and prior analyses with the current task showing a relatively better fit to a power function (Liu, Schunn, Fiez, & Libertus, in prep). The beta (the compression of estimates with increasing quantity) and sigma (the amount of variability in estimates) parameters of these estimation curves were used as measures of participants' *Integration Accuracy* and *Integration Precision*, respectively. The estimation curves were also applied to the Number Decision task and Number Comparison task data to account for estimation biases that participants may have when looking at dot arrays using the alpha and beta parameters: adjusted estimate = α*original dot quantity$^{\beta}$.

2.3.2.4 Number Decision task.

To measure participants' *Integration Strength*, participants viewed a series of images consisting of either a double-digit numeral overlaid with a black dot array or a letter pair overlaid with a black dot array (see Figure 1). They were asked to judge whether the image text showed an Arabic numeral or a letter pair. Participants pressed "S" on a computer keyboard when the text was an Arabic numeral and "L" when the text was a letter pair. Each image was shown for 400 ms to prevent exact counting of the dot arrays, and participants were given 1.5 s to respond. The dot arrays could be one of five quantities (28, 42, 54, 67, and 79), each of which were paired with a range of Arabic numerals. Two versions of the task were used, which were identical except for the paired Arabic numerals (see Table 2); Version B was implemented to balance the number of trial types after adjusting for participants' estimation biases. Because there were no significant accuracy differences between the two task variations, they were combined for analyses. Participants first completed 10 practice trials with trial-by-trial feedback and were required to achieve 80% accuracy or higher before moving on to the rest of the task. There were 150 numeral trials and 150 letter trials, separated into 6 blocks of 50 problems each (α = .91). The dot quantities of the numeral trials were adjusted based on the Dot Estimation data, which affected whether a trial was classified as either a "match" trial (in which the Arabic numeral quantity matched the adjusted dot quantity) or a "mismatch" trial (in which the Arabic numeral quantity was less than or greater than the adjusted dot quantity shown). Only mismatch trials with an Arabic numeral and adjusted dot quantity that were near in ratio (defined as being between 1.15x to 1.35x in ratio) were included in analyses, based on prior findings that integration effects are most evident at these smaller ratios (Liu, Schunn, Fiez, & Libertus, in prep). Participants' judgment accuracies and response times on match and near mismatch trials were recorded. The Number Decision accuracy results replicated prior integration findings with the task (Liu, Schunn, Fiez, & Libertus, 2015), in which match accuracy (M = .95, SD = .12) was significantly greater than near mismatch accuracy after adjustment (M = .90, SD = 11), $F(1, 69)$ = 8.51, p = 01, d = .43, suggesting that symbolic integration was found in the average participant. However, the current study's response time results did not replicate prior response time findings (in which response time is faster for match trials than mismatch trials) and showed little difference between the two trial types. Thus, to maximize the integration strength effect and have enough variation for integration strength to be used as an individual differences measure, level of symbolic integration was calculated as the absolute difference in accuracy between match and near mismatch trials. Participants with a higher level of symbolic integration should have shown a greater difference in accuracy between match and mismatch trials, as they would be processing both the symbolic and non-symbolic stimuli (even though only the symbolic stimuli were meaningful to the task). This difference score was standardized and then averaged with participants' standardized symbolic integration score from the Number Comparison task to calculate a combined integration strength score (described below) for each participant that was used in the individual difference analyses. Participants' average match or near mismatch accuracies more than two standard deviations away from the group average were winsorized before calculating their symbolic integration strength (6 participants for match accuracy, 4 participants for near mismatch accuracy).

Table 2. Dot quantities and range of symbolic numerals paired with each dot quantity in the Number Decision Task variations

Dot Quantity	Range of Paired Symbolic Numerals in Version A	Range of Paired Symbolic Numerals in Version B
28	11 – 76	13-28
42	15 – 81	21-42
54	21 – 86	33-54
67	25 – 91	35-67
79	31 – 96	44-79

2.3.2.5 Number Comparison task.

To provide a second measure of symbolic integration strength and measures of symbolic and non-symbolic representational acuity, a number comparison task was given. Participants were shown two quantities, presented sequentially on a computer screen, and asked to indicate whether the first or second quantity was larger by pressing the “S” or “L” key, respectively. The first quantity was shown for 0.4 s, followed by a fixation cue (“#”) for 0.3 s, followed by the second quantity for 0.7 s, followed by a response period of 1.3 s. The quantities were in either symbolic (Arabic numeral) format or non-symbolic (dot array) format. The two formats were combined to create three different types of trials: symbolic (i.e., two Arabic numerals), non-symbolic (i.e., two dot arrays), or mixed (i.e., one Arabic numeral and one dot array). Participants completed 10 practice symbolic comparison problems with trial-by-trial feedback provided and were required to receive 80% accuracy or higher to move on to the main task. The main task consisted of 50 problems (α = .94). Accuracy and response time for each trial was collected, and only response times on correct trials were used in analyses. Accuracy scores and response times for each trial type greater than two standard deviations from the group average were winsorized (4 participants for symbolic trial accuracy, 4 participants for symbolic trial response time, 3 participants for non-symbolic trial accuracy, 6 participants for non-symbolic trial response time, and 4 participants for mixed trial response time).

Participants' *Integration Strength* was measured by calculating the difference between the response time on correct mixed trials and correct non-symbolic trials (similar to the measure from Lyons, Ansari, & Beilock, 2012), after adjusting for bias in the encoding of non-symbolic quantity with the Dot Estimation task results. The bias adjustment affected whether mixed trials were considered correct or not, as dot quantities that were larger in quantity than their paired Arabic numeral before adjustment could become smaller in quantity after adjustment, and vice versa. If a participant has a higher level of integration, then the addition of the more precise Arabic numeral in the mixed trials should support faster comparisons compared to the non-symbolic trials; in other words, participants with a faster response time on mixed trials compared to non-symbolic trials demonstrate a higher level of integration than participants with a slower response time on mixed trials compared to non-symbolic trials.

As a sanity check within the currently collected data, the overall Number Comparison integration effect replicated what other studies using this task found (e.g., Lyons et al., 2012), in which the response times for mixed trials (M = 984.3, SD = 229.1) were slower on average than symbolic trials (M = 892.4, SD = 167.9) and non-symbolic trials (M = 891.8, SD = 205.7), and the response time differences between mixed and symbolic trials and mixed and non-symbolic trials were significantly greater than 0, $t(77) = 5.27$, $p < .001$ (mixed - symbolic), $t(74) = 3.82$, $p < .001$ (mixed - non-symbolic). Thus, there is evidence that the estrangement effect replicated and that this data could form the basis of an individual differences measure. While accuracy was collected, it was not used to calculate integration strength (besides determining whether a trial was correct or not and used in the response time calculation) because the integration effect in accuracy was small with little variation, making it a poor individual difference measures compared to the response time calculation.

This difference score was standardized and then averaged with the symbolic integration strength score from the Number Decision task to create a combined integration strength score that was used for analyses. Interestingly, the Number Decision and Number Comparison integration scores were uncorrelated ($r(70) = .002$). This lack of correlation may reflect high levels of measurement noise at the individual level, and therefore combining them will be critical to decrease that noise for use in individual different analyses. However, they may also capture different kinds of integration (e.g., automatic vs. deliberate integration). Therefore, we also examined each measure in isolation in the analyses and report whether analyses changed when using only the Number Decision or Number Comparison strength measures instead.

Participants' *Symbolic Representational Acuity* and *Non-Symbolic Representational Acuity* was measured by standardizing each participant's mean accuracy (removing trials with response times of under 100 ms to ensure full processing of the stimuli) and mean response time on correct trials, and then averaging the standardized accuracy and standardized response time scores to calculate an acuity measure. Accuracy and response time were used instead of Weber fractions or ratio effects because they have been found to be more reliable measures of representational acuity (Inglis & Gilmore, 2014). Symbolic acuity was measured using only the symbolic trials, and non-symbolic acuity was measured using only the non-symbolic trials (with dot quantities adjusted for bias using the Dot Estimation task results).

2.3.2.6 Forward and Backward Digit Span task.

The forward and backward digit span tasks (Weschsler, 1944) were used to measure individual differences in participants' verbal *Working Memory Capacity*. For the forward digit span task, participants were verbally told a series of single-digit numbers (e.g., "8, 2, 5"), and then asked to repeat those numbers to the experimenter. The same procedure was followed for the backward digit span, except that participants were asked to repeat the numbers back in reverse order. The first sequence was three digits long. Sequences progressively increased by one digit every time the participant could successfully repeat two sequences of a given length, up to a maximum sequence length of 12 digits (although no participant reached a sequence length higher than 8). Participants' forward and backward digit spans were separately recorded as the longest sequence length (across two sequences) that the person could successfully repeat forward or backward, respectively. Participants who could not repeat two consecutive three-digit sequences were automatically given a working memory score of three. The forward and backward digit spans were then combined to calculate the participants' overall working memory capacity, as the combined score was significantly more reliable than the individual forward and backward scores. Combined scores more than two standard deviations away from the group average were winsorized (1 participant).

2.3.2.7 Abbreviated Mathematics Anxiety Rating Scale (A-MARS).

The A-MARS (Alexander & Martray, 1989; $\alpha = .95$) asked participants to rate their level of anxiety in 25 situations that involve mathematics (e.g., studying for a math test, reading a receipt, walking into a math class). Ratings ranged from "Not at all" to "Very much" and were scored on a scale of 1-5, respectively. Participants' final *Math Anxiety* score equaled the sum of their ratings across the 25 items. Scores of participants with accuracies more than two standard deviations from the group average were winsorized (2 participants).

2.3.2.8 Demographics questionnaire.

Participants completed a demographics questionnaire at the end of the session. The questionnaire asked for their level of confidence with math (by rating their skill level relative to 100 people of their age), their quantitative scores on the SAT/ACT, gender, age, race and ethnicity (White and Asian defined as non-minority, and Black/African-American, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, and Hispanic/Latinx defined as minority), school and education level, and number of years since their last math class.

2.3.3 Procedure

Participants completed the Applied Problems test and the Forward and Backward Digit Span tasks verbally with the experimenter. They then completed the Dot Estimation task, the Math Facts task, the Number Decision task, and the Number Comparison task on the computer. The session concluded with the Abbreviated Mathematics Anxiety Rating Scale and the demographics questionnaire.

2.4 RESULTS

2.4.1 Investigating Correlations Between Measures

Correlations were run between the cognitive, affective, and demographics measures to understand how the various measures related to one another. Table 3 shows the correlations for all participants, and Table 4 shows the correlations split by skill level (Mid vs. Low). Because some participants were missing data for specific measures due to technical issues with some of the tasks, the *n* for each correlation varied from 63 to 78 in the combined data, from 38 to 50 in the Mid group, and from 24 to 28 in the Low group.

Table 3. Pearson correlation matrix for all variables for all participants combined

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Applied problems														
2. Math facts	.54**	–												
3. Integration accuracy	.05	.09												
4. Integration precision	-.22†	-.36**	.38**											
5. Integration strength	-.17	-.10	.13	.15										
6. Symbolic acuity	.24*	.25*	-.13	-.21†	.03									
7. Non-symbolic acuity	.13	.18	.06	-.08	-.32**	.46**								
8. Working memory	.45**	.42**	-.10	-.17	-.03	.18	.17							
9. Math anxiety	-.45**	-.50**	.11	.32**	.11	-.43**	-.22†	-.28*						
10. Math confidence	.39**	.36**	.08	-.18	-.13	.26*	.36**	.01	-.54**					
11. Gender	.13	-.21†	.08	.02	-.03	-.15	-.31*	-.12	.23†	-.30*				
12. Age	-.17	-.11	-.23†	-.04	.15	-.03	-.19	.08	-.18	.01	-.04			
13. Minority status	-.21†	-.24*	-.01	.06	-.05	-.07	-.14	-.11	.10	-.05	-.04	.03		
14. Education level	.16	.06	-.18	-.01	-.11	.15	.14	.01	-.16	.21†	-.02	.18	-.11	
15. Time of last math class	-.14	-.19	-.25*	.06	.07	-.004	-.12	-.03	-.16	-.09	.01	.80**	-.01	.04

Note. ** $p < .01$, * $p < .05$, † $p < .10$

2.4.1.1 Correlations with math outcomes, all participants included.

The two math outcome measures (Math Facts and Applied Problems) were positively correlated, showing that basic addition and subtraction fluency and complex math are closely related competencies and can be treated as two measures of a more general mathematics ability construct. However, the correlations were not so high as to make the measures redundant outcomes.

Both Math Facts and Applied Problems shared correlations with many other measures to similar degrees. Both math outcomes were associated with higher symbolic acuity, higher working memory, lower math anxiety, and higher self-reported math confidence. More precise symbolic integration and being a non-minority also correlated with higher Math Facts accuracy.

2.4.1.2 Correlations among predictors, all participants included.

Within the symbolic integration measures (Integration Accuracy, Precision, and Strength), higher integration accuracy was associated with lower integration precision and having more recently taken a math class. Higher precision in symbolic integration was also related to lower math anxiety (and lower integration strength when using only the Number Decision task). Finally, stronger symbolic integration was also associated with lower non-symbolic acuity, but this is likely because the integration score was partially calculated using the non-symbolic trials that made up the non-symbolic acuity measure. This correlation was not significant when using only the Number Decision task but trended in the same direction.

For the representational acuity measures (Symbolic and Non-Symbolic Acuity), participants with higher acuity in one generally showed higher acuity in the other. Both types of acuity correlated with higher math confidence, and higher symbolic acuity also correlated with lower math anxiety. Men were more likely to have higher non-symbolic acuity, but not symbolic acuity, than women.

Table 4. Pearson correlation matrix for all variables separately for Mid-Skill (top value) and Low-Skill (bottom value) groups. Green colored cells are variables significant for both Mid- and Low-Skill, orange colored cells are significant for only Mid-Skill, and blue colored cells are significant for only Low-Skill. Bolded cells signify significantly different correlations between groups

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Applied problems	–													
2. Math facts	.59** .46*	–												
3. Integration accuracy	.04 .11	**.26† -.31**	–											
4. Integration precision	-.19 -.33†	-.30* -.51**	.39** .34†	–										
5. Integration strength	-.35* -.01	-.20 .08	.25† -.06	**.38* -.26**	–									
6. Symbolic acuity	.28* .15	.31* .11	**-.32* .26**	-.35* .08	**-.32* .37†**	–								
7. Non-symbolic acuity	.25† .03	.24 .08	-.05 .27	-.22 .21	-.42** -.17	.45** .51**	–							
8. Working memory	.50** .37*	.38** .53**	-.02 -.28	-.11 -.29	-.06 .17	.24 .07	.28† -.02	–						
9. Math anxiety	-.48** -.44*	-.54** -.41*	.22 -.11	.29† .39†	.16 .06	-.51** -.29	-.18 -.29	-.23 -.36†	–					
10. Math confidence	.47** .21	.49** .10	.03 .21	-.17 -.22	**-.40* .15**	.27† .22	.42** .33†	.03 -.05	-.58** -.47*	–				
11. Gender	-.16 .33†	-.26† -.20	.02 .22	-.07 .21	.004 -.09	-.10 -.25	-.20 -.41*	-.15 -.13	.28† .22	-.47** -.20	–			
12. Age	-.24 .07	-.08 -.16	-.38* .07	-.02 -.09	.31† -.05	.02 -.11	-.18 -.23	.12 .003	-.12 -.32	-.13 .30	.004 -.12	–		
13. Minority status	-.23 -.27	-.23 -.30	-.04 .06	.04 .12	-.10 -.01	-.11 -.02	-.13 -.14	-.07 -.19	.13 .07	-.15 .09	.04 -.19	.08 -.07	–	
14. Education level	.09 .21	.05 .04	-.20 -.13	-.01 -.001	-.19 -.04	.16 .11	.16 .14	-.01 .03	-.10 -.23	.25 .13	.02 -.10	.37* -.14	-.17 -.04	–
15. Time of last math class	-.22 -.02	-.22 -.12	**-.46** .18**	.08 .01	.21 -.11	.02 -.05	-.15 -.09	-.03 -.02	-.06 -.32	-.15 .04	-.01 .04	.80** .81**	.03 -.11	.21 -.24

Note. ** $p < .01$, * $p < .05$, † $p < .10$

2.4.1.3 Correlations with math outcomes, split by skill level.

To investigate how prior math skill level affected the associations between measures, the correlations were also calculated within the subsamples based on skill level (Mid, Low; see Table 4, left). For the math outcome measures, Math Facts and Applied Problems still positively correlated for both skill groups. Both Mid-Skill and Low-Skill participants also showed correlations between higher Math Facts accuracy and higher integration precision, higher working memory, and lower math anxiety. Higher Applied Problems accuracy was also related to higher working memory and lower math anxiety.

However, there were also several correlations unique to each skill group (see Figure 2 for scatterplots of the significant correlations in the Mid-Skill group and Figure 3 for scatterplots of the significant correlations in the Low-Skill group). In the Mid-Skill group, higher Math Facts accuracy was associated with higher symbolic acuity and higher reported math confidence. Notably, for the Applied Problems test, Mid-Skill participants showed significant correlations with lower integration strength (only for the combined score, though the individual Number Decision and Number Comparison scores trended in the same direction), higher symbolic acuity, and higher math confidence, while the Low-Skill participants showed no significant correlations. These results suggest that the Mid-Skill group may be supported more by number cognition resources such as representational acuity when solving math problems, though integration between these two representational types may actually be detrimental to performance on more complex math.

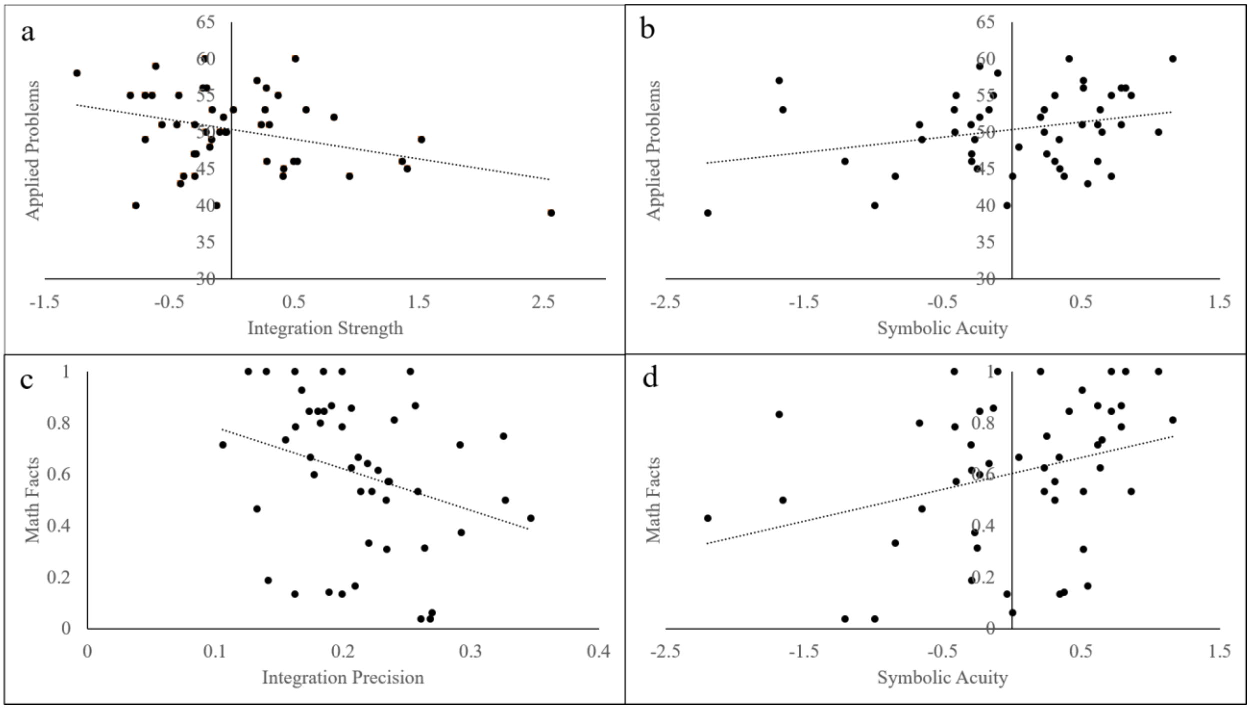

Figure 2. Scatterplots of correlations between a) Integration Strength and Applied Problems, b) Symbolic Acuity and Applied Problems, c) Integration Precision and Math Facts, and d) Symbolic Acuity and Math Facts in the Mid-Skill group

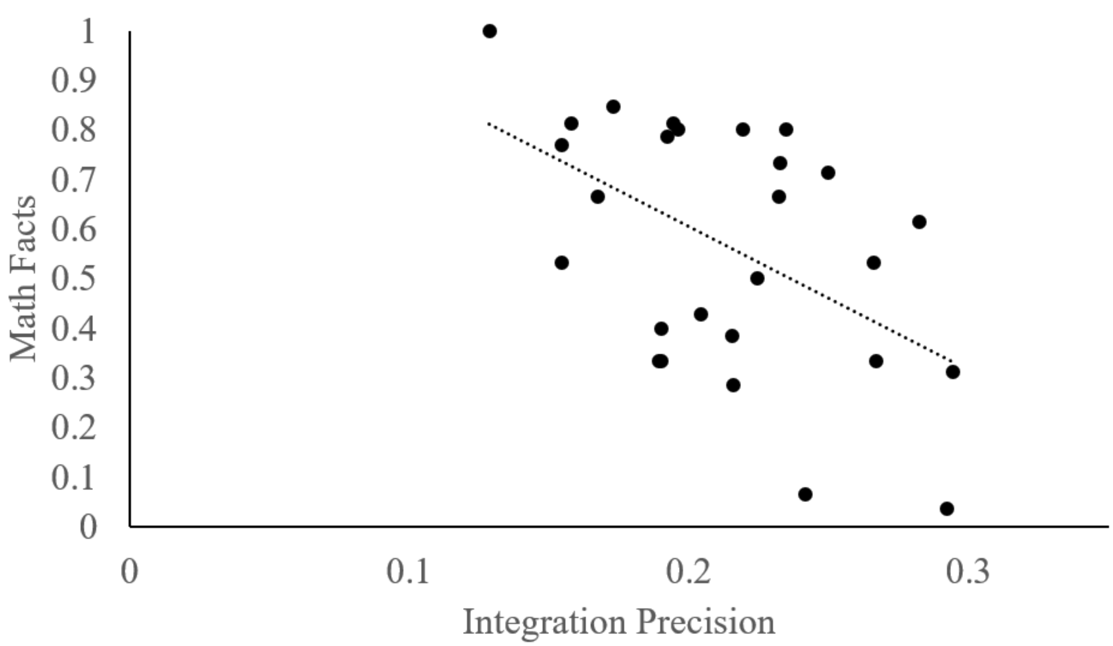

Figure 3. Scatterplot of correlations between Integration Precision and Math Facts in the Low-Skill group

2.4.1.4 Correlations among predictors, split by skill level.

For the Mid-Skill group only (see Table 4 upper right), higher integration accuracy correlated with lower integration precision, lower symbolic acuity, and being younger. Higher integration precision also correlated with lower integration strength (for the combined score and Number Decision score only) and higher symbolic acuity, while stronger symbolic integration was correlated with lower symbolic acuity (for the combined score only, with the Number Decision score trending in the same direction), lower non-symbolic acuity (for the combined score and Number Comparison task, though again likely conflated by these measures using the same non-symbolic trials), and lower math confidence (for the combined and Number Decision score, with Number Comparison trending in the same direction). The Low-Skill group showed no unique correlations compared to the Mid-Skill group.

In representational acuity, symbolic acuity and non-symbolic acuity correlated in both the Mid-Skill and Low-Skill groups. Only the Mid-Skill group showed symbolic acuity correlating with lower math anxiety and non-symbolic acuity correlating with higher math confidence.

2.4.2 Predicting Procedural Math Fluency and Complex Math Performance

To determine which measures independently predicted performance on Math Facts and Applied Problems, multiple linear regressions were calculated to predict Math Facts and Applied Problems. Only the cognitive and affective mechanisms were included as predictors (Integration Accuracy, Integration Precision, Integration Strength, Symbolic Acuity, Non-Symbolic Acuity, Working Memory, Math Anxiety, Math Confidence) because the demographic factors were unlikely to be direct causes of math performance. Analyses were again run separately between the Mid-Skill and Low-Skill groups, given the correlational differences seen between the two skill groups. Because the results from a forced model including all eight variables would likely be unstable due to the high number of predictors for the amount of data available, forward and backward regressions (with an entry probability of .05 and removal of .10) were first run to predict Math Facts and Applied Problems, and the significant predictors from these were input into a forced regression. Standardized betas for all three models, separated by skill level, are shown in Table 5 and Table 6.

Table 5. Multiple regression standardized betas predicting Math Facts accuracy with a forward regression, backward regression, and forced regression

	Mid Skill			Low Skill		
Variables	Forward	Backward	Forced	Forward	Backward	Forced
R^2	.39	.58	.54	.46	.46	.41
Int. Accuracy	-	.37**	.49**	-	-	-
Int. Precision	-	-.32*	-.29*	-.44*	-.44*	-.39*
Int. Strength	-	-.22†	-.12ns	-	-	-
Symbolic Acuity	-	-	-	-	-	-
Non-Symbolic Acuity	-	-	-	-	-	-
Working Memory	.28*	.30*	.29*	.40*	.40*	.41*
Math Anxiety	-.53**	-.50**	-.49**	-	-	-
Math Confidence	-	-	-	-	-	-

Note. ** $p < .01$, * $p < .05$, † $p < .10$, ns = not significant

Table 6. Multiple regression standardized betas predicting Applied Problems accuracy with a forward regression, backward regression, and forced regression

	Mid Skill			Low Skill		
Variables	Forward	Backward	Forced	Forward	Backward	Forced
R^2	.47	.47	.49	.20	.37	.39
Int. Accuracy	-	-	-	-	.36†	.31ns
Int. Precision	-	-	-	-	-.39†	-.33hs
Int. Strength	-.30*	-.30*	-.22†	-	-	-
Symbolic Acuity	-	-	-	-	-	-
Non-Symbolic Acuity	-	-	-	-	-	-
Working Memory	.43**	.43**	.44**	-	.41*	.36ns
Math Anxiety	-.45**	-.45**	-.38**	-.44*	-	-.15ns
Math Confidence	-	-	-	-	-	-

Note. ** $p < .01$, * $p < .05$, † $p < .10$, ns = not significant

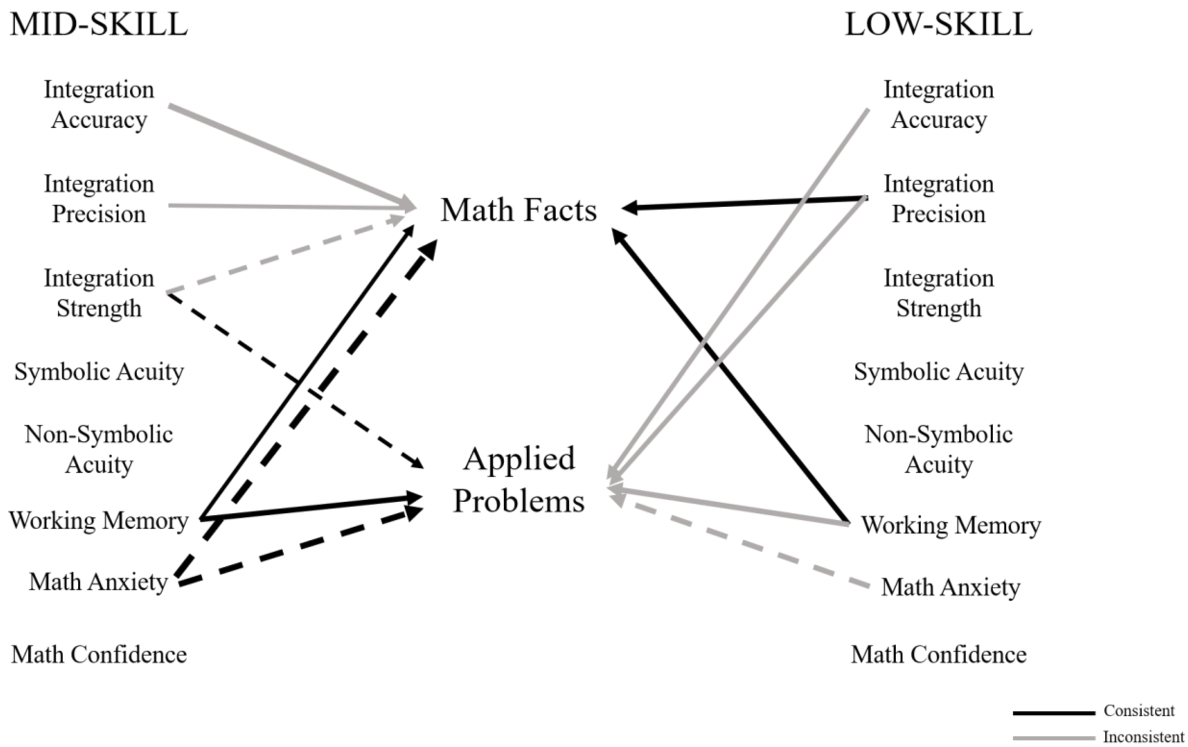

Figure 4. Multiple regressions predicting Math Facts and Applied Problems, separated by skill. Line width represents strength of each predictor. Black lines denote consistently found predictors, light gray lines denote inconsistently found predictors, solid lines denote positive predictors, and dotted lines denote negative predictors

Figure 4 shows the results of the multiple regressions by skill level. In the forward model (R^2 = .39, $F(2, 34) = 11.0$, $p < .001$), only higher working memory and lower math anxiety significantly predicted Math Facts accuracy. In the backward model (R^2 = .58, $F(5, 31) = 8.48$, $p < .001$), more accurate integration, more precise integration, marginally weaker symbolic integration, higher working memory, and lower math anxiety predicted Math Facts. Finally, the forced regression, which included integration accuracy, integration precision, integration strength, working memory, and math anxiety (R^2 = .54, $F(5, 36) = 8.30$, $p < .001$), showed that all except integration strength were significant predictors.

In contrast, the Low-Skill group's forward model (R^2 = .46, $F(2, 21) = 8.97$, $p = .002$) and backward model (R^2 = .46, $F(2, 21) = 8.97$, $p = .002$) both ended with higher integration precision and higher working memory as significant predictors. The final forced model (R^2 = .41, $F(2, 24) = 8.48$, $p = .002$), using integration precision and working memory as predictors, found that both variables were significant. Thus, the Mid-Skill group appeared to rely more heavily on cognitive and affective resources, as all three aspects of symbolic integration and math anxiety significantly predicted Math Facts performance (though symbolic integration strength was less consistent as a predictor), while the Low-Skill group only shared integration precision and working memory as predictors that supported procedural fluency.

Predictors for the Applied Problems test also varied depending on the skill group. For the Mid-Skill group, the forward model (R^2 = .47, $F(3, 33) = 9.67$, $p < .001$) and backward model (R^2 = .47, $F(3, 33) = 9.67$, $p < .001$) both found consistent predictors in lower symbolic integration strength, higher working memory, and lower math anxiety. Note that when only the Number Decision task integration strength score was used in the regressions, only higher working memory and higher math confidence predicted higher Applied Problems, while the combined integration strength score and Number Comparison integration strength score did not differ in results. The forced model that included combined integration strength, working memory, and math anxiety as predictors (R^2 = .49, $F(3, 38) = 12.2$, $p < .001$) showed that marginally lower integration strength, higher working memory, and lower math anxiety significantly predicted Applied Problems accuracy.

In the Low-Skill group, the forward model (R^2 = .20, $F(1, 22) = 5.37$, $p = .03$) ended with only lower math anxiety as a significant predictor. The backward model (R^2 = .37, $F(3, 20) = 3.90$, $p = .02$) ended with marginally higher integration accuracy, marginally higher integration precision, and higher working memory as predictors. The forced model included integration accuracy, integration precision, working memory, and math anxiety as predictors (R^2 = .39, F(4, 19) = 2.97, p = .05), but none of the predictors reached significance. Thus, in the Mid-Skill group, resources appear to shift away from integration as math becomes more complex, instead relying more on working memory or math anxiety. Meanwhile, the Low-Skill group is now relying more on integration in addition to working memory and the influence of math anxiety, but the lack of significance among the predictors suggests that some other resource is more important for their complex math performance.

2.5 DISCUSSION

In the current study, we investigated how symbolic integration accuracy, integration precision, integration strength, symbolic representational acuity, non-symbolic representational acuity, working memory, math anxiety and math confidence, and demographic variables predicted procedural and applied math performance, as well as how these factors related to one another. Regardless of people's initial math skill level, higher working memory and lower math anxiety related to both basic procedural math and complex applied math. More precise symbolic integration was also associated with better procedural math performance. Notably, interesting differences were found when looking at people with higher math skill separately from those with lower math skill. The Mid-Skill group showed that more precise symbolic representations and higher math confidence correlated with higher procedural math performance, and weaker symbolic integration strength, more precise symbolic representations, and higher math confidence correlated with better applied math performance. Meanwhile, the Low-Skill group showed no such correlations between performance on either type of math and any of the cognitive resources that were investigated.

This pattern, in which the Mid-Skill group showed stronger (though not significantly different) relationships between math performance and cognitive foundational resources than the Low-Skill group, was mostly consistent with what was found in our multiple regressions, particularly in more complex math. In the Mid-Skill group, more accurate and more precise symbolic integration, higher working memory, and lower math anxiety predicted higher procedural math performance. However, a shift in resources was seen in Mid-Skill participants when doing more applied math; while higher working memory and lower math anxiety were still significant predictors of applied math performance, now weaker symbolic integration was more important than the other two aspects of symbolic integration. In contrast, in the Low-Skill group, only more precise integration and higher working memory dependably contributed to procedural math performance across all regression models. Meanwhile, the predictors for applied math were less reliable; more accurate and precise integration, higher working memory, and lower math anxiety showed up in separate regression models as significant predictors, but none were consistently found across all models and even became non-significant when forced together into a model.

Together, these results show that for simpler math, all participants relied to some extent on their ability to access quantities from symbolic numbers and their capacity to hold information in working memory. The higher skilled participants also appeared to recruit additional resources that further refined their symbolic integration, as well as math anxiety (unique even when including working memory, suggesting that math anxiety did not affect math performance purely through co-opting working memory resources). Complex math is where notable differences began to appear when looking separately at the two math skill levels. Higher-skilled participants' applied math performance now benefitted from having weaker symbolic integration, the opposite of their reliance on symbolic integration accuracy and precision in simpler procedural math. While the Low-Skill group showed inconsistent predictors for applied math, the predictors were the same as the resources that Mid-Skill participants had used for procedural math (i.e., integration accuracy, integration precision, working memory, math anxiety). This shift away from symbolic integration-based resources when doing basic versus complex math may be the primary distinction between higher-skilled and lower-skilled math participants. Further, the lack of consistent predictors for the Low-Skill group may mean that the Low-Skill group did not have the declarative knowledge required for the applied problems and could therefore not meaningfully utilize any cognitive resources during math problem solving.

2.5.1 Symbolic Integration and Math Performance

Unlike prior studies, we found no strong association between non-symbolic acuity and either procedural math or applied math (e.g., Dehaene, 1992; Fazio et al., 2014; Feigenson et al., 2004; Gallistel & Gelman, 1992; Halberda et al., 2012; Halberda et al., 2008; Holloway & Ansari, 2009; Libertus et al., 2013; Lourenco et al., 2012; Piazza et al., 2004; Schneider et al., 2016). There were significant correlations (though only in the Mid-Skill group) between symbolic acuity and both math outcomes, which fits well with studies showing a much stronger relationship between symbolic acuity and math performance for adults compared to non-symbolic acuity (e.g., Fazio et al., 2014; Schneider et al., 2016) . However, symbolic acuity did not significantly predict math performance in any regression model. Instead, symbolic integration accuracy, precision, and strength won out as the primary predictors of math. Other studies have found that symbolic integration precision and strength can fully mediate the relationship between non-symbolic representational acuity and symbolic arithmetic (Jang & Cho, 2018; Libertus et al., 2016; Wong et al., 2016). Similarly, our results suggest that symbolic integration may be more important than symbolic representational acuity for math performance as well.

Inconsistent with most prior symbolic integration studies, our results suggest that symbolic integration may not always be helpful for math performance. Specifically, integration accuracy and precision may support more basic procedural math but may actually be detrimental for more complex math problems. While the predominant theory suggests that symbolic representations are mapped onto non-symbolic representations to give meaning to numeric symbols, an alternative account is that the non-symbolic and symbolic representational systems are largely separate and develop independently (Bulthe, De Smedt, & Op de Beeck, 2014; M. Le Corre & S. Carey, 2007; Lyons et al., 2012). In this case, numeric symbols would gain meaning through order relations between symbols. While our study cannot provide strong evidence for or against either of these theories, our findings that more accurate and precise symbolic and non-symbolic mappings predicted better procedural math performance, and that weaker symbolic integration predicted better applied math performance, may be evidence that symbolic-non-symbolic connections are most important for learning and supporting early mathematics, while developing symbolic-symbolic connections are more important for more advanced math where understanding of non-symbolic quantities are less helpful (e.g., as number quantities increase or math operations become more complicated).

It should be noted that our symbolic integration strength measure was created from two individual integration strength measures that had no correlation with each other. While results were very similar regardless of the integration strength measure used, one exception occurred with the forward and backward regressions predicting Applied Problems performance, in which integration strength and math anxiety were only significant predictors when using the combined or Number Comparison scores. The low correlations and slightly changing results could suggest that the Number Decision and Number Comparison tasks are measuring separate aspects of symbolic integration (in addition to integration accuracy and precision). The Number Decision task differs from the Number Comparison task in the kind of judgments being made during the task. While the Number Comparison task asks participants to make numerical comparison judgments that involve both symbolic and non-symbolic quantities, the Number Decision task asks participants to make non-numerical judgments on symbolic numerals when non-symbolic quantities are present (but irrelevant to the task). In this sense, the Number Comparison task could be viewed as a measure of deliberate integration, in which participants must purposefully process symbolic and non-symbolic quantities together, while the Number Decision task could be viewed as a measure of automatic integration, in which participants do not need to process either the symbolic or non-symbolic quantities shown but may still be influenced by their presence, with the extent of this automatic influence being the measure of integration. Because both of these aspects still fall under integration, the combined score used in the current study provides a comprehensive measure of general symbolic integration, but future studies should investigate whether different kinds of symbolic integration strength exist and how each relates to math performance.

2.5.2 The Unique Contributions of Working Memory and Math Anxiety to Math Performance

Prior studies have suggested that math anxiety impacts math performance by co-opting working memory resources (Ashcraft, 2002) or by preventing people from pursuing math opportunities (e.g., Hembree, 1990; Tobias, 1978). However, in the current study, math anxiety was found to uniquely contribute to both procedural and applied math performance in addition to working memory, suggesting that math anxiety does not affect math performance solely through its influence on working memory capacity. Further, math anxiety did not show a high correlation with participants' time since last math class, making it unlikely that math anxiety only reflected an avoidance of math experiences. The question remains as to how math anxiety affected math performance, separate from working memory or math avoidance.

One possibility comes from prior findings that one's math anxiety levels can influence the relationships between cognitive resources and math performance. For example, Braham and Libertus (2018) found that only highly math anxious individuals showed a positive relationship between ANS acuity and performance on the Applied Problems test. In our case, math anxiety may be acting as a proxy for more basic cognitive resources, such as symbolic acuity (note the relatively high correlation between math anxiety and symbolic acuity), that did not predict math performance in any of our regression models. That is, people's math anxiety may be indicative of differences in the development of their cognitive resources for math use during prior math learning, which may have affected which predictive foundational resources were found in the current study.

2.5.3 Limitations of the Current Study

While the current study provided a detailed look at many foundational math mechanisms and how they relate to procedural and applied math performance, there were still several limitations and open questions to address through future studies. One potential weakness was the use of the Number Comparison non-symbolic trials to calculate both symbolic integration strength and non-symbolic acuity. It is possible that one would find a unique contribution of non-symbolic acuity, separate from symbolic integration strength, were the two measures less related. However, it should be noted that the correlation for the two measures, while significant, was not so high as to suggest that they were redundant (r = .3), so results are unlikely to meaningfully change.

In addition, there are still a couple of other mechanisms not included in the current study that have been shown to potentially contribute to math performance. We investigated the acuity of symbolic and non-symbolic quantities by using cardinality judgments (i.e., which of two quantities is greater?), but recent studies have found that using symbolic and non-symbolic ordinality judgments (i.e., are these numerals or quantities in numerical order?) lead to significantly different patterns (e.g., Goffin & Ansari, 2016; Lyons & Beilock, 2011; Lyons & Beilock, 2013; Lyons, Vogel, & Ansari, 2016). For example, people are typically slower and less accurate on cardinality tasks when the ratios between the compared numbers are small compared to large ratios; while this same effect was seen with non-symbolic ordinality judgments, the opposite was actually found with symbolic ordinality judgments, in which people were more accurate and faster when quantities were closer (Lyons & Beilock, 2013). Individual differences in symbolic number ordering abilities also uniquely predicted both procedural and complex math (Goffin & Ansari, 2016; Lyons & Beilock, 2011), and it fully mediated the relationship between non-symbolic acuity and complex math performance (Lyons & Beilock, 2011). Thus, ordinality is another mechanism worth investigating in future studies. Symbolic ordinality in particular may be important, given our findings suggesting that weak symbolic integration, and therefore symbolic-symbolic connections, are more useful for complex math.

Several studies have also suggested that other components of executive function besides working memory, such as inhibitory control and cognitive flexibility, may mediate the ANS and math achievement association (e.g., Clark, Sheffield, Wiebe, & Espy, 2013; Fuhs & McNeil, 2013; Gilmore et al., 2013). For example, Fuhs and McNeil (2013) found that a significant relationship between ANS acuity and mathematics ability became non-significant after controlling for inhibitory control. Inhibitory control also appears to uniquely predict children's math scores in preschool after controlling for working memory capacity and cognitive flexibility (Espy et al., 2004), meaning that the inclusion of inhibitory control in particular could affect our results showing relationships between working memory capacity and math performance. However, other studies have found no correlations between inhibitory control and ANS acuity (Keller & Libertus, 2015), and that associations between non-symbolic acuity and math ability remain significant (though weaker) even after controlling for inhibitory control (Chen & Li, 2014; Keller & Libertus, 2015). While some studies have found that cognitive flexibility no longer relates to math performance when controlling for inhibitory control (Espy et al., 2004), a meta-analysis by Yeniad and colleagues (2013) showed that cognitive flexibility is associated with math performance throughout children's development. Further, cognitive flexibility may be particularly important in supporting the shift in resources seen in the current study. It is also possible that each executive function (working memory, inhibitory control, cognitive flexibility) contributes to specific math components; for example, Cragg and Gilmore (2014) proposed that working memory may be most important for acquiring new math facts and storing interim steps for procedural strategies, inhibitory control may be most important for learning new math facts and concepts by suppressing superficially similar but incorrect responses to problems (e.g., inhibiting 6 when asked 3x3) or inhibiting the use of more basic and automatic strategies, and cognitive flexibility may be most important for learning new procedures and concepts by switching between procedures and strategies as needed. Thus, inhibitory control and cognitive flexibility may be other factors worth investigating in more detail in future studies. This could also provide insight into how inhibitory control and cognitive flexibility relate to math abilities, as these mechanisms are still unknown.

Finally, there is a potential limitation in the method used to categorize participants as Mid-Skill and Low-Skill. In the current study, participants who scored at or above the 50^{th} percentile on the math SAT or ACT were placed into the Mid-Skill group, while participants who scored below the 50^{th} percentile were placed into the Low-Skill group. However, standardized tests, including the SAT and ACT, have been criticized as inaccurate predictors of math preparation, especially for non-Asian minorities and women. Non-Asian minorities and women regularly score lower on average on these standardized tests than non-minorities and men, respectively (CollegeBoard, 2017). A potential reason for this persistent gap is "stereotype threat" (Steele, Spencer, & Aronson, 2002), or the worry that one will conform to negative stereotypes about their social group – in this case, that minorities and women are bad at math. Similarly to math anxiety, stereotype threat is thought to cause ruminations that co-opt working memory resources needed for math performance (Maloney, Schaeffer, & Beilock, 2013). While taking standardized tests, the testing context of an academic environment may bring to mind these math stereotypes, thereby decreasing test scores on these sections of the test (Walton &

Spencer, 2009). The current studies' samples were majority female, meaning that stereotype threat may have caused the SAT and ACT scores to underestimate the actual math skill level of our participants. Thus, our two skill groups may be more similar in math skill than proposed. The Mid- and Low-Skill groups did not significantly differ across any measure but Applied Problems (and even that was a relatively small difference of 3 points on average). It is possible that participants viewed the experimental sessions as lower in pressure and non-academic, which would not activate stereotype threat, leading to higher scores on the math measures than on the SAT or ACT. We also took precautions to minimize stereotype threat as much as possible (e.g., only collecting demographics information at the end of the study to avoid bringing attention to participants' gender or race/ethnicity), which may have further increased their math scores compared to the standardized tests. However, both the Mid- and Low-Skill groups were mostly women, so one would expect stereotype threat to equally affect both groups such that the relative skill differences between the two groups would be maintained. Additionally, we still found differences in the correlations and regressions in the two groups even though they did not differ substantially in any cognitive measure, suggesting that while both groups' cognitive resources were equally developed, they utilized the resources during math performance in meaningfully different ways. Future studies may still want to use other measures, or even several measures, of initial math skill to see if the current studies' results still hold.

2.5.4 Conclusion

In sum, the current study found notable differences in the foundational resources that relate to procedural and complex math performance in Mid-Skill and Low-Skill people. Mid-Skill people appear to generally rely upon cognitive resources, such as symbolic integration accuracy and precision, for procedural math and then switch away from integration for more complex math. Meanwhile, Low-Skill people showed relationships between symbolic integration resources and both procedural and complex math. The development of using these foundational resources and the ability to switch resources depending on the type of math being done may be prime targets for improving math performance in adults who are struggling with math.

3.0 STUDY 2

In Study 1, we investigated the cognitive and individual difference mechanisms that related to math performance. Specifically, we examined how symbolic integration accuracy, integration precision, integration strength, symbolic representational acuity, non-symbolic acuity, working memory, math anxiety, math confidence, and several demographics variables correlated with procedural math and applied math performance. Further, we looked at how these relationships varied depending on people's overall math skill level, providing insight into the foundational math resources available to people with higher math skill compared to those with lower math skill.

In addition to understanding the mechanisms associated with math performance, it is equally important to determine whether interventions that target these mechanisms can effectively improve math performance. Many recent studies have attempted to improve math performance by acting through various cognitive mechanisms, with varying levels of success. We discuss relevant studies and the varying amounts of transfer that they obtained in more detail below (see Table 7 for a summary).

Table 7. Summary of the methods and findings of prior math training studies

Author & Year	Participant Ages	Intervention	Measures Showing Growth	Measures Showing No Effect
Bajic, Kwak, & Rickard, 2011	Adults	Limited addition and subtraction	Identical addition and subtraction problems Flipped order addition	Flipped order subtraction Problems with different operations
Rickard et al., 1994	Adults	Limited multiplication and division	Problems with same numbers and operations	Problems with any changed elements
Park & Brannon, 2013	Adults	Approximate arithmetic	Symbolic arithmetic	Vocabulary test
Au, Jaeggi, & Buschkuehl, 2018	Adults	Approximate arithmetic	Symbolic arithmetic Non-symbolic acuity Symbolic acuity	-
Park & Brannon, 2014	Adults	Approximate arithmetic	Symbolic arithmetic	-
		Approximate numerical comparison	-	Symbolic arithmetic
		Spatial working memory	-	Symbolic arithmetic
		Symbolic number ordering	-	Symbolic arithmetic
Park et al., 2016	3-5 years	Approximate arithmetic	Symbolic arithmetic	Vocabulary test Short-term memory Executive function
Szkudlarek & Brannon, 2018	3-5 years Low-achieving	Approximate arithmetic	Counting Magnitude Ordinality	Fact retrieval Numeral identification
Hyde, Khanum, & Spelke, 2014	6-7 years	Approximate arithmetic	Symbolic addition	-
		Approximate numerical comparison	Symbolic addition	-
Khanum et al., 2016	6 years	Approximate addition	Symbolic arithmetic Symbolic number line task	Non-symbolic acuity
Dillon et al., 2017	2-12 years	Approximate arithmetic	Non-symbolic math performance	Symbolic math performance
Sella et al., 2016	4-6 years	Number Race game (non-symbolic & symbolic comparison)	Mental calculation Spatial number mapping Semantic number representations	Counting Number naming Letter recognition
Siegler & Ramani, 2008, 2009	4-5 years Low-income	Number line board game	Numerical magnitude comparison Number line estimation Numeral identification More accurate later arithmetic	Counting
van Herwegen et al., 2018	3 years	Non-symbolic comparison	Symbolic math Non-symbolic math Non-symbolic acuity	Letter recognition
		Symbolic number line	Symbolic math Non-symbolic math	Letter recognition

Author & Year	Participant Ages	Intervention	Measures Showing Growth	Measures Showing No Effect
Maertens et al., 2016	5 years	Number comparison	Non-symbolic acuity Symbolic arithmetic Number knowledge	Number line performance
		Number line estimation	Symbolic arithmetic	Number comparison
Kallai, Schunn, Ponting, & Fiez, 2011	Adults	Mental computation	Addition & subtraction problems Complex symbolic math Symbolic acuity Automatic addition	Math fact retrieval Non-symbolic acuity
Liu, Kallai, Schunn, & Fiez, 2015	Adults	Mental computation	Addition & subtraction problems Complex symbolic math Symbolic acuity	-
Lindskog, Winman, & Poom, 2016	Adults	Symbolic arithmetic	Symbolic arithmetic	Non-symbolic acuity
Sullivan, Frank, & Barner, 2016	5-7 years	Abacus	Symbolic arithmetic	Non-symbolic acuity

3.1 TRAINING STUDIES AND MATH PERFORMANCE

3.1.1 Retrieval-Based Training Studies Find Near Math Transfer

Several studies have attempted to improve math performance through the memorization and consequent fast retrieval of simple math facts, and several popular mathematical training interventions used in classrooms follow these retrieval-based methods (e.g., FASTT Math, ExploreLearning Reflex). While these retrieval-based methods are useful in improving the number of simple math facts that can be rapidly retrieved (e.g., Scholastic Inc. 2005), other studies have shown that performance gains only transfer to practiced, or very similar, problems (Bajic, Kwak, & Rickard, 2011; Rickard, Healy, & Bourne, 1994). For example, Bajic, Kwak, and Rickard (2011) trained high-achieving undergraduates on limited sets of single- and double-digit addition and subtraction problems. Training improvements fully transferred to addition and subtraction test problems that were identical to trained problems, and there was also strong transfer to addition (not subtraction) test problems when number order was flipped (e.g., trained on "7+3", tested on "3+7"). However, there was no transfer to test problems in which the operation changed from training (e.g., trained on "10-3", tested on "3+7") or when both order and operation changed (e.g., trained on "10-7", tested on "3+7"). Similar limits in transfer were found when college students practiced limited sets of single- and double-digit multiplication and division problems instead of addition and subtraction. Strong transfer was only found when test problems had the same numbers and operations as the trained problems, but little transfer if elements were changed (Rickard et al., 1994). Thus, while retrieval-based training studies can improve math performance, the extent of improvement is severely limited by the range of practice problems.

3.1.2 Approximate Arithmetic Training Targeting the Approximate Number System Leads to Mid Transfer

Many studies with children and adults have found significant symbolic math improvements using training interventions that target the approximate number system (ANS) – the non-symbolic number system used to understand and process quantities without symbols, and which is thought to provide meaning to numeric symbols through mapping. Considerable success has been found with approximate arithmetic training programs, in which participants practice arithmetic operations using non-symbolic quantities (i.e., dot arrays). For example, Park and Brannon (2013) trained high-achieving undergraduate students on approximate addition and subtraction problems using single- and double-digit quantities. Two dot arrays were shown going behind an occluder, and participants were asked to either compare the sum or difference of the occluded array to a third array, or to choose which of two new arrays was equal to the sum or difference of the occluded array. After training, participants were able to complete significantly more symbolic multi-digit addition and subtraction in a given amount of time than before the training. Similar results were found with preschoolers, such that approximate arithmetic improved later performance on both symbolic and non-symbolic math problems (Park, Bermudez, Roberts, & Brannon, 2016). Interestingly, the effects were strongest for low-achieving children (though only in informal math abilities, such as counting, understanding of numerical magnitude, and understanding of ordinality), while children with high initial math scores benefitted more from a game involving the identification of numbers and letters (Szkudlarek & Brannon, 2018). However, approximate arithmetic does not always lead to transfer to symbolic arithmetic. In one large-scale study of preschool and elementary school children in India, approximate arithmetic and geometry training only improved non-symbolic math performance (which persisted for at least one year after training), but not symbolic math performance (Dillon, Kannan, Dean, Spelke, & Duflo, 2017).

Several studies have replicated and expanded on initial findings with approximate arithmetic to determine the underlying mechanisms that may be driving these training effects. For example, Au, Jaeggi, and Buschkuehl (2018) replicated the improvements on symbolic arithmetic in adults after approximate arithmetic training. In addition, they found preliminary evidence that non-symbolic and symbolic number representations became more acute after training, with a larger improvement in symbolic representations. Park and Brannon (2014) also conducted follow up studies in which they randomly assigned adult participants to training in approximate arithmetic, approximate numerical comparison (i.e., comparing the quantities of two non-symbolic dot arrays), spatial working memory, or symbolic number ordering (i.e., arranging sets of Arabic numerals in ascending or descending order). Symbolic arithmetic improvements were only shown after the approximate arithmetic training. The authors concluded that the lack of transfer after approximate numerical comparison showed that simply involving the ANS is not enough for symbolic arithmetic improvements; rather, it is the shared mental transformations involved in manipulating the approximate and symbolic quantities that underlie the observed transfer effects. Thus, only the approximate arithmetic training, which trained the manipulation of non-symbolic numerical representations, would show improvements in symbolic arithmetic. However, this distinction between approximate arithmetic and approximate comparison was not replicated in younger children: Hyde, Khanum, and Spelke (2014) found that elementary school children trained using approximate addition or approximate numerical comparison both performed better on a symbolic addition test. Mapping between non-symbolic and symbolic representations may also play a role in the symbolic arithmetic improvements. For example, Khanum and colleagues (2016) found that elementary school children also improved in a symbolic number line task (in addition to symbolic arithmetic improvements) after approximate addition training, though the effects were inconsistent.

3.1.3 Symbolic and Non-Symbolic Training Interventions Find Far Transfer

While approximate arithmetic training has shown mid-levels of transfer, in which the primary difference between the trained and tested tasks are the problem formats (i.e., training on non-symbolic arithmetic problems transferring to symbolic arithmetic performance), other studies have found far levels of transfer in which the trained and tested tasks differ more substantially. For example, several studies have found math improvements after training with basic number comparison or number line tasks. In the case of number comparison tasks, one study with preschoolers used the Number Race game, which focuses on improving number representations and the mapping between non-symbolic and symbolic number representations through non-symbolic and symbolic comparison games. Children who trained using the Number Race game showed large improvements in mental calculation and spatial mapping of numbers, and smaller improvements in their semantic representation of numbers (Sella, Tressoldi, Lucangeli, & Zorzi, 2016). In the case of number line tasks, Siegler and Ramani (Siegler & Ramani, 2008, 2009) asked low-income preschoolers to play a board game, in which children moved their game token along a linear board game by the number of spaces indicated on a spinner. Children were also required to say the numbers on the board's spaces as they moved (e.g., if they were at space 3 and spun a 2, they would say "4, 5"). After playing for approximately an hour, preschoolers showed improved performance on numerical magnitude comparison and number line estimation, and also produced more accurate answers and smaller magnitude errors in later arithmetic training.

There is also evidence that number comparison and number line training may improve math through unique mechanisms. In one five-week training program, preschoolers were randomly assigned to either a non-symbolic training program focused on estimating non-symbolic quantities or a symbolic training program focused on recognizing, counting, and ordering symbolic numerals. Both training conditions led to significant improvements in symbolic and non-symbolic math and non-symbolic acuity, and these improvements remained for at least six months (van Herwegen, Costa, Nicholson, & Donlan, 2018). Another study with elementary school children also found that number comparison training and number line estimation training (both using a mix of non-symbolic and symbolic trials) showed equal improvements in symbolic arithmetic. Notably, number comparison training did not improve number line performance and number line training did not improve number comparison performance (Maertens, De Smedt, Sasanguie, Elen, & Reynvoet, 2016). These findings suggest that there may be several separate mechanisms through which training can improve math performance.

Most training studies have focused only on improvements in procedural math performance, but there is also potential for complex math improvements using relatively simple interventions. Kallai, Schunn, Ponting, and Fiez (2011) found far transfer to untrained complex mathematics problems using a program that encouraged the development of meaningful number representations. Adults were trained to perform quick and accurate mental computation (i.e., the process of performing arithmetic operations without external devices (Sowder 1988)), which is thought to build flexible knowledge about numerical symbols and their associated quantities (Markovits & Sowder, 1994; Reys, 1984; Sowder, 1992; Thompson, 1999). Participants solved multi-digit addition and subtraction problems over several training sessions. Problems rarely repeated, meaning that participants could not memorize answers to specific problems and rely on retrieval practices (Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997). Problems were also horizontally formatted and limited in time to further encourage participants to utilize whole number quantity processing in place of rote retrieval strategies. The training included immediate feedback, high uncertainty, and rewards for correct responses, all of which have been shown to modulate a basal ganglia learning system that is involved in cognitive skill learning and representational change (see Tricomi & Fiez 2008 for a review). The mental computation training group greatly improved on the training task as expected. More notably, the training group also greatly improved on a complex math test that required skills beyond the addition and subtraction that had been trained, including reasoning about ratios and proportions, algebraic equations, and probabilities. They also showed significant improvements in symbolic representational acuity. The control conditions, a test-retest group and an active control group that was trained to select and type double-digit numbers based on a colored symbol cue (a numerically-meaningless task that controlled for exposure to and typing of numerical stimuli), showed no improvements on the complex math test. This suggests that the mental computation training, not raw exposure to double-digit numbers, was the basis for mathematical transfer.

In a follow-up study by Liu, Kallai, Schunn, and Fiez (2015), these far transfer effects were found to be robust when isolated features of the original training were removed, and when a more varied online Amazon Mechanical Turk (MTurk) population outside of college students and recent graduates was tested. In one study, three variations of the mental computation training were created that removed either the immediate feedback, high uncertainty, or rewards for correct responses (i.e., the features that modulate the basal ganglia learning system); this was done by removing the feedback given after each trial, making the problems significantly easier and increasing the response time limit, or removing additional monetary rewards for correct responses. Only the core mental computation component of Kallai and colleagues' (2011) original training stayed consistent across all variations. Despite these changes, the three training groups still significantly improved on the complex math test and symbolic representational acuity after training compared to a test-retest control group. In a second study, an estimation-based version of the training task was used, in which participants were encouraged to use holistic number processing to estimate the solutions to double-digit addition and subtraction problems instead of calculating the exact answer. As participants became more accurate at estimating, the estimation window of accepted responses grew smaller to gradually encourage more exact number processing. The participants also varied broadly in education level (from high-school/GED level to master's degrees) and in age (from 20 to 49 years, with an average age of 30 years), but were self-selective into a mathematics study and thus may have been relatively strong mathematically. Once again, participants in the training group significantly improved on a complex math test and symbolic acuity compared to a control test-retest group.

Finally, math improvements after training may not always be related to the ANS. In a longitudinal study of 205 elementary school children, Sullivan, Frank, and Barner (2016) trained students to use a physical abacus for math and eventually to perform math using only their mental representations of the abacus. The training led to improvements in procedural arithmetic abilities, but there was no associated improvement in non-symbolic acuity. Further, non-symbolic acuity and integration between non-symbolic and symbolic numbers did not significantly predict math performance when measures of general intelligence, verbal working memory, and mental rotation performance were controlled. Similarly, another study trained adults on symbolic arithmetic fluency and found that participants improved in addition, subtraction, multiplication, and division performance, but not in non-symbolic acuity (Lindskog, Winman, & Poom, 2016). This again speaks to the possibility of multiple mechanisms through which math performance can improve.

3.2 THE CURRENT STUDY

While past training studies have found promising results on effectively improving math performance, there are still many open questions left regarding math interventions. Perhaps two of the most pressing are the "why" and "who" behind math improvements. One open question regards the unique mechanisms that explain changes in math outcomes. Similar to prior correlational studies about math performance, many prior training studies have only investigated a couple of potential mechanisms in isolation for improving math performance. Few studies have simultaneously compared many mechanisms at once as foundations for change, and it is therefore unknown which mechanisms uniquely contribute to improvements in math performance when other mechanisms are also taken into account. For example, fluency training may also have reduced mathematics anxiety, which was actually responsible for the observed transfer. Additionally, most studies have only examined changes in procedural math performance, not complex math performance, making it unclear whether different levels of math are supported by different mechanisms.

Second, while several studies have looked at training effects in both high- and low math-achieving children, adult studies have primarily involved undergraduates from highly-ranked colleges and universities. Given the high number of adults who struggle with math and may not have developed the appropriate math resources during prior learning experiences, it is important to determine whether training effects generalize in adults regardless of initial math skill level and if they do not, which characteristics determine whether participants benefit from training or not. Therefore, it is important to examine training effects in at-risk adults and to generally examine what predicts which students are likely to benefit from training intervention. This issue is generally consistent with the broad recent calls for more personalized learning interventions.

In the current study, we address these questions by investigating changes in procedural and complex math after an estimation training task. The training task was previously found to be successful in producing transfer across multiple studies by Liu et al. (2015); this was one of the few prior training tasks that found improvements in complex math performance. We examine pre- and post-training changes in terms of the same set of mechanisms examined in Study 1:

symbolic integration accuracy, integration precision, integration strength, symbolic and non-symbolic representational acuity, working memory, math anxiety, and math confidence. As with Study 1, few training studies have looked at changes in different aspects of symbolic integration, so the current study also provides more detail about how these integration aspects uniquely contribute to math performance. Further, the study involved the low and mid skill level participants from Study 1, who were generally lower in math than have been studied in the past, and they also showed differential use of cognitive resources for math. By comparing participants who complete the training against control participants who only complete the pre- and post-tests, we aim to answer three main questions:

1) Which participants make progress during training, and what characteristics distinguish those who make progress from those who do not?
2) Does exposure to training lead to pre-post changes in any math outcomes or mechanism measures beyond what is seen in the Control condition?
3) Within the Training group, which changes in mechanism measures explain positive change in procedural and applied math?

3.3 METHOD

3.3.1 Participants

Seventy-eight adults (who also participated in Study1) participated in the study. Participants were recruited by advertising in local colleges and universities in the metropolitan Pittsburgh area and in the University of Pittsburgh's Pitt+Me research participant registry. Thirty-seven participants were randomly assigned to the Training condition (age *M* = 24.2, *SD* = 5.7, range = 18-44; 22 female, 11 male, 1 non-binary, 3 unreported; verbal SAT *M* = 587, *SD* = 88, 20 unreported; 23 White, 3 Asian, 7 Black/African-American, 1 Black & Pacific Islander, 2 Hispanic/Latinx, 3 unreported; 32 with at least some post-secondary education, 3 unreported), and 41 participants were randomly assigned to the Control condition (age *M* = 25.1, *SD* = 5.7, range = 18-35; 28 female, 11 male, 2 unreported; verbal SAT *M* = 591, *SD* = 68, 22 unreported; 29 White, 3 Asian, 7 Black/African-American, 1 Hispanic/Latinx, 2 unreported; 37 with at least some post-secondary education, 2 unreported). Participants were required to have a quantitative SAT or ACT score below the 80th national percentile rank (620 on the SAT, 25 on the ACT) and to major in non-quantitative majors to prevent ceiling effects. Participants were compensated US$10 for each session, plus a bonus of US$80 for completing the full study; Training participants were able to earn a maximum of US$150 across seven sessions, while Control participants were able to earn a maximum of US$100 across two sessions. All participants gave informed verbal consent before participating in the study. Two Training participants were excluded from analyses for attending fewer than three training sessions, and another two participants (1 Training, 1 Control) were excluded because they showed low effort on several tasks.

3.3.2 Materials

3.3.2.1 Estimation training task.

During each training session, participants completed five sets of double-digit addition problems and five sets of double-digit subtraction problems, with 40 problems per set. Problems were generated by randomly choosing two double-digit operands from the range of 11 to 88 (excluding decades, e.g., 20, 30, 40), allowing over 10,000 possible combinations for problems that were very unlikely to be repeated, ensuring that participants were not simply memorizing answers. Operands in the problem were shown sequentially. The first operand was shown for 0.5 s, followed by an operation symbol (+ or –) for 0.25 s, followed by the second operand for 0.5 s. Participants then saw a blank screen and had 2 s to type in their response to the problem on a computer keyboard. The sequential presentation and brief response time were meant to encourage the use of holistic number representations over exact counting or calculations that were focused on individual digits.

To further encourage holistic number representations, participants were only required to answer with approximate answers. In the beginning of training (difficulty level 1), they were required to answer within five of the exact answer (e.g., if the exact solution was '100', then any response from 95 to 105 would be correct). If participants achieved 70% or higher accuracy on a set of problems, then the difficultly level increased by one level. Each difficulty level decreased the accepted window of correct responses by one, down to a window of answering within one of the exact answer at difficulty level 5 (e.g., if the exact solution was '100', then any response from 99 to 101 would be correct). The decreasing estimation window was designed to support the refinement of participants' number representations, as more exact estimates were needed as difficulty went up. Further, this approach to estimation-based training allowed the training program to adapt to any heterogeneity in participants' starting mathematical abilities, allowing all participants to experience difficult training, which is thought to maximize learning effects, while still staying within the targeted double-digit operands. Per-trial feedback was given in the form of green checkmarks for correct responses (with more checkmarks indicating a response closer to the exact answer, up to a maximum of six checkmarks) and a red 'X' for incorrect responses to promote quick improvement in the accuracy of their estimates. At the end of each set, participants were also shown their percent accuracy for the block to provide a benchmark for participants and to motivate them to improve their scores.

3.3.2.2 Applied Problems test.

The *Applied Problems* sub-test (pre-post correlation = .78) was used to measure change in complex math performance. It was identical to the sub-test used in Study 1. Participants were given one of two forms of the test during the pre-test session and the other form during the post-test session. The form order was randomly assigned in a counterbalanced order across participants so that approximately half of the participants received Form A at pre-test and the other half received form B; specifically, at pre-test, 16 Training participants received Form A, 20 Training participants received Form B, 21 Control participants received Form A, and 19 Control participants received Form B. Accuracy scores at pre-test and post-test more than two standard deviations from the group average were winsorized (1 participant at pre, 3 participants at post).

3.3.2.3 Math Facts test.

The *Math Facts* test (pre-post correlation = .76), administered at pre-test and post-test to measure changes in procedural math performance, was identical to the test used in Study 1. Addition and subtraction problems were randomly generated at both pre-test and post-test from numbers ranging from 2 to 99. As with Study 1 analyses, only double-digit carry addition and subtraction trials were used because of ceiling effects in the single-digit and non-carry trials, and trials with response times faster than 100 ms were also removed to exclude trials in which participants had not fully processed the stimuli. Scores that were more than two standard deviations away from the average were winsorized (3 participants at pre, 1 participant at post).

3.3.2.4 Dot Estimation task.

The Dot Estimation task was used to measure change in *Integration Accuracy* (pre-post correlation = .80) and *Integration Precision* (pre-post correlation = .67), and to adjust for participants' systematic biases in their estimates of non-symbolic quantities. The task was identical to that used in Study 1.

3.3.2.5 Number Decision task.

The Number Decision task was used to measure *Integration Strength* (pre-post correlation = .20) and was identical to the task used in Study 1. As a reminder, only mismatch trials with quantities at near ratios to each other (between 1.15x to 1.35x in ratio) were included in analyses. Integration strength was measured as the absolute difference in accuracy between match trials and these near mismatch trials, with a larger difference indicating stronger symbolic integration. The difference score was standardized and averaged with participants' standardized symbolic integration score from the Number Comparison task to calculate a combined integration score at pre-test and post-test. Average match or near mismatch accuracies more than two standard deviations away from the group average were winsorized before calculating their symbolic integration strength (6 participants for match accuracy and 4 participants for near mismatch accuracy at pre, 7 participants for match accuracy and 4 participants for near mismatch accuracy at post).

3.3.2.6 Number Comparison task.

The Number Comparison task, used to measure change in Integration Strength, Symbolic Representational Acuity, and Non-Symbolic Representational Acuity, was also identical to the task used in Study 1. Accuracy scores and response times for each trial type greater than two standard deviations from the group average were winsorized (4 participants for symbolic trial accuracy, 4 participants for symbolic trial response time, 3 participants for non-symbolic trial accuracy, 6 participants for non-symbolic trial response time, and 4 participants for mixed trial response time at pre; 4 participants for symbolic trial accuracy, 4 participants for non-symbolic trial accuracy, 4 participants for symbolic trial response time, 6 participants for non-symbolic response time). Only response times on correct trials were used in analyses. As with Study 1, *Integration Strength* (pre-post correlation = .29) was measured as the response time difference between mixed trials (comparing one symbolic and one non-symbolic quantity) and the non-symbolic trials (comparing two non-symbolic quantity), after adjusting for bias in non-symbolic quantity estimates with the Dot Estimation results; having a faster response time on mixed trials compared to non-symbolic trials indicated stronger symbolic integration strength compared to having a faster response time on non-symbolic trials compared to mixed trials. Again, the difference score was standardized and averaged with the symbolic integration score from the Number Decision task to calculate a combined integration strength score at pre-test and post-test. Because, similar to Study 1, the Number Decision and Number Comparison integration scores were uncorrelated ($r(63) = .003$), we looked at both the combined integration strength score as well as each individual measure in analyses and report whether results changed depending on the integration strength measure used. *Symbolic Representational Acuity* (pre-post correlation = .64) and *Non-Symbolic Representational Acuity* (pre-post correlation = .37) were calculated by standardizing mean accuracy (removing trials with response times under 100 ms) and mean response times on correct trials on symbolic and non-symbolic comparison trials, respectively, and then averaging the standardized accuracy and response time.

3.3.2.7 Forward and Backward Digit Span task.

The forward and backward digit span tasks were identical to those used in Study 1 to measure individual differences in participants' verbal *Working Memory Capacity*. As a reminder, participants who could not repeat two consecutive three-digit sequences were automatically given a working memory score of three. The forward and backward digit spans were summed to represent participants' overall working memory capacity (combined score pre-post correlation = .53). Combined scores more than two standard deviations away from the group average were winsorized (1 participant at pre, 4 participants at post).

3.3.2.8 Abbreviated Mathematics Anxiety Rating Scale (A-MARS).

The A-MARS was used to measure change in reported *Math Anxiety* (pre-post correlation = .89) and as an individual difference measure. The task was identical to that used in Study 1. Scores of participants with accuracies more than two standard deviations from the group average were winsorized (2 participants at pre, 4 participants at post).

3.3.2.9 Demographics questionnaire.

Participants completed a demographics questionnaire at the post-test session. The questionnaire asked participants to rate their confidence with math (by rating their skill level relative to 100 people of their age). They were also asked to provide their quantitative scores on the SAT/ACT, gender, age, race and ethnicity (with White and Asian defined as being a non-minority in math contexts, and Black/African-American, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, and Hispanic/Latinx defined as being a minority in math contexts), school and education level, and number of years since their last math class.

3.3.3 Procedure

All participants completed a pre-test and a post-test session. During these pre-test sessions, all participants took the Applied Problems test and the Forward and Backward Digit Span tasks verbally with the experimenter, and then the Dot Estimation task, Math Facts task, the Number Decision task, and the Number Comparison task on the computer. Participants ended the session by completing the Abbreviated Mathematics Anxiety Rating Scale. Participants in only the Training condition also participated in three to five 1 h Estimation Task training sessions between the pre-test and post-test sessions (with the number of training sessions dependent on participants' scheduling constraints: 28 participants completed all five days of training, 5 completed four days, and 2 completed 3 days). During the post-test session, all participants repeated the tasks from the pre-test session and also filled out the demographics questionnaire at the end of the session.

3.4 RESULTS

3.4.1 Did Participants Make Progress During Training?

To determine whether participants were able to make progress within the Estimation Training task regardless of movement in training difficulty levels, Training participants' performance on the first day of training was compared to their performance on the last day of training, separately for addition and subtraction problems. Specifically, we calculated a distance score as the absolute difference between participants' response and the exact correct response for each trial, and then compared the median distance on the first day of training against the median distance on the last day of training. Repeated-measure ANOVAs were run separately on the median distance scores for addition and subtraction, using Training Day (first, last) as a within-subjects factor. The median distances significantly decreased with large effect sizes for both addition ($F(1, 33) = 34.7$, $p < .001$, $d = .89$) and subtraction ($F(1, 32) = 34.5$, $p < .001$, $d = .61$) trials. On average, the distance for addition trials decreased by approximately 2.0, and the distance for subtraction problems decreased by approximately 0.8.

Changes in median distance also did not depend upon participants' initial math skill levels before starting training. As in Study 1, participants were categorized as Mid- or Low-Skill based on whether they scored between the 50^{th} and 80^{th} national percentile on the SAT or ACT (Mid-Skill) or under the 50^{th} percentile (Low-Skill). We then re-ran the addition and subtraction repeated-measures ANOVAs, using Training Day (first, last) as a within-subject measure and Skill (Mid, Low) as a between-subjects measure. There were no significant Training Day X Skill interactions for either addition ($F(1, 32) = 0.11$, $p = .74$, $\eta^2_p = .004$) or subtraction ($F(1, 31) = 3.20$, $p = .08$, $\eta^2_p = .09$), showing that mean amount of improvement did not differ between the Mid- and Low-Skill groups.

We also investigated progress on the training task based on the difficulty levels reached by participants. In other words, progress was measured as whether participants made sufficient progress in overall accuracy on the training task to be moved up to harder difficulties on the training task. Table 8 shows the percentage of participants at each difficulty level (levels 1-5) at the end of each training day. Most participants increased in difficulty level across days, though there were still a substantial number of participants who were unable to reach higher levels of difficulty even after five days. At the same time, a few participants were able to attain level 3 even within the first training session, though they still had additional levels of growth to attain in subsequent training days. Thus, the adaptive procedure was able to make the task doable and challenging for the great majority of the participants throughout training.

Table 8. The percentage of participants at each difficulty level at the end of each training day, separated by addition and subtraction (modal level within each day shown in gray)

Addition		Day 1	Day 2	Day 3	Day 4	Day 5
	Level 1	76%	57%	38%	32%	21%
	Level 2	18%	23%	32%	29%	25%
	Level 3+	6%	20%	29%	39%	54%
Subtraction						
	Level 1	70%	37%	21%	19%	15%
	Level 2	10%	31%	35%	26%	15%
	Level 3+	20%	31%	44%	55%	70%

3.4.2 Which Participants Made Progress During Training?

To look more closely at the characteristics that distinguish people who made progress on the training program versus those who did not, we categorized 16 participants who reached at least difficulty level 3 on both addition and subtraction as Progress participants, and 18 participants who were not able to reach these difficulty levels as Little Progress participants. Note that both subgroups completed the same number of days of training, so the difference is not one of opportunity. Within participants who completed the full five days of training, 13 were categorized as Progress, while 14 were categorized as Little Progress. Similarly, within participants who only completed three or four days of training, 3 were categorized as Progress and 4 were categorized as Little Progress. Difficulty level progress also did not depend upon initial math skill level. Using the Mid-Skill and Low-Skill categories described above, a chi-square test was run using Skill (Mid, Low) and Progress (Progress, Little Progress) as variables.

The relation between these two variables was not significant (χ^2(1, N = 34) = 1.23, p = .32), showing that general skill level did not determine progress in difficulty level.

Multiple linear regressions were run using the cognitive and individual difference measures at pre-test (pre Integration Accuracy, pre Integration Precision, pre Integration Strength, pre Symbolic Acuity, pre Non-Symbolic Acuity, pre Working Memory, pre Math Anxiety, and pre Math Confidence) to predict difficulty level progress on the Estimation Training task. Because of the high number of predictors for the number of data points available, a forward and backward regression were run on the data (with an entry probability of .05 and removal probability of .10), and then significant predictors from these two models were entered into a forced regression model. Table 9 lists the standardized betas of the significant or marginal regression variables. In the forward regression (R^2 = .23, F(1, 24) = 7.15, p = .01), only higher Working Memory emerged as a significant predictor of progress. Meanwhile, the backward regression (R^2 = .42, F(3, 22) = 5.27, p = .01) ended with higher Working Memory, marginally weaker Integration Strength, and marginally higher Math Confidence as predictors of training progress. Finally, the forced model that included Integration Strength, Working Memory, and Math Confidence (R^2 = .33, F(3, 24) = 3.98, p = .02) found that Integration Strength and Working Memory significantly predicted training progress. However, in the regressions using only the Number Decision integration strength score or the Number Comparison integration strength score, only Working Memory showed as a significant predictor. Thus, working memory capacity appears to be the primary determinant on whether participants will be able to make progress on the Estimation Training, though Integration Strength and Math Confidence may also contribute to a lesser extent.

Table 9. Standardized betas of each pre-test variable predicting progress on the Estimation Training task in the Forward, Backward, and Forced regression models

	Forward	Backward	Forced
R^2	.23	.42	.31
Integration Strength	-	-.35†	-.34*
Working Memory	.48*	.55**	.48**
Math Confidence	-	.31†	.28†

Note. ** p < .01, * p < .05, † p < .10

3.4.3 Does Exposure to Training Lead to Pre-Post Changes in Math Outcomes or Mechanisms?

Although there was considerable variation in the amount of progress made on the Estimation Training task, we wanted to examine whether simple exposure to the training task led to significant pre-post changes in any math outcome or mechanism measure, beyond any changes seen in the Control condition. Table 10 shows each measure and the means and standard deviations at pre- and post-test for the Training and Control groups.

Table 10. Pre- and post-test means (and standard deviations) of the Training and Control condition for each measure

Group/ Time	Applied Problems	Math Facts	Int. Accuracy	Int. Precision	Int. Strength	Symbolic Acuity	Non-Symbolic Acuity	Working Memory	Math Anxiety
Training									
Pre	49.5 (6.0)	.64 (.28)	.75 (.13)	.20 (.04)	-.25 (.49)	-.06 (.77)	.16 (.67)	9.6 (1.2)	71.3 (15.7)
Post	49.5 (5.6)	.66 (.26)	.73 (.13)	.22 (.08)	.11 (1.02)	-.09 (.57)	-.05 (.73)	9.9 (1.4)	60.6 (15.6)
Control									
Pre	48.7 (4.7)	.61 (.27)	.75 (.13)	.21 (.05)	.02 (.83)	.08 (.61)	.04 (.59)	9.8 (1.4)	62.6 (17.5)
Post	50.2 (4.9)	.61 (.27)	.74 (.13)	.21 (.04)	-.06 (.80)	.19 (.61)	.09 (.69)	10.0 (1.5)	57.4 (19.6)

A repeated-measures ANOVA that included all math outcome and mechanism measures (Applied Problems, Math Facts, Integration Accuracy, Integration Precision, Integration

Strength, Symbolic Acuity, Non-Symbolic Acuity, Working Memory, Math Anxiety) was run using Time (pre, post) as a within-subjects factor and Condition (Training, Control) as a between-subjects factor. Of the nine measures, only Math Anxiety showed a significant Time X Condition interaction ($F(1, 51) = 6.02$, $p = .02$, $\eta^2_p = .11$). Both the Training condition ($F(1, 21) = 25.3$, $p < .001$, $\eta^2_p = .55$) and Control condition ($F(1, 30) = 22.0$, $p < .001$, $\eta^2_p = .42$) significantly decreased in Math Anxiety from pre-test to post-test, though the Training condition dropped twice as much. On average, the Training condition's Math Anxiety dropped by 11 points, while the Control condition's Math Anxiety dropped by only 5 points. For reference, 5 points on the Math Anxiety scale is equivalent to the difference between feeling very anxious during a math situation and feeling not at all nervous during that situation (though, because Math Anxiety is a sum score across 25 questions, it may also be smaller decreases in anxiety across several math situations). While simple exposure to the training task does not significantly improve either math outcome, these results suggest that the training task can still benefit participants by decreasing their math anxiety, regardless of the progress they make on the training task.

3.4.4 Which Mechanism Changes Explain Positive Changes in Procedural Math or Applied Math Performance?

Given that the training program did not generally improve either procedural math or complex math, we also examined which underlying mechanisms explain cases of positive change within each of the two math outcomes. Residual scores were used rather than difference scores, which suffer from regression-to-the-mean artifacts; for each measure, a regression was run using the pre-test score of that measure to predict the post-test score of that measure. The residuals from those regressions for each math outcome measure were then used as the dependent variables for separate forward and backward regressions (with an entry probability of .05 and removal probability of .10) with all of the residuals of each mechanism measures as well as Math Confidence and Training Progress (the Progress vs. Little Progress categorization, based on difficultly level) as possible predictors. Finally, the significant predictors from these regressions were entered into a forced regression. Note that the number of participants included in each regression model sometimes differed because some participants were missing data on some measures and were therefore removed listwise from the forward and backward regressions, but not necessarily from the forced regressions. Table 11 shows the intercorrelations between the residuals entered into the multiple regressions. Figure 5 summarizes the predictors for changes in Math Facts and Applied Problems within the Training condition.

Table 11. Pearson correlation matrix for all residuals used for the Math Facts and Applied Problems multiple regressions within the Training condition

	1	2	3	4	5	6	7	8	9	10
1. Applied problems	–									
2. Math facts	.21	–								
3. Integration accuracy	-.54**	-.09	–							
4. Integration precision	-.37*	-.13	.54**	–						
5. Integration strength	.10	.26	-.08	.41*	–					
6. Symbolic acuity	-.08	-.15	-.07	-.03	-.23	–				
7. Non-symbolic acuity	.30	.11	-.17	-.48**	-.36*	.45**	–			
8. Working memory	-.03	.27	-.14	-.002	.03	.46*	.40*	–		
9. Math anxiety	-.29	.24	.33	-.19	-.37†	.05	.16	.20	–	
10. Math confidence	.19	.03	-.18	.004	.16	-.19	.06	.11	-.34	–
11. Training progress	.28	.33†	-.19	-.19	.07	.08	.21	.23	.23	.40*

Note. ** $p < .01$, * $p < .05$, † $p < .10$

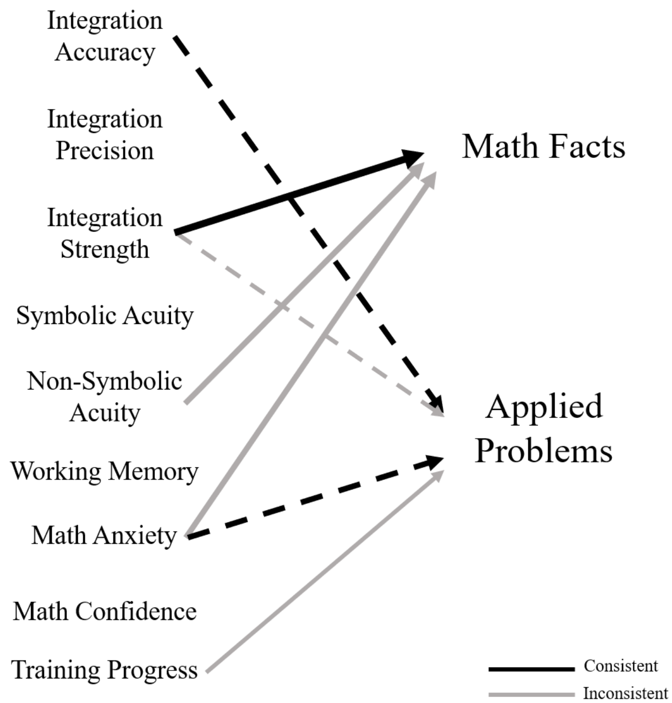

Figure 5. Multiple regressions connecting underlying mechanism changes to changes in Math Facts and Applied Problems in the Training condition. Line width represents strength of each predictor. Black lines denote consistently found predictors, light gray lines denote inconsistently found predictors, solid lines denote positive predictors, and dotted lines denote negative predictors

3.4.4.1 Explaining change in Math Facts performance.

Table 12 lists the standardized betas for each variable in the regressions predicting growth in Math Facts accuracy. No predictor reached the entry probability of .05 for the forward model. If the entry probability threshold is slightly raised to .1, then Training Progress becomes a marginal predictor. The backward model (R^2 = .43, $F(3, 17) = 4.26$, p = .02) ended with changes in Integration Strength, Non-Symbolic Acuity, and Math Anxiety as significant predictors of changes in Math Facts. When Integration Strength, Non-Symbolic Acuity, and Math Anxiety are entered into a forced regression (R^2 = .24, $F(3, 19) = 1.96$, p = .15), only Integration Strength remains as a marginal predictor of Math Facts change (though when the Number Decision integration strength score is used instead, then Symbolic Acuity and Working Memory emerge as the only marginal predictors, while the Number Comparison integration strength score reveals no significant predictors). Together, these results suggest that increases in Integration Strength, increases in Non-Symbolic Acuity, or, oddly, an increase in Math Anxiety, may predict positive changes in procedural math. However, these results were inconsistent across models, appearing to depend heavily on the integration strength measure used, and no single variable met the forward regression entry threshold, suggesting that no individual mechanism impacted procedural math performance much. Instead, changes in a combination of factors may be necessary for improvement to also be seen in Math Facts performance.

Table 12. Multiple regression standardized betas predicting Math Facts accuracy with a forward regression, backward regression, and forced regression

Variables	Forward	Backward	Forced
R^2	-	.43	.24
Residuals			
Integration Accuracy	-	-	-
Integration Precision	-	-	-
Integration Strength	-	.80**	.54†
Symbolic Acuity	-	-	-
Non-Symbolic Acuity	-	.62*	.37ns
Working Memory	-	-	-
Math Anxiety	-	.51*	.36ns
Covariates			
Math Confidence	-	-	-
Training Progress	-	-	-

Note. ** p < .01, * p < .05, † p < .10, ns = not significant

3.4.4.2 Explaining change in Applied Problems performance.

Table 13 shows the standardized betas of each variable that predicts growth in Applied Problems accuracy. The forward regression (R^2 = .26, $F(1, 20)$ = 7.09, p = .02) showed that decreases in Integration Accuracy predicted positive change in Applied Problems accuracy. The backward regression (R^2 = .60, $F(4, 17)$ = 6.32, p = .003) similarly found that decreases in Integration Accuracy predicted Applied Problems improvement, along with decreasing Integration Strength, decreasing Math Anxiety, and more Training Progress. However, only decreasing Integration Accuracy and decreasing Math Anxiety remain as marginal predictors when entered into the forced regression, while Integration Strength and Training Progress are no longer significant (R^2 = .39, $F(4, 19)$ = 3.09, p = .04). Additionally, using the Number Decision integration strength measure leads to only Non-Symbolic Acuity showing as a significant predictor, while the Number Comparison integration strength measure leads to only decreases in Integration Accuracy as a marginal predictor. Thus, decreases in aspects of symbolic integration (most consistently in accuracy, though strength may also contribute) or decreases in math anxiety may explain positive changes in complex applied math. Training progress may also play a role, though it again appears inconsistently across models compared to other variables.

Table 13. Multiple regression standardized betas predicting Applied Problems accuracy with a forward regression, backward regression, and forced regression

Variables	Forward	Backward	Forced
R^2	.26	.60	.39
Residuals			
Integration Accuracy	-.51*	-.56**	-.36†
Integration Precision	-	-	-
Integration Strength	-	-.44*	-.28ns
Symbolic Acuity	-	-	-
Non-Symbolic Acuity	-	-	-
Working Memory	-	-	-
Math Anxiety	-	-.60**	-.39†
Covariates			
Math Confidence	-	-	-
Training Progress	-	.55**	.32ns

Note. ** p < .01, * p < .05, † p < .10, ns = not significant

3.5 DISCUSSION

The current study examined the characteristics that distinguished participants who were able to make progress on our Estimation Training task, the pre-post changes in math outcomes and underlying mechanisms after training, and the changes in mechanisms that explained positive changes in procedural and applied math. As a brief summary, we found that higher working memory capacity was the primary predictor of progress in the Estimation Training task. Regardless of training progress, the Training condition showed significantly decreased math anxiety levels, and the decrease was approximately twice as much as the decrease seen in the Control condition. However, no overall improvements were found in either procedural or applied math performance. Increases in integration strength appeared to be the most consistent predictor of improvements in procedural math, while decreases in integration accuracy and math anxiety explained improvements in complex math performance. We now discuss each of these findings.

Working memory has generally been associated with better task performance in prior studies (e.g., Bull et al., 2008; McClelland et al., 2006; St Clair-Thompson & Gathercole, 2006), including math (Clark et al., 2010). Unfortunately, our findings that working memory predicts training progress suggest that those participants who would likely already perform better on math are also more likely to do well during training. If our finding generalizes to other training interventions, then this means that any achievement gaps based on working memory are further exacerbated by interventions being more effective for individuals with higher working memory, leaving those with lower working memory even farther behind. However, it is possible that working memory is particularly important for our training program because it required participants to hold double-digit quantities for computation in working memory. Future studies can test whether working memory also predicts progress in other training tasks that do not have as heavy working memory requirements, or if other factors can also contribute to training progress.

Interestingly, our study did not replicate the complex math or symbolic acuity improvements previously found with the Estimation Training task (Kallai et al., 2011; Liu, Kallai, et al., 2015). The symbolic acuity improvements were already somewhat inconsistent in prior studies, as some studies found improvements in symbolic comparison task accuracy, while others showed improvements in symbolic comparison response times. Because prior studies only included non-symbolic acuity, symbolic acuity, and math fact retrieval as other measures, it is possible that the changes in symbolic number comparison were not actually indicative of representational acuity changes. Instead, the training may have been improving integration between symbolic and non-symbolic representations, or improving symbolic-symbolic relationships separate from the ANS. In the current study, changes in symbolic integration strength and accuracy appeared to explain changes in procedural and applied math, respectively, lending some evidence that symbolic integration and symbolic acuity were conflated in prior studies.

This also suggests that the Estimation Training may not be refining symbolic or non-symbolic representations, as was previously hypothesized; instead, the training may be helping participants to divorce symbolic and non-symbolic number representations and rely primarily on symbolic strategies. There is some evidence for this explanation based on the limited strategy data collected in the current study, as participants were less likely to use strategies that involved the visualization of number lines on the last day of training compared to the first day of training. It would be worthwhile for future studies to investigate strategy changes after the Estimation Training in more detail. In addition, the inclusion of a symbolic ordinality measure, which focuses on symbol-symbol relationships, may also provide insight into whether people are learning to utilize their symbolic representations more effectively after training.

The greater math skill variance in the current sample may also be a cause of the lack of training results. Both Kallai's study (Kallai et al., 2011) and Liu's study (Liu, Kallai, et al., 2015) included participants whose SAT scores placed them within the 75^{th} to 93^{rd} percentile of college bound seniors. Meanwhile, the current study was restricted to participants in the 80^{th} percentile or below, with almost 40% of participants being below the 50^{th} percentile. Indeed, several adjustments had to be made to the training task to accommodate the lower math skill of the current sample, such as lowering the accuracy threshold needed to move up in difficulty level during the training task (from 90% accuracy in prior studies to 70% in the current study). It is possible that many of these lower-skilled participants did not have the requisite declarative knowledge to succeed on the complex math test (e.g., formulas needed for more difficult geometry problems on the Applied Problems test), so refining underlying cognitive mechanisms with the training would have little effect on their complex math performance.

Although our study does not match prior improvements in complex math or symbolic representational acuity, it does show promise for decreasing math anxiety and adds to existing literature on math anxiety interventions. It is unlikely that the decreases in math anxiety were caused by any changes in number representations, given that there were no training effects on symbolic integration or representational acuity and that the Control condition also decreased in math anxiety to a lesser extent. Rather, it may be the additional experience with math in a relatively casual context that led to lower math anxiety, based on other studies that have found significant decreases in math anxiety with greater exposure to math (Jansen et al., 2013; Supekar, Iuculano, Chen, & Menon, 2015). The feedback provided after each trial may have also contributed to changes in math anxiety. Jansen and colleagues (2013) found that elementary school children experienced less math anxiety, reported higher math competence, and improved their math performance when they were given positive feedback (regardless of actual performance) during a math task. Change in math anxiety also did not appear to depend upon the amount of positive feedback given, though the lowest amount of positive feedback given in their study was 60% correct. Thus, in the current study, it is possible that both exposure to math and seeing positive feedback for correct responses (in our study, the green checkmarks) helped participants feel less negatively about math. One question is how long this math anxiety decrease lasts after training. If longer lasting, then it is plausible that the training intervention could indirectly influence math ability by making participants feel more confident about their math abilities, encouraging them to participate in more math experiences in the future (and several participants mentioned a desire to continue refreshing their math skills after the study). While even transient decreases in math anxiety may be beneficial for math performance by alleviating working-memory burdens caused by anxious ruminations, the ideal would be to permanently decrease math anxiety to support math learning over longer periods of time, and to determine the kinds of math anxiety interventions that are most effective.

Even though the Training condition did not show significant improvements across all participants in either math outcome, we were still able to determine that the more selectively obtained increases in integration strength explained improvements in procedural math, while decreases in integration accuracy and math anxiety explained improvements in applied math, though the results were somewhat inconsistent across models. This indicates a potential shift in cognitive resources, in which simpler procedural math improvements rely more on non-symbolic representations (shown by improvements in symbolic integration strength and, weakly, non-symbolic acuity explaining procedural math improvements), while complex math improvements are benefitted by a lower reliance on non-symbolic representations (shown by decreases in integration accuracy and strength explaining complex math improvements). Thus, similarly to findings in Study 1, the ability to shift away from integration of symbolic and non-symbolic representations and relying primarily on symbolic representations depending on the type of math being done may be key to improving math performance.

3.5.1 Limitations of the Current Study

There are some limitations in the current study that future studies should address. First, as with Study 1, the two symbolic integration strength measures showed virtually no correlation with each other. However, the strength measure used for analysis caused considerably more divergent results than were seen in Study 1. This again brings up the question as to whether the two integration strength measures are gauging two separate aspects of symbolic integration strength, and what those separate aspects might be (e.g., deliberate vs. automatic integration). It is clear that whatever these aspects are can have significant implications for findings, and it is therefore greatly important for studies to investigate integration strength more closely in the future.

Second, we found that no mechanism measures met the threshold for the forward regression model predicting Math Facts performance, though some variables were able to enter the model when the threshold was slightly decreased. This suggests that multiple changes must co-occur to produce Math Facts improvements or that change in the mechanism measures included in the current study simply do not predict a large amount of the variance in Math Facts improvement. It is possible that we are missing other mechanisms that could be significant predictors of procedural math performance, such as symbolic or non-symbolic ordinality. Future studies should include more mechanisms that could contribute to math performance improvements to further refine understanding of the unique mechanisms that underlie math improvements.

3.5.2 Conclusions

The current study showed that the complex math improvements found previously from the Estimation Training task do not hold for more at risk populations. Further, it refines our understanding of the characteristics that contribute to training task progress (i.e., why the findings did not extend to this population), it added a new benefit (i.e., that it alleviated math anxiety), and it provided new insights into explaining growth in procedural and complex math performance. Specifically, working memory is a significant predictor of training progress, exposure to math and positive feedback may significantly decrease math anxiety, and the ability to shift between non-symbolic and symbolic math resources may be important to improving performance on procedural and applied math, respectively. Understanding more about the mechanisms underlying procedural and applied math performance improvements, as well as potential barriers against training progress, can allow us to greatly improve the effectiveness of training studies for struggling adult learners in the future.

4.0 GENERAL DISCUSSION

The current studies investigated the cognitive and individual difference resources that underlie procedural and applied math performance in two ways. Study 1 examined differences in underlying math resources based on people's initial skill level. Study 2 evaluated the effectiveness of an estimation training task, including the predictors of training progress, and looked at the mechanism changes that explained improvements in procedural and applied math.

Across Study 1 and Study 2, there was consistent evidence of a shift in cognitive resources depending on the complexity of math being done. In Study 1, the primary distinction found between the Mid- and Low-Skill groups was in the cognitive resources used for complex math performance. While both skill groups relied on similar resources of integration precision and working memory for procedural math, divergence was seen in complex math. The Low-Skill group continued to use the same integration and working memory resources for complex math that they had used for procedural math, while performance in the Mid-Skill group was now predicted by weaker symbolic integration. Study 2 followed a similar pattern, such that improvements in procedural math were explained by increases in symbolic integration resources and non-symbolic acuity, while improvements in complex math were explained by decreases in symbolic integration resources. It appears that simpler procedural math relies more upon resources that involve non-symbolic representations in some way, whether directly through non-symbolic representational acuity or through the mapping between symbolic and non-symbolic representations. In contrast, a greater separation between symbolic and non-symbolic representations, shown through weaker symbolic integration or less accurate mapping between representations, predicts better performance in complex math.

The procedural math results are consistent with prior findings relating stronger symbolic integration with better math performance (e.g., Jang & Cho, 2018; Wilson et al., 2009; Wong et al., 2016). Symbolic integration also significantly contributed to math performance, beyond any contributions of symbolic or non-symbolic representational acuity, matching studies that showed symbolic integration fully mediating the non-symbolic acuity and math performance relationship. Thus, the mapping between symbolic and non-symbolic representations appears to be more important for math performance than the acuity of either individual representation type. This also provides additional evidence for representational overlap theories of symbolic integration, in which symbolic and non-symbolic number representational systems are connected and numeric symbols are given meaning through their mapping onto non-symbolic representations.

However, our results showing that *weaker* symbolic integration predicts complex math performance suggests that the relationship between symbolic integration and math performance may depend upon the type of math being done. During complex math, it may be more helpful to think of symbolic numbers in terms of other symbolic numbers, rather than relate it back to non-symbolic quantities. This is similar to the alternative theory of symbolic integration that states that symbolic and non-symbolic number systems may act independently (Bulthe et al., 2014; Mathieu Le Corre & Susan Carey, 2007; Lyons et al., 2012), such that symbolic math relies upon symbol-symbol relationships rather than non-symbolic representations. Further, the fact that weaker integration predicts better performance suggests that integration may not only be unnecessary for complex math performance, but actually detrimental as well.

Typically, the theories of representational overlap and symbolic estrangement have been presented as mutually exclusive; that is, symbolic and non-symbolic representations are either closely related, or not at all. The relationship between symbolic and non-symbolic representations may instead be situational; that is, the extent to which non-symbolic representations support symbolic math depends upon the type of math being performed. When performing or learning simpler math, then having a concrete grounding of the number's quantity may be particularly helpful for processing and manipulating numbers. In contrast, when performing problems that require more complex or abstract operations or larger quantities, then this concrete foundation is no longer helpful and may distract from the task at hand. To fully understand this effect, future studies should determine how to quantify this representational shifting ability and whether it can explain either procedural or complex math abilities.

Examining the role of other mechanisms that were not included in the present studies, such as symbolic ordinality and inhibitory control, may also be another step toward understanding this representational shift. Symbolic ordinality is thought to be a measure of symbolic-symbolic relations (e.g., Goffin & Ansari, 2016; Lyons & Beilock, 2011; Lyons & Beilock, 2013; Lyons et al., 2016), so understanding its explanatory power for math performance may reveal whether higher-skilled individuals are indeed relying primarily on symbolic representations rather than mapping onto non-symbolic representations during more complex math. The association between inhibitory control and math performance (Clark et al., 2013; Fuhs & McNeil, 2013; Gilmore et al., 2013) may also be another indicator of this representational shift, if inhibitory control is measuring in part the process of stopping automatic integration of symbolic and non-symbolic representations so that one can focus on symbolic-symbolic relations between numbers instead.

The differences seen between the Mid- and Low-Skill groups also imply that the Low-Skill group may just be lagging behind in their development of certain resources, such as the ability to rely primarily on symbolic representations for problem solving. Given their heavier reliance on non-symbolic representations for both procedural and complex math performance, future studies should investigate whether it is better to train lower math-skilled individuals to divorce their symbolic and non-symbolic representations, or to continue fostering the non-symbolic representations that they already use instead. It is possible that there is a performance ceiling when relying primarily on non-symbolic representations, or even a point where it becomes detrimental (assuming it hurts complex math performance). This also brings up the question of how to train people to be able to rely on non-symbolic representations only in specific situations, as non-symbolic representations still appear to be useful for procedural math.

We also recommend that future studies continue to look in depth at the various aspects that make up the construct of symbolic integration. Our findings add to our understanding of symbolic integration by showing that integration accuracy, precision, and strength should be considered as separate aspects of symbolic integration, as all three showed separate and unique contributions to math performance. However, we again note the limitations of our current combined strength measure, and more research should be done to determine precisely how to measure symbolic integration strength, and whether even more facets of symbolic integration exist that were not included in the current studies. It is clear that integration is more complex than typically measured in the past, and thus important to determine precisely how to measure it and its full impact on math.

Working memory is another factor worth investigating further in relation to math. In Study 1, working memory significantly predicted both procedural and applied math performance regardless of skill level. Further, working memory predicted the amount of progress that participants made on the Estimation Training task. People with lower working memory are already unlikely to perform as well as people with higher working memory (e.g., Bull et al., 2008; Clark et al., 2010; McClelland et al., 2006; St Clair-Thompson & Gathercole, 2006); if working memory is also a predictor of intervention progress and effectiveness, then higher working memory individuals are likely to simply continue improving, while lower working memory individuals continue to be left behind. It is possible that working memory is not a predictor of progress for every training intervention, but our findings still emphasize the importance of examining the factors that influence training progress, and to design around these barriers so that training interventions have more reach and impact, especially for the people who need the intervention most.

The present results contribute to our current understanding of the cognitive resources and individual differences that underlie procedural and complex math performance. Importantly, this work investigated the mechanisms of math performance in a population of adults who struggle with math and are more at risk for dropout. We find evidence of a representational shift that differentiates lower math-skilled people from higher math-skilled people, such that simpler forms of math may be supported more by the use of non-symbolic representations, while more complex math may be supported more by the use of symbolic representations in higher skilled individuals. Specific recommendations from the present results include further investigation of this representational shift and its relation to math ability, as well as ways to train this ability should it be a fruitful method for improving math performance. One should also consider barriers that prevent math training interventions from being effective to ensure that those who struggle most with math can benefit equally from training as those who are already more proficient. These findings provide a positive step toward furthering our ability to increase math competency in adults and to increase their chances for future success.

BIBLIOGRAPHY

Alexander, L., & Martray, C. R. (1989). The development of an abbreviated version of the Mathematics Anxiety Rating Scale. *Measurement and Evaluation in Counseling and Development*, *22*(3), 143-150.

Andersson, U., & Ostergren, R. (2012). Number magnitude processing and basic cognitive functions in children with mathematical learning disabilities. *Learning and Individual Differences*, *22*(6), 701-714. doi:10.1016/j.lindif.2012.05.004

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience, 9*(4), 278-291. doi:10.1038/nrn2334

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181-185.

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *The Journal of Higher Education, 77*(5), 886-924. doi:10.1080/00221546.2006.11778948

Au, J., Jaeggi, S. M., & Buschkuehl, M. (2018). Effects of non-symbolic arithmetic training on symbolic arithmetic and the approximate number system. *Acta psychologica, 185*, 1-12. doi:10.1016/j.actpsy.2018.01.005

Ayres, P. L. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology*, *26*(2), 227-248. doi:10.1006/ceps.2000.1051

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47-89. doi:10.1016/S0079-7421(08)60452-1

Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*(2), 255-270. doi:10.1016/j.econedurev.2009.09.002

Bajic, D., Kwak, J., & Rickard, T. C. (2011). Specificity of learning through memory retrieval practice: The case of addition and subtraction. *Psychonomic Bulletin & Review, 18*(6), 1148-1155. doi:10.3758/s13423-011-0151-4

Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis: Which high schools produce the nation's dropouts? Where are they located? Who attends them?* Retrieved from Baltimore, MD: https://files.eric.ed.gov/fulltext/ED484525.pdf

Balfanz, R., McPartland, J., & Shaw, A. (2002). Reconceptualizing extra help for high school students in a high standards era. *The Journal for Vocational Special Needs Education, 25*(1), 24-41.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*(1), 71-92.

Braham, E. J., & Libertus, M. E. (2018). When approximate number acuity predicts math performance: The moderating role of math anxiety. *PloS one, 13*(5), e0195696. doi:10.1371/journal.pone.0195696

Brankaer, C., Ghesquière, P., & De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PloS one, 9*(4), e93565. doi:10.1371/journal.pone.0093565

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

Bulthe, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage, 15*(87), 311-322. doi:10.1016/j.neuroimage.2013.10.049

Bynner, J., & Parsons, S. (1997). *Does numeracy matter? Evidence from the National Child Development Study on the Impact of Poor Numeracy on Adult Life*. Retrieved from London: https://files.eric.ed.gov/fulltext/ED406585.pdf

Campbell, J. I. D., & Charness, N. (1990). Age-related declines in working-memory skills: Evidence from a complex calculation task. *Developmental Psychology, 26*(6), 879-888. doi:10.1037/0012-1649.26.6.879

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta psychologica, 148*, 163-172. doi:10.1016/j.actpsy.2014.01.016

Clark, C. A., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child development, 84*(2), 662-677. doi:10.1111/j.1467-8624.2012.01854.x

Clark, C. A. C., Pritchard, V. E., & Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Developmental Psychology, 46*, 1176-1191. doi:10.1037/a0019672

CollegeBoard. (2017). *SAT Suite of Assessments Annual Report*. Retrieved from New York, NY: https://reports.collegeboard.org/sat-suite-program-results

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education, 3*(2), 63-68. doi:10.1016/j.tine.2013.12.001

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition, 162*, 12-26. doi:10.1016/j.cognition.2017.01.014

CUNY Office of Institutional Research. (2015a). *Initial remedial needs*. Retrieved from http://www.qcc.cuny.edu/oira/factbook.html

CUNY Office of Institutional Research. (2015b). *Pass rates of selected math courses: Fall 2014*. Retrieved from http://www.qcc.cuny.edu/oira/factbook.html

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*(2), 48-55. doi:10.1016/j.tine.2013.06.001

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1-2), 1-42. doi:10.1016/0010-0277(92)90049-N

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*: Oxford University Press.

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*. Retrieved from www.unicog.org

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*(3), 487-506. doi:10.1080/02643290244000239

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic -- An fMRI study. *Cognitive Brain Research, 18*(1), 76-88. doi:10.1016/j.cogbrainres.2003.09.005

Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., & Duflo, E. (2017). Cognitive science in the field: A preschool intervention durably enhances intuitive but not formal mathematics. *Science, 357*(6346), 47-55. doi:10.1126/science.aal4724

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*(1), 465-486. doi:10.1207/s15326942dn2601_6

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of experimental child psychology, 123*, 53-72. doi:10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in cognitive sciences, 8*(7), 307-314. doi:10.1016/j.tics.2004.05.002

Fischer, U., Moeller, K., Bientzle, M., Cress, U., & Nuerk, H.-C. (2011). Sensori-motor spatial training of number magnitude representation. *Psychonomic Bulletin & Review, 18*(1), 177-183. doi:10.3758/s13423-010-0031-3

Friso-van den Bos, I., Van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory aNd mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29-44. doi:10.1016/j.edurev.2013.05.003

Fuhs, M. W., & McNeil, N. W. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Developmental science, 16*(1), 136-148. doi:10.1111/desc.12013

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43-74. doi:10.1016/0010-0277(92)90050-R

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends in cognitive sciences, 4*(2), 59-65. doi:10.1016/S1364-6613(99)01424-2

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Current Directions in Psychological Science, 22*(1), 23-27. doi:10.1177/0963721412469398

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., . . . Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PloS one, 8*(6), e67374. doi:10.1371/journal.pone.0067374

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition, 115*(3), 394-406. doi:10.1016/j.cognition.2010.02.002

Goffin, C., & Ansari, D. (2016). Beyond magnitude: Judging ordinality of symbolic number is unrelated to magnitude comparison and independently relates to individual differences in arithmetic. *Cognition, 150*, 68-76. doi:10.1016/j.cognition.2016.01.018

Grubb, W. N., Boner, E., Frankel, K., Parker, L., Patterson, D., Gabriner, R., . . . Wilson, S. (2011). *Understanding the "crisis" in basic skills: Framing the issues in community colleges*. Retrieved from https://www.edpolicyinca.org/publications/understanding-%E2%80%9Ccrisis%E2%80%9D-basic-skills-framing-issues-community-colleges

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences, 109*(28), 11116-11120. doi:10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665-668. doi:10.1038/nature07246

Hart, S. A., Logan, J. A., Thompson, L., Kovas, Y., McLoughlin, G., & Petrill, S. A. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology, 108*(2), 181-193.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21*(1), 33-46. doi:10.2307/749455

Hendel, D. D. (1980). Experiential and affective correlates of math anxiety in adult women. *Psychology of Women Quarterly, 5*(2), 219-230. doi:10.1111/j.1471-6402.1980.tb00958.x

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of experimental child psychology, 103*(1), 17-29. doi:10.1016/j.jecp.2008.04.001

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology, 26*(3), 339-366. doi:10.1080/01443410500341056

Hubber, P. J., Gilmore, C., & Cragg, L. (2014). The roles of the central executive and visuospatial storage in mental arithmetic: A comparison across strategies. *Quarterly Journal of Experimental Psychology, 67*(5), 936-954. doi:10.1080/17470218.2013.838590

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General, 123*(3), 284-296.

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition, 131*(1), 92-107. doi:10.1016/j.cognition.2013.12.007

Imbo, I., Duverne, S., & Lemaire, P. (2007). Working memory, strategy execution, and strategy selection in mental arithmetic. *Quarterly Journal of Experimental Psychology, 60*(9), 1246-1264. doi:10.1080/17470210600943419

Imbo, I., Vandierendonck, A., & Rammelaere, S. d. (2007). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *Quarterly Journal of Experimental Psychology, 60*(5), 708-731. doi:10.1080/17470210600762447

Indow, T., & Ida, M. (1977). Scaling of dot numerosity. *Perception & Psychophysics, 22*(3), 265-276. doi:10.3758/BF03199689

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review, 18*(6), 1222-1229.

Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *NeuroImage, 36*, 993-1003. doi:10.1016/j.neuroimage.2007.03.051

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*(3), 1221-1247. doi:10.1016/j.cognition.2007.06.004

Jang, S., & Cho, S. (2018). The mediating role of number-to-magnitude mapping precision in the relationship between approximate number sense and math achievement depends on the domain of mathematics and age. *Learning and Individual Differences, 64*, 113-124. doi:10.1016/j.lindif.2018.05.005

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24*, 190-197. doi:10.1016/j.lindif.2012.12.014

Kallai, A. Y., Schunn, C. D., Ponting, A. L., & Fiez, J. A. (2011). *Improving foundational number representations through simple arithmetical training*. Paper presented at the Society for Research on Educational Effectiveness, Evanston.

Keller, L., & Libertus, M. E. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Frontiers in Psychology, 22*(6), 685. doi:10.3389/fpsyg.2015.00685

Khanum, S., Hanif, R., Spelke, E. S., Berteletti, I., & Hyde, D. C. (2016). Effects of Non-Symbolic Approximate Number Practice on Symbolic Numerical Abilities in Pakistani Children. *PloS one, 11*(10), e0164436. doi:10.1371/journal.pone.0164436

Krueger, L. E. (1982). Single judgments of numerosity. *Perception & Psychophysics, 31*(2), 175-182. doi:10.3758/BF03206218

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics, 35*(6), 536-542. doi:10.3758/BF03205949

Larson, L. M., Pesch, K. M., Surapaneni, S., Bonitz, V. S., Wu, T.-F., & Werbel, J. D. (2015). Predicting graduation: The role of mathematics/science self-efficacy. *Journal of Career Assessment, 23*(3), 399-409. doi:10.1177/1069072714547322

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395-438. doi:10.1016/j.cognition.2006.10.005.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition, 105*(2), 395-438.

LeFevre, J.-A., DeStefano, D., Coleman, B., & Shanahan, T. (2005). Mathematical cognition and working memory. In J. I. D. Campbell (Ed.), *Handbook of Mathematical Cognition*. New York, NY: Psychology Press.

Leibovich, T., & Ansari, D. (2016). The symbol-grounding problem in numerical cognition: A review of theory, evidence, and outstanding questions. *Canadian Journal of Experimental Psychology, 70*(1), 12-23. doi:10.1037/cep0000070

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of experimental child psychology, 116*(4), 829-838. doi:10.1016/j.jecp.2013.08.003

Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2016). The Precision of Mapping Between Number Words and the Approximate Number System Predicts Children's Formal Math Abilities. *Journal of experimental child psychology, 150*, 207-226. doi:10.1016/j.jecp.2016.06.003

Lindskog, M., Winman, A., & Poom, L. (2016). Arithmetic Training Does Not Improve Approximate Number System Acuity. *Frontiers in Psychology, 25*(7), 1634.

Liu, A. S., Kallai, A. Y., Schunn, C. D., & Fiez, J. A. (2015). Using mental computation training to improve complex mathematical performance. *Instructional Science, 43*(4), 463-485. doi:10.1007/s11251-015-9350-0

Liu, A. S., Schunn, C. D., Fiez, J. A., & Libertus, M. E. (2015). *Symbolic integration, not symbolic estrangement, for double-digit numbers*. Paper presented at the 37th Annual Meeting of the Cognitive Science Society, Austin, TX.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & cognition, 22*(4), 395-410.

Lourenco, S. F., & Bonny, J. W. (2017). Representations of numerical and non-numerical magnitude both contribute to mathematical competence in children. *Developmental science, 20*(4). doi:10.1111/desc.12418

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences of the United States of America, 109*(46), 18737-18742. doi:10.1073/pnas.1207212109

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology, 141*(4), 635-641. doi:10.1037/a0027248

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256-261. doi:10.1016/j.cognition.2011.07.009

Lyons, I. M., & Beilock, S. L. (2013). Ordinality and the nature of symbolic numbers. *Journal of Neuroscience, 33*(42), 17052-17061. doi:10.1523/JNEUROSCI.1775-13.2013

Lyons, I. M., Vogel, S. E., & Ansari, D. (2016). On the ordinality of numbers: A review of neural and behavioral studies. *Progress in brain research, 227*, 187-221. doi:10.1016/bs.pbr.2016.04.010

Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., & Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: Does it matter? *Learning and Instruction, 46*, 1-11. doi:10.1016/j.learninstruc.2016.08.004

Maloney, E. A., Schaeffer, M., W., & Beilock, S. L. (2013). Mathematics anxiety and stereotype threat: Shared mechanisms, negative consequences and promising interventions. *Research in Mathematics Education, 15*(2), 115-128. doi:10.1080/14794802.2013.797744

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PloS one, 6*(9), e23749. doi:10.1371/journal.pone.0023749

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*, 471-490. doi:10.1016/j.ecresq.2006.09.003

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of experimental child psychology, 74*(3), 240-260. doi:10.1006/jecp.1999.2516

Meece, J. L., Wigfield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *Journal of Educational Psychology, 82*(1), 60-70. doi:10.1037/0022-0663.82.1.60

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37*(3), 591-606. doi:10.1016/j.paid.2003.09.029

Minturn, A. L., & Reese, T. W. (1951). The effect of differential reinforcement on the discrimination of visual number. *The Journal of Psychology, 31*, 201-321. doi:10.1080/00223980.1951.9712804

Moore, C., & Shulock, N. (2010). *Divided We Fail: Improving Completion and Closing Racial Gaps in California's Community Colleges*. Retrieved from Sacramento, CA: https://edsource.org/wp-content/uploads/old/divided_we_fail_final.pdf

Noel, M. P., Desert, M., Aubrun, A., & Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Memory and Cognition, 29*(1), 34-42.

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction, 23*, 125-135. doi:10.1016/j.learninstruc.2012.08.004

Odic, D., Im, H. Y., Eisinger, R., Ly, R., & Halberda, J. (2015). PsiMLE: A maximum-likelihood estimation approach to estimating psychophysical scaling and variability more reliably, efficiently, and flexibly. *Behavioral Research Methods*. doi:10.3758/s13428-015-0600-5

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of experimental child psychology, 152*, 278-293. doi:10.1016/j.jecp.2016.07.011

Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological science*, *24*(10), 2013-2019. doi:10.1177/0956797613482944

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training : An investigation of underlying mechanism. *Cognition, 133*(1), 188-200. doi:10.1016/j.cognition.2014.06.011

Pelavin, S. H., & Kane, M. (1990). *Changing the odds: Factors increasing access to college*. New York: College Entrance Examination Board.

Piazza, M., Izard, V., Pinel, P., Bihan, D. L., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547-555. doi:10.1016/j.neuron.2004.10.014

Pina, V., Castillo, A., Kadosh, R. C., & Fuentes, L. J. (2015). Intentional and automatic numerical processing as predictors of mathematical abilities in primary school children. *Frontiers in Psychology, 6*(375). doi:10.3389/fpsyg.2015.00375

Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., . . . Haase, V. G. (2014). In how many ways is the approximate number system associated with exact calculation? *PloS one, 9*(11), e111155. doi:10.1371/journal.pone.0111155

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110-122. doi:10.1016/j.lindif.2009.10.005

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology, 32*, 146-159. doi:10.1016/j.appdev.2011.02.005

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of experimental child psychology, 91*(2), 137-157. doi:10.1016/j.jecp.2005.01.004

Reynvoet, B., & Sasanguie, D. (2016). The Symbol Grounding Problem Revisited: A Thorough Evaluation of the ANS Mapping Account and the Proposal of an Alternative Account Based on Symbol-Symbol Associations. *Frontiers in Psychology, 13*(7), 1581.

Rickard, T. C., Healy, A. F., & Bourne, L. E. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1139-1153. doi:10.1037/0278-7393.20.5.1139

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music, and working memory training. *Current Directions in Psychological Science, 26*(6), 515-520. doi:10.1177/0963721417712760

Saxe, K., & Braddy, L. (2015). *A Common Vision for Undergraduate Mathematical Sciences Programs in 2025*. Retrieved from https://www.maa.org/programs-and-communities/curriculum%20resources/common-vision

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & de Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental science*, 1-16. doi:10.1111/desc.12372

Schuchardt, K., Maehler, C., & Hasselhorn, M. (2008). Working memory deficits in children with specific learning disorders. *Journal of Learning Disabilities, 41*(6), 514-523. doi:10.1177/0022219408317856

Sella, F., Tressoldi, P., Lucangeli, D., & Zorzi, M. (2016). Training numerical skills with the adaptive videogame "The Number Race": A randomized controlled trial on preschoolers. *Trends in Neuroscience and Education, 5*(1), 20-29. doi:10.1016/j.tine.2016.02.002

Shapiro, D., Dundar, A., Huie, F., Wakhungu, P. K., Yuan, X., Nathan, A., & Bhimdiwali, A. (2017). *Completing College: A National View of Student Completion Rates - Fall 2011 Cohort* (14). Retrieved from Herndon, VA: https://nscresearchcenter.org/wp-content/uploads/SignatureReport14_Final.pdf

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental science, 11*(5), 655-661. doi:10.1111/j.1467-7687.2008.00714.x

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games--but not circular ones--improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*(3), 545-560. doi:10.1037/a0014239

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

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 34, pp. 379-440). San Diego, CA: Academic Press.

Sullivan, J., Frank, M. C., & Barner, D. (2016). Intensive Math Training Does not Affect Approximate Number Acuity: Evidence From a Three-Year Longitudinal Curriculum Intervention. *Journal of Numerical Cognition, 2*(2). doi:10.5964/jnc.v2i2.19

Supekar, K., Iuculano, T., Chen, L., & Menon, V. (2015). Remediation of Childhood Math Anxiety and Associated Neural Circuits through Cognitive Tutoring. *Journal of Neuroscience, 35*(36), 12574-12583. doi:10.1523/JNEUROSCI.0786-15.2015

Symonds, W. C., Schwartz, R., & Ferguson, R. F. (2011). *Pathways to Prosperity: Meeting the challenge of preparing young Americans for the 21st century*. Retrieved from Cambridge, MA: https://dash.harvard.edu/handle/1/4740480

Szkudlarek, E., & Brannon, E. M. (2018). Approximate Arithmetic Training Improves Informal Math Performance in Low Achieving Preschoolers. *Frontiers in Psychology, 9*, 606. doi:10.3389/fpsyg.2018.00606

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2014). Cognitive components of a mathematical processing network in 9-year-old children. *Developmental science, 17*(4), 506-524. doi:10.1111/desc.12144

Tobias, S. (1978). *Overcoming Math Anxiety*. Boston: Houghton Mifflin.

U.S. Department of Education. (2015). *2015 Mathematics & Reading Assessments*.

van Herwegen, J., Costa, H. M., Nicholson, B., & Donlan, C. (2018). Improving number abilities in low achieving preschoolers: Symbolic versus non-symbolic training programs. *Research in Developmental Disabilities*, *2018*(77). doi:10.1016/j.ridd.2018.03.011

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological science*, *20*(9), 1132-1139. doi:10.1111/j.1467-9280.2009.02417.x

Weschsler, D. (1944). *The Measurement of Adult Intelligence*. Baltimore: Williams and Wilkins.

Whyte, J. C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology*, *44*(2), 588-596. doi:10.1037/0012-1649.44.2.588

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, *3*(4), 224-234. doi:10.1111/j.1751-228X.2009.01075.x

Wong, T. T.-Y., Ho, C. S.-H., & Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: The mediating role of number-numerosity mappings. *Contemporary Educational Psychology*, *46*, 208-217. doi:10.1016/j.cedpsych.2016.06.003

Yeniad, N., Malda, M., Mesman, J., an IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*, 1-9. doi:10.1016/j.lindif.2012.10.004

Zhang, Y., Chen, C., Liu, H., Cui, J., & Zhou, X. (2016). Both non-symbolic and symbolic quantity processing are important for arithmetical computation but not for mathematical reasoning. *Journal of Cognitive Psychology*, *28*(7), 807-824. doi:10.1080/20445911.2016.1205074