Investigating the Incremental Validity of Cognitive Variables in Early Mathematics Screening

Authors

Ben Clarke, Ph.D., Lina Shanley, Ph.D., Derek Kosty, Scott K. Baker, Ph. D., Mari Strand Cary, Ph. D., Hank Fien, Ph. D., Keith Smolkowski, Ph. D.

Full Reference

Clarke, B., Shanley, L., Kosty, D., Baker, S. K., Cary, M. S., Fien, H., & Smolkowski, K. (2018). Investigating the incremental validity of cognitive variables in early mathematics screening. *School Psychology Quarterly, 33*(2), 264–271. doi: 10.1037/spq0000214

Funding Source

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant **R305A080699** awarded to the Center on Teaching and Learning at the University of Oregon.

Publication History

Received April 12, 2016Revision received May 15, 2017Accepted May 18, 2017

Abstract

The purpose of this study was to investigate the incremental validity of a set of domain general cognitive measures added to a traditional screening battery of early numeracy measures. The sample consisted of 458 kindergarten students of whom 285 were designated as severely at-risk for mathematics difficulty. Hierarchical multiple regression results indicated that Wechsler Abbreviated Scales of Intelligence (WASI) Matrix Reasoning and Vocabulary subtests, and Digit Span Forward and Backward measures explained a small, but unique portion of the variance in kindergarten students' mathematics performance on the Test of Early Mathematics Ability – Third Edition (TEMA-3) when controlling for Early Numeracy Curriculum Based Measurement (EN-CBM) screening measures ($R^2_{change} = .01$). Furthermore, the incremental validity of the domain general cognitive measures was relatively stronger for the severely at-risk sample. We discuss results from the study in light of instructional decision-making and note the findings do not justify adding domain general cognitive assessments to mathematics screening batteries.

Keywords: Mathematics, Screening, Number Sense, Cognitive

Investigating the Incremental Validity of Cognitive Variables in Early Mathematics Screening

Research has indicated that a successful early start in mathematics is crucial to supporting long-term academic success (Duncan et al., 2007). Longitudinal studies of mathematics development have shown that early mathematics deficits are difficult to overcome and often compound over time and that kindergarten mathematics skills, measured upon school entry are predictive of mathematics performance at the end of first grade and well beyond (Bodovski & Farkas, 2007; Hanich, Jordan, Kaplan, & Dick, 2001; Morgan, Farkas, & Wu, 2009) . Based on these findings, there has been a call for the implementation of Response to Intervention systems to support early mathematics learning in elementary schools (Gersten et al., 2009). Consequently, there has been a good deal of effort invested in both early mathematics intervention development and creating early mathematics screeners that can quickly and effectively identify students in need of additional academic support.

Research conducted over the last 10+ years of assessment development suggests that brief academic measures based on the principles of curriculum based measurement (CBM) are generally effective for identifying students that lack strong foundational mathematics skills and who may be at risk for future mathematics difficulties (Fuchs et al., 2007; Gersten et al., 2012). In early mathematics, these screeners target critical number sense skills that have been found to be predictive of later mathematics achievement (Fuchs et al., 2007) .

In a review of early numeracy screeners, Gersten et al. (2012) noted that measures of early mathematics often tap a specific aspect of number knowledge and associated skills (NRC, 2009) and that commonly found measures in screening batteries include tasks to assess understanding of magnitude comparison and strategic counting . Both of these measures may also be indicative of a student's underlying fluidity and flexibility in working with number (Gersten & Chard, 1999) . Magnitude comparison measures typically require students to compare the quantity of numbers and identify the number in a pair or set that is greater or greatest. Strategic counting measures, often referred to as missing number measures, require students to identify the missing number from a sequence of numbers (e.g., 4 __ 6) with the position of the number missing and the length of the sequence of numbers varying across different screening batteries (Gersten et al., 2012). Measures assessing the constructs of magnitude comparison and strategic counting may be reliant on the development of a mental number line used to facilitate the student's engagement in a range of tasks include making increasingly finite judgments regarding magnitude and calculation (Booth & Siegler, 2006).

Other tasks commonly found in early mathematics screening batteries include measures of oral or rote counting and numeral identification (Fuchs et al., 2007). While these measures have demonstrated strong psychometric properties, measures assessing these areas may be more indicative of school readiness skills rather than mathematics specific skills (Methe et al., 2011). Correlations for early numeracy screeners with future mathematics performance, typically measured approximately one academic to one calendar year later, range from .50 to .70 (Gersten et al., 2012). Additional analyses suggest varying levels of accuracy in determining student at-risk status for both screening batteries and individual measures. For example, Seethaler and Fuchs (2010) found a measure of quantity discrimination correctly classified students at-risk for mathematics disability with approximately 90% sensitivity and 66% specificity whereas Fuchs and colleagues (2007) found 69% sensitivity and 70% specificity for a measure of number identification. In part, the varied findings on classification accuracy can be attributed to how researchers define at-risk status (Methe et al., 2011).

While the field of early mathematics screening has advanced greatly in the past decade and the use of screening measures in the field is becoming more common (Methe et al., 2011) calls to move the research forward are increasing (Gersten et al., 2012; Methe et al., 2011; VanDerHeyden, 2010). Researchers have advocated for expanding early screening batteries to include measures to assess informal number sense and other key conceptual areas such as measurement (Methe et al., 2011) and to assess students' ability to apply their understanding of number in varied contexts (e.g., word problems; Locuniak & Jordan, 2008) . Researchers have also advocated for examining the role of a wider array of variables such as student attentiveness and persistence and specific domain general cognitive process variables that have demonstrated relationships with mathematics achievement (DiPerna, Lei, & Reid, 2007). In addition, despite advances in providing information based on ROC analyses (e.g. sensitivity and specificity), relatively few studies have provided data anchored to more probabilistic or Bayesian approaches leading to calls for researchers to provide data on metrics such as positive and negative predictive power that are more readily understandable and useful for instructional decision making (VanDerHeyden, 2010).

Moreover, an increased focus has been paid to whether or not screening could indicate a priori students who may not respond to a research based intervention in either tier 1 or tier 2 (Compton et al., 2012; Vaughn, Denton, & Fletcher, 2010). While there are a number of potential reasons for non-response, one possibility is that students with domain general cognitive skill deficits may be more likely to be non-responsive to generally effective tier 2 instruction in contrast to students who have sufficient domain general cognitive skills. Thus, the inclusion of domain general cognitive measures into screening batteries may better help educators identify a priori potential non-responders and enable the provision of more targeted and intensive intervention services than typically provided in tier 2.

A wide body of research suggests that a variety of domain general cognitive processes including nonverbal problem solving, phonological processing, rapid automatized naming, visual-spatial skills, and working memory are valuable predictors of mathematics difficulties (Fuchs et al., 2005; Geary, Hoard, & Hamson, 1999; Swanson & Jerman, 2006), and thus show potential for utility as early screeners of student risk status. In fact, one recent study found that 49% of the variance in early numerical skills in students with a mean age of 5.3 could be explained by a combination of measures assessing working memory, processing speed, phonological abilities, and intelligence (Passolunghi, Lanfranchi, Altoè, & Sollazzo, 2015). Likewise, studies of the domain general cognitive correlates of mathematics performance for third-grade students found significant associations between mathematics skills and a variety of domain general cognitive abilities with domain general cognitive skills explaining 32–52% of the variance in various mathematics tasks (Fuchs et al., 2006).

Similarly, longitudinal investigations of the relationships between domain general cognitive skills and mathematics performance have demonstrated that working memory, processing speed, and visual spatial skills measured in kindergarten explain variance in later mathematics performance (Barnes et al., 2014; Fuchs et al., 2005). Processing speed, language skills, and rapid automatized naming have been shown to be (a) correlated with mathematics achievement (Mazzocco & Myers, 2003), and (b) effective in discriminating between typical and at-risk learners in the early grades (Cirino, Fuchs, Elias, Powell, & Schumacher, 2015). Initial studies (e.g. Mazzocco & Thompson, 2005; Seethaler, Fuchs, Fuchs, & Compton, 2012) have begun to examine the contribution of specific domain general cognitive factors in early mathematics screening batteries.

Because there is good deal of evidence that domain general cognitive skills are highly correlated with mathematics achievement (Fuchs et al., 2006; Mazzocco & Myers, 2003), the purpose of this study was to explore the incremental validity of a set of selected domain general cognitive assessments focused on verbal knowledge, abstract visual processing, auditory memory, and working memory when added to an early numeracy screening battery. Explorations of this nature may enable the design of more effective screening batteries to identify potential non-responders to tier 2 interventions. Domain general cognitive measures administered in the current study were selected based on a handful of key considerations including (a) previous research findings demonstrating a link to numeracy, (b) ease of administration, and (c) the ability to discriminate between discrete domain general cognitive processes (i.e., language knowledge, abstract reasoning, and basic working memory skills). Specifically, this study investigated the following research questions:

1. To what extent did domain general cognitive measures (WASI Vocabulary, WASI Matrix Reasoning, Digit Span Forward, Digit Span Backward) add incremental validity to a battery of early numeracy curriculum based measures (EN-CBM) predicting student mathematics achievement (TEMA-3)?

2. How did the relationship between domain general cognitive measures and mathematics achievement vary as a function of student risk status when controlling for EN-CBM performance?

Our primary research question was designed to examine incremental validity because we wanted the frame for examining results to focus on the value that domain general cognitive measures would add to the instructional decision making process for schools. That is, does the addition of the domain general cognitive measures to an early numeracy screening battery enable better decision making for schools within the context of the increased resources required to administer, score, and interpret domain general cognitive measures? Although there is strong evidence in support of the ability of early numeracy skills to predict mathematics achievement (Aunio & Niemivirta, 2010; Gesten et al., 2012), for our primary research question we hypothesized that domain general cognitive measures would explain additional variance in Test of Early Mathematics Ability – Third Edition (TEMA -3) scores above and beyond EN-CBM scores based on previous findings of strong correlations between working memory, verbal IQ, processing skills and mathematics achievement (Passolunghi & Lanfranchi, 2012). For our second research question, we hypothesized that severe at-risk status would moderate the relationship between CBM screening measures and global mathematics performance such that foundational early numeracy skills (i.e. number identification) would be more closely associated with mathematics performance for the severe at-risk sample and that the relationship between cognitive skills and mathematics performance would differ based on severe at-risk status.

Method

Data presented in this manuscript were collected as part of a larger study (Clarke et al., 2016) examining the efficacy of a Tier 2 kindergarten mathematics intervention program in 29 classrooms across two districts in the Pacific Northwest during the 2009-10 school year. Participating school districts were suburban with district-wide eligibility for free and reduced lunch ranging from 44-50%. All students in the study were administered the math achievement measures (TEMA and EN-CBM) at pretest (September). Students scoring below the 42nd

percentile on the TEMA were then administered the domain general cognitive measures in January prior to the start of the intervention (Methe et al., 2011).

Participants

Students. All participating students from the larger study (Clarke et al., 2016) who scored at or below the 42nd percentile on a global mathematics measure (i.e. Test of Early Mathematics Ability; TEMA-3) and had complete data on all mathematics and cognitive variables were included in this study resulting in a sample of 458 students. Of these 458 students, 58 were selected to receive additional mathematics intervention aligned to the core curriculum used in their regular mathematics instruction (Methe et al., 2011); however, because the current study aimed to explore relations between various screening measures (i.e., domain general and early numeracy measures) administered prior to intervention delivery, intervention effects were not considered here. In all, thirty seven percent of the students in the sample were White, 24% American Indian or Alaskan Native, 2% Asian American or Pacific Islander, 2% African American, 4% Other, and 31% were missing race information. Fifty one percent were male. Ten percent received special education services. Based on findings from the Morgan et al. (2009) longitudinal study which found that students who were in the 10th percentile for mathematics achievement at both entrance and exit from kindergarten had a 70% chance of remaining at or below the 10th percentile 5 years later, we classified students who performed below the 10th percentile on the TEMA-3 as severely at-risk for mathematics difficulty (*n* = 285).

Measures

Participants completed Early Numeracy – Curriculum Based Measures (EN-CBM; Clarke & Shinn, 2004), Test of Early Mathematics Ability (TEMA-3; Ginsburg & Baroody, 2003), Weschler Abbreviated Scales of Intelligence (WASI) Matrix Reasoning and Vocabulary subtests, and Digit Span Forward and Backward measures. Raw scores were used in all analyses, unless otherwise noted.

Early Numeracy – Curriculum Based Measures (EN-CBM; Chard et al., 2005; Clarke & Shinn, 2004). The EN-CBM assesses students' procedural fluency on four measures (each timed for one minute). The *Oral Counting (OC)* measure requires students to orally rote count as high as possible without making an error. The *Number Identification Measure (NI)* requires students to orally identify numbers between 0 and 10 when presented with a set of printed number symbols. The *Quantity Discrimination Measure (QD)* requires students to name which of two visually presented numbers between 0 and 10 is greater. The *Missing Number Measure (MN)* requires students to name the missing number from a string of three numbers between 0 and 10. Concurrent and predictive validities for the EN-CBM measures range from .46 to .72. Reliabilities, including test-retest, alternate form, and inter-rater, for all four measures are strong and range from .78 to .99.

Test of Early Mathematics Ability – Third Edition (TEMA-3; Ginsburg & Baroody, 2003). The TEMA-3 is a norm-referenced individually administered measure of early mathematics for children ages 3 to 8 years 11 months. It measures mathematics skills related to counting, number facts and calculations, and related mathematical concepts. Test authors report alternate-form reliability of .97 and test-retest reliability ranges from .82 to .93. Concurrent validity with other criterion measures of mathematics is reported as ranging from .54 to .91. Standard TEMA-3 scores were used in this study.

Wechsler Abbreviated Scale of Intelligence (WASI) Subtests: Matrix Reasoning and Vocabulary (Wechsler, 1999). WASI Vocabulary measures word knowledge, verbal concept formation, and fund of knowledge. The Vocabulary subtest requires examinees to name objects presented visually in pictures and define words that are presented both visually and orally. WASI Matrix Reasoning assesses visual information processing and abstract reasoning skills. In the Matrix Reasoning subtest, examinees are prompted to select a response option that completes a matrix or series of visual items. Test-retest reliabilities are .76 for matrix reasoning and .85 for vocabulary. Internal consistency for the 6-year-old sample is reported as .96 for matrix reasoning and .87 for vocabulary. Concurrent validity ranged from .66 to .88.

Digit Span-Forward and Backward. Based on the fairly standard scope of these assessments (i.e., Wechsler Intelligence Scale for Children, Revised; Wechsler, 1974), a researcher-designed measure modeled on typical digit span tasks was administered to all participants in this study. Digit Span-Forward is designed to assess auditory memory and has a total of eight, two-part items with strings of digits in increasing lengths from two digits to eight digits. Digit Span-Backward is designed to assess auditory working memory and has a total of seven, two-part items with strings of digits in increasing lengths. After the assessor presents an example, each set of digits is presented verbally and the examinee is prompted to repeat the digits back to the examiner. Each item is dichotomously scored based on either correct or incorrect recall of the digits. Average reliability for digit span tasks across the 6.5 to 15.5 age range is .78 (Wechsler, 1974) and test-retest reliability falls in the .30 - .50 range (Baker, 1993). Various sources report that Digit Span-Forward is correlated with reading recognition tasks (r = .45; Wechsler, 1974) and mathematics achievement (.32 - .36; Baker, 1993), and somewhat correlated with the general IQ assessments at age six (r = .11; Wechsler, 1974). Digit Span-Backward is also correlated with reading recognition tasks (r = .31 - .37) and mathematics achievement (.30 - .37; Baker, 1993).

Procedures

Data collection. Measures were individually administered to students by trained research staff with experience collecting similar data for other research projects. Assessors were required to reach inter-rater reliability coefficients of .90 during training sessions and trial field administration. Standardized, scripted testing protocols were utilized for data collection, and all data collectors participated in shadow scoring activities to ensure reliability.

Statistical Analysis

Univariate descriptive analyses were performed on measures of domain general cognitive and mathematics performance. Pearson's *r* correlation coefficients were used to examine the covariation among the study variables. A hierarchical multiple regression model was used to address our primary research question. Specifically, the extent to which domain general cognitive scores predicted TEMA-3 scores above and beyond EN-CBM scores for students at-risk for mathematics difficulty was examined, and the extent to which the incremental validity of the domain general cognitive scores varied as a function of severe at-risk status was also examined.

The hierarchical multiple regression model involved five blocks. In Block 1, the severe at-risk indicator (i.e., dichotomous variable of severe risk status) was entered to obtain interpretable R^2_{change} statistics for subsequent blocks because it was, by definition, correlated with the outcome measure. In Block 2, EN-CBM scores were entered to evaluate relations between EN-CBM scores and TEMA-3 scores and the total percent of variance in TEMA-3 scores explained by EN-CBM scores after controlling for severe at-risk status. In Block 3, the interactions between EN-CBM scores and the severe at-risk indicator were entered to determine whether relations between EN-CBM scores and TEMA-3 scores varied as a function of severe at-risk status. In Block 4, domain general cognitive scores were entered to evaluate the incremental validity of each domain general cognitive measure and the total percent of variance in TEMA-3 scores explained by domain general cognitive scores above and beyond EN-CBM scores.

Finally, in Block 5, the interactions between the domain general cognitive scores and the severe at-risk indicator were entered to determine whether the domain general cognitive measures were differentially associated with TEMA-3 scores as a function of severe at-risk status. All measures and interactions were entered in each step, regardless of previous patterns of statistical significance to allow for estimation of total variance explained by the set of measure and/or interactions in each block. The R^2_{total} and R^2_{change} statistics are reported in the model results to describe the proportion of variance in the dependent variable captured by the independent variables in each block. In addition to R^2 statistics, Akaike's Information Criterion (AIC) was used to compare the models produced by the five blocks of the hierarchical regression (Burnham & Anderson, 2002) . For each block, ΔAIC is reported. The change in AIC is the difference between the AIC for any given block and the minimum AIC among the set. Lower AIC values indicate greater parsimony and fit, or an optimal balance between under- and over-fitted models. Values of ΔAIC of two or less indicate competitive models, whereas values greater than 10 suggest a model with little support in contrast to the best-fitting model. All analyses were conducted using maximum likelihood estimation with SAS (SAS Institute Inc., 2009). As such, all available information from each case was utilized in the analyses.

Missing data. Five hundred and four students were initially screened for inclusion in the current study. Of those 504 students, 458 (91%) had complete data on all study variables. Listwise deletion methods were employed and these analyses included cases with complete data only, however there is reason to believe that missing data did not bias the results of this study. The prediction models included between 91% of the reference student sample. Whereas a large proportion of missing data can have a sizable impact on results (e.g., Smolkowski, Danaher, Seeley, Kosty, & Severson, 2010), analyses with less than 9% missing data should be relatively unbiased (Schafer & Graham, 2002).

Results

Descriptive statistics and correlations between the study variables are displayed in Table 1. Correlations between all variables were statistically significant ($p < .05$) and ranged from .12 to .68. As expected, the academic screening measures (EN-CBMs) were moderately correlated with the mathematics criterion measure (TEMA-3), $r = .45 – .68$. Of the CBMs, the missing number subtest had the lowest average score and the weakest relationships with the domain general cognitive measures, $r = .15 – .27$. The domain general cognitive measures had slightly weaker correlations with the TEMA-3, $r = .25 – .46$, and WASI matrix reasoning subtest demonstrated weak relationships with many of the other variables, $r = .12 – .32$.

The results of the hierarchical multiple regression analyses are presented in Table 2. The severe at-risk indicator was entered in Block 1 to obtain interpretable R^2_{change} statistics for subsequent blocks. Block 2 of the multiple regression demonstrated that the EN-CBM measures explained 9.2% of unique variance in TEMA-3 scores after controlling for severe at-risk status among students at-risk for mathematics difficulties (F-change = 39.33, $p < .001$). Oral counting and number identification subtests were most closely associated with TEMA-3 scores ($\beta = 0.20$, $p < .001$ and $\beta = 0.23$, $p < .001$, respectively).

The interactions between EN-CBM scores and the severe at-risk indicator in Block 3 explained an additional 3.1% of the variance in TEMA-3 scores (*F*-change = 15.09, $p < .001$), with oral counting ($\beta = 0.20$, $p < .001$), number identification ($\beta = 0.25$, $p < .001$), and missing number scores ($\beta = -0.13$, $p < .001$) having statistically significant unique interactions with severe at-risk status. The domain general cognitive measures in Block 4 explained an additional 1.1% of unique variance in TEMA-3 scores above and beyond the other predictors in the model (*F*-change = 5.44, $p < .001$), with WASI vocabulary and matrix reasoning scores demonstrating statistically significant relations with TEMA-3 scores ($\beta = 0.10$, $p < .001$ and $\beta = 0.05$, $p = .024$, respectively). The interactions between domain general cognitive scores and the severe at-risk indicator in Block 5 explained an additional 0.5% of unique variance in TEMA-3 scores (*F*-change = 2.44, $p = .046$), with WASI matrix reasoning having a statistically significant interaction with severe at-risk status ($\beta = 0.25$, $p = .016$). Information criteria reported in Table 2 indicate that Blocks 4 and 5 achieved the most optimal balance between parsimony and fit, each with similar approximating abilities (ΔAIC = 2.07 and 0.00, respectively).

For comparison purposes, a post hoc regression analysis was conducted to evaluate the amount of unique variance in TEMA-3 scores explained by domain general cognitive measures above and beyond EN-CBM scores among students severely at-risk for mathematics difficulty. The results indicated that the domain general cognitive measures explained an additional 4.7% of unique variance in TEMA-3 scores among the severely at-risk subgroup (*F*-change = 5.33, $p < .001$). The domain general cognitive measures explained a relatively greater amount of unique variance in TEMA-3 scores among the severely at-risk subgroup compared to the complete at-risk sample (4.7% versus 1.1%).

Discussion

This study sought to explore the extent to which predicting student performance in mathematics could be improved by adding a set of domain general cognitive measures to a set of traditional early numeracy screening measures and the extent to which student risk (i.e. severe at-risk) moderated the relationship between domain general cognitive skills and mathematics performance. Results from the early numeracy measures indicate that oral counting and number identification scores positively predicted TEMA-3 scores for both at-risk and severely-at risk students. However, the associations were stronger among severely at-risk students compared to at-risk students. Furthermore, missing number scores were positively associated with TEMA-3 scores among at-risk students, but negatively associated with TEMA-3 scores among severely at-risk students.

Previous findings from studies of early numeracy academic screeners and recommendations for implementation have emphasized the use of more complex academic screeners (e.g., strategic counting) to assess risk (Fuchs et al., 2007; Gersten et al., 2012; Methe et al, 2011). However, results from this study suggest that screeners focused on foundational skills may do a better job of distinguishing risk amongst students at the lower level of the performance spectrum. In part, this is problematic because while research studies have focused on evaluating how measures work with all students, the reality is that when making screening decisions, of critical importance for decision-making is how well a measure is able to discern amongst a pool of students who are at-risk. The use of screeners focused on more basic skills (e.g., numeral identification) may help alleviate floor effects found with more advanced screeners (e.g., discriminating quantities) for at-risk students and provide a method to identify between levels of risk within a lower performing sample.

Of particular interest for this study was examining whether or not the inclusion of key domain general cognitive variables would have value in predicting student mathematics performance above and beyond traditionally used early numeracy measures. Domain general cognitive measures were able to explain a small, but unique portion of the variance in kindergarten students' mathematics performance above and beyond achievement measures, and the incremental validity of the domain general cognitive measures was relatively stronger for the severe at-risk sample as compared to the entire at-risk sample. It is possible that this distinction arose from characteristics that are somewhat unique to the overall at-risk kindergarten population included in this sample. That is, because kindergarten-aged students experience a wide range of preschool instruction and bring vastly different academic experiences to school (Bernstein, West, Newsham, & Reid, 2014; Werthheimer, Moore, Hair, & Croan, 2003), it is likely that the at-risk sample included a range of students including those who demonstrated weaker initial mathematics achievement simply because they had not received any prior mathematics instruction. In the current study, these students were combined with students whose difficulties in mathematics may have been related to an underlying disability or even attention or self-regulation deficiencies. Thus, the heterogeneous nature of the at-risk sample and their presumably varying cognitive profiles, contrasted with the severe at-risk sample, who demonstrated very little basic number sense and were more likely to have an underlying disability was likely related to the findings presented here.

However, we strongly emphasize that in practice the relatively small amount of additional variance (less than 5%) needs to be weighed against the very real "costs" including time and money incurred by schools to add domain general cognitive measures to screening batteries. For example, inclusion of domain general cognitive measures to screening batteries requires the expertise to administer and interpret a new set of tests and while many educators may be able to administer academic skill measures, it may fall upon school psychologists to administer and interpret findings from relative complex domain general cognitive assessments. It should be noted that answers to these types of questions may vary based on how different stakeholders value the information provided by domain general cognitive measures and the costs associated with gaining that information. This challenge is particularly relevant given that first order factors or profiles from comprehensive intelligence tests often fail to provide additional incremental validity after accounting for general intelligence (Canivez, 2013: Glutting et al., 2006). That is, using a measure of a domain general cognitive skill such as working memory is in fact a measure of general intelligence and any variance accounted for by inclusion of such a measure is explained by general intelligence. It may be more parsimonious to either use a general measure of intelligence and/or note that any subtests used are functioning as measures of general intelligence. Issues related to the role of general intelligence and the current results of this study strongly suggest that the inclusion of domain general cognitive measures to early numeracy screening batteries, at present, lacks justification in light of the resources that schools would have to expend to collect and analyze this information and the value that would be provided in terms of more effective instructional decision making related to provision of services and the treatment utility of additional assessment in informing instruction.

Limitations and Future Research

A few key limitations should be kept in mind when interpreting results. First, the findings and implications discussed here are contingent on the assumption that the delay between academic and domain general cognitive testing did not have a significant impact on domain general cognitive scores. Due to the high variance found in mathematics achievement and lack of stability when operationalizing mathematics learning disabilities (e.g. Chong & Siegel, 2008; Geary, Bailey, & Hoard, 2009) caution should be exercised when interpreting results from a study that uses student achievement at one time point to establish mathematics risk. Second, while the battery of domain general cognitive measures utilized in this study represented a number of key constructs it was not inclusive of all domain general cognitive variables associated with mathematics achievement (e.g., processing speed). Additionally, variable correlations revealed moderately strong relations between scores on the WASI Vocabulary subtests and both Digit Span tasks, which suggests that there may be a good deal of overlap in the variance explained by each measure which may be explained by the overall role of general intelligence. Furthermore, the current study employed a conservative analytic approach based on the exploratory nature of this study, which may have limited the amount of explainable variance in mathematics achievement. Thus, targeted analyses aimed at exploring associations between particular domain general cognitive skills and mathematics outcomes (see Fuchs et al., 2006) may find stronger individual relationships and explain more variance than those in the analyses conducted here. Future research should focus on assessing an array of domain general cognitive variables, both individually and in combination, and their relationship with mathematics achievement. Such work should be undertaken with appropriate caution given the limited reliability of subtest and difference scores. Third, inter observer agreement was calculated during training when it is likely to be highest instead of in the field during the course of the study and assessors who administered the domain general cognitive assessments were not blind to the fact that the selected participants had scored below a specified threshold on the TEMA, which may have impacted their administration of the measures. Lastly, the results from the study are specific to the sample, which was limited to a subsample of those who participated in the larger intervention study and caution should be taken when generalizing findings. To that end, additional research with different samples is needed.

While questions of use in school settings remain, we do feel the results of the study warrant further exploration within the research community. The field should continue to investigate different mathematical constructs that could be included in screening batteries. Insights by mathematics educators as to key understandings in early numeracy may provide an array of constructs to explore as early screeners above and beyond the typical skills that have been studied in the past decade. In addition, future research should focus on exploring the value added of domain general cognitive measures to early numeracy screening batteries administered at singular (e.g., Fall) and multiple time points (e.g., Fall and Winter), and their relationships with mathematics outcomes at the end of kindergarten or beyond. This is of particular importance within the context of Response to Intervention (RtI) models of service delivery where increased interest is being paid to identifying students who may be non-responders to tier 2 instruction prior to providing that support. Greater accuracy in identifying that subset of non-responders would ensure that the intensity of intervention services provided did not require the student failing their way through the system to obtain tier 3 support. As research in this area emerges, schools should cautiously evaluate new evidence and consciously weigh decisions to modify existing screening protocols.

A growing field of research in beginning reading has begun to explore the use of screening students into different levels of instructional support and indicates promise for improving student outcomes (Al Otaiba et al., 2014). To that end, exploring the types of screeners, both academic and cognitive, that work best for subsets of students as part of a multiple gating system seem logical next steps for the field to take as we attempt to more accurately identify students in need of additional services in mathematics.

References

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81*, 11–27. doi: 10.1177/0014402914532234

Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and Individual Differences, 20*, 427–435. doi: 10.1016/j.lindif.2010.06.003

Baker, P. C. (1993). *NLSY child handbook: A guide to the 1986-1990 National Longitudinal Survey of Youth child data*. Columbus, Ohio: Center for Human Resource Research, Ohio State University.

Barnes, M. A., Raghubar, K. P., English, L., Williams, J. M., Taylor, H., & Landry, S. (2014). Longitudinal mediators of achievement in mathematics and reading in typical and atypical development. *Journal of Experimental Child Psychology, 119*, 1–16. doi: 10.1016/j.jecp.2013.09.006

Bernstein, S., West, J., Newsham, R., & Reid, M. (2014). *Kindergartners' skills at school entry: An analysis of the ECLS-K*. Princeton, NJ: Mathematica Policy Research, Inc. Retrieved from http://www.mathematica-mpr.com/~/media/publications/pdfs/earlychildhood/kindergarten_skills_School_entry.pdf

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *Elementary School Journal, 108*, 115–130. doi: 10.1086/525550

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*, 189–201. doi: 10.1037/0012-1649.41.6.189

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer - Verlag.

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds & V. Schwean (Eds.), *The Oxford Handbook of Child Psychological Assessment* (pp. 84–112). New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199796304.013.0004.

Chard, D. J., Clarke, B., Baker, S. K., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30*, 3–14. doi: 10.1177/073724770503000202

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology, 33*, 300–317. doi: 10.1080/87565640801982387

Cirino, P. T., Fuchs, L. S., Elias, J. T., Powell, S. R., & Schumacher, R. F. (2015). Cognitive and mathematical profiles for different forms of learning difficulties. *Journal of Learning Disabilities*, *48*, 156–175. doi: 10.1177/0022219413494239

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Strand Cary, M. (2016). Examining the efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities, 49*, 152–165. doi: 10.1177/0022219414538514

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234–248.

Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., . . . Bouton, B. D. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities, 45*, 204–216. doi: 10.1177/0022219412442151

DiPerna, J. C., Lei, P.-W., & Reid, E. E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the early childhood longitudinal study--kindergarten cohort. *Journal of Educational Psychology, 99*, 369–379. doi: 10.1037/0022-0663.99.2.369

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

Fuchs, L. S., Fuchs, D., Compton, D. L., Bryant, J. D., Hamlett, C. L., & Seethaler, P. M. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children, 73*, 311–330. doi: 10.1177/001440290707300303

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

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

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology, 74*, 213–239. doi: 10.1006/jecp.1999.2515

Gersten, R. M., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools* (Practice Guide Report No. NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/rti_math_pg_042109.pdf

Gersten, R. M., & Chard, D. J. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33*, 18–28. doi: 10.1177/002246699903300102

Gersten, R. M., Clarke, B., Jordan, N., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children, 78*, 423–445. Retrieved from http://cec.metapress.com/content/B75U2072576416T7

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability- third edition (TEMA-3)*. Austin, TX: ProEd.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the wisc-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education, 40*, 103–114. doi: 10.1177/00224669060400020101

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93*, 615–627. doi: 10.1037/0022-0663.93.3.615

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*, 451–459. doi: 10.1177/0022219408321126

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia, 53*, 218–253. doi: 10.1007/s11881-003-0011-7

Mazzocco, M. M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research and Practice, 20*, 142–155. doi: 10.1111/j.1540-5826.2005.00129.x

Methe, S. A., Hojnoski, R., **Clarke, B.**, Owens, B. B., Lilley, P. K., Politylo, B. C., . . . Marcotte, A. M. (2011). Innovations and future directions for early numeracy curriculum-based measurement. *Assessment for Effective Intervention, 36*, 200–209. doi: 10.1177/1534508411414154

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306–321. doi: 10.1177/0022219408331037

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Washington, DC: National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=12519

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology, 82*(1), 42–63. doi: 10.1111/j.2044-8279.2011.02039.x

Passolunghi, M. C., Lanfranchi, S., Altoè, G., & Sollazzo, N. (2015). Early numerical abilities and cognitive skills in kindergarten children. *Journal of Experimental Child Psychology, 135*, 25–42. doi: 10.1016/j.jecp.2015.02.001

SAS Institute Inc. (2009). *SAS/STAT ®9.2 user's guide*. Cary, NC: SAS Institute Inc.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177. doi: 10.1037//1082-989X.7.2.147

Seethaler, P. M., & Fuchs, L. S. (2010). The predictive utility of kindergarten screening for math difficulty. *Exceptional Children, 77*, 37–59. doi: 10.1177/001440291007700102

Seethaler, P. M., Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). Predicting first graders' development of calculation versus word-problem performance: The role of dynamic assessment. *Journal of Educational Psychology*, *104*, 224–234. doi: 10.1037/a0024968

Smolkowski, K., Danaher, B. G., Seeley, J. R., Kosty, D. B., & Severson, H. H. (2010). Modeling missing binary outcome data in a successful web-based smokeless tobacco cessation program. *Addiction*, *105*, 1005–1015. doi: 10.1111/j.1360-0443.2009.02896.x

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis fo the literature. *Review of Educational Research, 76*, 249–274. doi: 10.3102/00346543076002249

VanDerHeyden, A. M. (2010). Determining early mathematical risk: Ideas for extending the research. *School Psychology Review, 39*, 196–202. Retrieved from http://go.galegroup.com/ps/i.do?id=GALE%7CA233050812&v=2.1&u=s8492775&it=r&p=AONE&sw=w&asid=48b7aa017f7f5df80da301dc1fa509f3

Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools, 47*, 432–444. doi: 10.1002/pits.20481

Wechsler, D. (1974). *Wechsler intelligence scale for children-revised*. New York: Psychological Corporation.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence (WASI) subtests: Matrix reasoning and vocabulary*. San Antonio, TX: Psychological Corp.

Wertheimer, R. F., Moore, K. A., Hair, E. C., & Croan, T. (2003). *Attending kindergarten and already behind: Astatistical portrait of vulnerable young children* (Child Trends Research Brief Publication No. 2003-20). Washington, DC: Child Trends. Retrieved from http://www.childtrends.org/wp-content/uploads/2003/01/Fatherhood-DADS-Survey-Framework.pdf

Table 1.

Descriptive Statistics and Correlations for All Study Variables

Variable	1	2	3	4	5	6	7	8	*M* (*SD*)
1. Oral counting									13.6 (8.9)
2. Number identification	.41								17.2 (15.4)
3. Quantity discrimination	.45	.67							5.0 (6.2)
4. Missing number	.32	.51	.53						1.9 (3.2)
5. WASI vocabulary	.38	.37	.27	.16					6.2 (4.2)
6. WASI matrix reasoning	.16	.20	.14	.23	.21				9.1 (2.2)
7. Digit span forward	.39	.22	.25	.15	.43	.12			4.9 (1.9)
8. Digit span backward	.35	.31	.29	.27	.46	.32	.40		1.5 (1.4)
9. TEMA-3	.55	.68	.58	.45	.46	.25	.35	.40	77.6 (11.2)

Note. *M* = mean, *SD* = standard deviation. Correlations greater than .12 were statistically significant at $p < .01$; correlations greater than .20 were statistically significant at $p < .001$. All correlations were computed using pairwise deletion.

Table 2

Results of Sequential Multiple Regression Model Assessing Whether Cognitive Scores Predict TEMA-3 Scores Above and Beyond EN-CBM Scores

Predictor	Block 1		Block 2		Block 3		Block 4		Block 5	
	Coefficient (*SE*)	β	Coefficient (*SE*)	β	Coefficient (*SE*)	β	Coefficient (*SE*)	β	Coefficient (*SE*)	β
Intercept	89.21*** (0.51)		78.92*** (0.93)		83.54*** (1.21)		79.15*** (1.81)		84.00*** (2.41)	
Severe at-risk indicator	-18.59*** (0.65)	-0.80	-12.90*** (0.74)	-0.56	-19.76*** (1.36)	-0.85	-18.41*** (1.38)	-0.80	-26.18*** (2.92)	-1.13
Oral counting			0.26*** (0.04)	0.20	0.13** (0.04)	0.10	0.10* (0.04)	0.08	0.13** (0.05)	0.10
Number identification			0.17*** (0.03)	0.23	0.03 (0.03)	0.05	0.03 (0.03)	0.03	0.02 (0.03)	0.03
Quantity discrimination			0.05 (0.06)	0.03	0.14* (0.07)	0.08	0.17* (0.07)	0.10	0.15* (0.07)	0.08
Missing number			0.01 (0.10)	0.00	0.25* (0.11)	0.07	0.25* (0.11)	0.07	0.27* (0.11)	0.08
Oral counting × severe risk					0.30*** (0.07)	0.20	0.27*** (0.07)	0.18	0.23** (0.07)	0.15
Number identification × severe risk					0.28*** (0.05)	0.25	0.26*** (0.05)	0.24	0.26*** (0.05)	0.23
Quantity discrimination × severe risk					-0.26~ (0.13)	-0.08	-0.33* (0.13)	-0.10	-0.33* (0.13)	-0.10
Missing number × severe risk					-0.95*** (0.23)	-0.13	-0.91*** (0.22)	-0.13	-0.95*** (0.23)	-0.13

WASI vocabulary							0.26*** (0.07)	0.10	0.19~ (0.11)	0.07
WASI matrix reasoning							0.28* (0.12)	0.05	-0.06 (0.19)	-0.01
Digit span forwards							0.04 (0.16)	0.01	-0.21 (0.26)	-0.03
Digit span backwards							-0.09 (0.22)	-0.01	0.00 (0.34)	0.00
WASI vocabulary × severe risk									0.14 (0.15)	0.05
WASI matrix reasoning × severe risk									0.61* (0.25)	0.25
Digit span forwards × severe risk									0.43 (0.33)	0.10
Digit span backwards × severe risk									-0.23 (0.45)	-0.02
R^2_{total}	.645***		.737***		.768***		.779***		.784***	
R^2_{change}	.645***		.092***		.031***		.011***		.005*	

Note. *SE* = standard error, β = standardized regression coefficient.

~$p < .10$, *$p < .05$, **$p < .01$, ***$p < .001$.