2016 NCTM RESEARCH CONFERENCE
April 11-13 • San Francisco

Paper Title: A Practical and Powerful Screener of Middle School Mathematics Difficulties

Author(s): **Jessica Carrique**[1], **Nicole Hansen**[2], **Ilyse Resnick**[1], **Nancy I. Dyson**[1], **Ai Ye**[1] and **Nancy C. Jordan**[1], (1)University of Delaware (2)Fairleigh Dickinson University

Session Title: A Practical and Powerful Screener of Middle School Mathematics Difficulties

Session Type: Brief Research Report

Presentation Date: Tuesday, April 12, 2016

Presentation Location: San Francisco, California

Abstract

This study assesses number-related measures in elementary school as diagnostic tools for the prediction of later middle school mathematics difficulties. Predictor measures of students' fraction concepts, fraction procedures, and whole number multiplication fact fluency were administered to students (N = 342) in fourth and fifth grades. Using a framework adapted from Youngstrom (2014), we identify a reliable fraction screener as a strong diagnostic tool for the prediction of students at-risk for later mathematics difficulties. All items on the fraction screener are available to the public and accessible for use in the classroom. Practitioners can use the screener to identify students who are in need of supplemental support to attain desired mathematics benchmarks.

A Practical and Powerful Screener for Middle School Mathematics Difficulties

Proficiency in mathematics is important for students' success in science, technology, engineering, and mathematics (STEM) vocations (National Mathematics Advisory Panel [NMAP], 2008). Many students, however, do struggle in algebra, which is a gateway for achievement in STEM disciplines (NMAP, 2008). Fractions are foundational for learning algebra (e.g., Siegler et al., 2012), thus representing a crucial component of math education (NMAP, 2008). Facility with fractions also affects daily life functioning in areas such as managing personal finances and doing home repairs. Unfortunately, many students struggle to develop even a basic understanding of fractions (e.g., Bailey et al., 2012; Hansen, Jordan, & Rodrigues, in press) and are therefore at-risk for later mathematics difficulties. Reliable screeners help identify students at-risk for later mathematics difficulties who need support for future success.

Prior research has identified two number-related competencies that are highly predictive of mathematics achievement: fraction knowledge and whole number calculation fluency. Siegler et al. (2012) found that elementary students' fraction knowledge predicted their general mathematics achievement in high school, above and beyond contributions of working memory, nonverbal IQ, and family income. Similarly, Resnick et al. (2016) found a strong association between students' growth on a fraction number line estimation task from fourth through sixth grade and their performance on a standardized mathematics achievement test in the spring of sixth grade. Calculation fluency is also predictive of later mathematics performance; Jordan et al. (2013), for example, found addition fluency in third grade to be a strong predictor of general mathematics in fourth grade. Seethaler, Fuchs, Star, and Bryant (2011) assessed third-grade cognitive predictors of fifth-grade whole number and rational number computational skill, and students' early calculation fluency emerged as the strongest predictor of outcomes three grades later, above and beyond nonverbal reasoning, processing speed, and working memory. In the present study, we evaluate students' fraction knowledge and fluency on a whole number multiplication measure in fourth and fifth grades as potential screeners of later mathematics performance on a state test at the end of sixth grade.

Conceptual Framework

To assess the diagnostic accuracy of number-related measures given in fourth and fifth grades in the identification of children at-risk in mathematics, we utilized a framework outlined by Youngstrom (2014) for using receiver operating characteristic (ROC) curves for clinical decision-making. Here, we adapt this framework within a context of educational decision-making. For the present study, "diagnostic accuracy" refers to a measure's ability to accurately predict student membership into one of two groups: students who are likely to meet the sixth-grade math standards and students who are likely to *not* meet the sixth-grade standards. Figure 1 outlines the four steps involved in this conceptual framework.

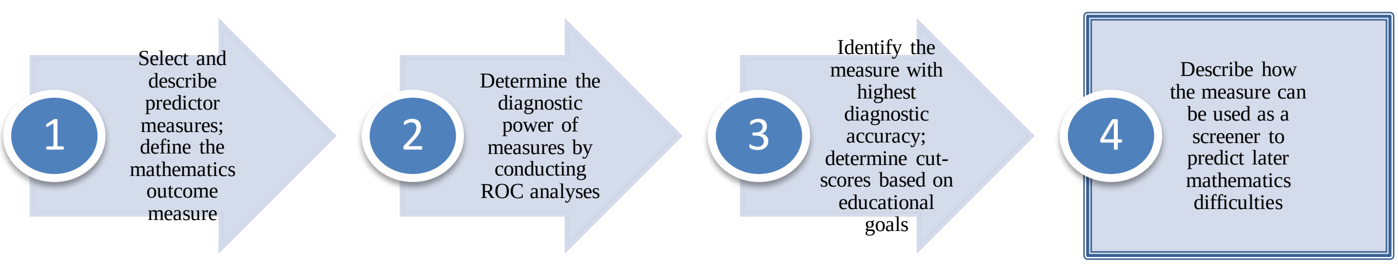

Figure 1. Conceptual framework

The first step of the framework requires the selection and description of predictor measures and one outcome measure, which will be further described below in the "Method." The second step introduces the statistical analysis used for evaluating predictor measures as potential diagnostic tools: receiver operating characteristic (ROC) curves. ROC curve analyses determine the diagnostic accuracy of a measure. In our study, we evaluate students' accuracy on three measures of fraction knowledge and one measure of whole number multiplication fact fluency in predicting whether or not students meet the mathematics standards in sixth grade. When investigating the utility of various measures as screeners, we must consider both sensitivity and specificity. Sensitivity is the proportion of students who do not pass the sixth grade test and are correctly identified as at-risk for later difficulties by a certain cut-score on a predictor measure. Specificity is the proportion of students who pass the sixth grade test and are correctly identified as *not* at-risk. ROC curve analyses plot the sensitivity of a measure (y-axis) against its specificity (1-specificity is represented on the x-axis). Therefore, if a measure accurately discriminates between students, its ROC curve will extend toward the upper left corner of the plot. That is, the measure will have *both* high sensitivity and high specificity. As a ROC curve approaches the upper left corner, the area under the curve (AUC) increases. Higher AUC values indicate higher accuracy in discriminating between students. Significant AUC values equal to or greater than .75 indicate good screeners for determining risk status (Cummings & Smolkowski, 2015). For example, an AUC of .75 means that the measure correctly places students 75% of the time. If predictors do not reach significant AUC values that equal or exceed .75, the measures cannot be considered as appropriate screeners of students' later achievement.

When AUC values of .75 are found, the third step is to identify the predictor measure with the greatest AUC value, meaning the measure with the highest diagnostic accuracy for predicting later math difficulties. The analysis also identifies a cut-score of student performance on the predictor measure that separates children into two groups: students who are likely to meet the math standards and students who are likely to be at-risk for math difficulties. Last, we describe how practitioners can use the measure for making predictions about their own students' later mathematics achievement.

A main advantage of using ROC curve analyses is that the selection of cut-scores is based on an educator's goals and/or the constraints of an educational intervention. For example, if an intervention demands extensive time and funds, the practitioner may need to more carefully identify students who are truly at-risk; this goal can be accomplished by maximizing the specificity value associated with a cut-score. However, if there are no reasons to limit the amount of students receiving the intervention, a cut-score can be selected that balances both sensitivity and specificity. We report a sensitivity threshold of ≥ 85%, which is a cut-score ideal for screening purposes (Jordan, Glutting, Ramineni, & Watkins, 2010).

Method

Participants

Children were recruited from nine elementary schools and followed from fourth grade through sixth grade. This study focuses on the 342 students for whom DCAS scores at the end of sixth grade were obtained. Demographic information is provided in Table 1 for students who did meet proficiency on the sixth grade math standards and students who did *not* meet the math standards.

Table 1

Demographic Information for Total Sixth-Grade Participants by Group

Characteristic	Met DCAS 6th Grade Math Standards (*n* = 227)	Did Not Meet DCAS 6th Grade Math Standards (*n* = 115)
Gender		
Male	43.6%	50.4%
Female	56.4%	49.6%
Race		
White	63.4%	38.3%
Black	29.1%	54.8%
Asian/Pacific Island	4.8%	1.7%
American Indian/Alaskan Native	2.7%	5.2%
Hispanic	20.3%	20.0%
Low Income	54.2%	71.3%
English Learner	10.6%	13.9%
Special Education	4.8%	17.4%
Mean Age (SD) (at start of fourth grade, in months)	117.4 (4.9)	119.9 (6.5)

Predictor Measures (administered in fall of fourth and fifth grades)

Fraction concepts.

NAEP fraction concepts. Released items (*n*=18 in fourth grade; 19 in fifth grade) from the National Assessments of Educational Progress (NAEP; U.S. Department of Education, 1990-2009) were used to measure fraction concepts. Items assessed concepts such as part-whole understanding and equivalence. Students earned one point for each correct response. The measure had a reliability of .78 in both grades. Items are publically available through the NAEP website (https://nces.ed.gov/nationsreportcard/).

Fraction number line estimation (FNLE). A computer number line task adapted from Siegler, Thompson, & Schneider (2011) was used to assess fraction magnitude estimation. Students estimated the location of nine fractions on a 0-1 number line and 19 fractions on a 0-2 number line. All estimations were combined to create a single score, which had high internal reliability (α = .91 in fourth grade; .98 in fifth grade). Scores were calculated as the mean percent absolute error (calculated by dividing the absolute value of the difference between the estimated position and actual position by the numerical range of the number line (1 or 2), multiplying by one hundred for each item, and averaging across all trials). Higher percent absolute error indicates poorer performance.

Fraction procedures. The fraction procedures measure, adapted from Hecht (1998), assesses fraction computation. In fourth grade, students received four addition and four subtraction items. All involved fractions with like-denominators and three involved mixed numbers. In fifth grade, the task included 26 items: six addition, six subtraction, nine multiplication, and five division items. Four items involved fractions with unlike denominators, and seven items included mixed numbers. At each time point, the measure was highly reliable (α = .95 in fourth grade; .84 in fifth grade).

Multiplication fluency. The Multiplication Fluency subtest of the Wechsler Individual Achievement Test (WIAT; The Psychological Corporation, 1992) assessed multiplication fluency. The WIAT is a paper and pencil task with forty single-digit multiplication problems. The score on this measure is the number of correct problems completed in one minute. Test-retest reliability was high (α = .90) in fourth and fifth grade.

Outcome Measure (administered in spring of sixth grade)

Standardized mathematics achievement test (DCAS). The DCAS is a high-stakes state assessment of students' attainment of the Common Core State Standards (CCSS). We used children's scores on the mathematics subtest of the DCAS (American Institutes of Research, 2012) at the end of sixth grade. This measure assesses numeric reasoning, algebraic reasoning, geometric reasoning, and quantitative reasoning. Scores are classified as a binary outcome for the present study: 0 (meets the sixth-grade standards) and 1 (does not meet the standards). Analyses were conducted to predict not meeting standards (1). Internal consistency is high (Cronbach alpha > .88).

Results

All predictor measures for both time points were significantly correlated (p <.001). Children who met the standards in sixth grade performed significantly better on these measures than children who did not meet the standards in sixth grade (p <.001). Approximately 33% of the sample did not meet the standards on the sixth grade standardized test.

In our fourth and fifth grade ROC curve analyses, we found that all predictors had strong diagnostic accuracy in predicting whether or not students met standards on the sixth-grade mathematics test, as indicated by curves extending close to the top left corner of the plots (Figure 2). On both plots, the curve for the NAEP fraction concepts measure has the highest area under the curve (AUC), meaning that the measure is more diagnostically accurate than the other predictors (AUC = .835 in fourth grade; .883 in fifth grade as reported in Table 2). We used a method proposed by Hanley and McNeil (1983) for evaluating whether the NAEP measure differed *significantly* from the other measures. The NAEP measure was significantly better than the fluency and procedures tasks (p < .001). The NAEP screener did not differ significantly from the number line task; however, unlike the number line task, the NAEP measure is easily accessible and easy to score. The NAEP measure is also more useful for teachers as a diagnostic tool because it is curriculum-based. By analyzing student errors, teachers have a good indicator as to where intervention instruction should begin. Thus, we selected it as the focus for subsequent analyses.

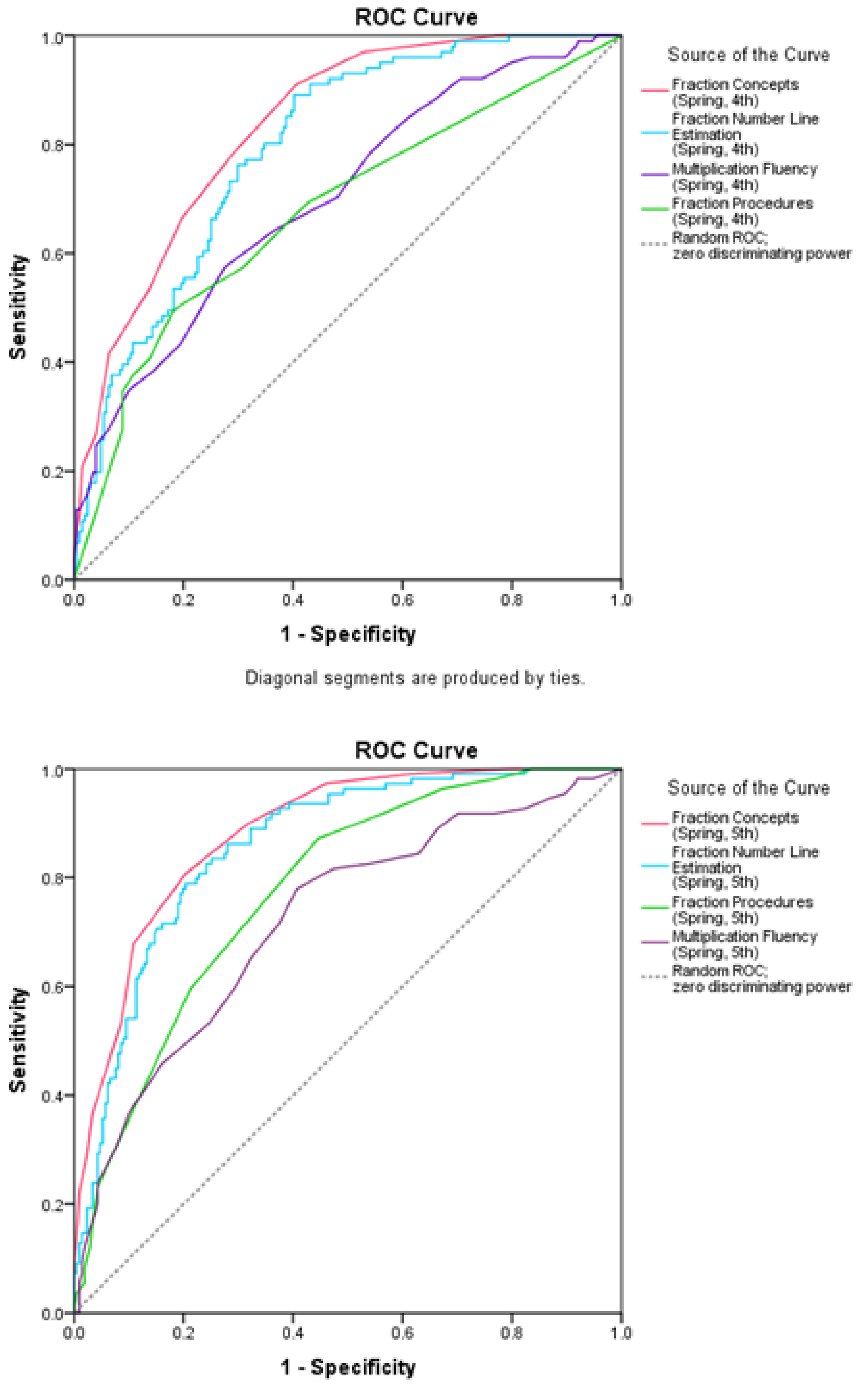

Figure 2. ROC curve analysis plots with fourth grade (top) and fifth grade (bottom) predictors. The fraction number line estimation task was reverse-coded to ease interpretation; higher scores indicate better performance.

Table 2

Diagnostic Utility Statistics of Predictor Measures

Measure	Overall *p* Values[a]	AUC	Std. Error	95% Confidence Interval	
				Lower	Upper
Fourth Grade					
NAEP Fraction Concepts	.001	.835[1]	.023	.791	.880
Fraction Number Line Estimation (FNLE)	.001	.796[1]	.025	.746	.846
Multiplication Fluency (WIAT)	.001	.702	.032	.640	.764
Fraction Procedures	.001	.679	.034	.613	.745
Fifth Grade					
NAEP Fraction Concepts	.001	.883[1]	.019	.847	.919
Fraction Number Line Estimation (FNLE)	.001	.859[1]	.021	.818	.900
Fraction Procedures	.001	.778[1]	.026	.727	.829
Multiplication Fluency (WIAT)	.001	.723	.030	.663	.782

Note. AUC = Area Under the Curve.

[a] Null hypothesis: true area = 0.5

[1] large effect size (Cummings & Smolkowski, 2015)

We selected cut-scores based on the 85%-sensitivity threshold, maximizing the sensitivity of the screener. For example, the sensitivity-based cut score for the fourth grade NAEP measure was 12.50, meaning that scores < 12.50 were indicators of students at-risk for later difficulties. The sensitivity value was .782, indicating that approximately 78% of students who did not pass the sixth-grade DCAS test scored *below* 12.50 on the fourth-grade NAEP measure. The specificity value is .711, meaning that approximately 71% of students who passed the sixth-grade DCAS test scored received a score *equal to or greater than* the 12.50 NAEP cut score. In fifth grade, analyses indicate that students with scores < 15.50 should be identified as at-risk; the sensitivity value is .899 and specificity is .683.

Educational Significance

Understanding of fraction concepts is central to overall mathematics achievement in sixth grade (e.g., Siegler et al., 2012). The present study demonstrates how a relatively brief, curriculum-based measure of fraction concepts can help identify students who are at-risk for later difficulties. The NAEP fraction concepts screener given in fourth and fifth grades is a powerful diagnostic tool and has real-world applicability; practitioners can access the items and apply findings from our study to interpret student scores. A practitioner can use the screener to identify students in need of an intervention, utilizing a sensitivity-based cut score to avoid missing students who truly are at-risk. Furthermore, practitioners can look at student responses on the fraction items to help identify which concepts (i.e., fraction comparison, fraction equivalence, etc.) are most problematic for these struggling learners and use this information to inform instruction.

References

American Institutes for Research (2012). DCAS 2011-2012 Technical Report. Retrieved from http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/111/Vol1_Annual_TechRep.pdf

Cummings, K. D., & Smolkowski, K. (2015). Bridging the Gap Selecting Students at Risk of Academic Difficulties. *Assessment for Effective Intervention*, *41*(1).

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*(3), 839-843. doi: 10.1148/radiology.148.3.6878708

Hansen, N, Jordan, N.C., & Rodrigues, J. (in press). Identifying persistent learning difficulties in fractions: A longitudinal study of student growth from third through sixth grade. *Contemporary Educational Psychology.* doi: doi:10.1016/j.cedpsych.2015.11.002

Hecht, S. (1998). Toward an information-processing account of individual differences in fraction skills. *Journal of Educational Psychology, 90*, 545-59. doi:10.1037//0022-0663.90.3.545

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sens screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39*(2), 181.

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology, 116*(1), 45-58. doi:10.1016/j.jecp.2013.02.001

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010) *Common Core State Standards for Mathematics*. Washington DC: Author.

Resnick, I., Jordan, N. C., Hansen, N., Rajan, V., Rodrigues, J., Siegler, R. S., & Fuchs, L. S. (2016). Developmental growth trajectories in understanding of fraction magnitude from fourth through sixth grade. *Developmental Psychology*.Advance online publication. doi:10.1037/dev0000102

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences, 21*(5), 536-542. doi:10.1016/j.lindif.2011.05.002

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691-697. doi:10.1177/0956797612440101

Technical Report for the Delaware Comprehensive Assessment System (DCAS). (2012). Delaware Department of Education.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress. (1990-2009). Mathematics assessment. Retrieved from http://nces.ed.gov/nationsreportcard.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, *39*(2), 204-221.