In press at Journal of Learning Disabilities, accepted for publication 2/22/2018

Quality of Explanation as an Indicator of Fraction Magnitude Understanding

Lindsay Foreman-Murray, M.A.

Peabody College, Vanderbilt University

One Magnolia Circle, Nashville, TN, 37212

lindsay.r.james@vanderbilt.edu

Lynn S. Fuchs, Ph.D

Peabody College, Vanderbilt University

One Magnolia Circle, Nashville, TN, 37212

lynn.fuchs@vanderbilt.edu

This research was supported in part by grant R324D130003 from the Institute of Education Sciences in the U.S. Department of Education to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences or the U.S. Department of Education.

Abstract

Student explanations of their mathematical thinking and conclusions have become a greater part of the assessment landscape in recent years. With a sample of 71 4th-grade students at-risk for mathematics learning disabilities, we investigated the relation between student accuracy in comparing the magnitude of fractions and the quality of students' explanations of those comparisons, as well as the relation between those measures and scores on a criterion test: released fraction items from the National Assessment of Educational Progress (NAEP). We also considered the extent to which reasoning and language contribute to the prediction. Results indicated a significant, moderate correlation between accuracy and explanation quality. Commonality analyses indicated explanation quality accounts for little variance in NAEP scores beyond what is accounted for by traditional measures of magnitude understanding. Implications for instruction and assessment are discussed.

Quality of Explanation as an Indicator of Fraction Magnitude Understanding

With the passage of the Individuals with Disabilities Education Act in 1997 and subsequent reauthorization in 2004, schools became newly accountable for ensuring that students with learning disabilities (LD), meet the same standards as typically developing students, in part by requiring that students with LD participate in high-stakes testing (Thurlow & Johnson, 2000). These requirements may benefit students with risk for or identified with LD in some ways, by increasing expectations, promoting the hiring of better-qualified teachers, and encouraging use of evidence based practices (Vannest, Mahadevan, Mason, & Temple-Harvey, 2009).

However, these changes have had not substantially improved the learning outcomes of students with high incidence disabilities, as evidenced by the results of the 2015 National Assessment of Education Progress (NAEP), on which only 16% of these students scored “Proficient” or above on the fourth-grade mathematics assessment (Nations Report Card, 2017). Poor performance has led to increased likelihood of grade retention, as many states use high-stakes test scores as a gate to promotion (Allensworth, 2005; Roderick, Anthony, Brian, Easton, & Allensworth, 1999). A single non-promotion at the conclusion of any grade from eight to 12 doubles a student’s likelihood of dropping out of school (Rumberger & Larson, 1998).

There has been a recent trend in mathematics assessment, accelerated by the Common Core State Standards (CCSS) and subsequent versions of College- and Career-Ready Standards (CCR), toward requiring students to provide explanations as an indicator of their understanding of mathematics work. Figure 1 shows an example of this type of item. This shift reflects a belief that explanation quality is a more accurate indicator of student understanding than more traditional assessment response formats, reflecting conceptual rather than procedural understanding of mathematical concepts (Glaser, 2015; Kilpatrick, Swafford, & Findell, 2001;

Matthews & Rittle-Johnson, 2009; Niemi, 1996). While on a given assessment only one or two items may reflect this trend, the shift toward an answer format with unknown implications for students with risk for or identified with LD nonetheless warrants investigation.

Common Core State Standards are divided into the Standards for Mathematical Content, which delineate the knowledge and procedural skill mathematically proficient students are expected to attain at each grade level, and the Standards for Mathematical Practice, which describe the ways in which students should engage with mathematics as they progress through elementary, middle, and high school and depth of understanding. The content standards and practice standards are connected by practitioners in lessons, linking practical mathematical engagement with the 'understanding' standards outlined in the content standards. It is notable that the practice standards do not explicitly require students to provide written explanations for their mathematical work or conclusions. Instead, the practice standards implicitly suggest the need for explanations by requiring that students "construct viable arguments… justify their conclusions, communicate them to others, and respond to the arguments of others" (CCSS, 2013, p. 6-7).

The lack of an explicit requirement that students provide written explanations for their mathematical work in CCSS indicates this form of assessment is advisable only insofar as it provides a more nuanced or accurate assessment of students' understanding. Despite the movement toward the use of explanation measures on high-stakes tests, few studies have examined the relation between traditional measures of mathematical understanding and the quality of students' explanations. Our search of the literature indicates that the use of self-explanation is widely recommended as a means of assessment despite little empirical support.

Further, little is known about the predictors of student success on assessments requiring mathematical explanations.

The present study extends the literature in these areas, focusing on student explanations of fraction magnitude understanding, a key measurement interpretation topic at fourth grade, and focusing on a sample of students with risk for or identified with LD. In this study, we sought information related to three primary areas of inquiry. First, we determined which of our measures account for variability in explanation quality. Second, we assessed the relation between the quality of student explanations and their accuracy in comparing fraction magnitudes. Finally, we examined the predictive strength of explanation quality relative to other indicators of success on a criterion measure.

In this introduction, we provide background information on three bodies of research underpinning our investigation. First, we summarize research into the use of self-explanation as an instructional tool for deepening conceptual understanding. This relation underlies the movement toward using explanation as an indicator of student learning. Next, we discuss research on the use of explanation as an assessment device to index student understanding of mathematical concepts. Then, we provide an overview of research examining the role of cognitive predictors of mathematical and the development of fraction knowledge.

Self-Explanation as an Instructional Tool

Self-explanation as an instructional technique has been incorporated in mathematics instruction since at least the 1980s (Kelley, 2011). Self-explanation in instruction occurs when students generate explanations to aid in making sense of new information (Chi, 2000; Rittle-Johnson, 2006). For example, a student might explain his or her procedure for solving eight minus two by saying, “The first number is eight, and then two. I have to count down from the first number by the second number, so that leaves six.” These explanations may be spontaneously generated, elicited without support by a teacher, or elicited with support during instruction (Fuchs et al., 2016).

Guidelines issued by The National Council of Teachers of Mathematics (NCTM) recommends that teachers encourage students to use language to express mathematical ideas (2000). Self-explanation is believed to aid students in mastering diverse mathematical skills and concepts including number representation, number words, the base-10 system, decimal place value, and the connection between physical or graphic representations of mathematics problems and their numerical representations (Kilpatrick et al., 2001). Whitenack and Yackel (2002) recommend having students explain their work aloud to classmates as a way of deepening collective understanding and demonstrating different approaches to solving problems. Use of self-explanation has also been recommended as a means of helping students to retain their learning (Kilpatrick et al., 2001).

Research provides mixed support for self-explanation as an instructional tool. Rittle-Johnson (2006) demonstrated that self-explanation can help students develop conceptual understanding and transfer mathematical skills to unfamiliar problems in third through fifth-grades, although effects were not stronger than other instructional conditions. Rittle-Johnson’s recent work indicates that, while self-explanation is generally effective for promoting learning in some domains, its use limits learning in other areas: inhibiting the acquisition of certain types of knowledge even as it promotes the acquisition of others (Rittle-Johnson & Loehr, 2016). McEldoon and colleagues (2013) established the benefits of explanation in promoting conceptual and procedural understanding for elementary students with lower levels of mathematics understanding, but as with Rittle-Johnson, student outcomes were not stronger than with other forms of instruction.

We located one study investigating the use of self-explanation as an instructional tool for students with risk for or identified with LD. Fuchs and colleagues (2016) compared three conditions: supported self-explaining embedded within a fractions multicomponent intervention, word-problem solving (to control for fractions instructional time) embedded within the same multicomponent fractions intervention, and a business-as-usual control group. The sample comprised students at risk for mathematics LDs. Students in the explanation condition were explicitly taught to generate explanations of their solutions to fraction comparison problems. In the word-problem condition, schema-based approaches to solving different types of word problems were taught.

Fuchs and colleagues (2016) found that students taught to provide high-quality explanations were more accurate in comparing fraction magnitude than both students in the word-problem condition (and compared to students in the control group). Students in the explanation condition also produced higher-quality explanations. These results indicate the efficacy of supported self-explaining in improving the quality of students' explanations and enhancing fraction magnitude understanding.

Explanations as an Assessment Tool

Many studies have made use of self-explanation as a measure of understanding (e.g., Zhang et al., 2013), but without investigating the relation between the quality of students' explanations and other indicators of understanding. Studies examining the strength of explanation quality in predicting student performance on other measures of fraction knowledge have consistently found a relation between the two.

Niemi (1996) tested measures of explanation quality and justification. For the explanation measure, students explained fractions to an imagined audience, using pictures to support their explanations. The quality of the explanations was scored according to a six-point rubric. The justification measure asked students to solve fraction problems and then justify their answers using pictures and words. Students earned one point for correctly solving the problem and one point each for including a verbal or graphic justification. The quality of the justifications was not assessed. Students were classified as belonging to groups of high, moderate, or low representational fluency. Students belonging to the high-fluency group produced stronger explanations than their lower-fluency peers, with particularly strong results in their explanation of conceptual knowledge. The high-fluency group also produced more justifications than their peers with lower fraction fluency.

Niemi (1996) also examined correlations between the explanation elements, justifications, and outside measures of mathematics understanding (teacher ratings and performance on a criterion measure). The author found moderate correlations between the external measures and students' likelihood of producing a justification, and weak to moderate correlations between external measures and the explanation quality ratings.

Niemi's (1996) results suggest that explanation and justification measures accurately indicate students' fraction understanding. However, the binary scoring (present/absent) of the justification measure does not allow for analysis of the quality of those justifications. It also does not provide the opportunity to assess the relation between the accuracy of students' problem solving and the quality of their justification.

Nicolaou and Pitta-Pantazi (2014) examined the relation between student understanding of fractions and facility with definitions and explanations about fractions, as well as arguments about and justifications for answers to fraction problems. In the analysis, students were grouped into three levels of fraction understanding (low, medium, and high) using latent class analysis. Results indicated that only students in the highest level of fraction understanding were proficient in defining and explaining fraction concepts and in justifying their answers to problems involving fractions. These suggest promise for explanations in distinguishing strong understanding from moderate or low levels of mastery, but the study design does not permit nuanced assessment at lower levels of student understanding. As with Niemi's 1996 study, the uncoupling of explanations from traditional measures of fraction understanding prohibits analysis of the relation between explanation quality and performance in solving fraction problems.

More recent work has investigated the relation among students' writing skill, computational skill, and mathematical writing. Hebert and Powell (2016) found that fourth graders have difficulty using mathematical vocabulary to express mathematical ideas, and called for instruction to directly address vocabulary words (e.g., for fractions, *numerator, denominator, equal parts*) related to mathematics to aid students in producing mathematically accurate writing. In a related study, Powell and Hebert (2016) examined correlations among general writing ability, mathematical writing ability, and computational skill. Student performance on the computational and general writing tasks was moderately correlated with the mathematical writing task, suggesting the two types of writing do not represent the same skill set, and students cannot transfer skills from the writing tasks to computational skill.

These findings indicate that mathematical writing may not be the best indicator of conceptual understanding, as it requires a set of explanatory skills and mathematical vocabulary that students with risk for or identified with LD are poorly prepared to leverage. The authors conclude that students may require specific instruction in mathematics writing to be successful on assessment items requiring the use of writing to express mathematical ideas. Therefore, other measures might be more reliable indicators of student understanding.

Research has established that question format is an important determinate of performance of students with risk for or identified with LD. In a 2012 study of third-grade students, Powell (2012) demonstrated that students with mathematics difficulties are more successful in answering multiple-choice items than constructed response items, when the question construct was controlled in the analysis, even though a correction procedure to account for the effect of guessing was employed in scoring the assessment. These results indicate that for students in elementary grades with mathematics difficulties, multiple-choice rather than constructed response items lead to stronger performance on assessments. One hypothesis following from this finding is that the added demands of using written language to express mathematical ideas may place a burden on students who are at risk for LD and whose language ability is limited. Another is that requiring students to explain their answers calls upon reasoning skills leveraged differently in answering other types of questions.

Cognitive Predictors of Mathematical Achievement

To pursue the relation between language, reasoning, and explanation quality further, we administered measures related to these cognitive processes. This line of inquiry is supported by a number of studies investigating the role of a variety of cognitive processes in predicting mathematical development (e.g., Fuchs et al., 2013, 2014, 2016; Seethaler et al., 2011; Hansen et al., 2015; Jordan et al., 2013). Language comprehension likely plays a significant role in predicting explanation quality, as students rely on language skills both to understand verbal instruction from teachers and to articulate their explanations. The role of language comprehension in mathematical development has been demonstrated (Fuchs et al., 2013; Jordan et al., 2013). Reasoning ability is critical in supporting students to make connections between mathematical concepts and representations and in problem solving. The relation between reasoning ability and mathematical development has been established in longitudinal studies (Fuchs et al., 2013; Seethaler et al., 2011). While these cognitive processes are critical to answering other types of test items, the way students leverage them may be different for different problem types.

Summary of Present Study's Purpose and Hypothesis.

The primary purpose of the present study was to investigate the strength of explanation quality as a measure of magnitude understanding for fourth-grade students with or at risk for LD. We included students qualifying as "at risk" in an attempt to capture data from students with mathematics difficulty who may not be identified as having a specific learning disability in mathematics, but may nonetheless experience the difficulties faced by students with LD in explaining their work. This strategy increases the likelihood of including students who may later be identified with LD, but have not yet been referred for evaluation or who are receiving some instruction in the second tier of RTI for mathematics.

First, to investigate predictors of explanation quality, we explored the relation between the quality of students' explanations and their language and reasoning skills. We hypothesized that language would be the stronger predictor, given the demands written response formats make on language. Next, we examined the relation between explanation quality and students' accuracy in comparing fraction magnitudes. We hypothesized a moderate correlation between these two measures of understanding, based on the results of Powell and Hebert's 2016 study.

Finally, to investigate the relation among explanation quality, more traditional measures of magnitude understanding, language ability, and reasoning, we ran a complete commonality analysis using performance on the NAEP as the outcome. We hypothesized that explanation quality is a weaker predictor of NAEP scores than traditional measures of magnitude understanding based on Niemi's (1996) findings. Given the language demands inherent in writing explanations, we expected language ability to be more strongly predictive of the quality of students' explanations than a measure of reasoning, and that these two measures would account for a moderate proportion of the variance in explanation quality.

Method

Participants

Participants were drawn from a parent study (Fuchs et al., 2016), conducted with 236 children from 52 classrooms in 14 schools in a southeastern metropolitan district. The parent study investigated the effects of teaching fourth-grade students to provide explanations on their mathematics performance. Participants were identified as at-risk for mathematics difficulty based on scoring below the 35th percentile on a broad-based calculations assessment at the start of the parent study (Wide Range Achievement Test-4 [WRAT]; Wilkinson & Robertson, 2006). Fifteen students who scored below the 9th percentile on both subtests of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) were excluded from the parent study sample.

Participants in the present analysis were the 71 students who had qualified as at-risk for mathematics difficulty and had been randomized into the control group's parent study. The mean WRAT standard score for these students was 85.41 (*SD* = 7.56); the mean WASI was 93.27 (*SD* = 11.40). The sample was 52% male; 19% were English-language learners (ELL), and 13% were receiving special education services. African-American students made up 43% of the sample; non-Hispanic white students, 21%; Hispanic students, 30%; the other 6% were of other race/ethnicities. There were no significant differences between the sample for the present study and that of the parent study on pre-intervention performance or demographics.

Screening Measures and Cognitive Predictors of Outcome

With WRAT-4-Math Computation (Wilkinson & Robertson, 2006), students solve calculation problems of increasing difficulty; alpha for the sample of the parent study was .87. The test was administered to groups of students by a research assistant (RA). The WASI (Weschler, 1999) is a measure of general cognitive ability composed of two subtests. With *WASI Vocabulary* (Weschler, 1999), students identify pictures and define words. As per Zhu (1999), split-half reliability is .86-.87. With *WASI Matrix Reasoning* (Weschler, 1999), students to choose between provided options that best complete a visual pattern. Zhu reports reliability at .94. These measures were administered to students individually by a RA. These were used to index students' skills with vocabulary and reasoning, for the purpose of examining cognitive predictors of performance on measures of fraction magnitude understanding. All three measures were double-scored by two different RAs, with disagreements resolved by consultation with a project coordinator.

Measures of Fraction Understanding

The measure of students' magnitude understanding and ability to explain their answers was based on performance on *Explaining Fraction Magnitude Comparisons* from the *Fraction Battery-revised* (Schumacher, Namkung, Malone, & Fuchs, 2013). The subtest includes nine items, each of which consists of two components. First, students place a greater or less than sign between two fractions. Second, students use written words and a drawing to explain which fraction was the greater or lesser magnitude. Items are evenly divided between same-numerator, same-denominator, and different-numerator/different-denominator problems. For each item, students can earn a point for accurately comparing the fractions (maximum score = 9), and three points for the quality of the explanation, as follows: one point for indicating that the numerator represents the number of parts in a fraction, one point for indicating that the denominator represents the size of the parts, and one point for including an accurate drawing comparing fraction magnitudes.

Examples of student responses are included in Figures 2 and 3. Figure 2 shows a response earning 2 points for explanation quality ("Same number of parts but forths [sic] are bigger"). Responses were not penalized for misspellings. This response earned an additional point for a drawing (two units of the same size, each with 3 parts shaded, with one divided into 4 parts and one divided into 6 parts). Figure 3 shows a 0-point response in which the student used whole-number logic in the explanation ("6 is bigger than 4") and provided no drawing. Student responses generally conformed to the length and depth of these examples.

Despite the potential for students to guess correctly on this measure given that there are only three possible answers (>, <, =), we found significant differences between the control and experimental groups on this measure in the parent study, lending credibility to the measure. Also, similar comparing tasks are commonly used in the research literature as an index of fractions magnitude understanding (Fuchs et al., 2013; Geary, Nicholas, Li, & Sun; 2017; Rinne, Ye, & Jordan, 2017). Alpha on this sample was .68 for accuracy and .81 for quality. Two coders worked independently to score the tests, with 100% agreement for accuracy of magnitude comparisons and 99.1% agreement for explanation quality, computed point by point. Discrepancies were resolved through discussion.

The criterion measure of fractions understanding was based on performance on 19 released items from the 1990-2009 NAEP. This includes easy, medium, and hard items from the fourth-grade assessment and easy items from the eighth-grade assessment. Eight items assess part-whole understanding (given a circle divided into five parts, the student shades 2/5), nine items assess measurement understanding (given several lists of fractions, the student circles the answer choice showing the fractions in order from least to greatest), and one question asks students how many fourths make a whole. Multiple-choice questions comprise 11 items, written response three items, placing a mark on a number line two items, shading a portion of a shape one item, short-answer one item, and one item presented as an open response which students completed by writing numbers, shading a shape, and explaining their answer. The maximum possible score for this section is 25. Alpha for the sample was .79.

Results

What Accounts for the Quality of Students' Explanations?

We investigated the contribution of two measures of cognitive skill, WASI Vocabulary and Matrix Reasoning, to the variability in explanation quality. To test the independent contributions of these two predictors to explanation quality (EQ), we ran a regression analysis using WASI Vocabulary as an indicator of language ability and WASI Matrix Reasoning as an indicator of reasoning ability, and predicting EQ. Table 1 shows results.

We found that language and reasoning accounted for a comparable proportion of the variance in EQ, indicating that language skill is not primarily responsible for students' success or difficulty in providing high-quality explanations. In further analysis (see Table 2), language was a stronger predictor of comparison accuracy (CA) than EQ. This suggests that language ability is not a stronger determinant of success on the EQ measure than for the more traditional measure of magnitude understanding. This ran contrary to our hypothesis that language ability would be a significant and greater contributor to student success on this measure that reasoning ability.

What is the Relation between Comparison Accuracy and Explanation Quality?

Table 3 provides means, *SD*s, and correlations among the predictors and NAEP. Raw and standard scores are shown where applicable. To address our second research question, we considered correlations between the two components of the *Explaining Fraction Magnitude Comparisons* measure: EQ and CA. Due to the skewness of EQ variable, we used Spearman's rank order correlation in all analyses involving that variable. First, we investigated the descriptive statistics for the two components. The mean for CA was 5.08 (maximum score = 9; *SD* = 2.41). For EQ, the mean was 1.27 (maximum score = 27; *SD* = 2.55). The mean for EQ was expected to be low for this sample due to the unfamiliarity of the measure for the students, who received no specific support in learning how to write explanations for their work. By contrast, evidence from the study from which these data were drawn (Fuchs et al., 2016) indicates that students explicitly taught to write explanations for their solutions to magnitude comparison problems during instruction showed substantially higher scores in explanation quality than did this sample. As hypothesized, the correlation between the two components was moderate and positive ($r = .51$, $p < .001$), supporting the view that the two measures of understanding are related but not synonymous, and represent related but different constructs.

What Predicts NAEP Scores?

To address our third research question, we conducted a complete commonality analysis, a method of investigating results that accommodates collinear variables within multiple regression analyses (Nimon, 2010; Siebold & McPhee, 1979). Commonality analysis partitions the variance accounted for in a regression model into non-overlapping parts accounted for by each predictor variable and each combination of variables, allowing researchers to identify the unique contribution of each variable, as well as the shared variance between variables. The analysis involves running multiple hierarchical regression models with predictor variables entered in all possible subsets of orders, with the researcher recording the independent contribution of each variable to the regression effect.

Table 4 shows the results of regression analyses predicting NAEP with the four indicator variables; Table 5 shows results of the commonality analysis, which provides estimates of the unique and shared variance associated with each predictor individually and in combination with each other predictor or group of predictors of NAEP. In the second column, we report the proportion of total variance accounted for by the predictor(s). In the third column, we report the percentage of explained variance accounted for by the predictor(s). To obtain this percentage, we took the proportion of total variance explained by the predictor(s) and divided it by the total explained variance across predictors, then multiplied by 100.

The purpose here was to examine the performance of EQ as a predictor of NAEP scores relative to CA, a more traditional measure of magnitude understanding, as well as its relation to language and reasoning, the two measures of cognitive processing. The unique variance accounted for by EQ was 1.17%, by far the smallest of the four predictors. Language was the strongest predictor, uniquely accountable for 15.37% of explained variance. CA and reasoning accounted for a similar proportion of explained variance, with 11.97% and 10.36% respectively. Together, EQ and CA account for another 7.01% of the explained variance, suggesting that the predictive value of EQ is largely duplicated by CA.

In explaining NAEP scores, the four predictors accounted for 42.1% of the variance, $F(3, 71) = 12.02$, $p < .001$. Three variables were uniquely predictive of NAEP scores; only EQ did not make a unique contribution. Language and comparison accuracy were the strongest contributors. The relative weakness of explanation quality as a predictor of NAEP scores confirmed our hypothesis.

Discussion

The main purpose of this study was to test the strength of explanation quality as a measure of magnitude understanding for at-risk fourth-graders in comparison to traditional measures. We focused on explanations because of the requirement on high-stakes assessments that students explain or justify their solutions to mathematical problems, premised on the belief that explanations better reflect conceptual knowledge than traditional measures of magnitude understanding (Niemi, 1996). We also investigated the predictive power of traditional measures of magnitude understanding and cognitive processes hypothesized to contribute to performance on mathematical assessments.

What is the Relation between Comparison Accuracy and Explanation Quality?

We found that students were more successful in comparing fraction magnitudes than in providing quality explanations for their work, and that there was a moderate, positive correlation between the two measures, indicating that the measures reflect student understanding differently. These data are consistent with findings by Powell (2012), who showed students with disabilities were more successful on assessments when items were presented in multiple-choice format than when they took the form of constructed response. These results indicate that the way students with risk for or identified with LD are asked to demonstrate their understanding plays a role in determining their success. Our results similarly support the idea that students may be less successful when asked to demonstrate their understanding using words and drawings instead of choosing a symbol to represent a magnitude difference.

Further research is necessary to parse this difference. One possibility, often cited by those favoring the use of explanations in assessment (Niemi, 1996), is that the comparison measure allows students to rely primarily on procedural knowledge to correctly solve problems. A procedural approach to comparing fraction magnitudes often seen in classrooms in the school district where the study took place (Malone & Fuchs, 2017) is cross-multiplying, or "the butterfly method," which allows students to arrive at the correct solution without applying any conceptual understanding of fractions. Students using this method could accurately compare fraction magnitudes, but would be unlikely to provide a high-quality explanation for their work due to limited conceptual understanding.

An alternative explanation for the disparity in student performance on these two measures is a difference in the skills required to successfully complete the different problem types. Explanations require students to call upon mathematics vocabulary, which may be lacking in classroom instruction. For example, if students have not discussed fractions in terms of important constructs and vocabulary (e.g., *numerator, denominator, equal parts*), they are likely incapable of calling upon relevant vocabulary to produce explanations (Hebert & Powell, 2016). This may lead to inaccurate or missing justifications for their answers. Relatedly, if students have not received instruction directly targeting mathematical writing, they may be unable to leverage the vocabulary they do understand to create high-quality explanations. Finally, using written language in any subject area likely poses a challenge for students with limited writing skill.

What is the Relation between Language and Reasoning in Predicting Explanation Quality?

Our analyses showed language and reasoning accounted for a comparable proportion of the variance in the quality of students' explanations, and that language is a stronger predictor of comparison accuracy than explanation quality. These results indicate that students' difficulty in providing high-quality explanations is not primarily a function of language skill. Our data suggest that students specifically with poor language may not be uniquely disadvantaged when it comes to assessment items requiring explanations. Further research is warranted into the cognitive processes associated with explanation quality to determine if there are other viable mediators of student performance on these measures.

It is notable that the measure of language skill used in this analysis is domain-general, and does not specifically address mathematical vocabulary. Future research investigating the relation of mathematical vocabulary and mathematical writing skill to explanation quality would provide a more detailed view of the contribution of those qualities. These outcomes support the recent argument by Powell and Hebert (Hebert & Powell, 2016; Powell & Hebert, 2016) that students require targeted instruction in mathematical writing and vocabulary to be successful on tasks requiring them to leverage these skills, and that general mathematical and writing instruction is not sufficient.

Does Explanation Quality Predict Performance on a Criterion Assessment?

The commonality analysis revealed the relative weakness of explanation quality as a predictor of performance on NAEP. Comparison accuracy, language, and reasoning were all uniquely predictive of NAEP scores, accounting for substantially higher percentage of the variability, and leaving explanation quality as the only predictor in the model that was not uniquely predictive of the outcome.

The strength of the comparison accuracy measure in contrast to the weak predictive power of explanation quality is notable, as fraction magnitude comparisons are already widely used to measure magnitude understanding (Fuchs et al., 2013, 2014, 2016; Hansen et al., 2015; Powell & Hebert, 2016). The movement toward explanation quality as a better indicator of student understanding, as reflected in success on the criterion measure, is not supported by these data, which indicate that the existing measure better reflects students' mathematical skill.

The predictive strength of the language and reasoning measures is not surprising, as these are powerful sources of domain-general cognitive processing (see, for example, Harvey & Miller, 2017; Purpura, Hume, Sims, & Lonigan, 2011; Tobia, Bonifacci, & Marzocchi, 2016; Ribeiro, Cadime, Freitas, & Viana, 2016), as is demonstrated in their use within tests of intellectual ability. Language and reasoning would be expected to contribute substantially to student outcomes on many assessments of mathematical understanding. As revealed in the commonality analysis, variance common to these measures and comparison accuracy accounts for the bulk of the explained variance in the model. By contrast, variance common to these predictors and explanation quality alone account for little explained variance. This indicates that domain-general measures of skill and a measure of magnitude understanding already in wide use account for the majority of the predictive power of the model.

Implications for Assessment and Instruction

Results from this investigation have several implications for assessment design and classroom instruction. First, assessors should be wary of including measures requiring the explanation of mathematical ideas without fully considering the constructs they intend to evaluate. Our results indicate that explanation items are likely to increase the difficulty of the assessment for students at risk for or diagnosed with LD without adding predictive power when it comes to criterion measures of mathematical understanding. While it is possible that explanation quality reflects conceptual understanding, studies by Niemi (1996) and Nicolau and Pitta-Pantazi (2014) suggest that measures requiring explanation are more effective at differentiating students with high levels of understanding from those with less, without offering a nuanced assessment of the conceptual understanding of students at the lower end of that spectrum.

We speculate that many explanation measures require high levels of mastery to successfully complete, as well as strong language and reasoning skill, effectively grouping students with lower levels of ability across these measures together. This grouping precludes differentiation of their grasp of the concepts underpinning mathematics, defeating the purpose of employing a measure targeting conceptual understanding. More traditional measures of magnitude understanding employed in new ways might offer a more detailed picture of what students with moderate or low levels of conceptual understanding have mastered. Use of number lines, ordering problems, and drawing tasks may allow students with limited language ability or burgeoning conceptual understanding to demonstrate that developing knowledge.

Our results suggest that the use of explanation measures on high-stakes tests is likely to be detrimental to students with risk for or identified with LD, because they register less success on these measures than on more traditional measures of fraction understanding. This echoes the work of Powell and Hebert (Hebert & Powell, 2016; Powell & Hebert, 2016) who showed that students are unable to transfer general writing ability to mathematical writing tasks, and they lack the mathematical vocabulary to be successful on explanation measures. While only one or two items on a single test may reflect this trend, the shift toward an answer format that disadvantages students with risk for or identified with LD is troubling. With the serious consequences of low achievement on high-stakes tests for this group (Rumberger & Larson, 1998; Thurlow & Johnson, 2000; Vannest, Mahadevan, Mason, & Temple-Harvey, 2009), further research into the benefits of using explanation test items is warranted.

As long as mathematical explanation is included as a problem type on high-stakes assessments (albeit few items such items may occur on any given assessment), teachers must prepare their students for success on this response format, perhaps by incorporating explicit mathematical vocabulary and writing instruction and practice into instructional routines. This may important not only to increase student performance on such measures but also to support students' conceptual understanding of the material.

As noted earlier, two studies (Rittle-Johnson, 2006; Rittle-Johnson & Loehr, 2016) reveal the limitations of elicited self-explanation as an instructional technique, in which students are prompted but *not explicitly* taught to generate self-explanations. On the other hand, a randomized control trial conducted by Fuchs and colleagues (2016) indicates that instructional time devoted to high quality explanations is a productive instructional activity, as long as an *explicit* form of self-explaining instruction is used. In Fuchs et al. (2016) students were randomized to three conditions: one focused on *supported* self-explaining, one on deepening conceptual understanding without supported self-explaining, and a control condition. On traditional measures of magnitude understanding and on measures of explanation quality, the performance of students in the supported self-explaining was stronger than that of students in the competing intervention condition (with both intervention conditions outperforming the control group). Data from this study suggest that supported self-explaining instruction, in which teachers explicitly engage students to create high quality explanations using mathematics vocabulary, represents a productive investment of teachers' instructional time.

Limitations

Before closing, we note several limitations. First, because the writing demands of the explanations quality measure were not extensive, it is possible that the measure did not adequately tap the construct of explanations. Future research using multiple measures of explanation quality to derive this construct would produce stronger evidence. Second, our language measure specifically measured vocabulary, rather than language comprehension, limiting conclusions that can be drawn about the relation of the boarder construct to explanation quality, and likewise, the reasoning measure specifically addressed non-verbal reasoning. Future research using a range of measures related to language and reasoning would provide a fuller and more accurate assessment of these relations. Further, while our sample size was too small to meaningfully leverage moderator analysis, inclusion of moderators of language ability, particularly ELL and special education status, would extend this line of work in important ways.

References

Allensworth, Elaine M. (2005). Dropout Rates After High-Stakes Testing in Elementary School: A Study of the Contradictory Effects of Chicago's Efforts to End Social Promotion. *Educational Evaluation and Policy Analysis, 27*, 341–364.

Chi, M. T. H. (2000). Self-explaining: The dual processes of generating inference and repairing mental models. In R. Glaser (Ed.), Advances in instructional psychology: Educational design and cognitive science (Vol. 5, pp. 161-238).

Common Core State Standards. (2013). Washington D.C.: National Governors Association Center for Best Practices & Council of Chief State School Officers.

Hansen, N., Jordan, N.C., Fernandez, E., Sigler, R.S., Fuchs, L.S., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development, 35*, 34-49. https://doi.org/10.1016/j.cogdev.2015.02.001

Hebert, M. A., & Powell, S. R. (2016a). Examining fourth-grade mathematics writing: features of organization, mathematics vocabulary, and mathematical representations. *Reading and Writing*, *29*, 1511-1537. https://doi.org/10.1007/s11145-016-9649-5

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*, 45-58. https://doi.org/10.1016/j.jecp.2013.02.001

Kilpatrick, J., Swafford, J., & Findell, B., (Eds.) (2001). Adding it up: Helping children learn mathematics. Washington, DC: National Academy Press (by the Mathematics Learning Study Committee/National Research Council).

Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Available from http://quantpsy.org.

McEldoon, K. L., Durkin, K. L., & Rittle-Johnson, B. (2013). Is self-explanation worth the time? A comparison to additional practice. *British Journal of Educational Psychology, 83*, 615-632. https://doi.org/10.1111/j.2044-8279.2012.02083.x

Fuchs, L.S., Malone, A.S., Schumacher, R.F., Namkung, J., Hamlett, C.L., Jordan, N.C., … Changas, P. (2016). Supported self-explaining during fraction intervention. *Journal of Educational Psychology, 108*, 493-508. https://doi.org/10.1037/edu0000073

Fuchs, L.S., Schumacher, R.F., Long, J., Namkung, J., Hamlett., C.L., Cirino, P.T., … Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105*, 683-700. https://doi.org/10.1037/a0032446

Fuchs, L.S., Schumacher, R.F., Sterba, S.K., Long, J., Namkung, J., Malone, A.S., … Changas, P. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology, 106*, 499-514. https://doi.org/10.1037/a0034341

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology, 109*(5), 680-693. http://dx.doi.org/10.1037/edu0000159

Glaser, R. (Ed.). (2015). *Advances in Instructional Psychology, Volume 5: Educational Design and Cognitive Science*. Routledge; New York, NY.

Harvey, H. A., & Miller, G. E. (2017). Executive function skills, early mathematics, and vocabulary in head start preschool children. *Early Education and Development, 28*(3), 290-307. https://doi.org/10.1080/10409289.2016.1218728

Kelley, D. (2011). The self-explanation effect when learning mathematics: A meta-analysis. *Society for Research on Educational Effectiveness*. Retrieved from https://eric.ed.gov/?id=ED518041

Kilpatrick, J., Swafford, J., & Findell, B. (2001). Adding It Up: Helping Children Learn Mathematics. Washington, DC: National Academy Press.

Malone, A.S. & Fuchs, L.S. (2017). Error patters in ordering fractions among at-risk fourth-grade students. *Journal of Learning Disabilities, 50*, 337-352. https://doi.org/10.1177/0022219416629647

Matthews, P. & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology, 104*, 1-21. https://doi.org/10.1016/j.jecp.2008.08.004

National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.

Nations Report Card. (2017, May). Retrieved from https://www.nationsreportcard.gov

Nicolaou, A.A. & Pitta-Pantazi, D. (2016). Hierarchical levels of abilities that constitute fraction understanding. *International Journal of Science and Mathematics Education, 14*, 757-776. https://doi.org/10.1007/s10763-014-9603-4

Niemi, D. Assessing conceptual understanding in mathematics: Representations, problem solutions, justifications, and explanations. *The Journal of Educational Research, 89*, 351-363. https://doi.org/10.1080/00220671.1996.9941339

Nimon, K. (2010). Regression commonality analysis: Demonstration of an SPSS solution. *Multiple Linear Regression Viewpoints, 36*, 10-17.

Perie, M., Grigg, W.S., and Dion, G.S. (2005). *The Nation's Report Card: Mathematics 2005* (NCES 2006–453). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, D.C.: U.S. Government Printing Office. Available from https://nces.ed.gov/nationsreportcard/pubs/main2005/2006453.asp

Perie, M., Grigg, W.S., & Donahue, P.L. (2005). *The Nation's Report Card: Reading 2005* (NCES 2006–451). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, D.C.: U.S. Government Printing Office. Available from https://nces.ed.gov/nationsreportcard/pubs/main2005/2006451.asp

Powell, S.R. (2012). High-stakes testing for students with mathematics difficulty: Response format effects in mathematics problem solving. *Learning Disability Quarterly, 35*, 3-9. https://doi.org/10.1177/0731948711428773

Powell, S.R. & Hebert, M.A. (2016b). Influence of writing ability and computation skill on mathematics writing. *The Elementary School Journal, 117*, 310-335. https://doi.org/10.1086/688887

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology, 110*, 647-658. https://doi.org/10.1016/j.jecp.2011.07.004

Ribeiro, I., Cadime, I., Freitas, T., & Viana, F. L. (2016). Beyond word recognition, fluency, and vocabulary: The influence of reasoning on reading comprehension. *Australian Journal of Psychology, 68*(2), 107-115. https://doi.org/10.1111/ajpy.12095

Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology, 53*(4), 713-730. http://dx.doi.org/10.1037/dev0000275

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*, 1-15. https://doi.org/10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., & Loehr, A. M. (2016). Eliciting explanations: Constraints on when self-explanation aids learning. *Psychonomic Bulletin & Review.* https://doi.org/10.3758/s13423-016-1079-5

Roderick, Melissa, Anthony S. Bryk, Brian A. Jacob, John Q. Easton, and Elaine Allensworth. (1999). *Ending Social Promotion: Results from the First Two Years.* Chicago, IL: Consortium on Chicago School Research.

Rumberger, R.W., & Larson, K.A. (1998). Student mobility and the increased risk of high school drop out. *American Journal of Education, 107*, 1-35. https://doi.org/10.1086/444201

Schumacher, R. F., Namkung, J. M., Malone, A., & Fuchs, L. S. (2013). *2013 Fraction Battery-revised*. Available from L.S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203.

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*, 536-542. https://doi.org/10.1016/j.lindif.2011.05.002

Siebold, D.R., & McPhee, R.D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analysis. *Human Communication Research, 5*, 355-363. https://doi.org/10.1111/j.1468-2958.1979.tb00649.x

Thurlow, M.L. & Johnson, D.R. (2000). High stakes testing of students with disabilities. *Journal of Teacher Education, 51*, 301-314. https://doi.org/10.1177/0022487100051004006

Tobia, V., Bonifacci, P., & Marzocchi, G. M. (2016). Concurrent and longitudinal predictors of calculation skills in preschoolers. *European Journal of Psychology of Education, 31*(2), 155-174. https://doi.org/10.1007/s10212-015-0260-y

Vannest, K.J., Mahadevan, L., Mason, B.A., Temple-Harvey, K.K. (2009). Educator and administrator perceptions of the impact of no child left behind on special populations. *Remedial and Special Education, 30*, 148-159. https://doi.org/10.1177/0741932508315378

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation. https://doi.org/10.1037/t15170-000

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources. https://doi.org/10.1037/t27160-000

Zhang, X., Clements, M.A. (Ken), Ellerton, N.F. (2015). Enriching student concept images: Teaching and learning fractions through a multiple-embodiment approach. *Mathematics Education Research Journal, 27*, 201-231. https://doi.org/10.1007/s13394-014-0137-4

Zhu, J. (1999). *WASI manual*. San Antonio, TX: Psychological Corporation.

Table 1
Regression Models Predicting Explanation Quality

Predictors	*B*	*SE*	*β*	*t(1, 71)*	*p*	R^2
Language (L)	0.08	0.04	0.21	1.84	0.07	0.05
Reasoning (R)	0.09	0.05	0.21	1.91	0.06	0.04
Total						0.11

Note: Language is *WASI Vocabulary* (Wechsler, 1999). Reasoning is *WASI Matrix Reasoning* (Wechsler, 1999);

Table 2
Regression Models Predicting Comparison Accuracy

Predictors	*B*	*SE*	*β*	*t(1, 71)*	*p*	R^2
Language (L)	0.11	0.04	0.27	2.46	<.05	0.07
Reasoning (R)	0.08	0.05	0.18	1.61	0.11	0.03
Total						0.12

Note: Language is *WASI Vocabulary* (Wechsler, 1999). Reasoning is *WASI Matrix Reasoning* (Wechsler, 1999);

Table 3

Means, Standard Deviations, and Correlations Among Predictors and NAEP (n = 71)

	Raw Score		*Standard Score*[a]		*Correlations*				
	M	(*SD*)	*M*	(*SD*)	*EQ*[b]	*CA*	*L*	*R*	*N*
Predictors									
Exp. Quality (EQ)*	1.27	2.41	NA						
Comp. Accuracy (CA)	5.08	2.55	NA		0.51**				
Language (L)	30.75	6.48	45.55	(0.8)	0.26*	0.30**			
Reasoning (R)	16.75	5.76	46.01	(9.77)	0.27*	0.23*	0.14		
Outcome									
NAEP (N)	12.61	3.86	NA		0.41**	0.55**	0.45**	0.43**	

Note: *Correlations are significant at ($p < .05$) **Correlations are significant at ($p < .001$). Language is *WASI Vocabulary* (Wechsler, 1999). Reasoning is *WASI Matrix Reasoning* (Wechsler, 1999); NAEP is based on performance on 19 released items from the 1990-2009 National Assessment of Educational Progress. [a] Standard scores for WASI Vocabulary and WASI Matrix Reasoning are *T* scores ($M = 50$; $SD = 10$). [b] To account for the skewness of the data, Spearman's rank order correlations were used for correlations with EQ

Table 4
Regression Models Predicting NAEP

Predictors	*B*	*SE*	*β*	*t(3, 71)*	*p*
Exp. Quality (EQ)	0.13	0.17	0.08	0.75	0.46
Comp. Accuracy (CA)	0.55	0.23	0.30	2.40	<0.05
Language (L)	0.17	0.06	0.28	2.72	<0.05
Reasoning (R)	0.15	0.07	0.23	2.23	<0.05

Note: Language is *WASI Vocabulary* (Wechsler, 1999). Reasoning is *WASI Matrix Reasoning* (Wechsler, 1999); NAEP is based on performance on 19 released items from the 1990-2009 National Assessment of Educational Progress

Table 5
Commonality Analysis for Predicting NAEP

Variable	*Proportion of explained variance*	*Percentage of explained variance*
Unique to:		
Explanation Quality (EQ)	0.005	1.17
Comparison Accuracy (CA)	0.050	11.97
Language (L)	0.065	15.37
Reasoning (R)	0.044	10.36
Common to:		
EQ + CA	0.029	7.01
EQ + L	0.003	0.59
EQ + R	0.003	0.58
CA + L	0.048	11.37
CA + R	0.048	11.28
L + R	-0.0009	-0.21
EQ + CA + L	0.034	8.01
EQ + CA + R	0.038	8.95
EQ + L + R	0.001	0.11
CA + L + R	0.024	5.59
EQ + CA + L + R	0.033	7.84
Total	0.422	100%

Sam did the following problems.

2 + 1 = 3

6 + 1 = 7

Sam concluded that when he adds 1 to any whole number, his answer will always be odd.

Is Sam correct? __________

Explain your answer.

Question ID: 2009-4M10 #11 M138201

Figure 1

(retrieved from https://nces.ed.gov)

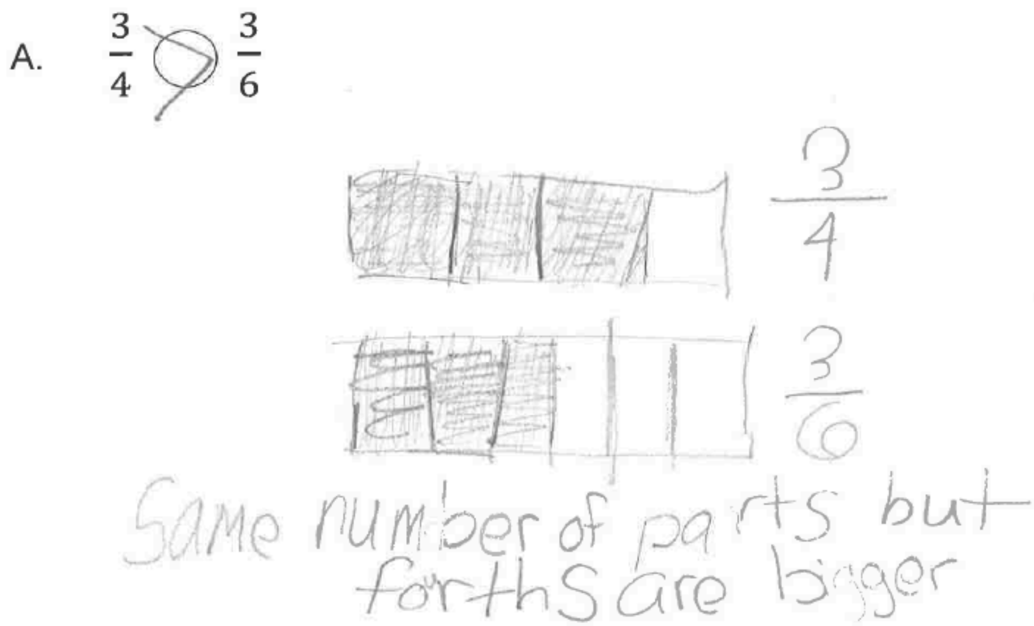

Figure 2

Figure 3