Comparing the Contribution of Teacher Versus Tutor Ratings of Inattentive Behavior in Predicting Mathematics Achievement

Amelia S. Malone and Lynn S. Fuchs

Remedial and Special Education published online 2 July 2014

DOI: 10.1177/0741932514539855

Article

Hammill Institute on Disabilities

Remedial and Special Education
1–9

DOI: 10.1177/0741932514539855
rase.sagepub.com
SAGE

Comparing the Contribution of Teacher Versus Tutor Ratings of Inattentive Behavior in Predicting Mathematics Achievement

Amelia S. Malone, PhD[1] and Lynn S. Fuchs, PhD[1]

Abstract

The purpose of this study was to assess the relative contribution of teacher and tutor ratings of inattentive behavior in two different instructional settings in predicting students' performance on fraction concepts and whole-number calculations. Classroom teachers rated each student's attentive behavior in a whole-class setting and tutors rated each student's attentive behavior in an intensive small-group setting. The sample was 131 fourth-grade students identified as at risk for developing mathematics difficulties (MD) and randomly assigned to fraction tutoring. Teachers and tutors rated each student's inattentive behavior in the 10th week of the 12-week intervention period. On average, tutors rated students as more attentive than teachers. Also, tutor ratings had greater predictive power than teacher ratings on fraction concepts, but neither set of ratings uniquely predicted performance on whole-number calculations. Possible explanations for findings and directions for future research are discussed.

Keywords

elementary school, mathematics instruction, teacher ratings, tutor ratings, inattentive behavior

Competence in mathematics is essential for success in school and the workplace. However, many individuals struggle to reason quantitatively, which hinders advancement in society. According to the Every Child a Chance Trust (2009), individuals with mathematics difficulties (MD) are more likely than those without MD to experience hardship (e.g., require special education services, have a low paying job) throughout their lifetime. However, mathematics learning is complex, and many students persistently experience poor mathematics achievement. For example, according to the National Assessment of Education Progress (NAEP; 2013), 58% of 4th-grade students have mathematics skills below the proficient level. Moreover, 95% of students with MD in 5th grade continue to perform at the same level in 11th grade (Shalev, Auerbach, Manor, & Gross-Tsur, 2000).

As these statistics suggest, improving mathematics achievement is imperative. Approximately 5% to 8% of students experience difficulty learning mathematics (Geary, 2004), and inattentive behavior has been linked to MD (e.g., Benedetto-Nasho & Tannock, 1999; L. S. Fuchs et al., 2006). Self-regulated and goal-directed sustained attention is required for learning about and succeeding with complex mathematical problems. If students cannot attend to a task, they will fail to learn the content; if they fail the task because they do not understand the academic content, they are likely to become frustrated and inattentive. This can become a cyclical pattern of inattentive behavior and academic underachievement, which may exacerbate MD.

Previous research supports the link between teacher ratings of inattentive behavior and academic achievement. For example, Merrell and Tymms (2001) conducted a 2-year longitudinal study examining the relation between teacher ratings of inattentive behavior, hyperactivity, and impulsivity on academic achievement among 4,148 students aged 4 to 7 years. Teacher ratings of inattentive behavior were most closely related to underachievement in both reading ($d = -1.07, p \leq .01$) and mathematics ($d = -1.18, p \leq .01$) on assessments at the end of the longitudinal study. That is, the higher a student's inattentive behavior (measured by teacher ratings), the lower his or her achievement in both reading and mathematics.

Similarly, Breslau et al.'s (2010) longitudinal study correlated teacher ratings of inattentive behavior with mathematics and reading achievement across a span of 11 years. Teachers rated each student's inattentive behavior at age 6

[1]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Amelia S. Malone, Department of Special Education, Vanderbilt University, Box #228, Peabody College, Nashville, TN 37203-5721, USA.
Email: amelia.malone@vanderbilt.edu

and again at age 11 using the *Teacher's Report Form*, which assesses a wide range of attention difficulties consistent with the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association [APA], 1994) criteria for diagnosing attention-deficit/hyperactivity disorder (ADHD). Persistent inattentive behavior in these early elementary grades predicted academic underachievement in both reading and mathematics in the secondary grades. If teachers rated students as distractible at age 6 and at age 11, their academic performance was likely to continue to decline through the secondary grades. However, if ratings of inattentive behavior changed positively from age 6 to age 11, academic performance was also likely to change in a positive direction.

In addition, Duncan et al. (2007) analyzed six longitudinal studies investigating the relation between school readiness predictors and later academic achievement and found that entry-level mathematics skills and inattentive behavior (measured by teacher ratings) were the most powerful predictors of later mathematics achievement. These findings are consistent with Barkley's (1997) view of the relation between inattentive behavior and academic difficulties, in which children with persistent inattention are at higher risk for low achievement. Inattentive behavior influences achievement because it disrupts student engagement in the classroom, which in turn influences students' ability to learn (Duncan et al., 2007).

More insight is needed on the role inattentive behavior plays in MD. One key issue in terms of teacher ratings of inattentive behavior is whether these ratings truly index inattentive behavior or whether they reflect teacher observations of students' general academic achievement. L. S. Fuchs et al. (2005) hypothesized that student inattentive behavior in the classroom may reflect a mismatch between instruction and ability level. If a student does not understand the academic content, he or she may become frustrated and inattentive, and teacher ratings may be a reflection of student response to this mismatch between instructional content and ability level. It is possible that teachers' perceptions of student achievement may cloud their ratings of inattentive behavior. As instruction becomes more aligned with students' needs, students likely become more capable of attending to the academic tasks. Therefore, it is plausible that ratings from an educator delivering more intensive small-group instruction designed to meet students' academic needs would reflect this increased attention, as opposed to ratings completed by the classroom teacher in a whole-class setting.

Understanding how ratings of inattentive behavior change as instruction becomes more aligned with student needs has important implications for designing and implementing intensive interventions in the Response to Intervention (RTI) framework. Within the standard treatment protocol approach of the RTI model, students who struggle at the classroom level (Tier I) receive more explicit and systematic instruction in a small-group setting to assist them in making adequate academic progress (Tier II). If students are capable of catching up to peers after receiving Tier II intervention, they return to the general education classroom. If, however, students do not make adequate progress, they progress to more intensive instruction (Tier III). As students move through the tiers, increased intensity is reflected in academic interventions that are longer, more individualized, teacher centered, explicit, and delivered by more highly trained instructors to smaller groups of students who have similar strengths and weaknesses (D. Fuchs & Fuchs, 2006).

In the multi-tiered model, the mismatch between instruction and ability level should be mitigated as intervention becomes more intensive, and students likely become more attentive because they are better able to self-regulate their behavior in a small-group setting (e.g., Schunk & Zimmerman, 1998). This increased attention should be reflected by ratings from a Tier II instructor (e.g., special educator, academic tutor, behavior specialist). The instructor would likely witness less inattentive behavior in a small-group setting when instruction is matched to students' learning needs, and therefore rate each student as more attentive. To better understand how ratings of inattentive behavior change when instruction becomes more intensive, we compared teacher ratings of inattentive behavior in the whole-class setting with tutor ratings of inattentive behavior during a Tier II fraction intervention (i.e., an intensive small-group setting). In this study, tutors were well-trained research assistants, who were full-time employees or graduate students at a local college of education.

We located only one study that investigated the difference between ratings of inattentive behavior as students receive more intense academic interventions. Strayhorn and Bickel (2002) found that, on average, tutors delivering instruction in a one-on-one setting rated students as more attentive than teachers rated students in a whole-class setting. Authors did not provide demographic information on tutors including what type of training tutors received or whether they were licensed educators. This study also did not assess whether such ratings have differential power in predicting academic outcomes. Greater predictive power for ratings of inattentive behavior from raters (e.g., tutors) in a more intensive small-group setting would support the notion that a mismatch between instruction and ability clouds teacher judgment in the whole-class setting. Also, in a more practical sense, if educators rating students during an intensive Tier II intervention have stronger power to predict academic achievement than teachers rating students in a whole-class setting, this may provide a better source for informing future diagnostic and/or remedial decisions for these at-risk students.

We were also interested in determining whether teachers and tutors had differential power in predicting performance on fraction concepts (proximal to tutoring) versus whole-number calculations (not related to tutoring). Across Cirino, Fletcher, Ewing-Cobbs, Barnes, and Fuchs (2007); L. S. Fuchs et al. (2006); Raghubar et al. (2009); and Seethaler and Fuchs (2006), teacher ratings of inattentive behavior uniquely predicted performance on 13 of 19 measures of whole-number knowledge. This included performance on small sums addition, large sums addition, small minuends subtraction, large minuends subtraction, estimation, and counting speed (Cirino et al., 2007); arithmetic, algorithmic computation, arithmetic word problems (L. S. Fuchs et al., 2006); accuracy, mathematics fact errors, and procedural "bugs" (Raghubar et al., 2009); and estimation skill (Seethaler & Fuchs, 2006). Teachers rated each student's inattentive behavior in a whole-class setting. We located no study that looked at the relation between teacher ratings of inattentive behavior in a whole-class setting and performance on fraction concepts. However, because previous research supports a relation between teacher ratings and mathematics performance (i.e., whole-number knowledge), we expected a similar relation between these ratings and performance on fraction concepts.

In the present study, teachers and tutors (i.e., research assistants) rated each student's inattentive behavior in the 10th week of a 12-week intervention on fractions. Teachers rated each student's inattentive behavior in a whole-class setting and tutors rated each student's inattentive behavior in an intensive small-group setting. These ratings were compared to determine whether tutors rated students as more attentive than teachers. That is, do ratings of inattentive behavior from an intensive small-group setting (i.e., from tutors) differ from ratings of inattentive behavior from a whole-class setting (i.e., from teachers)? We used these ratings to determine whether teacher and tutor ratings significantly predicted student outcomes on a measure of released fraction items from the NAEP and a measure of whole-number calculations. We also compared the predictability of teacher and tutor ratings to determine if ratings in different instructional settings (i.e., whole-class vs. intensive small-group instruction) affected the power in predicting students' performance on fraction concepts and whole-number calculations.

Three hypotheses guided the analysis. First, based on Strayhorn and Bickel's (2002) findings, we expected tutors (i.e., research assistants) to rate students as more attentive than teachers. Because tutors instructed students in a structured, small-group setting, students' attention likely increased because instruction was aligned with their academic needs. Also, the fraction intervention incorporated a stringent behavior management system, which held students accountable for on-task behavior and accurate work. These factors may contribute to students' increased attention in small-group tutoring, even if these students struggle to pay attention in the classroom because of motivational problems associated with a poor learning history. Attention issues should be mitigated when instruction is specifically designed to target students' instructional level and address their learning needs.

Second, we expected both teacher and tutor ratings of inattentive behavior would significantly predict performance on fraction concepts and whole-number calculations. Third, we anticipated that tutor ratings in a small-group setting would be more predictive than teacher ratings in a whole-class setting on fraction concepts (the focus of instruction in intervention) because the NAEP fraction items were proximal to the tutoring content. By contrast, because whole-number calculations were not the focus of tutoring content or the fourth-grade curriculum, we expected no difference in raters' predictiveness for whole-number calculations.

Method

Participants

The fourth-grade students were from 53 classrooms in 13 public elementary schools in a large school district in the Southeastern region of the United States. To select schools, we first contacted principals for permission to recruit teachers. Teachers voluntarily participated in the study. Parents provided written consent and students provided written assent to participate.

Students were identified as at risk for MD if they scored below the 35th percentile on the fourth edition of the *Wide Range Achievement Test* (WRAT-4; Wilkinson & Robertson, 2006); then, per study protocol, students who scored below the 9th percentile on both the Matrix Reasoning and Vocabulary subtests of the *Wechsler Abbreviated Scales of Intelligence* (WASI; Wechsler, 1999) were excluded. The sample comes from a larger study (see L. S. Fuchs, Schumacher, et al., 2013), in which at-risk students were randomly assigned to treatment or control. The results do not, however, overlap with those presented in that larger study. Also, the present analysis focused solely on data from the treatment group; that is, the control group's data were excluded because they did not have tutors to provide attention ratings.

Of the 145 treatment students, 12 moved before the end of the study and two had at least one piece of missing data. These 14 students were excluded in the data analysis, so the final sample was composed of 131 students. The 14 excluded students were not statistically different than the remaining students on gender, race, special education status, reduced/free lunch status, or performance on the NAEP fraction items and Double-Digit Addition pretest measure, $p > .05$. Demographics for the 131 students were as follows: 49% male; 53% Black, 26% White, 19% Hispanic, <2% biracial, and <1% Asian; <3% identified with a learning disability, <1% identified with a speech and language impairment, <1% identified with a comorbid behavioral and learning disability, <1% identified with autism or Asperger syndrome; and 78% received reduced/free lunch.

Classroom Teachers and Tutors

In all, 53 fourth-grade classroom teachers participated in the study. Teacher demographic data were not collected. Teachers were blind to the objectives of the study. The 12 tutors were not employees of the participating schools. They were trained research assistants, who were graduate students or employees at a local university. One of the tutors had a PhD in special education, two were doctoral students, and the remaining were master's degree students in the school of education. Tutors were trained on testing and tutoring protocols, but were blind to the objectives of the study. They had no background information on students prior to intervention.

Measures

Screening. Students were assessed on the WRAT-4 to determine risk status. In the WRAT-4, students solve 40 calculation problems progressing in difficulty: addition, subtraction, multiplication, and division with whole numbers, common fractions, decimals, ratios, and algebra. For the fourth-grade at-risk sample, performance largely reflected whole-number calculations. Coefficient alpha for the sample was .83.

Students who qualified to enter the study based on their WRAT-4 score then completed two subtests from the WASI to exclude students with intellectual disability. Matrix Reasoning measures nonverbal reasoning. Students fill in the missing piece of a pictorial matrix. For each item, students select one of five options at the bottom of the page that completes the incomplete matrix or series. The 30 items progress in difficulty and testing discontinues after the student makes four errors out of five consecutive responses. According to the publishers, the average internal consistency reliability is .92. Vocabulary measures expressive vocabulary knowledge. Students name pictures and define words. Testing discontinues after the student incorrectly defines five consecutive words. According to the publisher, the average internal consistency reliability is .89.

Rating inattentive behavior. The *Strengths and Weaknesses of ADHD-symptoms of Normal Behavior* (SWAN) is an 18-item Likert-type scale that rates students' inattentive and hyperactive behavior (Swanson et al., 2004), based on the ADHD criteria outlined by the fourth edition, text revision of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV-TR*; APA, 2000). The first nine items measure inattentive behavior; the second nine items measure hyperactivity/impulsivity. Each item is rated on a 7-point scale (*far below, below, slightly below, average, slightly above, above*, or *far above* their peers).

Because ratings of hyperactivity/impulsivity have not been found to be a unique predictor of mathematics performance in previous research (e.g., L. S. Fuchs et al., 2006; Merrell & Tymms, 2001; Raghubar et al., 2009), only results from the Inattentive Behavior subscale (nine items) were examined: (a) Give close attention to detail and avoid careless mistakes; (b) sustain attention on tasks or play activities; (c) listen when spoken to directly; (d) follow through on instructions and finish school work; (e) organize tasks and activities; (f) engage in tasks that require sustained mental effort; (g) keep track of things necessary for activities; (h) ignore extraneous stimuli; and (i) remember daily activities. Coefficient alpha for the Inattentive Behavior subscale for the sample was .95.

Outcomes. To measure fraction knowledge, we administered 18 released fraction items from the fourth- and eighth-grade NAEP (U.S. Department of Education, 2010). The 18 items comprise all fraction problems at the fourth-grade level and all easy fraction items at the eighth-grade level. Problems test basic fraction knowledge and address part–whole or fraction magnitude understanding. Coefficient alpha for the sample was .76. The content was related to tutoring, but items were not directly aligned with tutoring.

To measure whole-number knowledge, we administered the Double-Digit Addition (L. S. Fuchs, Hamlett, & Powell, 2003), which includes 20 double-digit addition problems, half of which require regrouping. Students have 3 min to complete the problems. Coefficient alpha for the sample was .77.

Procedure

In large groups, students completed WRAT-4 in the fall and completed NAEP fraction items and Double-Digit Addition in both the fall and spring. The WASI Matrix Reasoning and Vocabulary subtests were administered individually in the fall. Teachers and tutors filled out the SWAN rating scale for each tutored student in March of the spring semester during the 10th week of the 12-week fraction intervention. Both teachers and tutors were blind to the objectives of the study. Tutors had no previous relationship with students prior to the 12-week intervention (i.e., tutors did not work at their assigned schools and did not test students in their tutoring group during the testing phase).

Fraction tutoring consisted of 36 lessons, each lasting approximately 30 min. Tutoring occurred 3 times per week for 12 weeks, spanning from November to mid-March. The lesson activities focused on the measurement interpretation of fractions, part–whole relationships, comparing fractions, ordering fractions by magnitude, placing fractions on the number line, and fraction addition and subtraction. Each

Table 1. Means, Standard Deviations, and Correlations of SWAN Ratings and Outcome Variables.

	Treatment[a]		Correlations				
Variable	*M*	*SD*	2	3	4	5	6
1. Teacher SWAN	3.93	(1.30)	.46***	.29***	.44***	.27**	.32**
2. Tutor SWAN	4.26	(1.19)		.36***	.58***	.27**	.28**
3. NAEP pretest	8.35	(3.32)			.52***	.22*	.24**
4. NAEP posttest	14.46	(3.07)				.23**	.36**
5. DD addition pretest	14.92	(4.90)					.58**
6. DD addition posttest	17.59	(3.54)					

Note. SWAN = *Strengths and Weaknesses of ADHD-symptoms of Normal Behavior* (Swanson et al., 2004); NAEP = National Assessment of Education Progress (U.S. Department of Education, 2010); DD = Double-Digit (Fuchs, Hamlett, & Powell, 2003).
[a]*n* = 131.
**p* < .05. ***p* < .01. ****p* < .001.

tutoring group included three students. Students were randomly assigned to tutoring or control using stratified random assignment to ensure comparability of ability levels between groups. (This study only investigates treatment students' results.) Tutors were then assigned to a school. The same tutor taught the same students throughout the 12-week intervention. Tutors were trained in the beginning of the intervention in two 4-hr sessions. Weekly 2-hr meetings and updates were also provided. At each training session, trainers familiar with the lessons modeled the tutoring procedures. Tutors then practiced each lesson with peers and with guidance from trainers. As per L. S. Fuchs, Schumacher, et al. (2013), tutors implemented the tutoring procedures with strong fidelity.

At the beginning of tutoring, tutors outlined the behavior management system to students. Each session, tutors set a timer to beep at random times during the lesson. If all students in the group were on task (i.e., listening carefully, following directions, and working hard) when the timer beeped, each student earned a half dollar (pretend money). Students also had the opportunity to earn half dollars (and later quarter dollars) for individual work by completing "bonus" problems correctly. Each worksheet had two to four bonus problems, designated in advance to tutors, such that students were encouraged to work accurately on all problems. Students could spend their earned money (pretend money) at the "Fraction Store" at the end of each tutoring week. To know how much they could spend at the store, students had to use their fraction knowledge to figure out how many dollars their half dollars and quarter dollars summed to.

Data Analysis

We conducted *t* tests to determine whether teacher and tutor ratings of inattentive behavior differed in level. We then conducted two multiple regression and regression commonality analyses (Daniel, 1989) with teacher ratings and pretest scores as the independent variables and posttest scores (i.e., NAEP and Double-Digit Addition) as the dependent variable. This analysis allowed us to analyze the unique variance contribution of each predictor variable (pretest scores, teacher ratings, and tutor ratings) on posttest scores and the magnitude of predictability shared by each combination of the predictor variables. Alpha was set at .05 for the analyses.

Results

See Table 1 for means, standard deviations, and correlations. All correlations were statistically significant. Teacher (M = 3.93, SD = 1.30) and tutor (M = 4.26, SD = 1.19) ratings were significantly different, $t(130) = 2.85$, $p = .005$. See Table 2 for a summary of the multiple regression results. The commonality results are presented in text.

Fraction Concepts Outcome

NAEP pretest scores, teacher ratings, and tutor ratings together significantly predicted NAEP posttest scores, R^2 = .465, $F(3, 127) = 36.73$, $p < .001$. NAEP pretest uniquely predicted posttest scores, $\Delta R^2 = .095$, $F_{change}(1, 127) = 22.44$, $p < .001$. Teacher ratings accounted for less unique variance, $\Delta R^2 = .022$, $F_{change}(1, 127) = 5.15$, $p = .025$, than tutor ratings, $\Delta R^2 = .105$, $F_{change}(1, 127) = 24.90$, $p < .001$. Of the 47% of explained variance, the three predictor variables uniquely accounted for 48% of variance explained. The remaining 52% of explained variance was due to an overlap of the three predictor variables: 4% common to pretest and teacher ratings ($\Delta R^2_{\text{common}} = .017$), 15% common to pretest and tutor ratings ($\Delta R^2_{\text{common}} = .072$), 15% common to teacher and tutor ratings ($\Delta R^2_{\text{common}} = .068$), and 18% common to all three predictor variables ($\Delta R^2_{\text{common}} = .086$).

Whole-Number Calculation Outcome

Double-Digit Addition pretest scores, teacher ratings, and tutor ratings together significantly predicted Double-Digit

Table 2. Multiple Regression Results in Predicting Fraction Outcomes and Whole-Number Calculations Posttest Scores With SWAN Ratings.

Variable	*B*	*SE*	β	*t*
NAEP (fraction outcomes)				
(Constant)	(6.189)	0.830		7.46**
NAEP pretest	0.308	0.065	.334	4.74**
Teacher SWAN	0.397	0.175	.168	2.27**
Tutor SWAN	0.972	0.195	.378	4.99**
Double-Digit Addition (whole-number calculations)				
(Constant)	(9.482)	1.112		8.53**
DD addition pretest	0.377	0.054	.522	7.03**
Teacher SWAN	0.414	0.219	.152	1.89**
Tutor SWAN	0.200	0.239	.067	0.07**

Note. NAEP = National Assessment of Education Progress (U.S. Department of Education, 2010); SWAN = Strengths and Weaknesses of ADHD-symptoms of Normal Behavior (Swanson et al., 2004); DD = Double-Digit (Fuchs, Hamlett, & Powell, 2003).
*p < .05. **p < .01. ***p < .001.

Addition posttest scores, $R^2 = .370$, $F(3, 127) = 24.90$, $p < .001$. Double-Digit Addition pretest uniquely predicted posttest, $\Delta R^2 = .245$ $F_{change}(1, 127) = 65.96$, $p < .001$. However, teacher and tutor ratings did not contribute a statistically significant amount of unique variance to the model, $\Delta R^2 = .018$, $F_{change}(1, 127) = 3.77$, $p = .061$ and $\Delta R^2 = .003$, $F_{change}(1, 127) = 3.77$, $p = .403$, respectively. Of the 37% of explained variance, the three predictor variables uniquely accounted for 72% of variance explained. The remaining 28% of explained variance was due to an overlap of the three predictor variables: 8% common to pretest and teacher ratings ($\Delta R^2_{common} = .022$), 5% common to pretest and tutor ratings ($\Delta R^2_{common} = .019$), 3% common to teacher and tutor ratings ($\Delta R^2_{common} = 0.012$), and 12% common to all three predictor variables ($\Delta R^2_{common} = .044$).

Discussion

The present study had three purposes:

1. Determine whether tutors rated students as more attentive in an intensive small-group setting than teachers rated students in a whole-class setting;
2. Determine whether teachers and tutors significantly predicted student outcomes on a measure proximal (i.e., NAEP fractions) and distal (i.e., Double-Digit Addition) to tutoring; and
3. Determine whether tutor ratings in a small-group setting and teacher ratings in a whole-class setting differentially predicted student performance on fraction concepts and whole-number calculations.

Teacher Versus Tutor Ratings

As hypothesized, tutors (i.e., research assistants) rated students as more attentive than teachers, which is consistent with Strayhorn and Bickel's (2002) findings. That is, when students participated in a Tier II fraction intervention with instruction specifically designed to meet their academic needs, tutors rated them as more attentive. The intense instructional focus of the Tier II intervention likely influenced students' attention more strongly than instruction in the classroom, as reflected by higher tutor attention ratings.

One explanation for why tutors rated students as more attentive than did teachers centers on the fact that there were likely fewer distractions in the small-group setting than in the classroom. The tutor managed behavior for three students rather than an entire classroom of students. Tutors therefore had greater flexibility than classroom teachers did to address attention problems. The behavior management system used in the small-group setting also incentivized students to remain attentive and on task during the lessons. Students were motivated by tangible reinforcers to listen carefully, work hard, follow directions, and complete individual work accurately. Students were held accountable for their behavior at random intervals (with a timer) during tutoring. If students failed to self-regulate their behavior and remain on task, they did not earn fraction money (pretend money) to spend at the Fraction Store. Students also had fewer opportunities to become disengaged because every moment of tutoring was filled with a planned activity.

A second explanation for this finding is that teacher ratings in the whole-class setting may reflect a perceived mismatch between instruction and ability level that is reflected in attention ratings (L. S. Fuchs et al., 2005). This could also be the case for tutor ratings, but to a lesser degree because tutoring was designed to meet student needs. That is, it is possible that students became more attentive in a small-group setting but that some still struggled and exhibited inattentive behavior. However, inattention was likely less frequent in the small-group setting than observed in the classroom because of the alignment of instruction with students' learning needs in the tutoring environment.

Predicting Fraction Concepts

Both tutors and teachers significantly predicted performance on the fraction concepts outcome, but tutor ratings had greater predictive power than teacher ratings. Predictors with a large standard deviation, or great variability in possible scores, are more likely to be better predictors. However, in this study, the ratings with greater predictive power (tutor ratings) had less variability (for tutor ratings, $SD = 1.19$; for teacher ratings, $SD = 1.30$). Therefore, variability is not a viable explanation for why tutor ratings had more predictive power than teacher ratings for the fraction concepts outcome.

One reason why tutor ratings were a stronger predictor of the fraction concepts outcome than teacher ratings may be that students' attention was more academically profitable in this intensive setting because instruction was directly aligned with students' academic needs. The tutoring lessons incorporated explicit instruction and frequent reiteration of strategies, rules, and concepts, and there were few opportunities for students to become disengaged. Research suggests that students with learning difficulties and disabilities benefit more from explicit instruction than discovery learning (e.g., Kroesbergen & Van Luit, 2003). Explicit instruction likely contributed to students differentially profiting from attention in the small-group setting. As such, attention in this intensive small-group setting had a greater effect on fraction performance than attention in the whole-class setting.

The behavior management system could also factor into why tutors' ratings were more predictive of fraction concepts. L. S. Fuchs et al. (2008) found that stronger behavior management systems promoted greater learning among at-risk students when they compared mathematics performance growth among four cohorts of at-risk third-grade students. That is, students who participated in small-group tutoring with a similar behavior management system to the one used in the present study made greater academic gains than students who participated in small-group tutoring with a less well-defined behavior management system (i.e., no self-regulated learning strategies such as tangible reinforcement and checking student behavior at random intervals). Similarly, although not directly related to the present study (i.e., the present study did not target students with ADHD), Harris et al. (2005) found that students with ADHD made greater academic gains in spelling when they were required to self-regulate their attention rather than self-regulate their academic performance. Thus, the well-defined behavior management system may help to explain why ratings of at-risk students' attention were more closely associated with improved fraction performance in the small-group setting.

Predicting Whole-Number Calculations

Based on prior research (e.g., Cirino et al., 2007; L. S. Fuchs et al., 2006; Raghubar et al., 2009; Seethaler & Fuchs, 2006), we had also expected teacher and tutor ratings of inattentive behavior to predict performance on the whole-number calculations outcome. This was not the case. Neither teachers' nor tutors' ratings of student attention significantly predicted performance on whole-number calculations. This finding may be due to the fact that whole-number calculations are not a focus of the fourth-grade curriculum, nor were they the focus of the 12-week fraction intervention. L. S. Fuchs et al. (2006) and Seethaler and Fuchs (2006), who found a relation between teacher ratings of inattention and whole-number calculation skill, focused on third grade, where whole-number computation is still an instructional target. In fact, although we used the same whole-number measure in this study as was used in these two prior studies, the representative samples of third-grade students in those studies performed significantly better on the whole-number calculations measure than students in our at-risk fourth-grade sample, $d = 0.25$, 95% confidence interval (CI) = [0.04, 0.45].

At-risk fourth graders, who have been selected for fraction intervention based on low whole-number calculation skill, still have difficulty with whole-number calculations. Their whole-number calculation skill may suffer when the instructional emphasis shifts away from whole-number calculations. In fact, at-risk students in our sample made less progress on whole-number calculations than they did on fraction concepts. On the whole-number calculations outcome, students solved a mean of 15 of 20 problems correctly at pretest, compared with a mean of 18 of 20 problems correct at posttest, $d = 0.62$, 95% CI = [0.38, 0.87]. By contrast, students solved a mean of 8 of 18 fraction concepts problems correctly at pretest, compared with a mean of 14 of 18 problems correct on posttest, $d = 1.91$, 95% CI = [1.62, 2.20]. The more substantial improvement on fraction concepts versus whole-number calculations (33% vs. 15% improvement) may explain why pretest captured much more of the variance in the whole-number calculations outcome than was the case for the fraction concepts outcome (66% vs. 20% of the total variance explained). It could also be that students know little about fractions in the beginning of fourth grade, whereas whole-number calculations had been the focus of instruction in previous academic years. This helps to explain why there was less growth for whole-number calculations than fraction concepts and why neither teachers' nor tutors' ratings significantly predicted performance for whole-number calculations.

Implications for Practice

Although teacher ratings of inattentive behavior in a whole-class setting have been found to be a significant predictor of responsiveness to instruction (L. S. Fuchs, Fuchs, & Compton, 2013), results from the present study suggest that ratings from tutors in a small-group setting more strongly predict student performance on a proximal measure of performance and may, therefore, further differentiate students' academic needs. This has important practical implications considering that some students receiving Tier II instruction may be unresponsive to evidence-based instruction (L. S. Fuchs, Fuchs, & Compton, 2013). Because tutoring was specifically designed to match these at-risk students' needs, persistent inattentive behavior (as reflected by ratings in a small-group setting) may indicate the need for even more intensive and individualized instruction at Tier III. As such, small-group instructors may serve as a powerful tool to predict whether students should remain in Tier II or move up to Tier III to receive more differentiated instruction. This speaks to the importance of measuring students' ability to attend to academic tasks as instruction becomes more intensive and individualized. It makes sense that professionals (e.g., tutors) delivering intervention more aligned with student needs are more in tune to the differences between behavioral inattention stemming from academic frustration versus an inability to sustain attention due to other factors (e.g., medically diagnosable attention deficits).

In addition, the fact that students made much less progress on whole-number calculations than they did on fraction concepts highlights the importance of designing interventions to support students' foundational mathematics skills, even as the major focus of intervention addresses the pressing grade-level curricular priorities. This is especially important as the curricular focus in the United States ramps up with Common Core requirements (Powell, Fuchs, & Fuchs, 2013). As the curriculum becomes more rigorous and standards-based with Common Core, there will likely be even less time to focus on maintenance of previously taught material (e.g., whole-number calculations), which may be detrimental to struggling learners. Therefore, instruction must continue to address students' academic deficits even when the curriculum focus shifts to include more difficult mathematics concepts.

Limitations and Directions for Future Research

It is, of course, important to consider these findings in light of study limitations. First, although there was a statistically significant difference between teacher and tutor ratings of inattentive behavior, the effect size was 0.26, 95% CI = [0.02, 0.51]. It is unclear whether this difference is practically significant. In the present study, substantive differences between teachers and tutors may be reflected in their ratings. For example, teachers may have a more holistic view of their students as they witness students' academic and behavioral performance in a variety of settings. By contrast, tutors' perception of student behavior stemmed from a more restricted setting (i.e., in a small-group intensive academic intervention). These differences could have differentially affected the behavioral ratings. In addition, we lacked demographic information on teachers and tutors. Demographic characteristics such as differences in training and years in the field could also lead to different ratings.

A second limitation is that the study collected ratings of student attention only once. It would be interesting to investigate whether teachers' and tutors' ratings of inattentive behavior change as a function of time; that is, can tutors identify potential attention problems early on in intervention and does this perception persist throughout intervention, or can some of these perceived attention problems be remediated by structured and intense instruction designed to meet students' academic needs? Because the present study only collected attention ratings at one point in the intervention, these questions could not be addressed. Future research should consider collecting ratings at multiple points in the intervention.

Third, future researchers should replicate these findings with larger and different samples and include a longitudinal design to assess whether tutor ratings provide better long-term predictions of academic achievement. In addition, this study only indirectly examined students' attention and behavior. Questions remain about whether teacher and tutor ratings actually measure students' attention or whether they serve as a proxy for academic achievement. For example, would tutors also be more accurate predictors of student behavioral tasks (i.e., not just academic tasks)? Finally, we were unable to obtain reliability of ratings across raters. That is, tutors did not rate students' inattentive behavior in the classroom and teachers did not rate students' inattentive behavior during tutoring. Therefore, we cannot determine whether teachers and tutors would rate students similarly in the same environment.

Conclusion

In summary, prior research (e.g., Cirino et al., 2007; L. S. Fuchs et al., 2006; Raghubar et al., 2009; Seethaler & Fuchs, 2006) suggests that teacher ratings of inattentive behavior are a significant predictor of mathematics performance. However, questions remain about whether these ratings reflect a mismatch between instruction and ability level or whether they index true attention difficulties (i.e., medically diagnosable attention deficits) among students at risk for MD. The present study sought to address this question by comparing teacher versus tutor (i.e., research assistants) ratings of inattentive behavior during a 12-week Tier II fraction intervention. Tutors rated students as more attentive in an intensive small-group setting than teachers in a whole-class setting and tutor ratings were more predictive of student performance on fraction concepts (proximal to tutoring). This has important implications for practice in terms of providing valuable information for educators to modify instruction based on students' needs.

Authors' Note

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 or the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by Grant R324C100004 from the Institute of Education Sciences in the U.S. Department of

Education to the University of Delaware with a subcontract to Vanderbilt University and by Award R24 HD075443 and by Core Grant HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Vanderbilt University.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*, 65–94.

Benedetto-Nasho, E., & Tannock, R. (1999). Math computation, error patterns and stimulant effects in children with attention deficit hyperactivity disorder. *Journal of Attention Disorders*, *3*, 121–134.

Breslau, N., Breslau, J., Peterson, E., Miller, E., Lucia, V. C., Bohnert, K., & Nigg, J. (2010). Change in teachers' ratings of attention problems and subsequent change in academic achievement: A prospective analysis. *Psychological Medicine*, *40*, 159–166.

Cirino, P., Fletcher, J. M., Ewing-Cobbs, L., Barnes, M., & Fuchs, L. S. (2007). Cognitive arithmetic differences in learning difficulty groups and the role of behavioral inattention. *Learning Disabilities Research & Practice*, *22*, 25–35.

Daniel, L. G. (1989). *Commonality analysis with multivariate data sets* (Report No. 143). San Francisco, CA: American Educational Research Association.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Every Child a Chance Trust. (2009). *The long-term costs of numeracy difficulties*. Retrieved from http://www.everychildachancetrust.org/counts/index.cfm

Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93–98.

Fuchs, L. S., Compton, D., 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.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2013). Intervention effects for students with comorbid forms of learning disability: Understanding the need for nonresponders. *Journal of Learning Disabilities*, *46*, 534–548.

Fuchs, L. S., Fuchs, D., Compton, D., Powell, S. R., Seethaler, P., Capizzi, A., . . . 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.

Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., Hamlett, C. L., & Schatschneider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology*, *100*, 491–509.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). *Math Fact Fluency and Double-Digit Additional and Subtraction Tests*. (Available from L. S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203)

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P., . . .Chagas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology*, *105*, p. 683–700.

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

Harris, K. R., Friedlander, B. D., Saddler, B., Frizelle, R., Graham, S. (2005). Self-monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *The Journal of Special Education, 39,* 145–156.

Kroesbergen, E. H., & Van Luit, J. (2003). Mathematics interventions for children with special needs: A meta-analysis. *Remedial and Special Education*, *24*, 97–114.

Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity, and impulsiveness: Their impact on academic achievement and progress. *British Journal of Education Psychology*, *71*, 43–56.

National Assessment of Education Progress. (2013). *Nation's report card*. Washington, DC: U.S. Department of Education.

Powell, S. R., Fuchs, L. S., & Fuchs, D. (2013). Reaching the mountaintop: Addressing the Common Core standards in mathematics for students with mathematics difficulties. *Learning Disabilities Research & Practice*, *28*, 38–48.

Raghubar, K., Cirino, P., Barnes, M., Ewing-Cobbs, L., Fletcher, J., & Fuchs, L. S. (2009). Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities*, *42*, 356–371.

Schunk, D. H., & Zimmerman, B. J. (Eds.). (1998). *Self-regulated learning: From teaching to self-reflective practice*. New York, NY: Guilford Press.

Seethaler, P., & Fuchs, L. S. (2006). The cognitive correlates of computational estimation skill among third-grade students. *Learning Disabilities Research & Practice*, *21*, 233–243.

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Child & Adolescent Psychiatry*, *9*, 58–64. doi:10.1007/s007870070009

Strayhorn, J. M., & Bickel, D. D. (2002). Reduction in children's symptoms of attention deficit hyperactivity disorder and oppositional defiant disorder during individual tutoring as compared with classroom instruction. *Psychological Reports*, *91*, 69–80.

Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., . . . McCleary, R. (2004). *Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and SWAN rating scales*. Available from http://www.adhd.net/SNAP_SWAN.pdf

U.S. Department of Education. (2010). *National Assessment of Educational Progress (NAEP)*. Washington, DC: Institute of Education Sciences.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test* (4th ed.). Wilmington, DE: Wide Range.