Article

Hammill Institute on Disabilities

Remedial and Special Education
2021, Vol. 42(3) 182–192
© Hammill Institute on Disabilities 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0741932520923063
rase.sagepub.com
SAGE

The Influence of Behavior on Performance Within a Word-Problem Intervention for Students With Mathematics Difficulty

Sarah A. Benz, PhD[1] and Sarah R. Powell, PhD[2]

Abstract

Some elementary students may exhibit challenging externalizing or internalizing behaviors in addition to difficulty with mathematics. In this study, we explored the behavioral patterns of 441 third-grade students with and without mathematics difficulty (MD). Compared with students without MD, students with MD demonstrated higher rates of externalizing and internalizing behaviors. We then randomly assigned 162 third-grade students with MD to receive a 10-week word-problem intervention or to be in a business-as-usual comparison group. Within the word-problem intervention, students with MD who exhibited higher occurrences of externalizing behaviors performed significantly lower on a word-problem measure than students without as many occurrences of externalizing behaviors. Response to the word-problem intervention did not differ based on internalizing behavior patterns.

Keywords

behavior, learning difficulties, mathematics, word-problems

Students with challenging behaviors experience difficulty in all core academic subjects, but research has uncovered persistent difficulties with spelling and mathematics (Reid et al., 2004). Between 42% and 93% of students with challenging behaviors (e.g., emotional or behavioral disorders) also exhibit a mathematics difficulty (MD) and demonstrate mathematics performance 1–2 years below expected grade level (Epstein et al., 2005). For students with MD comorbid with challenging behaviors, interventions addressing both mathematics and behavior may be necessary.

Students With MD

Although not all students with disabilities have difficulty with mathematics, many of them do. Students with disabilities display uneven patterns of MD, ranging from 22% for students with speech impairments to 84% for students with intellectual disabilities (Blackorby et al., 2004). Approximately 5%–8% of school-age students have a mathematics learning disability with significant deficits requiring specially designed instruction (Devine et al., 2018; Geary, 2004). In the literature, these students may be referred to as experiencing a mathematics learning disability or dyscalculia (Skagerlund & Träff, 2016). Beyond a mathematics learning disability, approximately 25%–35% of school-age students persistently struggle with mathematics (Mazzocco, 2007). For the purposes of this article, we focus on students with persistent difficulty in mathematics and intentionally refer to this diverse group of learners as students with or at-risk for MD, which is common practice in special education (e.g., Jitendra et al., 2013). As such, our term may encompass students with mathematics disability, dyscalculia, or MD.

In this article, we focus on word-problem solving. Students with MD require explicit word-problem instruction because they demonstrate lower word-problem performance than students without MD (Fuchs et al., 2014). There are several approaches for teaching students with MD to solve word-problems. For example, teachers may use drawings and graphic organizers (van Garderen, 2007), metacognitive strategies (Rosenzweig et al., 2011), or provide schema instruction (Fuchs et al., 2008; Jitendra et al., 2009). Of these, schema instruction has the widest research base for improving the word-problem solving of students with MD (e.g., Fuchs et al., 2010; Jitendra et al., 2017; Powell et al., 2015; Xin et al., 2005). With schema instruction,

[1]American Institutes for Research, Austin, TX, USA
[2]The University of Texas at Austin, USA

Corresponding Author:
Sarah A. Benz, American Institutes for Research, 4700 Mueller Blvd., Austin, TX 78723, USA.
Email: sbenz@air.org
Associate Editor: Robin Ennis

students learn to recognize a word-problem as belonging to a specific schema and then employ strategies to solve the problem by schema.

Students With Challenging Behaviors

Students with or at-risk for challenging behaviors represent a group of students who may also experience difficulty in mathematics, especially word-problem solving, despite the use of evidence-based practices (Mulcahy & Krezmien, 2009). Not only do students with challenging behaviors demonstrate similar academic patterns to students with MD, they also exhibit unique and challenging behaviors. These challenging behaviors may hinder students' abilities to benefit from established mathematics interventions (Lane et al., 2008).

Researchers suggest that more than 30% of school-age students will experience a mental health difficulty during their educational career (i.e., anxiety, depression, or trauma; Forness et al., 2012). Approximately 5%–13% of the school-age population, however, will experience sustained emotional or behavioral difficulties that have long-term implications (Maggin et al., 2016). In this article, we use the term "challenging behaviors" to describe any student with an emotional, behavioral, or psychiatric disorder listed in the American Psychiatric Association (2013) diagnostic manual. Our term "challenging behaviors" may include students with a diagnosed disability (e.g., emotional disturbance [ED]) or students who display at-risk behaviors.

MD and Challenging Behaviors

Several research reviews have synthesized effective academic interventions for students with challenging behaviors (e.g., Hodge et al., 2006; Lane, 2004; Losinski et al., 2019; Mooney et al., 2003; Mulcahy et al., 2016; Ralston et al., 2014; Ryan et al., 2008; Templeton et al., 2008). In general, these reviews identified several academic interventions that included components of self-determination. For example, researchers studied components of self-regulation, self-monitoring, or self-evaluation and used academic skills as a primary or secondary outcome. Results demonstrated improved academic outcomes with interventions embedded with self-determination components.

Specific to mathematics, students with challenging behaviors have difficulty in word-problem solving (Bullis & Yovanoff, 2006). For example, Alter et al. (2011) examined a word-problem intervention with students in Grades 1 through 4 experiencing emotional or behavioral disorders. Results indicated all students increased the percentage of word-problems solved; however, none of the participants solved more than 50% of word-problems correctly after receiving intervention. Alter (2012) examined an eight-step checklist for solving word-problems with four students with challenging behaviors in Grades 4 and 5. Alter (2012) learned that all students improved their word-problem accuracy, and 75% of the students increased their rates of on-task behavior. Similarly, Peltier and Vannest (2016) investigated the effects of schema instruction on the word-problem performance of two Grade 4 students with challenging behaviors. The authors reported both students improved in word-problem accuracy from baseline. This collection of studies demonstrates the importance of providing word-problem intervention to students with challenging behaviors.

Purpose and Research Questions

Although researchers have conducted studies with students with challenging behaviors in the area of mathematics, little is known about the impact of student behavior on the efficacy of a word-problem intervention for students with MD comorbid with challenging behaviors. The purposes of the current study were to examine the behavioral patterns of students with and without MD and to investigate the efficacy of a word-problem intervention for students with MD with and without challenging behaviors. We asked the following research questions:

- **Research Question 1:** Are there differences in the behavioral patterns (i.e., externalizing behaviors or internalizing behaviors) of students with and without MD?
- **Research Question 2:** What is the efficacy of a word-problem intervention for improving the word-problem performance of third-grade students with MD?
- **Research Question 3:** For students with MD, is response to the word-problem intervention influenced by behavioral patterns? That is, do students who exhibit high rates of externalizing or internalizing behaviors respond differently to the word-problem intervention?

The present study occurred in the third year of a multi-year randomized control trial investigating the impact of a word-problem intervention with and without pre-algebraic reasoning instruction. We use the term "present study" for the study that occurred during the 2017–2018 school year, and we only provide results for the present study. We use the term "parent study" for the multi-year study, but we do not provide results from the parent study.

Method

Participants

Table 1 displays the demographics of the students with MD ($n = 132$) and without MD ($n = 309$) with behavioral data. Table 2 shows the demographics of students with MD with behavioral data in the word-problem intervention ($n = 74$) and business-as-usual control ($n = 50$). Chi-square analyses between the intervention and control groups yielded no

Table 1. Demographics and Descriptives by MD Status.

	Students without MD ($n = 309$)		Students with MD ($n = 132$)	
Demographic/descriptive	*n*	%	*n*	%
Demographics				
Female	143	46.3	77	58.3
Race or ethnicity				
African American	33	10.7	20	15.2
Asian	8	2.6	3	2.3
Caucasian	83	26.9	5	3.8
Hispanic/Latinx	170	55.0	93	70.5
Other	15	4.9	11	8.3
Dual-language learner status	131	42.4	79	59.8
Special education status	25	8.1	19	14.4
	M	*SD*	*M*	*SD*
Challenging behaviors				
Externalizing	2.01	3.18	3.73	3.43
Internalizing	0.84	1.48	1.77	2.25

Note. MD = mathematics difficulty.

significant differences in sex ($\chi^2 = 0.038$, $p = .846$), race or ethnicity ($\chi^2 = 1.070$, $p = .899$), dual-language learner status ($\chi^2 = 0.593$, $p = .441$), or special education status ($\chi^2 = 1.595$, $p = .207$).

Recruitment

In the present study, we recruited teachers from 13 elementary schools in one urban southwestern U.S. school district. School enrollment ranged from 378 to 502 students with 71%–94% of students reported as economically disadvantaged. Overall, all recruited teachers consented to participate (100%) for a total of 39 Grade 3 general education teachers. We screened 819 Grade 3 students from the classrooms of participating teachers. Of these 819 students, we received caregiver consent and student assent for 615 (75.1%) students, and teachers ($n = 36$) provided behavioral data on 441 (71.7%) of the consented students. Only 29 teachers provided behavioral data for every consented student in their classroom. Another seven teachers provided behavioral data but only for the students with MD (see next paragraph) in their classroom, and three teachers provided no behavioral data for any student in their classroom. We analyzed the behavioral data of the 441 students to investigate our first research question.

For our second and third research questions, we focused on students experiencing MD. Students with MD had to perform below the 13th percentile on a test of single-digit word-problems (Jordan & Hanich, 2000). The 13th percentile cut-off was determined based on grade-level scores from the first year of the parent study (Powell et al., 2020). We identified 230 students eligible for the present student and categorized those students as experiencing an MD. Of the 230 students with MD, we disqualified 68 for the following reasons: limited English language ($n = 24$), parent or student opted out of participation ($n = 12$), behavioral issues during pretesting such that tutoring would not have been possible ($n = 9$), too many students qualified in the same class for our research design ($n = 7$), student received special education services which already required frequent removal from core classroom instruction ($n = 6$), did not complete full pretesting battery ($n = 5$), student moved schools ($n = 4$), or severe disability ($n = 1$).

We randomized, blocking by classroom and school, 162 students with MD into one of two word-problem intervention groups (e.g., word-problem intervention or word-problem plus pre-algebraic reasoning intervention) or a business-as-usual control. We began with 98 students in the word-problem intervention and 64 students in the control group. We conducted two versions of the word-problem intervention but the content was nearly identical in terms of word-problems and intervention components; therefore, we treated the two versions as one word-problem intervention. At posttesting, 151 (93%) students with MD remained. Of the 11 students who left the study, nine students moved schools, one student went into protective custody, and one was suspended from school for 30 days. Overall attrition was calculated at 7%, while differential attrition was 6%.

Measures

Screening measure. We used single-digit word-problems (Jordan & Hanich, 2000) as the screener to identify students with MD. Students solved 14 one-step addition or subtraction word-problems. Students received one point for each

Table 2. Demographics and Descriptives of Intervention Participants With Behavioral Data.

	Word-problem intervention ($n = 74$)		Business-as-usual control ($n = 50$)	
Demographic/descriptive	*n*	%	*n*	%
Demographics				
Female	45	60.8	29	58.0
Race or ethnicity				
African American	12	16.2	6	12.0
Asian	1	1.4	1	2.0
Caucasian	3	4.1	1	2.0
Hispanic/Latinx	52	70.3	38	76.0
Other	6	8.1	4	8.0
Dual-language learner status	42	56.8	32	64.0
Special education status	13	17.6	5	10.0
	M	*SD*	*M*	*SD*
Challenging behaviors				
Externalizing	3.57	3.31	3.80	3.55
Internalizing	1.85	2.25	1.58	2.31
Mathematics measures				
Pretest single-digit word-problems	4.8	1.63	4.98	1.76
Posttest single-digit word-problems	7.95	3.09	7.36	2.98
Pretest double-digit word-problems	2.22	1.61	2.02	1.57
Posttest double-digit word-problems	5.49	3.67	2.44	2.46
Pretest open equations	5.09	3.54	5.32	3.58
Posttest open equations	10.74	5.91	8.98	6.07

correct mathematics answer for a maximum score of 14 ($\alpha = .72$).

Behavior measure. We administered the Student Risk Screening Scale—Internalizing Externalizing (SRSS-IE; Drummond, 1994; Lane, Menzies, et al., 2012; Lane, Oakes, et al., 2012) at the beginning of the present study. The SRSS-IE is a brief screening tool used to assess students' at-risk levels for externalizing and internalizing behaviors. The SRSS-IE included 12 items and required approximately 15 min for teachers to rate all students in their classroom. Teachers scored items on a 0- to 3-point scale, where 0 indicated *never* and a score of 3 indicated *frequently*. Subsequently, higher scores represented greater risk for externalizing or internalizing behaviors. Many studies have been conducted examining the reliability and validity of the SRSS. Researchers have found high internal consistency ($\alpha = .81–.82$), and the measure has been validated with elementary students (Lane, Oakes, et al., 2012; Menzies & Lane, 2012), students in urban elementary schools (Ennis et al., 2012; Oakes et al., 2010), and with students who are dual-language learners (Lane et al., 2014).

In addition to the SRSS-IE, classroom teachers completed a survey about the extent of supports students were provided for behavior. Specifically, teachers identified students with MD who received (a) individualized education program (IEP) services under the classification of ED, (b) behavioral supports through a 504 plan, or (c) behavioral intervention through multi-tiered system of supports (MTSS). Of the students with MD, teachers reported that three students (2%) had IEPs for ED and received specially designed instruction for behavior, two students (1%) received behavioral support through a 504 plan, and nine students (6%) received behavioral support through an MTSS framework. Students receiving behavioral support were equally distributed between the word-problem intervention and control groups.

Pretest measures. The pretesting battery consisted of three assessments: single-digit word-problems (administered before pretesting to determine MD eligibility but considered a pretest because it was administered before intervention), abbreviated double-digit word-problems, and open equations. Each examiner followed a pretesting protocol and was expected to read the protocol verbatim.

On the abbreviated version of double-digit word-problems (Powell & Berry, 2015), students solved double-digit addition and subtraction word-problems consisting of three total problems, three difference problems, and one change problem. Tutors read word-problems aloud and provided

time for students to solve the problem. Upon student request, each problem could be read an additional time. Students received one point for each correct mathematics answer and each correct label for a maximum score of 14 ($\alpha = .83$).

The open equations test (Powell, 2007) required students to solve 30 standard (e.g., 4 + __ = 5) and nonstandard (e.g., 4 + 3 = __ + 2) equations within 6 min. Tutors read the directions to the students and allowed the students to work. Students received one point for each correct answer for a maximum score of 30 ($\alpha = .88$).

Posttest measures. At the conclusion of Lesson 30 of the word-problem intervention (i.e., after completion of Units 1–3), the tutors administered a posttest to all students. We tested intervention students alongside business-as-usual students. Posttesting sessions mirrored the conditions in pretesting, in that the same protocol from pretesting was re-administered.

Procedures

Intervention. The word-problem intervention was an updated version of a schema-based intervention (Fuchs et al., 2014). As a note, the parent study evaluated the effects of 48 lessons in which students learned three schemas: total, difference, and change. The present word-problem intervention study featured only two schemas (e.g., total and difference) and included 30 one-on-one lessons implemented 3 times a week with each session lasting about 30 min. The intervention occurred during the school day at a time agreed upon with the teacher and the tutor. Tutors administered lessons using a provided lesson guide. Each lesson consisted of five activities, described in subsequent paragraphs.

Schemas. In the present study, we evaluated the use of two schemas within a word-problem intervention: total and difference. In total problems, students have parts that are put together for a total. The missing information from the word-problem (i.e., "X") could be the total or one of the parts. After students identified word-problems as fitting into the total schema, students used a total equation to organize word-problem information (i.e., P1 + P2 = T; where P is a part and T is the total). After an introductory unit, students started learning about total problems on Day 5 of the intervention. We introduced difference problems on Day 17 of the word-problem intervention. In difference problems, students learned to compare an amount that is greater and an amount that is less to find the difference. The missing information (i.e., "X") for difference problems could have been the amount that is greater, the amount that is less, or the difference. To successfully identify the greater amount, the lesser amount, and the difference, tutors taught students to locate a "compare sentence." This compare sentence featured a comparison. To organize the information from difference problems, students used the equation G − L= D, where G was the greater amount, L was the lesser amount, and D was the difference.

Intervention components. Tutors started each word-problem intervention session with a fluency-building activity in which students solved addition and subtraction flashcards (i.e., addends 0–9 or minuends of 0–19 and subtrahends from 0 to 9). With the flashcards, the student completed two separate 1-min trials. The tutor also provided immediate, corrective feedback to the student by reviewing counting-up strategies for any noted errors. At the end of the second 1-min timing, the tutor and student graphed the highest score. The total time on this activity was approximately 3 min with student graphing.

The second activity each day was either pre-algebraic reasoning instruction or a non-word-problem mathematics review on perimeter, area, fractions, money, order of operations, or telling time. We randomly assigned one-half of students in the word-problem intervention to receive pre-algebraic reasoning instruction while the other half received the mathematics review. Each activity lasted approximately 2–3 min. The different activities were part of the research focus of the parent study but did not lead to performance differences within the present study after implementation of 30 lessons.

The third activity consisted of tutor-led explicit instruction on solving word-problems for approximately 15–18 min. Days 1 through 4 consisted of a review of addition and subtraction skills using counting strategies, learning how to label and interpret data presented on charts and graphs, and learning a specific attack strategy called RUN. To RUN, the student learned to Read the problem, Underline the label and cross out irrelevant information, and Name the problem type (i.e., choose the correct schema to use) by asking questions about the problem.

On Days 5 through 16, the tutor taught about solving total word-problems in which parts are put together for a total. The missing information from the word-problem (i.e., "X") could be the total or one of the parts. Tutors introduced difference problems (i.e., greater and lesser amounts compared for a difference) on Day 17 and practiced a combination of total and difference problems through Day 30. The missing information (i.e., "X") for difference problems could have been the amount that is greater, the amount that is less, or the difference.

The fourth activity for each lesson was a schema sorting 1-min timed activity that allowed the students to practice identifying word-problem schemas (i.e., total or difference). This activity lasted approximately 2 min with feedback from the tutor. The final activity for each lesson was a brief, timed review of the lesson contents. The student had 1 min to answer up to nine single and double-digit addition or subtraction problems or write appropriate equations for

the two word-problem schemas. Then, the student had 2 min to complete a word-problem independently using the appropriate schema steps. After 2 min, the student received feedback from the tutor, which reinforced content mastery. In total, this final activity lasted 4 min.

The word-problem intervention incorporated a standardized token-based reward system. At the beginning of each lesson, the tutor reviewed the intervention rules (i.e., follow directions, stay seated, use a quiet voice, and be respectful) with the student. When students followed the rules, they received gold coins throughout the lesson. Typically, students earned between three and six coins per lesson. At the end of the lesson, students recorded the number of coins they received by coloring on their treasure map. They received a prize when they reached the treasure box on the map. Students typically earned one treasure chest prize per week.

Tutors. Overall, we hired and trained 14 graduate research assistants to conduct all screening, pretesting, tutoring, and posttesting. Each tutor had or was seeking a master's or doctoral degree in an education-related field. Tutors received approximately 20 hr of training, completed practice sessions, and completed a reliability check with a project manager before implementing any testing or tutoring in schools.

Intervention fidelity. Fidelity was collected throughout the intervention in two ways: direct observation and digital audio recording. The project manager conducted direct observation fidelity checks on 30% of tutoring sessions using a fidelity checklist that corresponded with each lesson (i.e., 30 different fidelity checklists). Intervention fidelity was considered acceptable at 90%. When fidelity was assessed below 90%, subsequent direct observations occurred until 90% was achieved. The overall fidelity of direct observations ranged between 79% and 100% with an average of 97.08%. Fidelity was also evaluated through audio recordings. Overall, we measured fidelity for 20% of the remaining lessons (i.e., lessons not directly observed) using the same checklist as the direct observation. The total fidelity for audio recordings was 98.5%.

Data Analysis

To investigate our first research question about differences in the behavioral patterns of students with and without MD, we used student externalizing and internalizing scores to compare the behavioral patterns of students with and without MD. We collected complete behavioral data from 441 students, and we conducted a one-way analysis of variance (ANOVA) for each type of behavior (i.e., externalizing and internalizing) comparing students with MD to students without MD.

For our second research question, we analyzed the data of 124 students with MD who (a) completed posttesting, (b) had complete demographic data, and (c) had complete behavioral data. To evaluate the efficacy of the word-problem intervention for students with MD, we calculated gain scores from pretest to posttest for each outcome measure and used ANOVAs to identify any significant differences between conditions (i.e., word-problem intervention versus business-as-usual control). For our third research question about response to the word-problem intervention based on behavioral patterns, we analyzed the data of 74 students with complete data who participated in the word-problem intervention. We used ANOVAs to identify significant differences between students with low and high risk for challenging behaviors, and we utilized regression models to determine if higher externalizing or internalizing scores predicted double-digit word-problem gain scores.

For all analyses, we used a significance threshold of $p < .05$ for interpretation of results. We calculated effect sizes (*ES*) using Hedges' *g* by subtracting mean values and dividing by the pooled standard deviation, as outlined by the What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education (2017).

Results

In this study, we examined the behavioral patterns of students with and without MD. Then, we investigated the efficacy of a word-problem intervention for students with MD with and without challenging behaviors. The following are the results of the present study.

Behavioral Patterns

With our first research question, we asked whether students with MD demonstrated a higher rate of occurrences of externalizing or internalizing behaviors compared with students without MD using the SRSS-IE. We collected 441 student behavioral patterns (i.e., 309 students without MD and 132 students with MD). Several teachers failed to return completed SRSS-IE rating scales for students with demographic data; therefore, we only analyzed the data of students with complete demographic information and complete behavioral patterns (see Table 1). According to the SRSS-IE, higher scores are consistent with higher risk status for challenging behaviors.

Results indicated significant differences in behaviors between students with and without MD. Students with MD demonstrated higher occurrences of externalizing behaviors, $F(1, 439) = 25.71$, $p < .001$ ($ES = .53$). Students with MD also demonstrated higher occurrences of internalizing behaviors, $F(1, 439) = 26.08$, $p < .001$ ($ES = .53$).

Efficacy of Word-Problem Intervention

We collected complete posttest data, behavioral data, and demographic data for 124 students with MD (see Table 2). Related to our categorization of MD, we note that, of the 18 students with a school-diagnosed disability, only nine of these students had a learning disability. This result shows the low rate of learning disability diagnosis by third grade. Because the majority of students classified as dual-language learners, we calculated the average Texas English Language Proficiency Assessment System (TELPAS) ratings for the word-problem intervention and control conditions. Based on recommendations from the TELPAS (i.e., 1 = *beginning*, 2 = *intermediate*, 3 = *advanced*, 4 = *advanced high*), the average score for the 42 dual-language learners in the word-problem intervention was a 2.04 ($SD = 0.66$), and the average score for dual-language learners in the control group ($n = 32$) was 1.71 ($SD = 0.68$).

We conducted a preliminary analysis to ensure assumptions for ANOVA were not violated (i.e., inspecting the data for abnormalities, ensure normality of distribution in the dependent variable, and to determine homogeneity of variances), and we determined no assumptions for ANOVA were violated. We detected no abnormalities in the data with the dependent variables normally distributed. We also conducted a test of homogeneity of variances to ensure that all groups had the same variance within the data. This assumption of ANOVA was not violated either. For single-digit word-problems, double-digit word-problems, and open equations, we did not identify significant differences in means at pretest. In addition, we calculated adjusted posttest means and these were not statistically different from unadjusted posttest means. Therefore, we used unadjusted posttest means in subsequent analyses for ease of interpretation.

At posttest, we noted no significance differences between the word-problem and control conditions on the single-digit word-problems, $F(1,122) = 1.10$, $p = .296$ ($ES = .19$) or open equations, $F(1, 122) = 2.60$, $p = .109$ ($ES = .29$). Although insignificant, we identified ESs favoring students in the word-problem intervention on both single-digit word-problems and open equations. We identified a statistically significant difference between conditions on double-digit word-problems, $F(1, 122) = 26.46$, $p < .001$, with an *ES* of .94 favoring students in the word-problem intervention. That is, students who participated in the word-problem intervention demonstrated significant improvement on a test of double-digit word-problems over students who did not receive the word-problem intervention.

Influence of Challenging Behavior

We compared the double-digit word-problems performance of the 74 students with MD who participated in the word-problem intervention. Before conducting the analysis, we coded whether students were considered at high-risk for externalizing or internalizing behaviors. The SRSS-IE considers a student at high-risk for externalizing behaviors at the elementary level if a student scored nine or above on the externalizing items (e.g., steal, lie, and cheat) and high-risk for internalizing behaviors (e.g., shy, depressed, and anxious) if a student scored four or above.

We identified 69 students considered at low risk for externalizing behaviors and five students considered at high risk for externalizing behaviors. On double-digit word-problems, the average score for low-risk students was 5.77 ($SD = 3.60$) and the average score for high-risk students was 1.60 ($SD = 2.07$). Using an ANOVA, we noted a significant difference between low-risk and high-risk students, $F(1, 72) = 6.483$, $p = .013$, with an *ES* of 1.17 favoring low-risk students. Results of a regression model, with double-digit word-problems posttest score as the outcome and externalizing behavior score (continuous) as the independent variable determined that, as externalizing scores increased by one point, students performed 2.98 points lower on double-digit word-problems, $F(1, 72) = 8.343$, $p = .005$.

We conducted similar analyses for internalizing behaviors. We identified 60 students at low risk for internalizing behaviors and 14 students at high risk. Low-risk students demonstrated a double-digit word-problems average score of 5.48 ($SD = 3.54$) and high-risk students demonstrated an average score of 5.50 ($SD = 4.29$). We calculated no significant difference between internalizing low-risk and high-risk students, $F(1, 72) = 0.000$, $p = .988$ ($ES = .01$).

Discussion

We conducted this study to determine whether students with MD demonstrated different externalizing and internalizing behavioral patterns compared with students without MD. After comparing students with and without MD, we focused on a subset of third-grade students with MD who participated in a randomized-control trial investigating the efficacy of a word-problem intervention. We investigated response to the word-problem intervention based on high or low risk for challenging behaviors.

Behavioral Patterns

In this study, we determined that students with MD displayed higher occurrences of externalizing and internalizing behaviors compared with students without MD. Specifically, students with MD had significantly higher externalizing and internalizing behaviors on the SRSS-IE as rated by their classroom teachers. This aligns with previous research, which suggested a minimum of 42% of students that demonstrate challenging behaviors exhibit evidence of

MD (Epstein et al., 2005). Therefore, it was expected that students with MD would demonstrate higher occurrences of at-risk behaviors.

Efficacy of Word-Problem Intervention

After understanding the differences in challenging behaviors for students with and without MD, we focused on the students with MD randomly assigned to receive word-problem intervention or to act as a business-as-usual comparison. Before determining how challenging behavior may influence performance in the word-problem intervention, we examined whether the intervention was efficacious. We determined students in the word-problem intervention outperformed students in the comparison group on the double-digit word-problems measure. These findings were consistent with previous iterations of the parent study, in which students in the word-problem intervention outperformed the comparison group on the outcome measures proximal to the intervention (Fuchs et al., 2008).

Challenging Behavior

After determining the efficacy of the word-problem intervention, we focused on the students with MD who actively participated in the word-problem intervention. We learned that students with MD with high occurrence rates of externalizing behaviors did not display the same pattern of gains as peers without high occurrence rates of challenging behaviors. Students with a low-risk status for externalizing behaviors ($n = 69$) outperformed students with high-risk status for externalizing behaviors ($n = 5$). Most notably, we found the greatest differences on double-digit word-problems, where students with low risk significantly outperformed students with high externalizing occurrence rates with an *ES* of 1.17. We also learned that, as externalizing behaviors increased by one point on the SRSS-IE, student word-problem gains decreased by 2.98 points.

Students with high occurrences of internalizing behaviors ($n = 14$) did not perform significantly different from students at low risk for internalizing behaviors ($n = 60$). Interestingly, the control groups for externalizing and internalizing behaviors did not display any significant differences. This confirms the hypothesis that students with externalizing types of challenging behaviors do not respond to academic interventions as well as students without externalizing behaviors and, in turn, may need more intensive intervention to see higher effects.

While a greater number of students with MD demonstrated high-risk internalizing behaviors (i.e., $n = 14$ internalizing vs. $n = 5$ externalizing), the posttest gains of students with high rates of internalizing behaviors did not significantly impact the outcome measures. Students with high occurrence rates of externalizing behaviors, however, demonstrated significantly lower word-problem performance when compared with students without high rates of externalizing behaviors. This establishes the need to intensify the word-problem intervention to ensure all students make expected growth, regardless of their behavioral challenges. Prior research demonstrated that students with challenging behaviors displayed greater achievement gaps as they progress through their academic careers (Wei et al., 2013). Therefore, intensifying interventions for students with MD and challenging behaviors becomes increasingly important to ameliorate future lack of response to intervention, especially as challenging behaviors become more prevalent as students approach middle and high school (Nelson et al., 2004).

Limitations

There are several limitations in the present study. First, we were unable to obtain behavioral data on 174 of the consented students with and without MD. We collected behavioral data from 29 full classrooms of third-grade students, but missing data from the other 10 classrooms, however, could have impacted the results of the findings. Second, as we identified students with MD, we removed some students because of behavioral challenges ($n = 9$) and very low English proficiency levels ($n = 24$) demonstrated during individual pretesting. Behavioral challenges included refusal to participate in pretesting and lack of cooperation with the tutor. If students demonstrated no understanding of test directions provided in English and could not provide English oral responses to simple questions, we confirmed with classroom teachers whether students had a minimum level of English proficiency to allow for participation in pretesting; in all suspected cases, teachers confirmed limited English proficiency. Our removal at pretest of these students with MD also could have impacted results. Another limitation was that the 50 students in the business-as-usual group did not necessarily receive individualized tutoring while the students in the word-problem intervention groups did receive such supplemental support. Students in the business-as-usual group may have received individualized tutoring from school personnel (and not research project tutors) but we did not collect such information. This study may have been more impactful if the comparison group had received some form of systematic individualized tutoring.

The SRSS-IE is primarily used as a behavioral screener in schools and is designed for administration 3 times per year to monitor student behavioral progress. In the present study, we collected SRSS-IE data only once, prior to the beginning of the word-problem intervention. Therefore, students identified with high occurrences of externalizing and internalizing behaviors at the beginning of the year may not necessarily be the same students identified as such in the middle of the school year (i.e., the end of the present

study). In addition, we identified very few students as having high externalizing ($n = 5$) and internalizing ($n = 14$) behaviors. We may have inadvertently removed students with high externalizing or internalizing behaviors during pretest (see previous paragraph) when we identified severe behavioral challenges that precluded students from participation in intervention. The lower number of students with high externalizing and internalizing behaviors may have impacted the results of the data due to the small number of students in each group. Finally, between screening and posttest, several students ($n = 7$) who scored high on the SRSS-IE for externalizing behaviors moved to other schools, had long-term suspensions, or displayed extreme behaviors during pretest. Although not surprising due to the somewhat behavioral-related transiency of students with challenging behaviors (Mattson et al., 2015), it is an important consideration.

Future Research

Future research should continue to explore the link between behavior and mathematics. First, research should be conducted in other areas of mathematics intervention research to identify if challenging behaviors impact more than word-problem-solving. In addition, future research should explore how to intensify word-problem interventions for students with higher occurrences of externalizing behaviors. This intervention was conducted one-on-one and was already considered a form of intensification (National Center on Intensive Intervention, 2016). Therefore, we should examine the results of (a) adjusting the word-problem intervention components, specifically increasing reinforcement rates or performing preference assessments to determine specific reinforcers, (b) increasing dosage, (c) using more three-dimensional representations, and (d) explicitly teaching transfer skills to increase opportunities for practice, or other recommendations from the Taxonomy of Intervention Intensity (Fuchs et al., 2017). Finally, future research should replicate this study in other geographical areas of the United States and beyond to examine whether similar results are observed.

Summary

In conclusion, the purpose of this study was to examine the intersection of behavior and mathematics, specifically the influence of behavior within a word-problem intervention for students with MD. We identified a significant difference in the behavioral patterns of students with MD and those without MD. In addition, students with MD in the intervention group outperformed students in the control group on all posttest outcomes but significant differences emerged on double-digit word-problems, a measure of word-problems similar to those from high-stakes assessments. Moreover, higher scores on the SRSS-IE for students with high occurrences of externalizing behaviors corresponded with significantly lower scores on the word-problem outcome measure compared with peers without high occurrences of externalizing behaviors. Students with co-occurring behavior and MD may require more intensive intervention to demonstrate similar results as their peers without behavioral challenges.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences.

Declaration of Conflicting Interestsv

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 R324A150078 from the Institute of Education Sciences in the U.S. Department of Education to the University of Texas at Austin.

References

Alter, P. (2012). Helping students with emotional and behavioral disorders solve mathematics word problems. *Preventing School Failure*, *56*, 55–64. https://doi.org/10.1080/1045988X.2011.565283

Alter, P., Brown, E. T., & Pyle, J. (2011). A strategy-based intervention to improve math word problem-solving skills of students with emotional and behavioral disorders. *Education and Treatment of Children*, *34*, 535–550. https://doi.org/10.1353.2011.0028

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Blackorby, J., Wagner, M., Cameto, R., Davies, E., Levine, P., Newman, L., . . . Sumi, C. (2004). *Engagement, academics, social adjustment, and independence: The achievements of elementary and middle school students with disabilities*. SRI International.

Bullis, M., & Yovanoff, P. (2006). Community employment experiences of formerly incarcerated youth. *Journal of Emotional and Behavioral Disorders*, *14*, 71–85. https://doi.org/10.1177/10634266060140020401

Devine, A., Hill, F., Carey, E., & Szücs, D. (2018). Cognitive and emotional math problems largely dissociate: Prevalence of developmental dyscalculia and mathematics anxiety. *Journal of Educational Psychology*, *110*, 431–444. https://doi.org/10.1037/edu0000222

Drummond, T. (1994). *The Student Risk Screening Scale-Internalizing and Externalizing (SRSS-IE)*. www.CI3T.org

Ennis, R. P., Lane, K. L., & Oakes, W. P. (2012). Score reliability and validity of the Student Risk Screening Scale: A psychometrically sound, feasible tool for use in urban elementary

schools. *Journal of Emotional and Behavioral Disorders*, *20*, 241–259. https://doi.org/10.1177/1063426611400082

Epstein, M. H., Nelson, J. R., Trout, A. L., & Mooney, P. (2005). Achievement and emotional disturbance: Academic status and intervention research. In M. H. Epstein, K. Kutash, & A. J. Duchnowski (Eds.), *Outcomes for children and youth with emotional and behavioral disorders and their families: Programs and evaluation practices* (2nd ed., pp. 451–477). Pro-Ed.

Forness, S. R., Freeman, S. F. N., Paparella, T., Kauffman, J. M., & Walker, H. M. (2012). Special education implications of point and cumulative prevalence for children with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *20*, 4–18. https://doi.org/10.1177/1063426611401624

Fuchs, L. S., Fuchs, D., & Malone, A. S. (2017). The taxonomy of intervention intensity. *Teaching Exceptional Children*, *50*, 35–43. https://doi.org/10.1177/0040059917703962

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., . . . Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to pre-algebraic knowledge? *Journal of Educational Psychology*, *106*, 990–1006. https://doi.org/10.1037/a0036793

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learning and Individual Differences*, *20*, 89–100. https://doi.org/10.1016/j.lindif.2009.09.003

Fuchs, L. S., Seethaler, P. M., Powell, S. R., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2008). Effects of preventative tutoring on the mathematical problem solving of third-grade students with math and reading difficulties. *Exceptional Children*, *74*, 155–173. https://doi.org/10.1177/001440290807400202

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

Hodge, J., Riccomini, P. J., Buford, R., & Herbst, M. (2006). A review of instructional interventions in mathematics for students with emotional and behavioral disorders. *Behavioral Disorders*, *31*, 297–311. https://doi.org/10.1177/019874290603100304

Jitendra, A. K., Harwell, M. R., Dupuis, D. N., & Karl, S. R. (2017). A randomized trial of the effects of schema-based instruction on proportional problem-solving for students with mathematics problem-solving difficulties. *Journal of Learning Disabilities*, *50*, 322–336. https://doi.org/10.1177/0022219416629646

Jitendra, A. K., Rodriguez, M., Kanive, R., Huang, J.-P., Church, C., Corroy, K. A., & Zaslofsky, A. (2013). Impact of small-group tutoring interventions on the mathematical problem solving and achievement of third-grade students with mathematics difficulties. *Learning Disability Quarterly*, *36*, 21–35. https://doi.org/10.1177/0731948712457561

Jitendra, A. K., Star, J., Starosta, K., Leh, J., Sood, S., Caskie, C., Hughes, C., & Mack, T. (2009). Improving seventh grade students' learning of ratio and proportion: The role of schema-based instruction. *Contemporary Educational Psychology*, *34*, 250–264. https://doi.org/10.1016/j.cedpsych.2009.06.001

Jordan, N. C., & Hanich, L. B. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*, *33*, 567–578. https://doi.org/10.1177/002221940003300605

Lane, K. L. (2004). Academic instruction and tutoring interventions for students with emotional/behavioral disorders: 1990 to present. In R. B. Rutherford, M. M. Quinn, & S. R. Mathur (Eds.), *Handbook of research in emotional and behavioral disorders* (pp. 462–486). Guilford Press.

Lane, K. L., Barton-Arwood, S. M., Nelson, J. R., & Wehby, J. (2008). Academic performance of students with emotional and behavioral disorders served in a self-contained setting. *Journal of Behavioral Education*, *17*, 43–62. https://doi.org/10.1007/s10864-007-9050-1

Lane, K. L., Menzies, H. M., Oakes, W. P., Lambert, W., Cox, M. L., & Hankins, K. E. (2012). A validation of the Student Risk Screening Scale for internalizing and externalizing behaviors: Patterns in rural and urban elementary schools. *Behavioral Disorders*, *37*, 244–270. https://doi.org/10.1177/019874291203700405

Lane, K. L., Oakes, W. P., Harris, P. J., Menzies, H. M., Cox, M. L., & Lambert, W. (2012). Initial evidence for the reliability and validity of the Student Risk Screening Scale for internalizing and externalizing behaviors at the elementary level. *Behavioral Disorders*, *37*, 99–122. https://doi.org/10.1177/019874291203700204

Lane, K. L., Richards-Tutor, C., Oakes, W. P., & Connor, K. (2014). Initial evidence for the reliability and validity of the student risk screening scale with elementary age English learners. *Assessment for Effective Intervention*, *39*, 219–232. https://doi.org/10.1177/1534508413496836

Losinski, M., Ennis, R. P., Sanders, S. A., & Nelson, J. (2019). Examining the evidence-base of mathematical interventions for students with emotional disturbances: A meta-analysis. *The Journal of Special Education*, *52*, 228–241. https://doi.org/10.1177/0022466918796200

Maggin, D. M., Wehby, J. H., Farmer, T. W., & Brooks, D. S. (2016). Intensive interventions for students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *24*(3), 127–137. https://doi.org/10.1177/1063426616661498

Mattson, B., Forlenza, V. C., Hamilton, S., & Cohen, C. (2015). *Summary of school and district strategies for transient/mobile students: Prepared for Pennsylvania Department of Education academic recovery liaisons* [A report from the Mid-Atlantic Comprehensive Center at WestEd]. WestEd.

Mazzocco, M. M. M. (2007). Defining and differentiating mathematical learning disabilities and difficulties. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 29–48). Paul H. Brookes.

Menzies, H., & Lane, K. L. (2012). Validity of the Student Risk Screening Scale: Evidence of predictive validity in a diverse, suburban elementary setting. *Journal of Emotional and Behavioral Disorders*, *20*, 82–91. https://doi.org/10.1177/1063426610389613

Mooney, P., Epstein, M. H., Reid, R., & Nelson, J. R. (2003). Status of and trends in academic intervention research for

students with emotional disturbance. *Remedial and Special Education*, *24*, 272–287. https://doi.org/10.1177/07419325030240050301

Mulcahy, C. A., & Krezmien, M. P. (2009). Effects of a contextualized instructional package on the mathematics performance of secondary students with EBD. *Behavioral Disorders*, *34*, 136–150. https://doi.org/10.1177/0741932515579275

Mulcahy, C. A., Krezmien, M. P., & Travers, J. (2016). Improving mathematics performance among secondary students with EBD: A methodological review. *Remedial and Special Education*, *37*, 113–128. https://doi.org/10.1177/0741932515579275

National Center on Intensive Intervention. (2016). *How to intensify academic interventions*. National Center on Intensive Intervention, Office of Special Education Programs, U.S. Department of Education.

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children*, *71*, 59–73. https://doi.org/10.1177/001440290407100104

Oakes, W. P., Wilder, K., Lane, K. L., Powers, L., Yokoyama, L., O'Hare, M. E., & Jenkins, A. B. (2010). Psychometric properties of the Student Risk Screening Scale: An effective tool for use in diverse urban elementary schools. *Assessment for Effective Intervention*, *35*, 231–239. https://doi.org/10.1177/1534508410379796

Peltier, C., & Vannest, K. J. (2016). Schema-based strategy instruction and the mathematical problem-solving performance of two students with emotional or behavioral disorders. *Education and Treatment of Children*, *39*, 521–544. https://doi.org/10.1353/etc.2016.0023

Powell, S. R. (2007). *Open equations*. The University of Texas at Austin.

Powell, S. R., & Berry, K. A. (2015). *Texas word problems*. The University of Texas at Austin.

Powell, S. R., Berry, K. A., & Barnes, M. A. (2020). The role of pre-algebraic reasoning with a word-problem intervention for third-grade students with mathematics difficulty. *ZDM Mathematics Education*, *52*, 151–163. https://doi.org/10.1007/s11858-019-01093-1

Powell, S. R., Fuchs, L. S., Cirino, P. T., Fuchs, D., Compton, D. L., & Changas, P. C. (2015). Effects of a multitier support system on calculation, word problem, and prealgebraic performance among at-risk learners. *Exceptional Children*, *81*(4), 443–470. https://doi.org/10.1177/0014402914563702

Ralston, N. C., Benner, G. J., Tsai, S. F., Riccomini, P. C., & Nelson, J. R. (2014). A best evidence synthesis of the effects of mathematics instruction for students with emotional and behavioral disorders. *Preventing School Failure*, *58*, 1–16. https://doi.org/10.1080/1045988X.2012.726287

Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education*, *38*, 130–144. https://doi.org/10.1177/00224669040380030101

Rosenzweig, C., Krawec, J., & Montague, M. (2011). Metacognitive strategy use of eighth-grade students with and without learning disabilities during mathematical problem solving: A think-aloud analysis. *Journal of Learning Disabilities*, *44*, 508–520. https://doi.org/10.1177/0022219410378445

Ryan, J. B., Pierce, C. D., & Mooney, P. (2008). Evidence-based teaching strategies for students with EBD. *Beyond Behavior*, *17*(3), 22–29.

Skagerlund, K., & Träff, U. (2016). Number processing and heterogeneity of developmental dyscalculia: Subtypes with different cognitive profiles and deficits. *Journal of Learning Disabilities*, *49*, 36–50. https://doi.org/10.1177/0022219414522707

Templeton, T. N., Neel, R. S., & Blood, E. (2008). Meta-analysis of math interventions for students with emotional and behavioral disorders. *Behavioral Disorders*, *16*, 226–239. https://doi.org/10.1177/1063426608321691

van Garderen, D. (2007). Teaching students with LD to use diagrams to solve mathematical word problems. *Journal of Learning Disabilities*, *40*, 540–553. https://doi.org/10.1177/00222194070400060501

Wei, X., Lenz, K. B., & Blackorby, J. (2013). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*, 154–165. https://doi.org/10.1177/0741932512448253

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education. (2017, October). *What Works Clearinghouse: Procedures handbook* (Version 4.0). https://whatworks.ed.gov

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *The Journal of Special Education*, *39*, 181–192. https://doi.org/10.1177/00224669050390030501