Didion, L., Jeppsen, C., Hoffman, L., & Smith, B. P. (2025). Extending Data Mountain to improve fluency for elementary students with or at-risk for reading disabilities. *Research in Special Education*, *2*, 1–24. https://doi.org/10.25894/rise.2614

RESEARCH IN SPECIAL EDUCATION

ALETHEIA SOCIETY

EMPIRICAL ARTICLE

Extending Data Mountain to Improve Fluency for Elementary Students with or At-Risk for Reading Disabilities

Lisa Didion*, Department of Special Education, University of Kansas, lisadidion@ku.edu

Charlotte Jeppsen, Department of Psychological and Quantitative Foundations, University of Iowa, charlotte-jeppsen@uiowa.edu

Lesa Hoffman, Department of Psychological and Quantitative Foundations, University of Iowa, lesa-hoffman@uiowa.edu

Benjamin P. Smith, Department of Psychological and Quantitative Foundations, University of Iowa, benjamin-p-smith@uiowa.edu

*Corresponding author

ABSTRACT

We conducted a randomized controlled trial to replicate the efficacy of a self-determination learning program, Data Mountain, to improve the oral reading fluency performance of students with or at-risk for reading disabilities. This study included 109 students in second through fifth grade enrolled in three rural public schools. Participants were randomized into one of three conditions: Data Mountain delivered individually, Data Mountain delivered in a small group, or a comparison condition. In the Data Mountain conditions, students evaluated their progress on oral reading fluency via a line graph, set goals on the number of words read correctly in one minute, and attributed

Submitted: 30 July 2024 **Accepted:** 29 March 2025 **Published:** 02 June 2025

success with strategy use. The original program was 15 lessons, and 15 new lessons were developed for the present study, for a total of 30. These newly developed lessons scaffolded learning, and students independently realized the self-determination process. Multilevel models were used to examine differences between Data Mountain students and those in the comparison condition, and results indicated that treatment effects varied by tutor and year. There were significant differences between Data Mountain and comparison students in the second year only. Delivery formats were also compared, and there were no differences found between students receiving Data Mountain individually and those receiving it in a small group. Pretest reading fluency scores moderated the effects—Data Mountain may be more effective for students with or at risk for reading disabilities with a higher level of reading proficiency. Discussion is centered around implications of COVID-19 and replication research.

Keywords: self-determination; self-monitoring; goal setting learning disabilities; reading

The coronavirus pandemic (COVID-19) caused unprecedented academic disruptions (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2021) and inconsistency in school engagement nationwide (Colvin et al., 2022). Before then, only approximately 35% of students could read at proficient levels by fourth grade (National Assessment of Educational Progress [NAEP], 2022). The pandemic reversed the incremental progress of previous decades (NAEP, 2022). Students with and without disabilities both experienced significant reading loss from pre- to post-pandemic (e.g., Betthäuser et al., 2022). Findings suggested that students, on average, lost half a year of learning (Patrinos et al., 2022). Some evidence indicated the largest impact was on already marginalized students, including students with learning disabilities (e.g., Ponce, 2021). Others suggested that students without disabilities suffered a somewhat greater loss of reading (Chatzoglou et al., 2023). Regardless of who lost more, in general, growth rates on oral reading fluency (ORF) post-COVID-19 slowed compared to before the pandemic (Domingue et al., 2022; Kuhfeld et al., 2022). As such, ORF could serve as an indicator of how regional and national emergencies (e.g., pandemics, political unrest, significant weather patterns) affect student learning.

Education stands at a crucial point, as the long-term impact of COVID-19 remains unknown, and students with disabilities are more than three years behind their peers in reading (Gilmour et al., 2019). The Matthew Effect (Stanovich, 2009) predicts that students who struggle with reading will not engage in reading tasks as often as proficient readers, reducing their opportunities to practice and, in turn, their overall growth in reading. Students with reading difficulties may experience a change in motivation, evidenced by an increased reluctance to read. Broadly, researchers associate motivation with reading ability (e.g., Logan et al., 2011; Morgan & Fuchs, 2007; Retelsdorf et al., 2011; Taboada et al., 2009; Toste et al., 2020), and the decline in motivation may begin as early as elementary school (Jacobs et al., 2002; Petscher, 2010; Spinath & Spinath, 2005; Stoel et al., 2001). Now more than ever, it is essential to recognize that improving motivation to read can greatly impact reading outcomes.

At a systematic level, schools should support the development of knowledge and skills alongside self-regulatory abilities (e.g., self-determination skills) and motivations (Darling-Hammond et al., 2020). Immediate attention is warranted to embed self-determination learning into academic routines and instruction to allow students to optimize their entire learning profile. Given the significant learning loss due to COVID-19 and its implications on ORF, we aimed to examine the extended effect of a self-determination learning program, Data Mountain, on the fluency performance of elementary students with or at risk for reading disabilities (RD).

SELF-DETERMINATION IMPROVES ACADEMIC OUTCOMES

Interventions that target self-determination enable students to maximize the learning opportunities available to them (Lazowski & Hulleman, 2016; Yaeger & Walton, 2011). Many authors have called for more experimental studies in education to better understand how to improve educational outcomes through motivation (see Blackwell et al., 2007; Lazowski & Hulleman, 2016; Maehr & Meyer, 1997; Midgley & Edelin, 1998; Pintrich, 2003). Lazowski and Hulleman (2016) examined 74 published and unpublished intervention studies grounded in motivation research (e.g., self-determination theory, attribution theory). They found that motivation interventions had an average effect size of $d = 0.49$ on educational outcomes (e.g., performance, behaviors). No significant differences were observed based on theoretical frameworks, participant characteristics, or the type of educational outcome as moderators of the effects. Notably, this effect size is considered substantial in the context of education research, in which effect sizes above 0.20 are regarded as large (Kraft, 2020). Despite the high impact of motivation interventions, few studies have been identified that focus on teaching self-determination skills (e.g., self-monitoring, goal setting) to students with RD to improve academic outcomes, especially for elementary-aged students (Didion et al., 2021; Konrad et al., 2007). Similarly, just 10% of the studies included in the meta-analysis conducted by Lazowski and Hulleman examined motivation interventions for elementary students. Moreso, replications of motivational research are rare and merit attention (Yaeger & Walton, 2011).

Students with disabilities express lower levels of self-determination relative to their peers, and researchers believe this is connected to limited opportunities to act as causal agents (Shogren et al., 2018). Causal Agency Theory defines self-determination as a "dispositional characteristic manifested as acting as the causal agent in one's life" (Shogren et al., 2015, p. 258). Being self-determined allows individuals to cause things to happen in their own lives as they work toward goals (Shogren et al., 2015). Causal Agency Theory identifies several abilities and skills as integral for developing self-determination, including choice-making, decision-making, goal setting and attainment, problem-solving, self-advocacy, self-management, and self-monitoring (Shogren et al., 2017). This paper specifically focuses on self-monitoring and goal setting with attainment plans. *Self-monitoring*, a subcomponent of self-management, involves assessing and recording the occurrence of a target behavior. *Goal setting with attainment plans* involves identifying achievable goals and developing objectives, steps, and actions necessary to bring the goal to fruition (Didion & Toste, 2021). We conceptualize attainment plans within attribution retraining. In this way, students select reading strategies to support their growth and consider if the strategy was helpful in reaching their personal best. They learn to attribute their success (or failure) to their use of strategy.

By repeatedly self-monitoring, setting goals, and strategizing attainment plans, individuals can directly associate their actions with their outcomes (Shogren et al., 2017). As students master challenging tasks, they recognize that their skills and strategies shape the results, increasing the likelihood they will engage with similar tasks in the future (Ryan & Deci, 2017; Shogren et al., 2017). For example, students can self-monitor their progress and set goals for the number of words read correctly per minute (WPM). This cyclical process requires them to repeatedly evaluate the success of the reading strategies they use to reach their goals (see Figure 1). When students see that a strategy supported their goal attainment, they are more likely to use that strategy in the future when faced with the same task (Shogren et al., 2017). Schools could include self-monitoring and goal setting alongside school-wide data collection procedures (e.g., multi-tiered systems of support, data-based decision-making, intensive intervention), turning progress monitoring sessions into a framework for students to learn self-determination skills alongside academics.

THE DATA MOUNTAIN PROGRAM

Data Mountain embeds self-determination learning within progress monitoring—students set goals, self-monitor goal achievement, evaluate progress over time, and assess their use of strategies. Students come to understand that variable

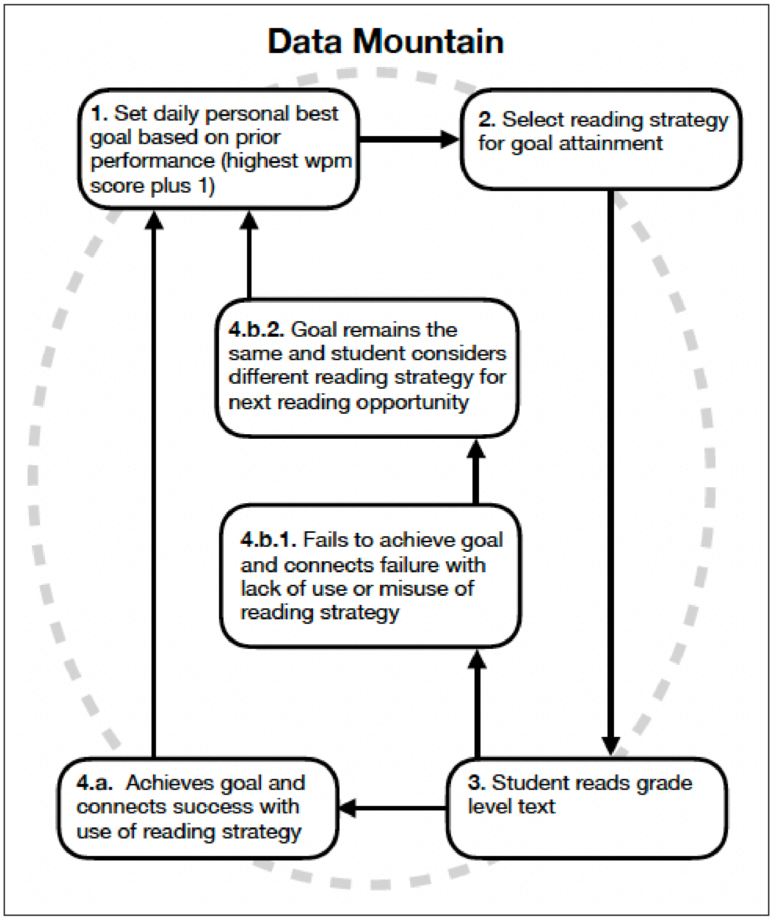

Figure 1:

Note. Through repeated practice with Data Mountain, students will self-monitor and set goals. They recognize the extent reading strategies attributed to their performance. This process is embedded with the context of ORF progress monitoring.

growth is normal and does not necessarily imply a lack of ability. They attribute reaching a daily reading goal to their successful use of an individualized reading strategy (e.g., sound out the word, phrase read, follow with your finger). In contrast, they associate not reaching their daily goal with their selection of the individualized reading strategy (e.g., already a strength, mismatch for their skillset). According to Causal Agency Theory, participation in Data Mountain allows students repeated opportunities to engage in the same reading fluency task and select reading strategies to achieve their daily WPM goal (see Figure 1). Students may become more internally motivated as they continuously self-monitor their achievement of personal best goals on a line graph across sessions. As causal agents, they observe that their performance is directly related to their own actions (i.e., the strategies they used; De Charms, 2013).

Initial studies of Data Mountain examined how this brief (approximately 6 min) intervention impacted ORF growth rate. Researchers found promising results, showing that after 15 sessions, the program improved the ORF growth rate of struggling readers in elementary grades (Didion & Toste, 2021; Didion et al., 2020) and post-secondary students with disabilities (Didion et al., 2024). In a pilot randomized controlled trial (Didion & Toste, 2021) with 81 students in grades 2–5, Data Mountain delivered individually (DM-I) or in small groups (DM-G) was compared to a control (COM). Results showed students in Data Mountain read 31 more words per minute by the end of the program, with growth rates of 2.09 WPM (DM-I), 1.65 WPM (DM-G), and 0.80 WPM (COM) *each session* when accounting for grade level, pretest fluency, and English Learner status.

Instructional Groupings

When comparing instructional groupings, DM-I students increased at a higher rate than DM-G students, although this difference was not statistically significant (Didion & Toste, 2021). Descriptive statistics revealed that DM-I students started the intervention with a lower average number of words per minute (M = 76.0) compared to DM-G (M = 77.9) and COM (M = 79.5). Interestingly, posttest data indicated that DM-I achieved a higher average WPM (M = 108.2) compared to DM-G (M = 103.6) and COM (M = 87.5). It was hypothesized that the lack of significance could be related to the limited duration of the intervention or limited power related to sample size.

The higher growth rate associated with the DM-I condition may stem from the intensive nature of individualized instruction (Vaughn & Wanzek, 2014). Students facing persistent reading challenges typically require tailored support to maximize their learning opportunities (Fuchs et al., 2014). In the pilot randomized controlled trial (Didion & Toste, 2021), the hypothesis was that DM-G students would show greater improvement due to the collaborative learning approach and peer support. Positive peer relationships are known to correlate with improved academic outcomes (Hamm & Zhang, 2010; Ladd & Herald, 2009) because they provide support, validation, modeling, and assistance (Bukowski et al., 2009) In Data Mountain, the peer model provides students the opportunity to have both a model of reading and the opportunity to discuss strategies to improve fluency efforts with their peers. Given that results showed no indication of enhanced effects with peer models in the pilot study, we predict that DM-I students will demonstrate a higher growth rate than DM-G students when given more time with the intervention.

Further investigation over an extended period is necessary to explore the moderating effects of delivery format and pretest performance. Peer modeling in the reading literature has demonstrated effectiveness in improving fluency among struggling readers (Chard et al., 2002; Stevens et al., 2017). Given the importance of collaboratively learning self-regulation skills with peers and teachers (Darling-Hammond & Friedlaender, 2008; Darling-Hammond et al., 2020), we should consider that students with lower baseline scores could benefit from peer models in a small group, potentially enhancing their performance. Conversely, moderation effects related to pretest ORF performance could further support the argument that intensive one-on-one instruction is crucial for our lowest performers.

Purpose

Psychological processes require time to affect change because they operate through experiences of success and failure (Yeager & Walton, 2011). Extending the time spent progressing through the causal agency cycle may lead to variations in growth rates over additional sessions beyond those observed across the 15 sessions in the intervention applied in previous studies. For the present study, the aim of adding 15 lessons was to enhance students' self-directed learning. In the initial 15 lessons, the process of self-monitoring and goal setting was primarily guided by teachers, limiting students from fully assuming the role of causal agents. Taking self-directed steps toward a goal allows for agentic actions (Shogren et al., 2015), promoting progress toward identified objectives (Shogren & Raley, 2022). When students actively engage in their intervention, they have the potential to apply this knowledge in new contexts and take credit for their successes, rather than attributing progress solely to instructor support (Yeager & Walton, 2011). Developing self-management skills is crucial for fostering self-determination and enabling students to direct their progress toward their goals (Shogren & Raley, 2022). Questions remain regarding the extent to which student performance can improve if provided with additional time to practice self-determination skills.

The current study examined the effect of adding 15 lessons to complete the Data Mountain program (30 sessions total). These lessons allow students to guide themselves through the intervention with scaffolded support. It was hypothesized that the additional sessions would increase the growth rate similar to the first 15 sessions, resulting in significant improvement among students with or at risk for RD. Due to the intensity of the one-on-one instruction, we hypothesized that the growth rate of students in the DM-I condition would be greater than that of the DM-G condition. We also considered the moderating effects of pretest ORF rates and participating in DM-I or DM-G. Therefore, we formulated four research questions (RQ):

1. Do second through fifth grade students with or at-risk for RD demonstrate increased ORF growth when participating in the extended Data Mountain program as compared to a comparison condition?
2. Does ORF grow at the same rate during Data Mountain sessions 16–30 as compared to sessions 1–15?
3. Do pretest ORF rates moderate the effects of Data Mountain on growth over time?
4. Does ORF grow at a higher rate when the extended Data Mountain program is delivered individually (DM-I) as compared to a small group (DM-G)?

METHOD

Participants

This study took place in three elementary schools in rural areas of a Midwestern state across two study years (one school in year 1 and two schools in year 2). Rural areas were defined as counties or clusters outside of metro areas with less than 25,000 people. Overall, the percentage of children achieving proficiency on statewide standardized reading tests in School 1 (year 1) was 57%, in School 2 (year 2) was 58%, and in School 3 (year 2) was 77% (the state average was 68%). Schools 1 and 2 were in the same district. The percentage of students eligible in Schools 1, 2, and 3 for free or reduced-price lunch was 57%, 58%, and 25%, respectively.

Due to COVID-19, during the school year before the start of the study (2020–2021) participating schools had used a mix of in-person, hybrid, and virtual learning, along with multiple transitions per year from the different modes of learning with mandated face coverings. For year 1 (2021–2022) in School 1, at the start of the study, the fall pretest occurred in the first months back to in-person learning. For year 2, Schools 2 and 3 had a full year of in-person learning before the study started.

First, a university research ethics review board approved the study. Then, we preregistered the study on the Open Science Framework (OSF; https://doi.org/10.17605/OSF.IO/TEXRY). Second, students with or at-risk for RD were recruited for potential inclusion by nominations from teachers, reading specialists, and administrators. Then, nominated students were screened using the Test of Word Reading Efficiency-Second Edition (TOWRE-2) and included if they scored below the 25th percentile on either of its subtests (see measures section). The final sample included 109 students (54% male). All schools used a Tier 1 curriculum that included decodable books and explicit phonics instruction. School 1 had 54 participants, School 2 had 31 participants, and School 3 had 24 participants. Across all schools, 27 students received special education services in reading, one student received accommodations under Section 504 (34 C.F.R. Part 104.4), 87 students received either Tier 2 or Tier 3 reading services, and 85 received free or reduced-price lunch. At baseline, the final sample included 29 students in second grade, 32 in third grade, 27 in fourth grade, and 21 in fifth grade. According to school administration, 71% of students were White, 6% were Black, 18% were Hispanic, 5% were multiracial, and 1% were Asian.

Participants at each grade level per each school were randomized into one of three conditions, which resulted in unbalanced group size across the three schools: Data Mountain delivered individually (DM-I, n = 41), Data Mountain delivered in a small group (DM-G, n = 35), and comparison (COM, n = 33). In addition to the type of intervention, students were randomly assigned to one of two tutors in School 1. Students in School 1 received the intervention in 2021 and students in Schools 2 and 3 and received the intervention in 2022. To account for group dependency in both initial ORF levels and the magnitude of growth, students were categorized into one of four clusters (unique combinations by school, year, and tutor) before conducting the analysis. See Table 1 for demographic data by tutor and year.

	YEAR 1 TUTOR 1			**YEAR 1 TUTOR 2**			**YEAR 2 TUTOR 2**			**YEAR 2 TUTOR 3**		
Special Education	40.74%			18.52%			16.13%			11.11%		
Tier 2	59.26%			81.15%			74.19%			42%		
Gender	70.37% male			62.96% male			48.28% male			50% male		
	DM-I	DM-G	COM	DM-I	DM-G	COM	DM-I	DM-G	COM	DM-I	DM-G	COM
TOWRE-2												
Pre												
M	78.71	67.5	71.1	78.2	74.75	71.5	73.17	76.1	78.38	81.38	72.88	81.88
SD	(4.57)	(4.25)	(3.21)	(2.55)	(3.34)	(2.33)	(3.35)	(4.34)	(2.34)	(2.58)	(3.81)	(3.0)
Post												
M	84.13	75.75	79.64	92.13	87.5	91	83.27	85.5	86.33	90.14	79.75	90.75
SD	(4.96)	(5.58)	(3.03)	(3.84)	(4.5)	(4.04)	(4.73)	(4.46)	(2.82)	(3.62)	(3.91)	(3.28)
FastBridge												
Pre												
M	77.88	44.75	67.73	87.07	59.75	37	58.27	65.9	78.67	90.63	55.75	93.25
SD	(16.49)	(16.6)	(9.3)	(11.33)	(13.64)	(2.31)	(10.78)	(11.85)	(12.38)	(18.74)	(13.52)	(12.99)
Session 1												
M	89	51.38	65.27	96.53	72.25	47.5	68.27	79.1	77.22	95.75	71.88	94.63
SD	(17.57)	(18.09)	(10.01)	(11.84)	(15.31)	(3.97)	(14.12)	(12.71)	(10.17)	(16.31)	(15.29)	(13.17)
Post												
M	114.25	77.38	92.91	132.27	107.38	87.25	98.36	107	104.33	147	108.38	130.38
SD	(18.91)	(20.73)	(10.35)	(13.81)	(14.72)	(6.3)	(14.72)	(12.34)	(8.61)	(17.23)	(19.33)	(11.11)
Total n	8	8	11	15	8	4	12	10	9	8	8	8

Table 1: Demographic Information and Average Oral Reading Fluency Performance by Cohort.

Note. DM-I = Data Mountain delivered individually; DM-G = Data Mountain delivered in a small group; COM = comparison group; TOWRE-2 = Test of Word Reading Efficiency, 2nd edition; TOWRE-2 reports scaled scores; FastBridge reports words read correct per minute.

Attrition Analysis

Overall, 204 students qualified for the study and 113 students returned consent forms. One student moved during the consenting process and a second student withdrew when teachers noted scheduling conflicts with other educational services. The remaining 111 students were randomly assigned and completed the pretest battery. Two DM-G students withdrew after the training lessons. Then, 109 students began the study, and 107 students completed the post-test battery. One DM-I student withdrew after session 5, and one DM-I student withdrew after session 8. Overall attrition was 3.6%. Differential attrition between Data Mountain and comparison was 5.2% and between DM-I and DM-G was 1.2%. Considering overall and differential attrition estimates, conservative attrition standards indicate low attrition and expected bias (What Works Clearinghouse [WWC], 2020). All 109 participants who began the study and had progress monitoring data were included in the analysis (see Figure S1 in supplemental materials for a CONSORT flow diagram).

Study Design and Procedures

First, students were administered the pre-test battery. Then, intervention activities (DM-I, DM-G, or COM) began. Students received up to 30 sessions across all conditions with a tutor. Depending on schedules, 2 or 3 sessions were delivered each week (see Didion & Toste, 2021, for rationale for varied sessions each week). After session 30, all students were administered the post-test battery.

Next, we outline the procedures for progress monitoring, the training lesson, and the 30 lessons of Data Mountain. More specific details about the first 15 lessons can be found in the pilot study (Didion & Toste, 2021). The only change to Data Mountain in the first 15 lessons in the present study was to include the terms *increase* and *decrease* instead of asking students to observe if their mountains (line graphs) "go up" or "go down." This was designed to contribute to their overall vocabulary and ability to communicate mathematics concepts (Powell et al., 2023).

Progress Monitoring Activities (DM-I, DM-G, COM)

For all 30 sessions, students in all conditions were presented with a randomly assigned grade-level ORF passage and instructed to read it for one minute. The tutor read directions from the standardized administration guide and then the student began reading aloud. At the end of one minute, the tutor calculated the WPM by counting the total number of words read correctly minus the errors made. Students in the COM condition were only progress monitored, and tutors did not share progress monitoring data with COM students, nor did they provide any reading instruction. Students in the DM-I and DM-G conditions had Data Mountain layered on top of their progress monitoring.

Data Mountain Activities (DM-I, DM-G)

Tutors used Data Mountain to teach self-monitoring, goal setting with plans for attainment, and evaluation of strategy use. The intervention began with one training lesson immediately after all students had completed the pre-test battery. All DM-I and DM-G students received the training lesson grounded in explicit instruction in the format of their intervention group—DM-I students completed the training lesson individually and DM-G students completed the training lesson in their small groups. The training provided steps for self-monitoring and goal setting and introduced vocabulary related to data analysis: data, trend, variability, goals, and line graphs. Students were introduced to the abstraction of progress with a line graph illustrated as a mountain—Data Mountain. They were taught that progress is measured by data, which can increase or decrease (variability) like peaks and valleys on a mountain. They learn that progress monitoring allows for observation of performance and evaluation to determine if data are increasing overall, like a mountain incline (trend). First, the tutor modeled scripted self-talk by sharing fictional data (i.e., running distances and times) with students via a line graph that drew attention to low-level data points (i.e., when little distance was covered). Program language was modeled around effort and strategy use to increase the data mountain (e.g., "I'll eat breakfast to increase my running speed"). The tutor also modeled how to graph a data point. Then, the students discussed data using a researcher-created worksheet that included a line graph and eight questions pertaining to a scenario of an elementary student's self-monitoring of math scores. The questions guided students to observe the line graph while considering changes to trend. The students practiced graphing data. Finally, the training lesson concluded with conversation about the students' individual ORF performance. Each student was presented with a graph of their baseline ORF data as collected during pretest (see Data Collection and Measures).

After the training lesson, students were administered 30 Data Mountain lessons two or three times per week. Please refer to previous work (Didion & Toste, 2021; Didion et al., 2020) for more information about the scope and sequence for lessons 1–15. For the present study, 15 new lessons were developed to allow students to self-direct the steps of the intervention with scaffolded support. See Figure S2 in the supplemental material for the script of the training lesson and examples of intervention lessons (Lesson 9; Lesson 20).

Broadly, in lessons 1–15, the tutor taught positive attributions by modeling positive self-talk and assisting students to generate their own self-motivated statements for use during reading tasks. This process involved using fictitious story vignettes depicting students experiencing reading difficulties, which were paired with discussions about the thoughts these students might have when facing challenging reading tasks. Drawing from strategies commonly taught by classroom teachers or specialists, students identified potential approaches that the fictitious student could use to overcome their challenges. The student and tutor discussed how to change negative thoughts into positive thoughts and link to strategies that the fictitious student could use to persevere. Program language emphasized that negative thoughts indicate that a task is personally difficult, but that negative thinking slows progress down, whereas positive thoughts remind us of strategies that help learners reach their goals. During lessons 8–15, the tutor and student identified individual strengths and weaknesses related to ORF. The student reflected on their prior reading strategy knowledge to create the personalized list tailored to their specific needs. When necessary, the tutor recommended commonly used strategies to support fluency, such as sounding out words, chunking, and tracking text with a finger. Tutors did *not* teach students how to use any reading strategies and only had discussions with students about their ideas to improve their WPM. Finally, in each lesson, students identified positive thoughts for use during ORF tasks. Prior to reading, students connected their positive thoughts to their selected strategy (e.g., "I know I can do this by phrase reading").

In the next step, each student reviewed their WPM line graph and identified their daily personal best goal. This goal was their highest WPM score plus one more word. Then, the tutor provided students with their assigned ORF passage with a pre-marked star intended to highlight where their goal is located. Before reading the ORF passage, the student was prompted to say a generated positive strategic statement three times (e.g., "I know I can do this by phrase reading"). After that, the tutor completed the progress monitoring procedures (i.e., read the standardized directions and the student then read the ORF passage for 1 min). Next, the WPM score was shared with the student, and they plotted their datum point on their paper line graph. After the datum was graphed, the tutor and students compared whether the new point increased, decreased, or remained stable compared to the previous session. Tutors used descriptive praise statements connecting success to strategy use and goal achievement (e.g., "you met your goal because you sounded out the words"). If a student did not achieve their goal, the student was reminded that personal best goals are not attained every day but that it is important to track performance data so that improvements can be observed over time. Each lesson ended with a reminder to use positive thoughts and strategies throughout the school day across difficult academic situations.

From Lessons 15–25, tutors reduced their prompts. They guided students through self-monitoring, goal setting, selecting a reading strategy, and evaluating progress. Students shared challenges they faced during personal reading tasks and discussed their thoughts. They brainstormed strategies for future use in similar situations. Students then "taught" the tutors to use Data Mountain. Tutors asked scripted questions such as "How do I set a daily goal?" "What do I do before I read?" or "How do I graph my new data point?" Students explained the steps, identified their goals, formulated positive statements about their strategies, graphed their new data points, and assessed their progress. During Lessons 25–30, students conducted sessions largely independently with minimal tutor support. Tutors provided visuals and prompts on an individualized basis. Some students completed Data Mountain steps without assistance, whereas others still required prompting support or reminders. Progress monitoring procedures remained tutor-led and unchanged.

Individualized and Small-Group Conditions

The procedures and script of Data Mountain were the same across all participants in the Data Mountain conditions (DM-I and DM-G). Because the DM-G groups included two to three grade-level peers, students were explicitly taught respectful conversation norms and given behavior expectations during the training lesson. During each session, students were monitored for progress individually while other students watched silently. Peers were instructed to identify the reading strategies the reader used. After all students in the group read their ORF passage, students simultaneously received their WPM score and worked alongside one another to place it on their individual graphs. Students were encouraged to compliment their peers on the strategies that they observed each other using.

Data Collection and Measures

ORF Progress Monitoring. The primary outcome of interest was change in ORF across 30 sessions. The Formative Assessment System for Teachers CBMReading (FastBridge Learning; Christ et al., 2014) provided the WPM score for each session. To help control for passage difficulty effects, the CBMReading ORF passages were randomized in an individual sequence for all participants across conditions. For second through fifth grade, interrater reliability estimates were greater than .97 and alternate form reliabilities ranged from .75 to .83 (Christ et al., 2014). See Table 1 for average score at baseline, session 1, and session 30 by cluster (year, tutor) and treatment group.

Pre-/Post-test Battery. For all students across conditions, these measures were administered by the tutor prior to the session 1 (baseline pre-test) and after session 30 (post-test). See Table 1 for average score on pre-test and post-test by cluster and treatment group. A mid-test battery was administered, and additional descriptive fluency (i.e., Test of Silent Reading Efficiency and Comprehension) and attribution data (i.e., Reading Attribution Scale) were collected at pre- and post-test. However, due to missing data in year 1, these data were not used in the analysis for the present study (see Table S1 in supplemental file).

Word Reading. The TOWRE-2 was administered as both a screener and a measure of word reading for pre- and post-test. It includes two subtests: Sight Word Efficiency assesses the number of real words in print that can be read accurately within 45 seconds, and Phonetic Decoding Efficiency measures the number of decodable nonsense words in print that can be accurately identified within 45 seconds. Developers reported overall average alternate-form reliabilities for subtests of .91 and .92, respectively (Torgesen et al., 2012).

Social Validity. A researcher-developed survey from the pilot studies was used to collect social validity data at post-test from all Data Mountain participants. This survey required students to rate how statements about the usefulness and likability of Data Mountain pertain to them using a 3-point picture scale: *not at all*, *sometimes*, and *always*.

Instructor Training

Tutor 1 was a female completing a master's degree in psychology; Tutor 2 was a male completing a master's degree in psychology; and Tutor 3 was the primary investigator (PI), a female who has a doctorate in special education. All three tutors implemented both intervention and comparison groups. Tutors 1 and 2 participated in a three-hour initial training session led by Tutor 3, which covered Data Mountain and progress monitoring procedures. The training included a review of the program's scope and sequence, demonstrations of various lesson types, and an overview of the procedural fidelity checklist. Tutors had opportunities to practice various parts of the lesson script with feedback. Tutors were instructed to study the script and practice delivering the lessons. Two weeks later, the first author conducted individual follow-up meetings to evaluate the tutors' fidelity using three randomly selected sessions, including progress monitoring. Tutors were required to achieve a procedural fidelity score of at least 90% before beginning the intervention with students. Feedback was provided as needed, and all instructors met the fidelity requirements before initiating any sessions.

Procedural Fidelity

Only 71% of students had consent to have their sessions recorded, which affected DM-G sessions if any student opted out of audio recording. The audio was uploaded to shared cloud storage daily and the PI reviewed the audio frequently to ensure that the procedures were implemented as intended. In total, 20% of the audio was reviewed. The fidelity checklist used in previous studies assessed behaviors related to specific conditions. Implementation fidelity was 100% across tutors. The checklist included a qualitative rating to score the tutors on their pacing, correction procedures, and promotion of positive behavior (3 = highly effective, 2 = somewhat effective, 1 = not effective). The average score was 3. See Figure S3 in supplemental file for a procedural fidelity checklist.

Research Design and Analytic Plan

The extent to which student growth in reading fluency (as measured by WPM) differed by intervention group was examined in a sample of 109 students measured up to 31 sessions (1 baseline session and 30 progress monitoring sessions). Residual maximum likelihood was employed within SAS PROC MIXED for estimating multilevel models in which sessions at level 1 were modeled as nested within students at level 2. The significance of fixed effects was assessed using Wald tests with Satterthwaite denominator degrees of freedom, and the significance of random effects was assessed using likelihood ratio tests (i.e., $-2\Delta LL$ with degrees of freedom corresponding to the number of new random effect variances and covariances). Linear combinations of model fixed effects were created and tested using the ESTIMATE command. Pseudo-R^2 effect sizes for the fixed slopes targeting each variance component were determined by calculating the proportion reduction in each model variance component relative to a baseline model without the fixed slopes of interest. In addition, specific effect sizes per slope in a partial-r metric were computed as $r = t$ / SQRT(t^2 + DDF), in which DDF is the denominator DF as given by the Satterthwaite method. Full results are given in the OSF supplemental materials (https://doi.org/10.17605/OSF.IO/TEXRY).

RESULTS

Preliminary Analyses

Initially, a null (i.e., empty) two-level model (i.e., with only a level-2 student random intercept and a level-1 occasion residual) was estimated to quantify the variability in WPM at each level. An intraclass correlation (ICC) indicated that 86.4% of the variance in WPM was attributable to student mean differences. Based on the pattern of model-estimated (i.e., saturated) means across sessions, fixed linear and quadratic effects of time were then included. Time was centered such that 0 indicated the end of the study, and thus all effects related to the intercept or linear time slope are conditional on the last session. The addition of these fixed linear and quadratic time slopes (creating a pattern of decelerating growth) accounted for 43.6% of the level-1 residual variance. Adding a random variance for the student linear time slope (and its covariance with the student random intercept) significantly improved model fit, $-2\Delta LL(2) = 413.39$, $p < .001$. However, the addition of a random variance for the student quadratic time slope (and its covariances with the student intercept and linear time slope) resulted in model nonconvergence, indicating that only a random intercept and random linear time slope were needed to describe individual student differences. Given the study length of up to 30 sessions, we also added an autoregressive residual correlation, which significantly improved model fit, $-2\Delta LL(1) = 7.89$, $p = .005$, with a correlation of $r = .06$ for the residuals of adjacent sessions.

We then examined differences in WPM trajectories across grades 2–5 as a control variable. There was a significant omnibus effect of grade on the intercept (i.e., predicted outcome at time 30), $F(3, 111) = 14.85$, $p < .001$, which explained 28.4% of the student random intercept variance. Except for grades 2–3 and 4–5, all pairwise comparisons indicated greater WPM in older grades. Although nonsignificant, the omnibus effect of grade on the linear time slope at time 30 explained 15.9% of the random linear time slope variance. Each grade had a significantly positive linear time slope (with no significant pairwise differences). Finally, there was a significant omnibus effect of grade on the quadratic time slope, $F(3, 970) = 2.75$, $p = .042$, which explained 0.2% of the level-1 residual variance (given the lack of a random quadratic time slope). The quadratic time slope was significantly negative in grades 2 and 4, nonsignificantly negative in grades 3 and 5, and significantly less negative in grade 3 than grade 2. Given these differences, we retained the effects of grade on the intercept, linear time slope, and quadratic time slope by using three contrasts to represent adjacent grade differences (with grade 2 as the reference for the fixed intercept, linear time slope, and quadratic time slope).

Treatment Effects

Initial exploration of the data suggested differences in treatment effects due to year/tutor cohort (model-based versions of which are shown in Figure 2). Therefore, treatment effects were allowed to interact with cohort. Specifically, trajectory differences by binary treatment condition (Data Mountain combined versus COM) and year/tutor cohort were examined, in which three contrasts distinguished: (a) Tutor 2 in year 1 (the reference) versus Tutor 2 in year 2, (b) Tutor 2 in year 1 versus Tutor 1 in year 1, and (c) Tutor 2 in year 2 versus Tutor 3 in year 2. We began by allowing all possible trajectory differences by binary treatment and cohort in combination. We then sequentially removed nonsignificant higher-order interactions, which included trajectory differences by treatment between the year-1 tutors. The final model parameters are given in Table 2, with findings shown in Figure 2. In terms of effect size, the effects of treatment condition and cohort explained an additional 6.8% of the student random intercept variance, 15.0% of the student random linear time slope variance, and 1.3% of the level-1 residual variance.

Here, we focus only on the new effects related to treatment and cohort (Table 2 rows 13–30, along with other model-implied fixed effects provided in the supplemental materials on OSF). First, with respect to RQ1, the binary treatment

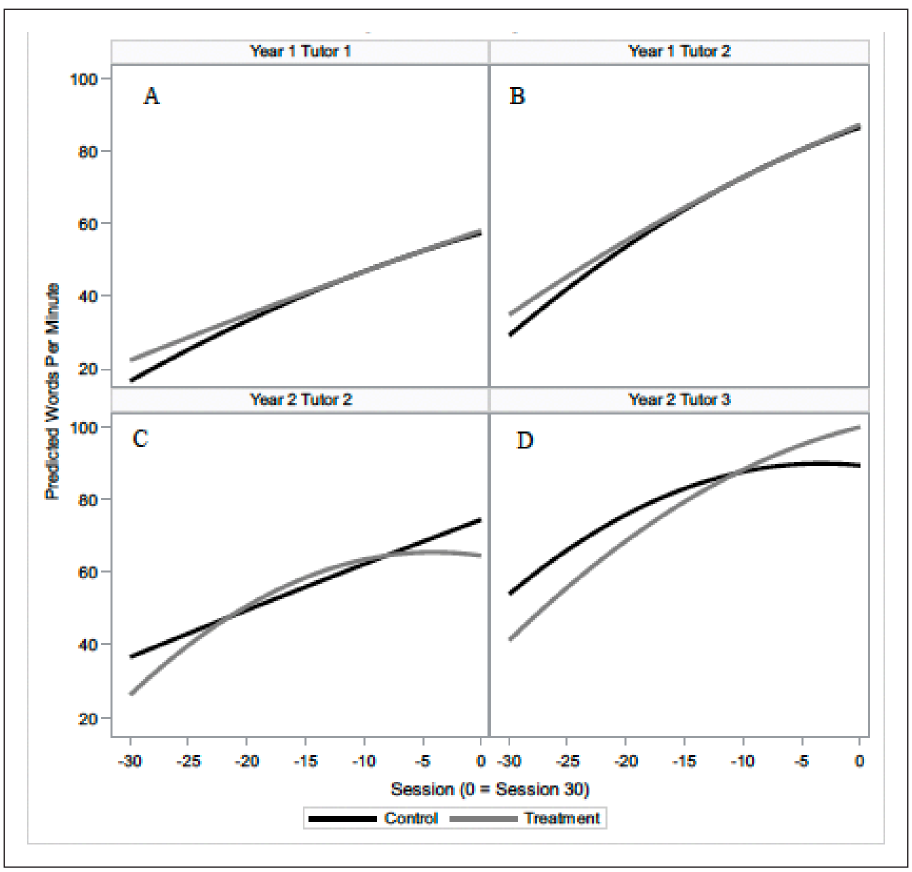

Figure 2:

Note. Growth trajectories by treatment group and cohort.

ROW	FIXED EFFECT	ESTIMATE	SE	p <	PARTIAL r
1	Intercept	86.305	12.848	.001	.556
2	Linear Time	1.104	0.384	.004	.111
3	Quadratic Time	–0.027	0.011	.012	–.082
4	Grade 3	32.199	10.294	.002	.292
5	Grade 4	32.608	10.235	.002	.301
6	Grade 5	–11.069	10.895	.312	–.099
7	Linear Time*Grade 3	0.235	0.345	.496	.025
8	Linear Time*Grade 4	–0.159	0.323	.623	–.018
9	Linear Time*Grade 5	–0.024	0.362	.947	–.003
10	Quadratic Time*Grade 3	0.019	0.009	.044	.064
11	Quadratic Time*Grade 4	–0.008	0.009	.384	–.028
12	Quadratic Time*Grade 5	0.006	0.010	.557	.019
13	Treatment	0.724	11.582	.950	.006
14	Linear Time*Treatment	0.195	0.337	.564	.023
15	Quadratic Time*Treatment	0.012	0.009	.201	.042
16	Year2	–12.013	17.052	.483	–.068
17	Linear Time*Year2	0.102	0.594	.863	.006
18	Quadratic Time*Year2	0.025	0.016	.120	.049
19	Treatment*Year2	–10.657	18.506	.566	–.056
20	Linear Time*Treatment*Year2	–1.879	0.661	.005	–.098
21	Quadratic Time*Treatment*Year2	–0.069	0.018	.001	–.121
22	Tutor 1	–29.068	10.830	.009	–.259
23	Linear Time*Tutor 1	–0.200	0.318	.531	–.025
24	Quadratic Time*Tutor 1	0.012	0.009	.184	.044
25	Tutor 3	14.905	17.903	.407	.080
26	Linear Time*Tutor 3	–1.536	0.652	.019	–.082
27	Quadratic Time*Tutor 3	–0.049	0.017	.006	–.087
28	Treatment*Tutor 3	20.470	21.579	.345	.090
29	Linear Time*Treatment*Tutor 3	2.797	0.798	.001	.120
30	Quadratic Time*Treatment*Tutor 3	0.068	0.021	.002	.100

Table 2: Results from Models Including Grade, Treatment, and Cohort.

group effects on growth varied by year/tutor cohort (Figure 2). In year 1, there was no effect of treatment on the intercept, linear time slope, or quadratic time slope (Table 2 rows 13–15), equivalently so across Tutors 1 and 2 (no interaction; Figure 2 panels A and B). Both DM and COM students in year 1 grew in WPM across the study, as reflected through a significantly positive linear time slope in each cohort, although their negative quadratic time slopes indicated that growth tapered off towards the study end (significantly so only for Tutor 2 comparison students; see supplemental materials on OSF). However, in year 2, treatment effects on growth over time were found differentially by tutor, as shown in Figure 2 panels C and D. For Tutor 2 in year 2, relative to COM students (Table 2 rows 16–18), DM students had a significantly less positive linear slope at time 30 paired with a significantly more negative quadratic time slope (Table 2 rows 20–21). In contrast, for Tutor 3 in year 2, DM students had a significantly more positive linear slope at time 30, but there was no treatment difference in the quadratic time slope (supplemental materials on OSF). Thus, DM students continued improving more under Tutor 3 than Tutor 2 (Table 2 rows 29–30).

With respect to RQ2, we consider our evidence of nonlinear change. Although the quadratic time slope was negative in every cohort of DM students, it was significant only in year 2 (equivalently so for both tutors in year 2). This result suggests diminishing returns of exposure to Data Mountain in later sessions.

Moderation by Oral Reading Fluency

With respect to RQ3, we examined how baseline ORF (TOWRE-2, centered at 75 near the sample mean) moderated growth trajectories by binary treatment group and year/tutor cohort. We began by allowing moderation of all possible trajectory differences by treatment and cohort in combination. We then sequentially removed nonsignificant higher-order interactions, which included all moderation differences between tutors. The final model parameters are given in Table 3 with findings shown in Figure 3. In terms of effect size, ORF moderation effects explained an additional 41.8% of the student random intercept variance, 2.0% of the student random linear time slope variance, and ~0% of the level-1 residual variance.

ROW	EFFECT	ESTIMATE	SE	*p*	PARTIAL *r*
1	Intercept	89.492	8.291	<.001	.747
2	Linear Time	0.782	0.406	.054	.076
3	Quadratic Time	−0.039	0.011	<.001	−.113
4	Grade 3	24.941	6.541	<.001	.353
5	Grade 4	32.412	6.694	<.001	.443
6	Grade 5	−8.503	6.976	.226	−.119
7	Linear Time*Grade 3	0.386	0.353	.274	.040
8	Linear Time*Grade 4	−0.342	0.341	.316	−.038
9	Linear Time*Grade 5	−0.059	0.374	.874	−.006
10	Quadratic Time*Grade 3	0.023	0.010	.015	.079
11	Quadratic Time*Grade 4	−0.011	0.009	.245	−.038
12	Quadratic Time*Grade 5	0.004	0.010	.677	.014
13	Treatment	−3.560	7.602	.641	−.049
14	Linear Time*Treatment	0.525	0.363	.149	.058
15	Quadratic Time*Treatment	0.024	0.010	.021	.077
16	Year2	−12.005	11.708	.308	−.099
17	Linear Time*Year2	0.371	0.658	.573	.020
18	Quadratic Time*Year2	0.035	0.018	.048	.063
19	Treatment*Year2	−1.613	12.502	.898	−.012
20	Linear Time*Treatment*Year2	−2.249	0.712	.002	−.113
21	Quadratic Time*Treatment*Year2	−0.083	0.019	<.001	−.137
22	Tutor 1	−14.397	6.948	.041	−.212

(Contd.)

ROW	EFFECT	ESTIMATE	SE	*p*	PARTIAL *r*
23	Linear Time*Tutor 1	–0.167	0.335	.619	–.020
24	Quadratic Time*Tutor 1	0.012	0.009	.192	.043
25	Tutor 3	9.181	12.189	.453	.072
26	Linear Time*Tutor 3	–1.117	0.694	.108	–.058
27	Quadratic Time*Tutor 3	–0.037	0.019	.047	–.064
28	Treatment*Tutor 3	14.710	14.283	.305	.097
29	Linear Time*Treatment*Tutor 3	2.350	0.820	.004	.102
30	Quadratic Time*Treatment*Tutor 3	0.058	0.022	.009	.084
31	ORF	1.241	0.708	.083	.173
32	Linear Time*ORF	–0.086	0.037	.022	–.086
33	Quadratic Time*ORF	–0.003	0.001	.003	–.097
34	Grade 3*ORF	0.252	0.623	.687	.040
35	Grade 4*ORF	–0.431	0.626	.493	–.071
36	Grade 5*ORF	1.957	0.622	.002	.301
37	Linear Time*Grade 3*ORF	0.033	0.033	.311	.038
38	Linear Time*Grade 4*ORF	–0.047	0.032	.138	–.057
39	Linear Time*Grade 5*ORF	0.015	0.033	.641	.017
40	Quadratic Time*Grade 3*ORF	0.002	0.001	.033	.069
41	Quadratic Time*Grade 4*ORF	–0.002	0.001	.048	–.065
42	Quadratic Time*Grade 5*ORF	0.000	0.001	.741	–.011
43	Treatment*ORF	1.565	0.582	.008	.262
44	Linear Time*Treatment*ORF	0.089	0.030	.003	.111
45	Quadratic Time*Treatment*ORF	0.002	0.001	.008	.087
46	Year2*ORF	–0.816	0.467	.084	–.169
47	Linear Time*Year2*ORF	0.045	0.026	.080	.063
48	Quadratic Time*Year2*ORF	0.002	0.001	.010	.083

Table 3: Results from Models Adding Baseline Oral Reading Fluency (ORF) as a Moderator.

Here we focus only on the new effects related to ORF moderation (Table 3 rows 31–48, along with other model-implied fixed effects provided in the supplemental materials). With respect to moderation of the intercept, greater baseline ORF was related to higher predicted WPM at time 30, significantly more so for DM students (Table 3 row 43), for whom intercept moderation was significant in both years, than for COM students, for whom intercept moderation was not significant in either year (supplemental materials). However, ORF moderation of growth varied additively by treatment group and year (Table 3 rows 44–45, 47–48). Analysis of simple effects (supplemental materials) indicated that significant moderation was found only in the year 1 COM students (Table 3 rows 32–33), in which greater baseline ORF was related to significantly less positive linear slopes and less negative quadratic time slopes. Corresponding interactions with treatment group (Table 3 rows 44–45) essentially reversed these moderation effects in DM students in both years. Likewise, corresponding interactions with year (Table 3 rows 47–48) weakened these moderation effects in year 2. Finally, independent of treatment and cohort, we found that ORF moderation differed by grade. The ORF benefit on

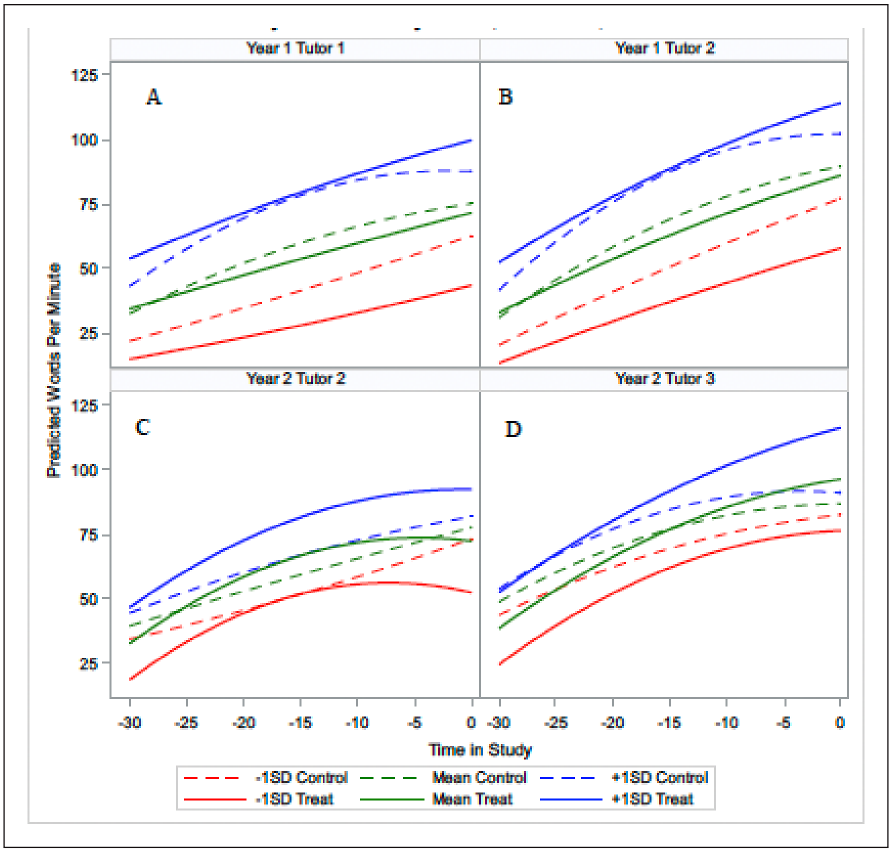

Figure 3:

Note. Growth trajectories by ORF performance moderated by performance on the TOWRE-2 by treatment group and cohort.

the intercept was stronger in fifth grade than in fourth grade (Table 3, row 36). The pattern of greater ORF relating to less positive linear time slopes was largely comparable across grades, whereas the pattern of greater ORF relative to more negative (more decelerating) quadratic time slopes was found in grades 2 and 4 (Table 3 rows 37–42).

Moderation by Treatment Delivery Format

With respect to RQ4, we also examined the extent of differences by delivery format (DM-I, DM-G). There were no significant differences between Data Mountain delivery formats (and no interactions with year/tutor cohort) on trajectories of growth in WPM. For these reasons, these results are not included (but are available upon request).

Social Validity

Finally, students rated eight items on a 3-point scale: not at all (1), sometimes (2), and always (3). DM students rated highest that "they read more words because they tried hard" (M = 2.75, SD = 0.71), "when they self-talk positively they could read more words" (M = 2.79, SD = 0.79), and that "they wanted to keep using their self-monitoring graph" (M = 2.74, SD = 0.84) The lowest-rated item was the "self-monitoring graph made them read faster" (M = 2.33, SD = 0.72).

DISCUSSION

Learning and self-determination skills are interrelated and should be taught simultaneously to maximize effectiveness. Schools often separate academic learning from the development of psychosocial skills (Darling-Hammond et al., 2020), such as self-determination. Integrating self-determination learning into existing school-wide systems (i.e., progress monitoring) is doable. Through the iterative process of progress monitoring, students can realize the direct impact of their actions on their performance, prompting them to make informed decisions based on this awareness. Engaging students in ongoing progress monitoring can enable them to understand the direct link between their actions and outcomes, empowering them to employ skills and strategies that foster success (Shogren et al., 2017). This allows the growth of students' agency and their ability to implement strategies that enhance achievement. Previous research on Data Mountain has illustrated this process (e.g., Didion et al., 2024; Didion & Toste, 2021; Didion et al., 2020). However, our current findings were nuanced and varied by tutor and year, further compounded by the challenges of returning to school post-COVID-19. We first discuss the primary effects of the extended version of Data Mountain on ORF growth. Then, we examine variations in growth patterns across different tutors and academic years. Additionally, we explore patterns in pretest ORF performance and the complexities introduced by COVID-19. Furthermore, we address the challenges posed by maternity leave in academia and its impact on the study. We conclude with a discussion of limitations, implications, and future directions for replication research.

Effects of Data Mountain

Although Data Mountain students significantly improved their WPM over 30 sessions, the significance of the difference in improvement in growth trajectories compared to students in the comparison condition depended on both the study year and tutor (Figure 2, Table 2). There were no differences between Data Mountain and comparison students in year 1 of the study. However, in year 2 of the study, Tutor 3's students in the Data Mountain intervention exhibited significantly more positive linear growth in WPM at the end of the study than comparison students. In contrast, Tutor 2's Data Mountain students showed significantly less linear growth in WPM over time and their progress tapered off more towards the end of year 2 compared to students in the comparison condition. These results indicate significant differences in the effect of both tutor and year on the efficacy of Data Mountain. Specifically, Data Mountain students continued improving more when under Tutor 3 than Tutor 2 in year 2 only.

Tutor 3 was the primary investigator (PI)—an experienced teacher who created the original program in her elementary classroom. Then, she integrated self-determination principles within Data Mountain, developed the current version, and tested it over the last 10 years. The PI is also well-versed in self-determination strategies and has published multiple papers on the topic (e.g., Didion et al., 2024; Didion & Toste, 2021; Didion et al., 2021; Didion et al., 2020). The Data Mountain script maintains a relatively open-ended approach when discussing whether a reading strategy supported students' growth. The PI facilitated high-quality discussions within the parameters of the procedural fidelity checklist. For example, she discussed with some students that achieving substantial growth makes it challenging to continue surpassing one's high score—likening it to climbing a mountain where progress becomes tougher near the summit due to fatigue. Moreover, in lessons 16–30 with reduced scaffolds, some students began discussing long-term goals and linking them to their short-term goals. The PI encouraged these conversations, fostering student initiative and promoting more self-directed goal setting rather than relying solely on teacher direction. Consequently, Tutor 3's students likely engaged more extensively in the self-determination processes compared to students taught by other tutors, which may have contributed to the sustained growth observed in students under the PI's guidance. However, the extent to which discussion centered on self-determination skills was not reflected in the procedural fidelity checklists (see Figure S3 in supplemental materials). Future Data Mountain research should include on-going training or coaching to enhance tutors' understanding of self-determination concepts to appropriately facilitate student conversations. Studies evaluating interventions that promote self-determination should adopt flexible procedural fidelity measures to accommodate in-depth discussions that enhance growth in self-determination (Shogren et al., 2020).

Across all studies of Data Mountain, a large increase from baseline to session 1 has been noted; on average Data Mountain students increase by 10 WPM while comparison students increase by 1 WPM (see Didion et al., 2024). In the current study, similar effects were found in both project years, such that from baseline to session 1, on average DM-I students increased by 8.96 WPM and DM-G students increased by 12.12 WPM, while COM students increased by 1.98 WPM. At the very least, this illustrates that simply showing students what their reading fluency goal is can greatly increase their rate of improvement. Teachers are encouraged to share progress monitoring data with students and to provide descriptive positive praise for any improvement (e.g., "you met your goal because you used your reading strategies").

Plateau at the Top of Data Mountain?

For Tutor 2, there was a significantly less positive slope in year 2 for the second half of the intervention, whereas the students under Tutor 3 (the PI) continued to grow. It was originally hypothesized that all students would continue to grow in sessions 16–30 similar to the first half of the intervention, but this was not observed. In fact, while it appears Tutor 2's groups had a higher growth rate at the beginning (see Figure 2), their growth flattened out and the benefits of exposure to Data Mountain decreased in later sessions. Students were still learning, but not as quickly as the beginning. While not within the scope of the current project, any instructor collecting progress monitoring data should intensify the intervention when they notice data leveling off on a student's graph (Fuchs et al., 2014). Multiple intervention adaptations may occur in an intervention's trajectory until students reach a predetermined goal (see Didion, 2024). Once student(s) show stagnant growth for 3–5 data points, the instructor could consider adjusting the intervention, such as continuing Data Mountain alongside an evidenced-based reading intervention.

Challenges with Return to Regular Schooling

The first year of the study marked the return to in-person schooling following the COVID-19 pandemic. These students had been in and out of lockdown and mandatory masking for the past year. COVID-19 had the largest impact on reading tests scores, particularly for students in grades 3–5 (Kuhfeld et al., 2022)—it disrupted foundational reading instruction. In the current study, students had to score below the 25th percentile to be included. We noticed that almost half (44%) of students performed above the 50th percentile at post-test in both Data Mountain and comparison conditions. In previous studies, students did not show this drastic growth from pre- to post-test on the TOWRE-2, with just under a fourth (23%) of students performing above the 50th percentile at post-test. This may suggest that although they met the inclusion criteria based on the TOWRE-2, students who qualified for the study may not have been students who we would have traditionally identified as at-risk for RD prior to the pandemic. Instead, their low scores could have been influenced by factors such as limited recent reading instruction and other pandemic-related challenges (e.g., anxiety, loss of loved ones). Their growth may have confounded the effects of the study.

Pretest ORF Scores Predicted WPM

We found that pre-test TOWRE-2 scores significantly moderated the effect of Data Mountain in year 1. Specifically, Data Mountain students who started with higher TOWRE-2 scores at the beginning of the study exhibited more growth and a higher predicted WPM at the end of the study (Table 3, Lines 43–45). This means that initial reading efficiency played a significant role in predicting later reading fluency for students using Data Mountain; students who start off with better reading skills benefitted more from Data Mountain over time, showing greater improvements in reading speed. Because this relationship was not observed in the comparison group, Data Mountain might be particularly effective for at-risk students who have a relatively higher level of reading proficiency at the start. This could imply that the intervention is less effective for students who start with lower reading skills and highlights the need for additional or different support for these students to improve their reading fluency (i.e., intensive intervention).

Delivery Format Illustrated No Differences

Possibly due to the many confounding factors already discussed, there were no differences in delivery format for any cohort. In the pilot study (Didion & Toste, 2021), we hypothesized that students in the small group would have an advantage due to peer models (Stevens et al., 2017). However, that pilot study revealed that students who received individual sessions had a higher growth rate than students in small group sessions, though differences were not significant. Therefore, it was hypothesized for the current study that the increased intensity of the 1:1 support would benefit these students over the small group sessions. This was not observed in our findings. As such, we continue to recommend instructors choose the delivery format that works best for their students' needs (i.e., intensive 1:1 support) or instructional time constraints (i.e., small group).

Study Limitations and Related Recommendations

We have already discussed the challenges posed by COVID-19 in this study. Other limitations include the procedural fidelity checklists' inability to capture differences in implementation levels between the PI and other tutors. Future procedural fidelity measures should encompass the complexity of developing self-determination and other variables that influence fidelity (i.e., adherence, participant responsiveness; Shogren et al., 2020). Check-in meetings with tutors or teachers should continually develop their understanding of self-determination so they can engage in meaningful conversations with students. Because the PI also served as a tutor and tutor effects were observed, the effects of interest needed to be stratified by year and tutor, thereby reducing overall statistical power. These results should be interpreted cautiously due to the study's limited sample size within each group, potentially impacting the reliability and generalizability of the findings.

It was not the original intent of the study for the PI to be a tutor. The original study was preregistered to be completed in two schools during one academic year. The PI was pregnant and had her baby prematurely during the fall of the first year. The small research team was unable to get the study off the ground in two schools in one year. During her six weeks of leave, the PI continued to complete fidelity checks of the audio, check in, provide feedback to tutors, and manage materials. These are common experiences for women during maternity leave (Byrne & Dillon, 1996; Maxwell et al., 2019). A second year was needed to have adequate power to examine effects and, given limited funding, the PI had to serve as a tutor. The pre-registration was adapted to reflect this need. Despite the challenges of maintaining high standards for research, mothers in academia also face challenges when returning to work after childbirth, including working in short segments, increased sense of urgency, lower expectations for self, inconsistent daycare (e.g., illness, school closures), and emotional effects (Craft & Maseberg-Tomlinson, 2015; Philipsen & Bostic, 2008), yet research metrics are not reflective of maternity leave (O'Brien & Hapgood, 2012; Maxwell et al., 2019). Institutes of higher education need to consider how to adequately support all faculty to conduct rigorous research, including parents.

Implications

Integrating Data Mountain alongside progress monitoring holds promise for enhancing formative assessment practices. Formative assessment can move beyond just data collection and serve as a supportive framework and guide for students in learning self-determination skills (Darling-Hammond et al., 2020). Research indicates that assessments emphasizing growth rather than fixed targets can bolster student motivation and engagement (Darling-Hammond & Cook-Harvey, 2018). Thus, embedding Data Mountain within progress monitoring not only facilitates ongoing assessment of student progress, but it may also foster a more motivating and supportive learning environment, ultimately improving student outcomes. Schools and districts should consider integrating self-determination skills within current system-level procedures. However, the findings suggest that Data Mountain may be most effective for students with higher initial reading proficiency, emphasizing the need for additional or more intensive support for struggling readers. The study also underscores the need for flexible procedural fidelity measures that capture the dynamic, individualized nature of self-determination-based interventions.

Future Directions

Replication research is crucial to validate and generalize the findings of this study on Data Mountain. Repeated studies can strengthen evidence of its effectiveness across diverse populations and settings and help refine implementation strategies to enhance student learning outcomes. This study's demographic sample from rural schools was predominantly White (71%), with smaller proportions of Black (6%) and Hispanic (18%) students, compared to the pilot, which was conducted in a large city and involved 5% White, 6% Black, and 83% Hispanic participants. Testing interventions across different demographics is vital to determine their efficacy under various conditions (Bryan et al., 2021).

Future studies should consider using teachers instead of tutors. When students' progress is monitored alongside Data Mountain by their reading teachers, they potentially could link the specific strategies that they are currently teaching to the strategy bank to enhance outcomes. Teachers would then notice stagnant data patterns and build intensive intervention within it. Next steps will enhance the training to include coaching teachers to use Data Mountain and develop teachers' sense of self-determination learning.

Future iterations of Data Mountain could incorporate language prompts that mirror the scaffolding provided by the PI. Students should be encouraged to persevere towards their goals, especially when facing challenges near the summit. Much like climbing a real mountain, when the trek is difficult, one requires essential supports like suitable gear, water, food, breathing techniques, and companions. In Data Mountain, students will need intensive intervention to reach their full potential. One enduring service we can offer students is teaching them how to self-navigate and remain determined through difficult tasks.

AUTHOR NOTE

This work was supported in part by grant #21–5530 from the Roy J. Carver Charitable Trust and by a grant funded by the Scanlon Center for School Mental Health. Nothing in the article necessarily reflects the positions or policies of the trust or the center, and no endorsement by it should be inferred.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTION

Lisa Didion: Conceptualization – Ideas; Funding Acquisition; Investigation; Methodology; Project Administration; Resources; Supervision; Writing – original draft; Writing – review & editing. **Charlotte Jeppsen**: Formal Analysis; Methodology; Visualization; Writing – original draft; Writing – review & editing. **Lesa Hoffman**: Formal Analysis; Methodology; Visualization; Writing – original draft; Writing – review & editing. **Benjamin P. Smith**: Investigation.

AUTHOR AFFILIATIONS

Lisa Didion
Department of Special Education, University of Kansas

Charlotte Jeppsen
Department of Psychological and Quantitative Foundations, University of Iowa

Lesa Hoffman
Department of Psychological and Quantitative Foundations, University of Iowa

Benjamin P. Smith
Department of Psychological and Quantitative Foundations, University of Iowa

REFERENCES

Betthäuser, B., Bach-Mortensen, A., & Engzell, P. (2022). A systematic review and meta-analysis of the impact of the COVID-19 pandemic on learning. *LIEPP Working Paper*, (134). https://doi.org/10.2139/ssrn.3878217

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

Bukowski, W. M., Motzoi, C., & Meyer, F. (2009). Friendship as process, function, and outcome. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 217–231). Guilford Press.

Byrne, A. P., & Dillon, N. K. (1996). *Academics don't have babies!: Maternity leave amongst female academics: IFUT survey report*. Irish Federation of University Teachers.

Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*(5), 386–406. https://doi.org/10.1177/00222194020350050201

Chatzoglou, P., Chatzopoulos, C., & Tsolou, O. (2023). The footprint of the COVID-19 pandemic in reading performance of students in the US with and without disabilities. *Research in Developmental Disabilities, 140*, 104585. https://doi.org/10.1016/j.ridd.2023.104585

Christ, T. J., Arañas, Y. A., Kember, J. M., Kiss, A. J., McCarthy-Trentman, A., Monaghen, B. D., & White, M. J. (2014). *Formative Assessment System for Teachers technical manual: CBMReading*. Formative Assessment System for Teachers.

Colvin, M. K., Reesman, J., & Glen, T. (2022). The impact of COVID-19 related educational disruption on children and adolescents: An interim data summary and commentary on ten considerations for neuropsychological practice. *The Clinical Neuropsychologist, 36*(1), 45–71. https://doi.org/10.1080/13854046.2021.1970230

Craft, C. M., & Maseberg-Tomlinson, J. (2015). Challenges experienced by one academic mother transitioning from maternity leave back to academia. *NASPA Journal About Women in Higher Education, 8*(1), 66–81. https://doi.org/10.1080/19407882.2015.1057166

Darling-Hammond, L., & Cook-Harvey, C. M. (2018). *Educating the whole child: Improving school climate to support student success*. Learning Policy Institute.

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97–140. https://doi.org/10.1080/10888691.2018.1537791

Darling-Hammond, L., & Friedlaender, D. (2008). Creating excellent and equitable schools. *Educational Leadership, 65*(8), 14.

De Charms, R. (2013). *Personal causation: The internal affective determinants of behavior*. Routledge. https://doi.org/10.4324/9781315008337

Didion, L. (2024). You did that!: Let data illustrate your effectiveness. *Teaching Exceptional Children.* https://doi.org/10.1177/00400599241260493

Didion, L., Bruno, L., Marshall, G., Immerfall, J. Kunkel, A., & McGinn (2024). Reaching the top of Data Mountain: Post-secondary students with disabilities use data to improve reading performance. *Career Development and Transition for Exceptional Individuals*. https://doi.org/10.1177/21651434241250326

Didion, L., & Toste, J. R. (2021). Data Mountain: Self-monitoring, goal setting, and motivation training to support the oral reading fluency of struggling readers in the elementary grades. *Journal of Learning Disabilities, 55*(5), 375–392. https://doi.org/10.1177/0022194211043482

Didion, L., Toste, J. R., & Benz, S. A. (2020). Self-determination to increase oral reading fluency for third grade students with and at risk for reading disabilities: Pilot and replication single-case designs. *Learning Disabilities Research and Practice, 35*(4), 218–231. https://doi.org/10.1111/ldrp.12234

Didion, L., Toste, J. R., Benz, S. A, & Shogren, K. A. (2021). Components of self-determination integrated within reading interventions for elementary students with learning disabilities: A research synthesis. *Learning Disabilities Quarterly, 44*(4), 288–303. https://doi.org/10.1177/0731948721989328

Domingue, B. W., Dell, M., Lang, D., Silverman, R., Yeatman, J., & Hough, H. (2022). The effect of COVID on oral reading fluency during the 2020–2021 academic year. *AERA Open, 8*, 23328584221120254. https://doi.org/10.1177/23328584221120254

Fuchs, D., Fuchs, L. S., & Vaughn, S. (2014). What is intensive instruction and why is it important? *Teaching Exceptional Children, 46*(4), 13–18. https://doi.org/10.1177/0040059914522966

Gilmour, A. F., Fuchs, D., & Wehby, J. H. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children, 85*(3), 329–346. https://doi.org/10.1177/0014402918817863

Hamm, J. V., & Zhang, L. (2010). School contexts and the development of adolescents' peer relations. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling and human development* (pp. 128–145). Routledge. https://doi.org/10.4324/9780203874844

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., 7 Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*(2), 509–527. https://doi.org/10.1111/1467-8624.00421

Konrad, M, Fowler, C. G., Walker, A. R., Test, D. W., & Wood, W. M. (2007). Effects of self-determination interventions on the academic skills of students with learning disabilities. *Learning Disability Quarterly, 30*(2), 89–113. https://doi.org/10.2307/30035545

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kuhfeld, M., Soland, J., Lewis, K., Ruzek, E., & Johnson, A. (2022). The COVID-19 school year: Learning and recovery across 2020–2021. *AERA Open, 8*, 23328584221099306. https://doi.org/10.1177/23328584221099306

Ladd, G. W., & Herald, S. L. (2009). Peers and motivation. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 337–362). Routledge. https://doi.org/10.4324/9780203879498

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research, 86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Logan, S., Medford, E., & Hughes, N. (2011). The importance of intrinsic motivation for high and low ability readers' reading comprehension performance. *Learning and Individual Differences, 21*(1), 124–128. https://doi.org/10.1016/j.lindif.2010.09.011

Maehr, M. L., & Meyer, H. A. (1997). Understanding motivation and schooling: Where we've been, where we are, and where we need to go. *Educational Psychology Review, 9*, 371–409. https://doi.org/10.1023/A:1024750807365

Maxwell, N., Connolly, L., & Ní Laoire, C. (2019). Informality, emotion and gendered career paths: The hidden toll of maternity leave on female academics and researchers. *Gender, Work & Organization, 26*(2), 140–157. https://doi.org/10.1111/gwao.12306

Midgley, C., & Edelin, K. C. (1998). Middle school reform and early adolescent well-being: The good news and the bad. *Educational Psychologist, 33*(4), 195–206. https://doi.org/10.1207/s15326985ep3304_2

Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children, 73*(2), 165–183. https://doi.org/10.1177/001440290707300204

National Assessment of Educational Progress (NAEP). (2022). *The Nation's Report Card: 2019 mathematics and reading assessments*. Retrieved from http://www.nationsreportcard.gov

O'Brien, K. R., & Hapgood, K. P. (2012). The academic jungle: Ecosystem modelling reveals why women are driven out of research. *Oikos, 121*(7), 999–1004. https://doi.org/10.1111/j.1600-0706.2012.20601.x

Patrinos, H. A., Vegas, E., & Carter-Rau, R. (2022). Learning loss during Covid-19: An early systematic review. *Prospects, 51*(4), 601–609. https://doi.org/10.1007/s11125-021-09517-9

Petscher, Y. (2010). A meta-analysis of the relationship between student attitudes towards reading and achievement in reading. *Journal of Research in Reading, 33*(4), 335–355. https://doi.org/10.1111/j.1467-9817.2009.01418.x

Philipsen, M. I., & Bostic, T. (2008). *Challenges of the faculty career for women: Success and sacrifice*. John Wiley & Sons.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667–686. https://doi.org/10.1037/0022-0663.95.4.667

Ponce, A. (2021). *Educational recovery now: LA's children and schools need a comprehensive plan*. https://greatpublicschoolsnow.org/wp-content/uploads/2021/03/EdRecoveryNow_Final_3-29-21.pdf

Powell, S. R., Bouck, E. C., Sutherland, M., Clarke, B., Arsenault, T. L., & Freeman-Green, S. (2023). Essential components of math instruction. *TEACHING Exceptional Children*, *56*(1), 14–24. https://doi.org/10.1177/00400599221125892

Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction, 21*(4), 550–559. https://doi.org/10.1016/j.learninstruc.2010.11.001

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications. https://doi.org/10.7208/chicago/9780226516829.001.0001

Shogren, K. A., Burke, K. M., Anderson, M. H., Antosh, A., LaPlante, T., & Hicks, T. (2020). Examining the relationship between teacher perceptions of implementation of the SDLMI and student self-determination outcomes. *Career Development and Transition for Exceptional Individuals, 43*(1), 53–63. https://doi.org/10.1177/2165143418809264

Shogren, K. A., & Raley, S. K. (2022). *Self-determination and causal agency theory: Integrating research into practice*. Springer Nature. https://doi.org/10.1007/978-3-030-74471-7

Shogren, K. A., Wehmeyer, M. L., & Palmer, S. B. (2017). Causal agency theory. In M. L. Wehmeyer, K. A. Shogren, T. D. Little, & S. J. Lopez (Eds.), *Development of self-determination through the life-course* (pp. 71–88). Springer. https://doi.org/10.1007/978-94-024-1042-6

Shogren, K. A., Wehmeyer, M. L., Palmer, S. B., Rifenbark, G. G., & Little, T. D. (2015). Relationships between self-determination and postschool outcomes for youth with disabilities. *The Journal of Special Education, 48*(4), 256–267. https://doi.org/10.1177/0022466913489733

Shogren, K. A., Wehmeyer, M. L., Shaw, L. A., Grigal, M., Hart, D., Smith, F. A., & Khamsi, S. (2018). Predictors of self-determination in postsecondary education for students with intellectual and developmental disabilities. *Education and Training in Autism and Developmental Disabilities*, *53,* 146–159.

Stanovich, K. E. (2009). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Journal of Education, 189*(1–2), 23–55. https://doi.org/10.1177/0022057409189001-204

Spinath, B., & Spinath, F. M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction, 15*(2), 87–102. https://doi.org/10.1016/j.learninstruc.2005.04.008

Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities, 50*(5), 576–590. https://doi.org/10.1177/0022219416638028

Stoel, R. D., Peetsma, T. T. D., & Roeleveld, J. (2001). Relations between the development of school investment, self-confidence, and language achievement in elementary education: A multivariate latent growth curve approach. *Learning and Individual Differences, 13*(4), 313–333. https://doi.org/10.1016/S1041-6080(03)00017-7

Taboada, A., Tonks, S. M., Wigfield, A., & Guthrie, J. T. (2009). Effects of motivational and cognitive variables on reading comprehension. *Reading and Writing, 22*, 85–106. https://doi.org/10.1007/s11145-008-9133-y

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *TOWRE-2 examiner's manual*. Pro-Ed.

Toste, J. R., Didion, L. A., Peng, P., Filderman, M. J., & McClelland, A. (2020). A meta-analytic review of the relations between motivation and reading achievement for K-12 students. *Review of Educational Research, 90*(3), 420–456. https://doi.org/10.3102/0034654320919352

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2021). *Global monitoring of school closures caused by COVID-19*. https://en.unesco.org/covid19/educationresponse#schoolclosures

Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with reading disabilities: Meaningful impacts. *Learning Disabilities Research & Practice, 29*(2), 46–53. https://doi.org/10.1111/ldrp.12031

What Works Clearinghouse (WWC). (2020). *WWC procedures and standards handbooks*. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-HandbookVer5.0AppIES-508.pdf

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research, 81*(2), 267–301. https://doi.org/10.3102/0034654311405999