

## Student Behavior Ratings and Response to Tier 1 Reading Intervention: Which Students Do Not Benefit?

Wilhelmina van Dijk, Christopher Schatschneider, Stephanie Al Otaiba & Sara A. Hart

**To cite this article:** Wilhelmina van Dijk, Christopher Schatschneider, Stephanie Al Otaiba & Sara A. Hart (2023): Student Behavior Ratings and Response to Tier 1 Reading Intervention: Which Students Do Not Benefit?, Journal of Research on Educational Effectiveness, DOI: [10.1080/19345747.2023.2194894](https://doi.org/10.1080/19345747.2023.2194894)

**To link to this article:** <https://doi.org/10.1080/19345747.2023.2194894>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 10 Apr 2023.



[Submit your article to this journal](#)



Article views: 404



[View related articles](#)



[View Crossmark data](#)



This article has been awarded the Centre for Open Science 'Open Data' badge.

## Student Behavior Ratings and Response to Tier 1 Reading Intervention: Which Students Do Not Benefit?

Wilhelmina van Dijk<sup>a</sup> , Christopher Schatschneider<sup>b</sup> , Stephanie Al Otaiba<sup>c</sup> ,  
and Sara A. Hart<sup>b</sup> 

<sup>a</sup>Department of Special Education and Rehabilitation Counseling, Utah State University, Logan, Utah, USA; <sup>b</sup>Florida Center for Reading Research and Department of Psychology, Florida State University, Tallahassee, Florida, USA; <sup>c</sup>Department of Teaching and Learning, Southern Methodist University, Dallas, Texas, USA

### ABSTRACT

Core reading instruction and interventions have differential effects based on student characteristics such as cognitive ability and pre-intervention skill level. Evidence for differential effect based on affective characteristics is scant and ambiguous; however, students with problem behavior are more often non-responsive to core reading instruction and intensive reading interventions. In this study, we estimated the range of students' behavior ratings in which a core reading instruction intervention was effective using a data set including 3,024 students in K-3. Data came from seven independent studies evaluating the Individualized Student Instruction (ISI) Tier 1 reading intervention and were pooled using integrative data analysis. We estimated Johnson–Neyman intervals of student behavior ratings that showed a treatment effect both at the within and between classroom level. ISI was effective in improving reading scores ( $b = 0.51$ ,  $p = .020$ ,  $d = 0.08$ ). However, students with very low or very high behavior ratings did not benefit from the approaches (range of behavior rating factor scores:  $-0.95$ – $2.87$ ). At the classroom level, students in classrooms with a higher average of problem behaviors did not benefit from ISI (average classroom behavior rating factor score:  $0.05$ – $4.25$ ). Results suggest differentiating instruction alone is not enough for students with behavior problems to grow in reading ability.

### ARTICLE HISTORY


Received 13 May 2022  
Revised 26 January 2023  
Accepted 4 March 2023

### KEYWORDS

Reading intervention;  
problem behavior; response  
to intervention

Becoming literate is an essential skill in today's information society. The basis for these skills is established in the early elementary grades. Reading is not a natural process and needs to be taught through explicit and systematic instruction (Foorman & Torgesen, 2001). Unfortunately, many students struggle learning to read. This is exemplified by

**CONTACT** Wilhelmina van Dijk  [willa.vandijk@usu.edu](mailto:willa.vandijk@usu.edu)  Department of Special Education and Rehabilitation Counseling, Utah State University, 2685 Old Main Hill, Logan, UT, USA.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19345747.2023.2194894>.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

the fact that about 39% of US fourth graders are reading below a basic level (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2022). Low reading proficiency places students at risk for several adverse life outcomes, such as not completing high-school (Hernandez, 2011). Since young students with higher reading skills show steeper growth across the early elementary years (Tang & Dai, 2021), and students that struggle early do not easily catch up with peers (Biancarosa & Snow, 2004), it is especially important to provide high-quality classroom instruction and targeted interventions to prevent and remediate difficulties in foundational reading skills such as phonological and phonemic awareness, decoding, and vocabulary.

Fortunately, research consistently shows high quality Tier 1 instruction and Tier 2 and 3 interventions geared toward struggling readers and readers with reading disabilities are effective in remediating deficits (e.g., Swanson et al., 2017; Wanzek et al., 2016, 2018) and that improvements continue over time (Suggate, 2016). However, while high quality Tier 1 instruction and early reading interventions haven proven effective for struggling readers, not all students respond to instruction and intervention to the same degree. For most instruction and intervention, there are individual differences in responsiveness, resulting in some children benefitting less than others (Al Otaiba & Fuchs, 2002; Fuchs & Fuchs, 2019). Unfortunately, there is little evidence allowing researchers and teachers to pinpoint for which children an instruction or intervention might be ineffective before starting. Students who do not respond well to instruction or intervention will fall even more behind and, because they have wasted precious time, need to work even harder to catch up. In other words, when instruction and interventions are not specific to students' needs, they are impeding reading progress instead of nurturing growth. In order to provide individualized instruction by choosing, implementing, and adjusting instruction or intervention to suit a specific student's needs (Hagan-Burke et al., 2013), it is important to understand the robustness of instruction and interventions by detailing the specifics of child characteristics by instruction (CxI) interactions (Connor et al., 2009; Fuchs & Fuchs, 2019). Knowing for which students a specific instructional approach or intervention works or does not work, allows for better individualized instruction leading to the greatest and fastest gain for all students.

This line of thought is a product of the theory on aptitude by treatment (ATI), originating in the late 1960s and further developed by Cronbach and Snow. ATI states that any aptitude, whether cognitive or affective, can moderate the effectiveness of instruction and intervention and that optimal learning occurs when instruction is adapted to the aptitudes of a learner (Cronbach & Snow, 1977). More broadly speaking, any child characteristic may affect instruction, whether these are skills or behaviors. One way to uncover these CxI interactions is through identifying preintervention characteristics that differentiate students who improved after an intervention from those that did not (Preacher & Sterba, 2019). Many researchers interested in CxI interactions have focused on cognitive factors such as preintervention characteristics. Phonological awareness skills, rapid naming speed, knowledge of the alphabetic principle, memory capacity (Al Otaiba & Fuchs, 2002; Hart et al., 2016), and pre-intervention skill level (e.g., Coyne et al., 2018) are all factors influencing response to intervention. Collectively, this research has shown that instruction

and intervention is less effective, or not effective at all, for students with low cognitive abilities.

Besides cognitive skills, researchers have also looked at behavioral factors as potential CxI interactions. Students with behavior problems, whether these are externalizing, internalizing, or hyperactive/inattentive, are more likely to have reading difficulties (DuPaul et al., 2016). Conversely, poor readers are more likely to also have behavior problems (Lin et al., 2013; Morgan et al., 2008). For students with existing behavior problems, learning to read may prove difficult for a variety of reasons, including the need to be focused, attentive to details, and on task. Students with hyperactivity/inattentive behaviors are likely to have a harder time with attention and focus. Students with externalizing behavior problems may lose instructional time by acting out as a means to escape instruction. During their acting out and the subsequent aftermath, they do not have to engage in the task and can avoid possible feelings of failure and embarrassment (Lin et al., 2013). Students with internalizing problems may experience loss of concentration due to negative thoughts (Grills-Taquechel et al., 2012). The potential of interference of problem behaviors on intervention effectiveness is evident. Profile analyses of non-responders to several interventions have demonstrated that students with problem behaviors are less likely to benefit from universal and intensive reading interventions than students without problem behaviors (Al Otaiba & Fuchs, 2006).

In addition to identifying which pre-intervention characteristics separated students that benefited from instruction or intervention from those who did not benefit, researchers can also examine CxI interactions using moderation analysis. To our knowledge, only four studies examined this interaction between problem behavior and intervention effectiveness specifically (i.e., Hagan-Burke et al., 2011; Hurry et al., 2018; Rabiner & Malone, 2004; Roberts et al., 2019), and the results are varied.

Hagan-Burke et al. (2011) examined if the influence of problem behavior on reading outcomes is moderated by the type of intervention provided. Kindergarten students at risk for reading difficulties received either a school designated intervention or a code-based early reading intervention. The results showed a relation between internalizing problems and reading outcomes was present, but this relation was not moderated by explicit and systematic code-based instruction. However, there was slight variation in kindergarten students' alphabet knowledge and decoding skills depending on their level of externalizing problems and hyperactivity. Hurry et al. (2018) examined if problem behaviors (i.e., emotional, conduct, hyperactivity, and a composite) moderated the effects of two interventions relative to a control group for a sample of first grade students at risk for reading difficulties. They found no statistically significant interactions for any of the reading measures, with the exception of hyperactivity which had a negative influence on the effect of phonological training on reading ability. However, the authors of this study did not account for classroom level clustering effects in their analyses, limiting the interpretability and validity of these findings.

Rabiner and Malone (2004) investigated the moderating relation of inattention on the effect of tutoring on word reading skills for first graders with high externalizing problems who were at risk for reading difficulties. They found a significant interaction between inattention and treatment; however, when accounting for other problem behavior (i.e., externalizing, internalizing, and hyperactivity) the moderation was not

significant. Similar to the Hurry et al. (2018) study, this study did not account for student clustering in the analysis, and it is unclear if interventionists were experienced enough to deal with the inattention and externalizing behaviors of these students. Finally, Roberts and colleagues (2019) examined the moderating relation of problem behavior on the effects of a multi-component reading intervention on struggling 4th and 5th grade readers. They found the interaction between the intervention and problem behavior subscale of Social Skills Improvement System–Rating Scale was statistically significant for the Gates MacGinitie Reading Test but not for the Woodcock Johnson Passage Comprehension subtest. When looking within the subscales of the Social Skills Improvement System–Rating Scale, only the externalizing subscale statistically significantly moderated both reading comprehension outcomes. Neither internalizing nor hyperactivity showed a statistically significant interaction. In sum, the outcomes from these studies did not show a clear pattern for any of the subtypes of problem behavior (i.e., externalizing, internalizing, or hyperactivity/inattention) or for any of the reading outcomes assessed. It is plausible this inconclusive evidence is a result of low power to detect significant interactions due to the relatively small sample sizes used in these studies ( $n$  ranging from 108 to 581). Additionally, all studies were limited to struggling readers which may have impacted the relation through range restriction.

## **The Current Study**

While the relation between reading difficulties and problem behavior has been solidly established, it is still unclear how problem behavior, especially as perceived by the classroom teacher, is related to intervention and instructional effectiveness. Likely, this is a result of the small number of studies explicitly probing this interaction effect and the low sample sizes and restricted ranges of reading abilities or problem behavior in these studies. In the current study, we aim to broaden the evidence by examining how perceived problem behavior moderates the relation between the effect of one core classroom reading instruction intervention, Individualized Student Instruction (Connor et al., 2007), on student reading outcomes in the early grades.

### ***Individualized Student Instruction***

Connor and colleagues established in earlier studies (Connor, Morrison, & Katch, 2004; Connor, Morrison, & Petrella, 2004) that students' reading skills at the beginning of an academic year influenced their growth in word reading, vocabulary and reading comprehension; however, the growth also depended on the amount of time teachers spent on code-focused or meaning-focused instruction. As a result of these exploratory, observational studies, Connor and colleagues conceptualized Individualized Student Instruction (ISI) as a Tier 1 intervention. ISI aimed to help early elementary teachers increase differentiation in their core reading instruction by adapting the amount of time they provided a type of instruction (code- or meaning-focused) in specific instructional groupings (e.g., whole class, small group, or individual instruction) based on each student's needs. By attending to CxI interactions and adapting instruction to student characteristics, reading growth should be optimized.

ISI has been described in detail previously (see, e.g., Al Otaiba et al., 2014; Al Otaiba et al., 2011; Connor et al., 2007; Connor, Morrison, Fishman, et al., 2011; Connor, Morrison, Schatschneider, et al., 2011) and we will provide a brief description here. Three main features distinguish the ISI classroom intervention: (a) a software program that supports data-based individualization by providing recommended amounts of code-focused and meaning-focused reading instruction for each student at various points in the school year; (b) extensive professional development in the use of the software program and in adapting instruction (e.g., type, grouping, intensity) to meet students' needs; and (c) coaching for literacy instruction in the classroom through bi-weekly classroom-based observations and support as well as monthly meetings as communities of practice (Al Otaiba et al., 2011; Connor et al., 2013). Classroom instruction under ISI supported teachers in providing their students with the appropriate amount of code- and meaning-focused instruction in either teacher-directed small group settings or independent student centers. Activities and instruction followed core reading curricula that were adapted to meet the needs of the students, and were supplemented with other sources, such as activities from the Florida Center for Reading Research.

Connor and colleagues conducted seven RCTs to establish the effectiveness of ISI. In these RCTs, classrooms were randomly assigned to a condition and the intervention was compared to business as usual (BAU) conditions, or alternative interventions (see Table 1 for specifics of each RCT). In the control conditions, teachers were still expected to differentiate instruction based on students' needs, as part of general classroom instruction. Results across the RCTs suggest that teachers trained and coached in ISI increased their use of differentiating instruction and that their students' reading skills increased compared to those in control conditions (Al Otaiba et al., 2014, 2016; Al Otaiba et al., 2011; Connor et al., 2007, 2013; Connor, Morrison, Fishman, et al., 2011; Connor, Morrison, Schatschneider, et al., 2011). In two of the studies, analyses suggested that students' growth increased as teachers provided code- and meaning-focused instruction close to the amounts recommended by the software (Connor et al., 2007, 2009), suggesting that attention to CxI interactions can increase growth in reading skills. Additionally, recent research on two of the ISI RCTs suggests teachers attended to more than just the ISI software when providing instructional grouping: ratings of problem behavior and perceived academic competence were also of influence. Teachers spent more time in small group instruction with students rated as higher in problem behavior, if they perceived these students to be more competent academically compared to students with higher problem behaviors but lower perceived academic competence (Toste et al., in press). Taken together, the results from previous studies suggest attending to

**Table 1.** ISI study characteristics.

Project	N (treatment-control)	Control group	Grade	Year
Project 1	362-279	BAU	K	2007-2008
Project 2	261-253	BAU	K	2008-2009
Project 5	410-394	BAU	1	2005-2006
Project 6	245-150	BAU	1	2007-2008
Project 7	196-284	Vocabulary	2	2007-2008
Project 8	176-183	Vocabulary	3	2008-2009
Project 9	279-232	Math	1	2008-2009

Note. BAU: Business as usual instruction.

students' cognitive skills increases reading growth, but perceived problem behavior might influence teachers' decisions on how to provide instruction.

### **Study Objectives**

The ISI intervention studies provide a unique opportunity to examine the moderation relation of perceived problem behavior on intervention effectiveness. Data from the seven RCTs are publicly available through Project KIDS (Hart et al., 2021a, 2021b) and we applied Integrative Data Analysis (IDA) to combine data from these seven intervention studies into one dataset. With this large and varied sample all receiving the same intervention, we examined if the effectiveness of Individualized Student Instruction is different dependent on teacher ratings of their students' behavior by examining the regions of significance of this interaction. Specifically, we aimed to answer the following research questions:

1. What is the overall effect of ISI across different intervention studies?
2. Are there ranges of behavior ratings for which the ISI classroom intervention is not effective?

### **Method**

To examine the differential effect of behavior ratings on the effectiveness of ISI, a comprehensive Tier 1 approach to early reading instruction, we used publicly available data from Project KIDS (Hart et al., 2021a, 2021b). Project KIDS (Daucourt et al., 2018; van Dijk, Norris, et al., 2022) combined item level achievement and behavioral data from eight independent randomized control trials evaluating the effect of ISI conducted in schools in a southeastern state in grades K-3 between 2005 and 2013 (Al Otaiba et al., 2011, 2014, 2016; Connor et al., 2007, 2013; Connor, Morrison, Fishman, et al., 2011; Connor, Morrison, Schatschneider, et al., 2011), with the intention of analyzing the data in novel ways. van Dijk, Norris, et al. (2022) provide extensive details about the design and procedures of the original studies, the procedures of Project KIDS' data collection, and each of the datasets.

For the current study, we used two IDA techniques to pool scores from the individual RCTs into one estimate. Generally, IDA can be defined as "the analysis of a single data set that consists of two or more separate samples that have been pooled into one" (Curran & Hussong, 2009, p. 83). While to date not often used in educational science (e.g., Jansen et al., 2020), IDA has been used in health research, epidemiology, and developmental psychology (e.g., Daucourt et al., 2018; Hornburg et al., 2017; Leijten et al., 2018). However, IDA is an ideal methodology for pooling educational intervention studies together because it capitalizes on between-study variability, for example variability that arises from differences in sampling techniques, the timeframes in which studies were conducted, overall study design, and measurement (Curran & Hussong, 2009). Additional advantages of IDA are (1) the increased statistical power associated with the larger sample, (2) the greater heterogeneity of the sample increasing generalizability to the population, (3) the higher occurrence of low base rate behaviors that allows for

subgroup analysis in the pooled sample, and (4) the stronger psychometric properties of measurement of a construct through pooling items (Curran & Hussong, 2009).

**Participants**

The complete Project KIDS data set includes 4,036 individual students. We used the complete data set to generate scaled scores and impute missing data, but the moderation analysis included only those students with teacher behavior ratings from seven projects. Students from a study in which two approaches to response to intervention were compared (Al Otaiba et al., 2014) were excluded from the analysis since both treatment and control group received the ISI intervention. The sample for the moderation analysis consisted of 2,683 students nested in 257 teachers in 29 schools and represents a diverse student body (49.8% female, 4.1% Latine, 0.2% Native American, 2.5% Asian, 44.4%, Black, 0.4% Pacific Islander, 45.1% White, and 3.4% multiracial). Table 2 provides specifics on participant characteristics. All original studies received IRB approval for their randomized control trials, and Project KIDS also had IRB approval to access and combine the original data samples and conduct further investigations.

**Measures**

**Problem Behavior**

Classroom teachers filled out the Social Skills Rating Scales (SSRS; Gresham & Elliott, 1990) during the Winter semester. The SSRS asks teachers to rate students’ social skills, academic competence, and problem behaviors. In this study, we used items of the

**Table 2.** Select participant characteristics.

Variable	Current sample				Intervention group				Control group			
	N	%	M	SD	N	%	M	SD	N	%	M	SD
Total	2,683				1,472	54.8			1,210	45.2		
Gender												
Male	1,346	50.2			763	51.8			583	48.2		
Female	1,337	49.8			709	48.2			627	51.8		
Ethnicity												
Latino/a	110	4.1			62	5.1			48	4.2		
Not Latino/a	2,557	95.3			1,397	94.9			1,159	95.8		
Race												
Native American/Alaskan Native	6	0.2			3	0.2			3	0.2		
Asian	66	2.5			35	2.3			31	2.6		
Black	1,190	44.9			682	46.3			508	42.0		
Hawaiian/Pacific Islander	10	0.4			6	0.4			4	0.3		
Caucasian/White	1,209	45.6			629	42.7			579	47.9		
Multiracial	92	3.5			53	3.6			39	3.2		
Other	33	1.2			16	1.1			17	1.4		
FRL												
Qualified	1,086	40.5			583	39.6			585	48.3		
Not qualified	827	30.8			444	30.2			383	31.7		
(Missing)	788	29.4			445	30.2			342	28.3		
Pretest reading	2,663		0.05	5.89	1,466		-0.04	5.82	1,196		0.17	6.00
Posttest reading	2,619		-0.19	6.13	1,436		0.01	6.20	1,182		-0.45	6.04
SSRS-PB	2,683		0.01	1.00	1,472		0.05	1.00	1,210		-0.05	1.01

Note. FRL: Free or reduced priced lunch; SSRS-PB: Social Scales Rating System Problem Behavior, Scores are factor scores. Scores for reading are factor scores.



Problem Behavior subscale. This subscale consists of 18 items on a 3-point scale (i.e., 0 = never, 1 = sometimes, and 2 = very often) addressing students' typical externalizing, internalizing, and hyperactive behavior. The reported internal consistency estimates on the SSRS teacher form for the problem behavior subscale is  $\alpha = 0.88$  (Gresham et al., 2011). After the development of the SSRS in the 1980s, results from several studies suggest the teacher form has criterion validity (Gresham & Elliott, 1990), discriminant validity (e.g., Ogden, 2003; Van der Oord et al., 2005), and construct validity (e.g., Elliott et al., 1988; Walthall et al., 2005).

### **Reading**

During the intervention studies, all participants were assessed by study staff with a battery of achievement and cognitive measures at the beginning, middle, and end of the academic year. For this study, we used a subset of reading achievement measures from the beginning and end of the year to estimate the effects of the treatments. Specifically, we included scores from the Letter-Word ID (LW), Word Attack (WA), Picture Vocabulary (PV), and Passage Comprehension (PC) subtests of the Woodcock-Johnson III Tests of Achievement (Woodcock et al., 2007) and scores from the Print Knowledge (PK) subtest of the Test of Preschool Early Literacy (TOPEL; Lonigan et al., 2007).

For LW, students are asked to name letters and read unknown words increasing in difficulty tapping into word recognition. The test-retest reliability estimates of the norming sample range between .90 and .96 and split half reliability estimates range between .88 and .99 (McGrew et al., 2007). During the WA subtest, students are asked to read nonsense words and highly infrequent words of increasing difficulty to gauge their decoding skills. One year test-retest reliability estimates range from .63 to .81 and split-half reliability estimates range from .78 to .94 (McGrew et al., 2007). The PV subtest assesses students' oral vocabulary and language development by asking them to name depicted objects. The test-retest reliability of the norming sample range between 0.70 and 0.81 and split half reliability ranges from .70 to .93 (McGrew et al., 2007). During the PC subtest, students are asked to read sentences or passages in which specific words are blanked out; students are to provide words without the help of a word bank. Alternate form reliability estimates range between .84 and .96 and split-half reliability estimates range from .73 to .96 (McGrew et al., 2007). The PK subtest of the TOPEL measures students' alphabet knowledge and knowledge about written language conventions, ranging from identifying which picture has a word in it to naming sounds of specific letters. The internal consistency of the TOPEL-PK is reported at  $\alpha = .96$  (Lonigan et al., 2007).

### **Data Analytic Plan**

#### **Scaling of Measures**

We first used IDA (Bauer & Hussong, 2009) to estimate latent scores on behavior and reading. The use of IDA helps to ensure that scores of participants from different projects are on the same scale. To estimate students' behavior rating, we used moderated non-linear factor analysis (MNLFA) to estimate factor scores for all students. In MNLFA, the parameters estimating factor loadings, means, and variance as well as

indicator intercepts, are allowed to vary as a function of moderator variables (Bauer & Hussong, 2009). This moderation removes possible differential item functioning and establishes scores on the same scale. Following procedures delineated by Curran et al. (2014), our model included project and student age (linear, squared, and cubed) as moderators. See Table S-1 for the final model. The composite reliability of the factor scores for our sample was 0.95. The MNLFA analysis was conducted in Mplus 7.3 (Muthén & Muthén, 1998–2002).

To scale students' reading ability, we used measurement invariance models. Our approach has been detailed in van Dijk, Schatschneider, et al. (2022), and we will provide a brief description here. We estimated reading ability as a single latent factor using the five reading assessments as indicators. First, we created random normal deviates for variables missing completely within a project (Widaman et al., 2013). We then used the free baseline approach to examine invariance across the projects. To avoid rejecting models with small differences that are significant due to sample size, we compared the sample value of the central  $\chi^2$  distribution to a critical value adjusted based on a non-centrality parameter representing close fit (MacCallum et al., 2006). This led to a partial strong invariance model for both pre- and posttest (Table S-2 contains the full models). Composite reliability within projects for the latent reading scores ranged from 0.82–0.93 at pretest and 0.73–0.94 at posttest. We extracted the factor scores for each student to use as variables in the final models. All measurement invariance modeling was conducted in R (R Core Team, 2022) with the *lavaan* package (Rosseel, 2012).

### **Missing Data**

Our approach to missing data included scores of those students with SSRS scores from the complete sample ( $n = 3,026$ ). Data were winsorized at the 99th percentile to reduce the influence of outliers. Because some students were missing scores on the latent reading assessments ( $n = 41$  at pretest and  $n = 100$  at posttest), we used multiple imputation (Rubin, 2004) to avoid case-wide deletion and enable the analyses to be done with a complete data set. We generated 5 imputations based on 25 iterations with the *mice* package (Buuren & Groothuis-Oudshoorn, 2011), taking the multi-level structure of the data into account. Table S-3 shows means and SD for the full Project Kids sample, the sample used for the present study, and of the imputed data sets; means and standard deviations across data sets and imputations are comparable.

### **Modeling Approach**

With the imputed data set, we ran hierarchical linear models (HLM), accounting for the nesting of students in classrooms and schools for the seven projects on ISI effectiveness ( $n = 2,683$ ). We analyzed means-as-outcomes models, separating the effect of the predictor variables into within and between components. For the variables of interest, we used a model building approach adding in the pre-intervention reading estimates first, then treatment variables, and finally the behavior variables and their interaction with treatment, comparing models using the *D2* likelihood-based method (Li et al., 1991), specifically tailored toward data from multiple imputations. We estimated the

interaction of treatment and student behavior both at the student and classroom level (Preacher & Sterba, 2019). The final model is as follows:

$$\begin{aligned}
 Y_{ijk} = & \tilde{\gamma}_{000} + \tilde{\gamma}_{100}TREAT_{ijk} + \\
 & \tilde{\gamma}_{200} \left( Read_{ijk} - \overline{Read}_{.jk} \right) + \tilde{\gamma}_{020} \left( \overline{Read}_{.jk} - \overline{\overline{Read}}_{.k} \right) + \tilde{\gamma}_{002} \overline{\overline{Read}}_{.k} + \\
 & \tilde{\gamma}_{300} \left( PB_{ijk} - \overline{PB}_{.jk} \right) + \tilde{\gamma}_{030} \left( \overline{PB}_{.jk} - \overline{\overline{PB}}_{.k} \right) + \tilde{\gamma}_{003} \overline{\overline{PB}}_{.k} + \\
 & \tilde{\gamma}_{400}TREAT_{ijk} * \left( PB_{ijk} - \overline{PB}_{.jk} \right) + \tilde{\gamma}_{040}TREAT_{ijk} * \left( \overline{PB}_{.jk} - \overline{\overline{PB}}_{.k} \right) + \\
 & \nu_{00k} + r_{0jk} + \varepsilon_{ijk},
 \end{aligned}$$

where  $Y_{ijk}$  is the individual student latent reading score post-intervention,  $\tilde{\gamma}_{000}$  the intercept and  $\tilde{\gamma}_{100}$  the overall treatment effect (the focal coefficient for research question 1). Regarding pre-intervention reading scores,  $\tilde{\gamma}_{200}$  represents the within classroom effect,  $\tilde{\gamma}_{020}$  the between classroom effect, and  $\tilde{\gamma}_{002}$  the between school effect. Similarly, for ratings of problem behavior,  $\tilde{\gamma}_{300}$  represents the within classroom effect,  $\tilde{\gamma}_{030}$  the between classroom effect, and  $\tilde{\gamma}_{003}$  the between school effect.  $\tilde{\gamma}_{400}$  represents the within classroom interaction coefficient between treatment and ratings of problem behavior and  $\tilde{\gamma}_{040}$  the between classroom interaction coefficient (these represent the focal coefficients for research question 2). Finally,  $\nu_{00k}$  represents the school level residual,  $r_{0jk}$  the classroom level residual, and  $\varepsilon_{ijk}$  the student level residual. All HLMs were estimated in R (R Core Team, 2020) with the *lme4* (Bates et al., 2015) and *mitml* (Grund et al., 2019) packages.

Finally, we probed the interaction terms of interest (i.e., treatment effect by behavior rating) by plotting the simple slope and 95% confidence band and calculating the range of significance through Johnson–Neyman (J–N) intervals (Johnson & Neyman, 1936). These intervals are useful to understand under which condition a moderating effect operates (Preacher et al., 2006; Preacher & Sterba, 2019), and have been specifically used in ATI research (Preacher & Sterba, 2019; Rogosa, 1981). The J–N technique involves plotting the simple slope of an interaction effect with 95% confidence bands (i.e., continuously plotted 95% confidence intervals). The confidence intervals at each point vary as a function of the moderator, and the upper and lower bounds are therefore not parallel to the simple slope (Preacher et al., 2006). After plotting the simple slope with its confidence bands, the values on the moderator where the confidence bands first exclude zero are calculated. The region within these two values constitutes *the region of significance* and includes all values of the moderator where the treatment effect is statistically significantly different from zero (Rogosa, 1981). One additional value of the J–N technique is that “a useful region of significance can be obtained even when the null hypothesis of parallel within-group regressions is not rejected.” (Rogosa, 1981, p. 83). In other words, the technique can help illuminate for which values of a moderator the difference between groups is significantly different from zero, even if the initial hypothesis test of an interaction was not statistically significant. The J–N plots were generated in R with code provided through <http://www.quantpsy.org/interact/hlm2.htm>. This website provides tools for plotting J–N intervals for interactions with both treatment effect and moderator happening at L1 (i.e., within classrooms) and for cross-level interactions (i.e., between classrooms).

## Results

About 55% of students were part of the treatment group. The treatment and control group did not differ in student demographics (Gender  $\chi^2 = 4.55$ ,  $p = .103$ ; Ethnicity  $\chi^2 = 4.69$ ,  $p = 0.320$ ; Race  $\chi^2 = 11.33$ ,  $p = .789$ , or eligibility for free or reduced priced lunch  $\chi^2 = 4.04$ ,  $p = .854$ ). The groups were equal in their pre-intervention reading skills ( $t_{2521.3} = 0.84$ ,  $p = .402$ ,  $d = -0.03$  [ $-0.11$ – $0.04$ ]) but differed on their scores on behavior rating scale (measured mid-year during the original studies) with the control group scoring slightly higher ( $t_{2571.1} = -2.6$ ,  $p = .010$ ,  $d = 0.1$  [ $0.02$ – $0.18$ ]).

The unconditional 3-level models showed about 3% of the variance due to school level factors, 11% due to classroom level factors, and the remaining 86% at the student level, suggesting a three-level model was appropriate. After accounting for students' preintervention reading ability, the treatment showed a statistically significant effect ( $b = 0.47$ ,  $p = .032$ ) and this addition was a significant improvement to the model conditional on reading ( $F_{1, 260.662} = 4.64$ ,  $p = .032$ ). Our next model included students' behavior rating, and its interaction with treatment. This model was an improvement to the previous model ( $F_{5, 662.945} = 11.656$ ,  $p < .001$ ). Table 3 provides model estimates for all models.

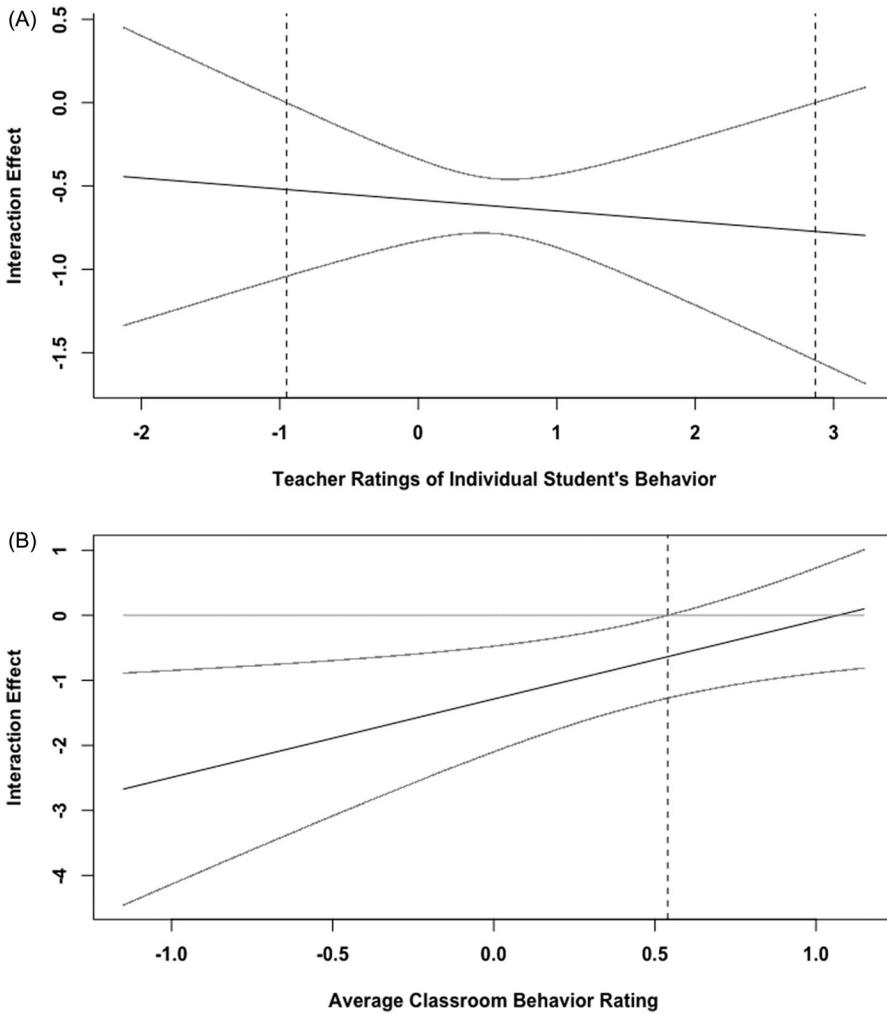
Our final model included preintervention reading skills, behavior ratings, and the interaction between behavior ratings and treatment, separated at the within and between classroom level. The results suggest the ISI intervention had a statistically significant effect on students' reading skills ( $b = 0.51$ ,  $p = .020$ ,  $d = 0.08$ ) after controlling for their prereading skills and behavior rating. Students' prereading skills were positively associated with their post-test scores, not only at the student level ( $b = 0.72$ ,  $p < .001$ ), but also at the classroom ( $b = 0.80$ ,  $p < .001$ ) and school level ( $b = 0.81$ ,  $p < .001$ ). That is, controlling for treatment and behavior ratings, students in higher performing classrooms at pretest had, on average, higher post-test scores, and students in higher performing schools at pretest had, on average, higher post-test scores. Unsurprisingly, students' behavior ratings had a significant negative relation to their post-test performance, both as individuals ( $b = -0.57$ ,  $p < .001$ ) and in classrooms ( $b = -1.18$ ,  $p = .007$ ). There was no difference in reading outcomes and behavior at the school level ( $b = -0.18$ ,  $p = .865$ ). The interaction term of treatment with student behavior ratings was not statistically significant at the individual student level ( $b = -0.08$ ,  $p = .621$ ), but was at the classroom level ( $b = 1.06$ ,  $p = .045$ ). Since these parameters are estimated at the mean level only, we estimated the Johnson–Neyman intervals to determine the range of scores on the behavior rating within which the treatment was significant (Preacher & Sterba, 2019). This range of scores was between  $-0.95$  and  $2.87$  at the individual level and between  $0.05$  and  $4.25$  for classrooms (see Figure 1). These results suggest that, compared to students in the control conditions with similar behavior ratings, students with either low or very high average behavior ratings did not increase their reading skills significantly from pre- to postintervention. Similarly, students in classrooms with higher average behavior ratings did not benefit from the intervention compared to students in similar classrooms in the control conditions.

**Table 3.** Fixed and random effects of hierarchical linear models.

Component	Unconditional model			Conditional on reading			Treatment effect			Interaction models		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Fixed												
Intercept	-0.21	0.30	0.47	-0.26	0.17	.135	-0.51	0.21	.014*	-0.54	0.21	.010*
Treatment							0.47	0.22	.032**	0.51	0.22	.020*
Reading-GMC				0.82	0.12	<.001***	0.83	0.12	<.001***	0.81	0.13	<.001***
Reading-SMC				0.83	0.07	<.001***	0.83	0.07	<.001***	0.80	0.07	<.001***
Reading-CMC				0.74	0.02	<.001***	0.74	0.02	<.001***	0.72	0.02	<.001***
SSRS-GMC										-0.18	1.06	.865
SSRS-SMC										-1.18	0.44	.007**
SSRS-CMC										-0.57	0.13	<.001***
Treatment*SSRS-CMC										-0.08	0.17	.621
Treatment*SSRS-SMC										1.06	0.53	.045*
Random												
School	1.27			0.33			0.33			0.35		
Classroom	4.08			1.78			1.71			1.75		
Student	32.94			15.68			15.67			15.31		

Note. GMC: centered at the grand mean; SMC: centered around the school mean; CMC: centered around the classroom mean.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .



**Figure 1.** 95% Confidence bands of the simple slope for treatment effect by student problem behavior. The dotted lines represent the values of factor scores on Social Skills Rating Scores-Problem Behavior subscale between which the treatment is significant. The x-axis spans only the observed values in our sample.

## Discussion

ATI theory states student learning can be optimized by adapting instruction and interventions to match a student’s cognitive or affective characteristics (Cronbach & Snow, 1977). Understanding which child characteristics interact with instruction and interventions can help teachers choose interventions and implement them in a way that will lead to highest gains for all students. It is clear that the effectiveness of comprehensive reading approaches in raising reading outcomes for early elementary students depends on their pre-intervention reading and cognitive ability (e.g., Coyne et al., 2018; Hart et al., 2016). However, previous research has provided mixed evidence of students’ behavior as a moderator to treatment effectiveness. To provide more specific evidence

on the moderating relation, our study used a large, combined sample of seven RCTs of the ISI intervention. The sample included students with wide ranges of both reading ability and behavior ratings. Similar to other studies using a composite reading score (Hurry et al., 2018; Rabiner & Malone, 2004), the interaction coefficient of problem behavior and treatment in our study was not statistically significant at the individual student level, suggesting mean student behavior ratings are not related to a differential effect of the ISI intervention. In order to obtain a more nuanced picture of the interaction, we modeled the simple slope with confidence bands and estimated the range of statistical significance of the treatment across behavior ratings with Johnson–Neyman intervals. These ranges suggest the ISI intervention was only statistically significantly effective for students with average behavior ratings. Roberts and colleagues (2019) also found above average behavior ratings moderated the treatment effect on measures of reading comprehension. While those authors tentatively contributed this pattern to low sample sizes, our results, based on a large sample, suggest that this might be a general pattern. Combined, these results support the idea that, besides cognitive skills, affective characteristics of students also influence response to treatment.

Additionally, the results from this study also suggest teachers' perceptions of student behavior relative to their classroom as a whole influences the effectiveness of reading instruction and intervention. The statistically significant interaction of teacher ratings and treatment at the classroom level suggests that ISI became increasingly effective as the average behavior ratings in a classroom increase. However, for classrooms with a relatively high average of behavior ratings, the difference with the control group was no longer statistically significant. None of the previous studies investigating the moderating relation of student behavior ratings on reading separated within and between classroom effects, and it is therefore unclear if our findings are common or anomalous. Recent ISI research suggests that teachers spent more time in small group instruction with students rated as higher in problem behavior, if they perceived these students to be more competent academically compared to students with higher problem behaviors but lower perceived academic competence (Toste et al., in press). However, in classrooms with higher average perceived problem behaviors, unproductive non-instructional time may be increased due to managing behavior, time that is taken away from general reading instruction (e.g., Day et al., 2015). If students from these classrooms are, instead, provided with high-quality reading instruction and interventions that include organizational and emotional support to engage in the material, students may learn more effectively (Foorman & Torgesen, 2001). To better understand how classroom dynamics influence the effectiveness of interventions, future work should routinely include the distinction of within and between classroom effects.

Our results suggest that students rated with very high behavior problems may need additional supports to be able to benefit from reading instruction and interventions. This finding is of importance because students with emotional and behavioral disorders increase their reading skills at a lower rate compared to students with reading disabilities (Anderson et al., 2001) and the rate decreases as the severity of the behavior challenges increases (Mellado de la Cruz et al., 2019). Failure to provide appropriate interventions and core reading instruction will place them at increased disadvantage. Students with problem behaviors not proficient in reading by third grade are four times

more likely to drop-out of school compared to their reading proficient peers (Hernandez, 2011), are more likely to become victims of bullying (Turunen et al., 2019), have low high-school completion rates (Bradley et al., 2008), take longer to graduate (Hakkarainen et al., 2016) and have poor post-school employment (Bradley et al., 2008; Hakkarainen et al., 2016). Fortunately, there are specific interventions for students with behavior problems that have been shown effective in increasing their reading abilities (Roberts et al., 2020) and it is important for schools to provide these more specialized interventions that take students' problem behaviors into account.

Additionally, the results also suggest students with very low behavior problems ratings did not benefit from the core reading instruction. It is possible this group of students included highly proficient readers that needed amounts of meaning-focused instruction that was beyond the reach of classroom teachers. Previous ISI research suggested greater reading gains when teachers provided students with amounts of instruction closer to the recommended amount (Connor et al., 2007, 2009). It is equally possible students with very low behavior ratings are weaker readers. Previous research has suggested teachers' perceptions of student behavior, such as motivation, work habits, and classroom behaviors, influence their expectations of students (e.g., Rubie-Davies, 2010; Timmermans et al., 2016). If teachers overestimated the reading skills of these students due to their compliant demeanors, they may not have provided enough individualization in code- or meaning-focused instruction. It is unclear what mechanism specifically underlies these results and future research might focus on this group of students in particular.

Besides providing more specialized instruction and intervention, an increased focus on behavior management seems pertinent given that students in classrooms with higher perceived average behavior problems did not benefit from the ISI intervention compared to control students in similar classrooms. Good behavior management is associated with better student achievement across academic subjects (e.g., Korpershoek et al., 2016; van Dijk et al., 2019) and especially, in the case of reading, for boys at risk for behavior disorders (Garwood & Vernon-Feagans, 2017). Unfortunately, teachers often believe behavior management is the most difficult part of their job (Reinke et al., 2011) and receive very little behavior management training (Freeman et al., 2014; Oliver & Reschly, 2010).

The results of this study should be considered in light of its limitations. First, to represent students' behavior, we relied on teacher ratings of their students' behavior taken at one timepoint during the individual randomized control trials. This approach also does not take the reciprocal relation between reading difficulties and problem behavior into account, but rather treats teachers' ratings of problem behavior as a relatively stable construct. It is possible these ratings were influenced by biases and teachers may have rated students relative to other students in their classroom (Dinnebeil et al., 2013). The inclusion of multiple sources of information on student behavior, such as teacher and parent ratings, could help stave off influence of teacher bias on behavior in future studies (Dinnebeil et al., 2013). Additionally, this study cannot distinguish between differential effects of the ISI intervention due to inadequate response to the intervention for students with higher rated problem behaviors and a lack of intervention provided as a result of actual problem behaviors. To get a more precise idea of how behavior might



impede intervention effectiveness, future studies might consider including direct observation methods, specifically how teachers are interacting with students rated having very high or very low problem behaviors during reading instruction. It is possible interactions differ based on behavior ratings, as these influence teachers' expectations of academic achievement (Rubie-Davies, 2010; Timmermans et al., 2016). Observations should also take into account how teachers are managing behavior in their classrooms, as this can lead to wasted instructional time or decreased efficiency of instruction. Since the original ISI studies suggest better results for teachers that provided code- and meaning-focused instruction closer to the recommended amounts (Connor et al., 2007, 2009), tying observations of student behavior and teachers' behavior management with observations of instruction could help us understand the mechanisms behind the non-response of both students with very low and very high ratings of problem behavior.

Second, our results are focused on effects during the intervention year only. Given the slower reading growth of students with behavior problems, it is possible the interaction becomes more pronounced over time and the range of non-responders might differ. Longitudinal models may provide a more detailed picture of how problem behaviors influence reading outcomes for students receiving interventions over time. Relatedly, we did not take grade level differences into account. The relation between problem behavior and academic achievement is recursive (e.g., Morgan et al., 2008). Students in higher grades may have different responses to interventions due to a compounding effect of not having learned to read in the previous years. This might lead to increased frustration, feelings of failure, and embarrassment (Lin et al., 2013). Future research might focus on disentangling these possible grade level effects.

## Conclusion

Providing high-quality, core classroom reading instruction and interventions are essential for students to become literate, but they do not work for all students. In order to ensure equal opportunities for all students in a society highly focused on information processing, it is essential to uncover the contextual factors that make reading instruction more or less effective. Teacher ratings of student behavior is one such contextual factor. Adapting instruction to match the individual needs of students is important to help students start their education with the best possibility for becoming successful.

## Open Research Statements

### *Study and Analysis Plan Registration*

There is no study and analysis plan registration associated with this manuscript.

### *Data, Code, and Materials Transparency*

The data (10.33009/ldbase.1620844399.85a0) and code (<http://ldbase.org/code/83324a11-e646-4f78-9a54-7726488b3a5f>) that support the findings of this study are openly available LDbase.org.

## **Design and Analysis Reporting Guidelines**

There is not a completed reporting guideline checklist associated with this manuscript.

## **Transparency Declaration**

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

## **Replication Statement**

This manuscript reports an original study.

## **Acknowledgements**

We thank the original project participants, research staff, and funding agencies. We also honor the late Dr. Carol Connor for her tremendous contributions to the science of reading interventions and child by instruction interaction literature. Dr. Connor was an original Co-Investigator of Project KIDS.

## **Funding**

This work is supported by Eunice Kennedy Shriver National Institute of Child Health & Human Development Grants R21HD072286, P50HD052120, and R01HD095193. Views expressed herein are those of the authors and have neither been reviewed nor approved by the granting agencies. Views expressed herein are those of the authors and have neither been reviewed nor approved by the granting agencies.

## **Open Scholarship**



This article has earned the Center for Open Science badges for Open Data. The data are openly accessible at <https://ldbase.org/datasets/e092873f-8c70-49de-afb9-61de8ad1d508> and <https://ldbase.org/datasets/a102b645-3eb3-40e9-891f-982611f62107>.

## **ORCID**

Wilhelmina van Dijk  <http://orcid.org/0000-0001-9195-8772>

Christopher Schatschneider  <http://orcid.org/0000-0002-1700-7685>

Stephanie Al Otaiba  <http://orcid.org/0000-0001-7125-3791>

Sara A. Hart  <http://orcid.org/0000-0001-9793-0420>

## **References**

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective?. *Journal of Learning Disabilities*, 39(5), 414–431. <https://doi.org/10.1177/00222194060390050401>

- Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention. *Remedial and Special Education, 23*(5), 300–316. <https://doi.org/10.1177/07419325020230050501>
- Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81*(1), 11–27. <https://doi.org/10.1177/0014402914532234>
- Al Otaiba, S., Folsom, J. S., Schatschneider, C., Wanzek, J., Greulich, L., Meadows, J., Li, Z., & Connor, C. M. (2011). Predicting first-grade reading performance from kindergarten response to Tier 1 instruction. *Exceptional Children, 77*(4), 453–470. <https://doi.org/10.1177/001440291107700405>
- Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal, 111*(4), 535–560. <https://doi.org/10.1086/659031>
- Anderson, J. A., Kutash, K., & Duchnowski, A. J. (2001). A comparison of the academic progress of students with EBD and students with LD. *Journal of Emotional and Behavioral Disorders, 9*(2), 106–115. <https://doi.org/10.1177/106342660100900205>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods, 14*(2), 101. <https://doi.org/10.1037/a0015583>
- Biancarosa, G., & Snow, C. E. (2004). *Reading next: A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York*. Alliance for Excellent Education.
- Bradley, R., Doolittle, J., & Bartolotta, R. (2008). Building on the data and adding to the discussion: The experiences and outcomes of students with emotional disturbance. *Journal of Behavioral Education, 17*(1), 4–23. <https://doi.org/10.1007/s10864-007-9058-6>
- Buuren, S. v., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.
- Connor, C. M., Lara, J. J., Crowe, E. C., & Meadows, J. G. (2009). Instruction, student engagement, and reading skill growth in Reading First classrooms. *The Elementary School Journal, 109*(3), 221–250. <https://doi.org/10.1086/592305>
- Connor, C. M., Morrison, F. J., Fishman, B., Crowe, E. C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science, 24*(8), 1408–1419. <https://doi.org/10.1177/0956797612472204>
- Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., Bayraktar, A., Crowe, E. C., & Schatschneider, C. (2011). Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly, 46*(3), 189–221. <https://doi.org/10.1598/RRQ.46.3.1>
- Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464–465. <https://doi.org/10.1126/science.1134513>
- Connor, C. M., Morrison, F. J., & Katch, L. E. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading, 8*(4), 305–336. [https://doi.org/10.1207/s1532799xssr0804\\_1](https://doi.org/10.1207/s1532799xssr0804_1)
- Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child × instruction interactions. *Journal of Educational Psychology, 96*(4), 682–698. <https://doi.org/10.1037/0022-0663.96.4.682>
- Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J., Lundblom, E., Crowe, E. C., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristics by

- reading instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness*, 4(3), 173–207. <https://doi.org/10.1080/19345747.2010.510179>
- Coyne, M. D., Oldham, A., Dougherty, S. M., Leonard, K., Koriakin, T., Gage, N. A., Burns, D., & Gillis, M. (2018). Evaluating the effects of supplemental reading intervention within an MTSS or RTI reading reform initiative using a regression discontinuity design. *Exceptional Children*, 84(4), 350–367.
- Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Irvington.
- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14(2), 81–100. <https://doi.org/10.1037/a0015914>
- Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., & Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in Integrative Data Analysis. *Multivariate Behavioral Research*, 49(3), 214–231. <https://doi.org/10.1080/00273171.2014.889594>
- Daucourt, M. C., Schatschneider, C., Connor, C. M., Al Otaiba, S., & Hart, S. A. (2018). Inhibition, updating working memory, and shifting predict reading disability symptoms in a hybrid model: Project KIDS. *Frontiers in Psychology*, 9, 238. <https://doi.org/10.3389/fpsyg.2018.00238>
- Day, S. L., Connor, C. M., & McClelland, M. M. (2015). Children's behavioral regulation and literacy: The impact of the first grade classroom environment. *Journal of School Psychology*, 53(5), 409–428. <https://doi.org/10.1016/j.jsp.2015.07.004>
- Dinnebeil, L. A., Sawyer, B. E., Logan, J., Dynia, J. M., Cancio, E., & Justice, L. M. (2013). Influences on the congruence between parents' and teachers' ratings of young children's social skills and problem behaviors. *Early Childhood Research Quarterly*, 28(1), 144–152. <https://doi.org/10.1016/j.ecresq.2012.03.001>
- DuPaul, G. J., Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Academic and social functioning associated with attention-deficit/hyperactivity disorder: Latent class analyses of trajectories from kindergarten to fifth grade. *Journal of Abnormal Child Psychology*, 44(7), 1425–1438. <https://doi.org/10.1007/s10802-016-0126-z>
- Elliott, S. N., Gresham, F. M., Freeman, T., & McCloskey, G. (1988). Teacher and observer ratings of children's social skills: Validation of the Social Skills Rating Scales. *Journal of Psychoeducational Assessment*, 6(2), 152–161. <https://doi.org/10.1177/073428298800600206>
- Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research and Practice*, 16(4), 203–212. <https://doi.org/10.1111/0938-8982.00020>
- Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. *Teacher Education and Special Education*, 37(2), 106–120. <https://doi.org/10.1177/0888406413507002>
- Fuchs, D., & Fuchs, L. S. (2019). On the importance of moderator analysis in intervention research: An introduction to the special issue. *Exceptional Children*, 85(2), 126–128. <https://doi.org/10.1177/0014402918811924>
- Garwood, J. D., & Vernon-Feagans, L. (2017). Classroom management affects literacy development of students with emotional and behavioral disorders. *Exceptional Children*, 83(2), 123–142. <https://doi.org/10.1177/0014402916651846>
- Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system (SSRS)*. American Guidance Service.
- Gresham, F. M., Elliott, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skills Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly*, 26(1), 27–44. <https://doi.org/10.1037/a0022662>
- Grills-Tauchel, A. E., Fletcher, J. M., Vaughn, S. R., & Stuebing, K. K. (2012). Anxiety and reading difficulties in early elementary school: Evidence for unidirectional- or bi-directional

- relations?. *Child Psychiatry & Human Development*, 43(1), 35–47. <https://doi.org/10.1007/s10578-011-0246-1>
- Grund, S., Robitzsch, A., & Luedtke, O. (2019). *mitml: Tools for multiple imputation in multilevel modeling*. <https://CRAN.R-project.org/package=mitml>
- Hagan-Burke, S., Coyne, M. D., Kwok, O-m., Simmons, D. C., Kim, M., Simmons, L. E., Skidmore, S. T., Hernandez, C. L., & McSparran Ruby, M. (2013). The effects and interactions of student, teacher, and setting variables on reading outcomes for kindergarteners receiving supplemental reading intervention. *Journal of Learning Disabilities*, 46(3), 260–277. <https://doi.org/10.1177/0022219411420571>
- Hagan-Burke, S., Kwok, O., Zou, Y., Johnson, C., Simmons, D., & Coyne, M. D. (2011). An examination of problem behaviors and reading outcomes in kindergarten students. *The Journal of Special Education*, 45(3), 131–148. <https://doi.org/10.1177/0022466909359425>
- Hakkarainen, A. M., Holopainen, L. K., & Savolainen, H. K. (2016). The impact of learning difficulties and socioemotional and behavioural problems on transition to postsecondary education or work life in Finland: A five-year follow-up study. *European Journal of Special Needs Education*, 31(2), 171–186. <https://doi.org/10.1080/08856257.2015.1125688>
- Hart, S. A., Al Otaiba, S., Connor, C. M., & Norris, C. U. (2021a). *Project KIDS item level data* [Data set]. <https://doi.org/10.33009/ldbbase.1620837890.bcf8>
- Hart, S. A., Al Otaiba, S., Connor, C. M., & Norris, C. U. (2021b). *Project KIDS total scores data* [Data set]. <https://doi.org/10.33009/ldbbase.1620844399.85a0>
- Hart, S. A., Piasta, S. B., & Justice, L. M. (2016). Do children’s learning-related behaviors moderate the impacts of an empirically-validated early literacy intervention? *Learning and Individual Differences*, 50, 73–82. <https://doi.org/10.1016/j.lindif.2016.07.005>
- Hernandez, D. J. (2011). *Double Jeopardy: How third-grade reading skills and poverty influence high school graduation*. Annie E. Casey Foundation.
- Hornburg, C. B., Rieber, M. L., & McNeil, N. M. (2017). An integrative data analysis of gender differences in children’s understanding of mathematical equivalence. *Journal of Experimental Child Psychology*, 163, 140–150. <https://doi.org/10.1016/j.jecp.2017.06.002>
- Hurry, J., Flouri, E., & Sylva, K. (2018). Literacy difficulties and emotional and behavior disorders: Causes and consequences. *Journal of Education for Students Placed at Risk (JESPAR)*, 23(3), 259–279. <https://doi.org/10.1080/10824669.2018.1482748>
- Jansen, M., Lüdtke, O., & Robitzsch, A. (2020). Disentangling different sources of stability and change in students’ academic self-concepts: An integrative data analysis using the STARTS model. *Journal of Educational Psychology*, 112(8), 1614–1631. <https://doi.org/10.1037/edu0000448>
- Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs*, 1, 57–93.
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students’ academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, 86(3), 643–680. <https://doi.org/10.3102/0034654315626799>
- Leijten, P., Raaijmakers, M., Wijngaards, L., Matthys, W., Menting, A., Hemink-van Putten, M., & Orobio de Castro, B. (2018). Understanding who benefits from parenting interventions for children’s conduct problems: An Integrative Data Analysis. *Prevention Science*, 19(4), 579–588. <https://doi.org/10.1007/s11121-018-0864-y>
- Li, K.-H., Meng, X.-L., Raghunathan, T. E., & Rubin, D. B. (1991). Significance levels form repeated p-values with multiply-imputed data. *Statistica Sinica*, 1(1), 65–92.
- Lin, Y.-C., Morgan, P. L., Hillemeier, M., Cook, M., Maczuga, S., & Farkas, G. (2013). Reading, mathematics, and behavioral difficulties Interrelate: Evidence from a cross-lagged panel design and population-based sample of US upper elementary students. *Behavioral Disorders*, 38(4), 212–227. <https://doi.org/10.1177/019874291303800404>
- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *TOPEL: Test of preschool early literacy*. Pro-Ed Austin.

- MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods*, 11(1), 19–35. <https://doi.org/10.1037/1082-989X.11.1.19>
- McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Technical manual. Woodcock-Johnson III normative update*. Riverside Publishing.
- Mellado de la Cruz, V., Al Otaiba, S., Hsiao, Y.-Y., Clemens, N. H., Jones, F. G., Rivas, B. K., Brewer, E. A., Hagan-Burke, S., & Simmons, L. E. (2019). The prevalence and stability of challenging behaviors and concurrent early literacy growth among Kindergartners at reading risk. *The Elementary School Journal*, 120(2), 220–242. <https://doi.org/10.1086/705785>
- Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities*, 41(5), 417–436. <https://doi.org/10.1177/0022219408321123>
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.
- Ogden, T. (2003). The validity of teacher ratings of adolescents' social skills. *Scandinavian Journal of Educational Research*, 47(1), 63–76. <https://doi.org/10.1080/00313830308605>
- Oliver, R. M., & Reschly, D. J. (2010). Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders*, 35(3), 188–199. <https://doi.org/10.1177/019874291003500301>
- Otaiba, S. A., Folsom, J. S., Wanzek, J., Greulich, L., Wasche, J., Schatschneider, C., & Connor, C. (2016). Professional development to differentiate kindergarten Tier 1 instruction: Can already effective teachers improve student outcomes by differentiating Tier 1 instruction? *Reading & Writing Quarterly: overcoming Learning Difficulties*, 32(5), 454–476. <https://doi.org/10.1080/10573569.2015.1021060>
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, 31(4), 437–448. <https://doi.org/10.3102/10769986031004437>
- Preacher, K. J., & Sterba, S. K. (2019). Aptitude-by-treatment interactions in research on educational interventions. *Exceptional Children*, 85(2), 248–264. <https://doi.org/10.1177/0014402918802803>
- Rabiner, D. L., & Malone, P. S. (2004). The impact of tutoring on early reading achievement for children with and without attention problems. *Journal of Abnormal Child Psychology*, 32(3), 273–284. <https://doi.org/10.1023/B:JACP.0000026141.20174.17>
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.Rproject.org/>
- Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, 26(1), 1–13. <https://doi.org/10.1037/a0022714>
- Roberts, G. J., Cho, E., Garwood, J. D., Goble, G. H., Robertson, T., & Hodges, A. (2020). Reading interventions for students with reading and behavioral difficulties: A meta-analysis and evaluation of co-occurring difficulties. *Educational Psychology Review*, 32(1), 17–47. <https://doi.org/10.1007/s10648-019-09485-1>
- Roberts, G. J., Vaughn, S., Roberts, G., & Miciak, J. (2019). Problem behaviors and response to reading intervention for upper elementary students with reading difficulties. *Remedial and Special Education*, 42(3), 169–181. <https://doi.org/10.1177/0741932519865263>
- Rogosa, D. (1981). On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within-group regressions. *Educational and Psychological Measurement*, 41(1), 73–84. <https://doi.org/10.1177/001316448104100108>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology*, 80(1), 121–135. <https://doi.org/10.1348/000709909X466334>
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81). John Wiley & Sons.

- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*(1), 77–96. <https://doi.org/10.1177/0022219414528540>
- Swanson, E. A., Stevens, E. A., Scammacca, N. K., Capin, P., Stewart, A. A., & Austin, C. R. (2017). The impact of tier 1 reading instruction on reading outcomes for students in Grades 4–12: A meta-analysis. *Reading and Writing, 30*(8), 1639–1665. <https://doi.org/10.1007/s11145-017-9743-3>
- Tang, X., & Dai, T. (2021). How do classroom behaviors predict longitudinal reading achievement? A conditional autoregressive latent growth analysis. *Early Childhood Research Quarterly, 54*, 239–251. <https://doi.org/10.1016/j.ecresq.2020.09.007>
- Timmermans, A. C., de Boer, H., & van der Werf, M. P. C. (2016). An investigation of the relationship between teachers' expectations and teachers' perceptions of student attributes. *Social Psychology of Education, 19*(2), 217–240. <https://doi.org/10.1007/s11218-015-9326-6>
- Toste, J. R., McLean, L., Peng, P., Didion, L., Filderman, M. J., Sparapani, N., & Connor, C. M. (in press). Do teacher perceptions of students' academic and behavioral skills influence time spent in smallgroup reading instruction? *The Elementary School Journal*. <https://doi.org/10.35542/osf.io/tk28j>
- Turunen, T., Kiuru, N., Poskiparta, E., Niemi, P., & Nurmi, J.-E. (2019). Word reading skills and externalizing and internalizing problems from grade 1 to grade 2—Developmental trajectories and bullying involvement in grade 3. *Scientific Studies of Reading, 23*(2), 161–177. <https://doi.org/10.1080/10888438.2018.1497036>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). (2022). *Reading assessment*. NAEP.
- Van der Oord, S., Van der Meulen, E. M., Prins, P. J. M., Oosterlaan, J., Buitelaar, J. K., & Emmelkamp, P. M. G. (2005). A psychometric evaluation of the social skills rating system in children with attention deficit hyperactivity disorder. *Behaviour Research and Therapy, 43*(6), 733–746. <https://doi.org/10.1016/j.brat.2004.06.004>
- van Dijk, W., Gage, N. A., & Grasley-Boy, N. (2019). The effect of behavior and classroom management on upper elementary students' mathematics achievement: A multilevel structural equation model. *Psychology in the Schools, 56*(7), 1173–1186. <https://doi.org/10.1002/pits.22254>
- van Dijk, W., Norris, C. U., Otaiba, S. A., Schatschneider, C., & Hart, S. A. (2022). Exploring individual differences in response to reading intervention: Data from Project KIDS (Kids and Individual Differences in Schools). *Journal of Open Psychology Data, 10*, 2. <https://doi.org/10.5334/jopd.58>
- van Dijk, W., Schatschneider, C., Al Otaiba, S., & Hart, S. A. (2022). Assessing measurement invariance across multiple groups: When is fit good enough? *Educational and Psychological Measurement, 83*(2), 482–505. <https://doi.org/10.1177/00131644211023567>
- Walthall, J. C., Konold, T. R., & Pianta, R. C. (2005). Factor structure of the Social Skills Rating System across child gender and ethnicity. *Journal of Psychoeducational Assessment, 23*(3), 201–215. <https://doi.org/10.1177/073428290502300301>
- Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities, 51*(6), 612–624. <https://doi.org/10.1177/0022219418775110>
- Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review, 28*(3), 551–576. <https://doi.org/10.1007/s10648-015-9321-7>
- Widaman, K. F., Grimm, K. J., Early, D. R., Robins, R. W., & Conger, R. D. (2013). Investigating factorial invariance of latent variables across populations when manifest variables are missing completely. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(3), 384–408. <https://doi.org/10.1080/10705511.2013.797819>
- Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2007). *Woodcock-Johnson III normative update*. Riverside Publishing.