Original Research

The Effect of Supplemental Reading Instruction on Fluency Outcomes for Children With Down Syndrome: A Closer Look at Curriculum-Based Measures

Exceptional Children
2022, Vol. 88(4) 421–441

DOI: 10.1177/00144029221081006
journals.sagepub.com/home/ecx
SAGE

Seth King[1], Derek Rodgers[2], and Christopher J. Lemons[3]

Abstract

Research supports the efficacy of intensive literacy instruction for children with moderate intellectual disabilities and Down syndrome (DS). However, much of the literature features measures closely aligned with evaluated interventions. Despite their increasing role in instruction, curriculum-based measures (CBM) are rarely featured in reading studies involving DS. Increasing the use of CBM in research has the potential to provide insight into the effectiveness of intervention and address concerns regarding the utility of approaches predicated on CBM. This single-case design study used CBM to examine the performance of children with DS (N = 17) who had largely exhibited gains on intervention-aligned measures following an intensive reading intervention. Results of multilevel modeling were mixed, with significant (p < .05) effects relegated to letter- and first-sound fluency. No more than 29% of participants met goals created using a procedure derived from CBM. Findings have implications for future studies and implementation of literacy interventions for children with DS.

Over the past 2 decades, the focus of reading instruction for children with moderate and severe intellectual disabilities (ID; i.e., IQ approximately ≤ 50; deficits in adaptive behavior; Boat & Wu, 2015) has shifted from functional or sight word reading to more advanced reading components necessary for independence (e.g., decoding, comprehension; Ahlgrim-Delzell & Rivera, 2015). Although children with moderate ID display lower performance in reading relative to their peers (Allor et al., 2014), research suggests this population benefits from systematic, direct literacy instruction (Dessemontet et al., 2019). In addition to moderate ID, the developmental profile of Down syndrome (DS)—a genetic disorder occurring in 14 of every 10,000 births—includes relative strengths in visual processing coupled with deficits in expressive language, phonological awareness, word attack, and other skills with the potential to attenuate otherwise effective reading instruction (Cologon et al., 2011; Grieco et al., 2015).

Predictors of literacy in children without disabilities have been well-established, with

[1]University of Lowa
[2]University of Nebraska-Lincoln
[3]Stanford University

Corresponding Author:
Seth King, Department of Teaching and Learning, College of Education, University of Iowa, Iowa City, IA 52246, USA.
Email: chris.lemons@stanford.edu

phonological awareness (Melby-Lervag et al., 2012), letter-sound knowledge (e.g., Clemens et al., 2020), and receptive language (e.g., Psyridou et al., 2018) all recognized as critical to the development of advanced reading skills. There is far less consensus regarding significant contributors to reading development in DS, with varying reports of IQ and other participant characteristics impeding direct comparisons of results (Dessemontet et al., 2019). Phonological awareness, for example, has been identified as a significant predictor of word-reading in some longitudinal and intervention studies (e.g., Lemons & Fuchs, 2010; Bird et al., 2000), yet unpredictive of reading outcomes in others (Burgoyne et al., 2012; Steele et al., 2013). Similar contradictory findings have emerged in regards to the contribution of letter-sound knowledge (e.g., Laws & Gunn, 2002 ; c.f. Steele et al., 2013) and age (Roch et al., 2019). Additional factors positively correlated with reading outcomes for individuals with DS include IQ and listening comprehension (Burgoyne et al., 2012; Roch et al., 2019). Teacher-level factors may also be at play, as an emerging body of literature suggests teacher experience and education level could be positively associated with student outcomes (Burroughs et al., 2019). Differences in observed outcomes may be linked to the characteristics of featured samples and outcome measures (Dessemontet & de Chambrier, 2015).

Notwithstanding disagreement regarding the influence of specific factors to the development of reading in individuals with DS, direct instruction appears to yield gains in terms of word reading, letter identification, and phoneme blending (e.g., Burgoyne et al., 2012, 2013). Assessing the impact of instruction is complicated, however, because of the various measures used to assess student progress. Studies involving children with DS often feature single-case design research predicated on measures closely aligned with the intervention (e.g., van Bysterveldt et al., 2010). Single-case design reading studies involving children with moderate ID and closely aligned measures yield effect sizes over 300% larger than group design studies featuring standardized measures (Dessemontent et al., 2019). On standardized measures, children with DS who possess limited reading skills (cf., relatively advanced readers; Lim et al., 2019) usually achieve more modest outcomes. In a wait-list control trial, Burgoyne et al. (2012) evaluated the effect of 20–40 weeks of intensive reading instruction for children with DS and moderate ID ($N = 57$; age 5–10 years). Intervention consisted of 1:1 text centered reading activities coupled with direct instruction in sight words and phonics, delivered concurrently with a language program during daily 40-min sessions. Compared to the control group, children who received 20 weeks of intervention exhibited small and moderate improvements on standardized measures including single-word reading ($d = 0.23$; $p = .02$), letter-sound knowledge ($d = 0.42$; $p < .00$), and phoneme blending ($d = 0.54$; $p = .02$).

That reading outcomes for children with moderate ID would vary based on measurement is not altogether surprising. As noted by Yoder et al. (2013), measures closely aligned with instruction—which commonly appear in single-case design research and practical settings—often indicate that students make substantial progress following an intervention. This performance, however, may reflect participants' familiarity with instructional materials or a narrow scope and sequence rather than their competence in a broader domain (i.e., overalignment; What Works Clearinghouse, 2020). Standardized measures showing less pronounced improvement potentially provide a more compelling indication of whether instruction prepares participants to perform highly generalized skills (e.g., reading) whose permutations cannot be anticipated by a single intervention. Fuchs et al. (2018) suggest intervention-aligned measures may nonetheless provide evidence of the effectiveness of teaching procedures and ultimately assist in the development of effective instructional programs. Differences in outcome measures could likewise reflect the insensitivity of standardized measures to student gains. Disparities in measures featured in reading studies raise questions regarding interventions for children with moderate ID and DS.

The use of assessment in reading for individuals with ID and DS is further complicated by the increasing application of curriculum-based measures (CBM; Lemons & Fuchs, 2010). Whereas intervention aligned measures typically relate to skills taught within a specific instructional unit and may not be comparable to measures employed across lessons, CBM—unlike standardized measures—are designed to be administered repeatedly across an extended period of time (e.g., first grade) as a means of assessing student performance (Tindal, 2013). Additionally, the technical adequacy of CBM (e.g., reliability), relative to many intervention-aligned measures, provides a higher degree of confidence when making decisions or determining the extent of student learning. Researchers (e.g., Hosp et al., 2014; Lemons & Fuchs, 2010) have found CBM provide a valid metric of reading performance for individuals with DS. When used in conjunction with intensive reading instruction, however, the progress of children with moderate ID on CBM appears to emerge slowly, with many participants requiring at least 15–20 weeks of instruction before demonstrating gains (e.g., Allor et al., 2010). Likewise, in a sample of 3,811 students with ID in Grades 3–8 and 11 assessed at the end of the school year, only 5.2% and 0.6% met measure-specific normative benchmarks on word or passage reading CBM, respectively (Lemons et al., 2013).

The extent to which CBM captures growth and responsiveness to instruction, relative to other measures, remains an important consideration in evaluating reading instruction for children with DS (Lemons & Fuchs, 2010). CBM is used as a universal screening and progress monitoring tool within school-wide multitiered systems of support (MTSS) frameworks designed to allocate educational resources through ongoing assessment (Van Meveren et al., 2020). A related use of CBM is data-based individualization (DBI; Fuchs et al., 2020). Although steps in the process vary, DBI involves (1) ascertaining the child's baseline level of performance; (2) establishing an instructional goal based on benchmark data or the child's current level of performance; (3) implementing an intervention; and (4) monitoring progress to determine if the goal should be adjusted (if performance exceeds expectations) or if further instructional changes are needed (Bailey & Weingarten, 2019). Instructors generally make decisions based on four consecutive CBM (Fuchs et al., 2014). Instructors using the process are advised to administer CBM on a weekly or bi-weekly basis. Approaches to instruction featuring CBM and DBI are increasingly recommended for children with intellectual and developmental disabilities (Lemons et al., 2019). Evidence regarding the application of DBI among individuals with DS remains limited, however. Of the 14 DBI studies assessed by Jung et al. (2018), none involved children with moderate ID or DS. Practitioners hoping to extend DBI to these populations consequently have little guidance from the research.

Likewise, reading intervention studies involving children with DS rarely feature CBM as outcomes measures (Dessemontent et al., 2019). Lemons et al. (e.g., 2017, 2018) evaluated the effect of a reading intervention on the early literacy skills of children with DS and moderate ID ($N = 13$) using single-case design and intervention-aligned measures exclusively. Results suggested 77% of participants ($n = 10$) responded to instruction. In a study involving the instruction of digraphs and blends for children with DS ($N = 4$), King et al. (2020) administered intervention-aligned measures and CBM. The intervention produced moderate baseline corrected Tau effects (Tau_{bc}; Tarlow, 2017) on intervention-aligned measures ($M = 0.54$; Range $= 0.34$–0.76). Performance on CBM was mixed, however, with moderate gains in letter-sound fluency (LSF; $M = 0.59$; range $= 0.51$–0.66) offset by minimal response on word identification fluency (WIF; $M = -\ 0.14$; range $= -\ 0.66$–0.38), oral reading fluency (ORF; $M = 0.03$; range $= -0.24$–0.27) and first sound fluency (FSF; $M = -0.33$; range $= -0.75$–0.41).

Integrating CBM into reading intervention studies for children with DS has the potential to provide insight into variables that moderate responsiveness to reading instruction (Lemons et al., 2013). The use of CBM could also provide practitioners and researchers with insight into how interventions for individuals with DS might perform in the context of MTSS and DBI. Typical single-case design studies often feature few participants, however, and were originally designed to accommodate behaviors that immediately respond to changes in contingencies (Hurtado-Parrado & Lopez-Lopez, 2015). Previous studies suggest relatively large populations of participants with DS and the ability to analyze delayed gains may be needed to explore responsiveness to reading intervention on CBM measures and their relationship with intervention-aligned measures (Lemons & Fuchs, 2010).

Integrating CBM into reading intervention studies for children with DS has the potential to provide insight into variables that moderate responsiveness to reading instruction.

Using rigorous, experimental single-case research designs, our research team has demonstrated functional relations between reading intervention and intervention-aligned measures of early reading skills across numerous studies. The purpose of the current work is to explore whether CBM also reflects improvements in early reading skills following the same intervention. Specifically, we examine the performance of a relatively large number of children with DS and moderate ID ($N = 17$) who generally exhibited gains on intervention-aligned measures in previous studies (e.g., Lemons et al., 2017, 2018) using concurrently collected literacy CBM. We then examine the effect of intervention using multi-level modeling. Specific research questions include: (1) What is the overall effect of instruction on distal measures of fluency related to literacy (e.g., LSF, ORF); (2) to what extent do specific participant (e.g., age, IQ) or treatment (e.g., number of sessions) variables moderate the effect of the intervention; and (3) what is the efficacy of intervention relative to objectives calculated using data-based procedures associated with DBI (e.g., Bailey and Weingarten, 2019)?

Method

Participants and Setting

Participants with DS were recruited from public and private schools in Tennessee and Pennsylvania following approval by school district and university Institutional Review Boards. Screening procedures were designed to ensure participants had the skills needed to benefit from beginning reading instruction roughly aligned with K-1 content without also demonstrating mastery of the material prior to intervention. Eligible individuals (a) had DS; (b) primarily communicated using spoken English; (c) exhibited minimal challenging behavior; and (d) demonstrated single letter-sound correspondence. Excluded children (a) had hearing or visual impairments; (b) read more than 20 words correct on two first-grade reading passages selected from the *Dynamic Indicators of Basic Early Literacy* (DIBELS) *Next* (Good et al., 2013); or (c) demonstrated knowledge $\geq$ 50% of the intervention scope and sequence during screening. We enrolled 17 children with DS and cooperating staff in the study. After receiving training from the research team, school staff administered instruction in a one to one arrangement at convenient times and locations in the participants' schools. Lemons et al. (2017, 2018) fully describe screening, researcher qualifications, and instructor training (see, too, King et al., 2020).

A description of the characteristics of participating children and school staff, as well as their involvement in previous studies, appears in Table 1. Data for 13 of participants' performance on intervention-aligned measures are included in published manuscripts (Lemons et al., 2017, 2018). The unpublished data for the additional four participants appears in a supplementary file (Supplemental Figure S1). The total sample included 11 boys and six girls with an average age of 9.34 years (SD = 3.17; range = 6–17 years). Average experience of school staff was 7.5 years (SD = 8.26; range = 1–35). Participants' typical special educators or paraprofessionals administered the intervention. Terminal degrees for staff included MA (n = 9), BA (n = 5), and high school diplomas (n = 3).

Design

The progression of participants through the intervention, described in previous work

Table 1. Summary of Participant and Interventionist Characteristics.

Participants	Demographics Gdr/Age/Grd	Intelligence[e]			Reading[f]				Intervention		Instructor Role/Edu/Exp
		Vrb	NVrb	CX	LID	WID	WATK	Comp	Sess/Wks	Mastery	
Jack[a]	M/6/1	9	0	41	0	0	0	0	41/12	1	SE/MA/35
Anna[a]	F/6/K	8	9	56	14	4	0	2	39/11[g]	8	Para/BA/9
Lilli[a]	F/7/1	18	2	51	13	4	2	1	30/12[g]	8	SE/MA/7
Alex[a]	M/8/1	14	9	48	14	1	0	2	29/14[g]	0	Para/AA/3
Miguel[a]	M/8/2	6	6	40	9	0	0	2	64/10	8	SE/MA/5
Craig[a]	M/8/K	27	15	67	16	2	0	3	26/8	8	Para/HS/3
Robert[a]	M/8/2	7	1	40	16	3	0	3	27/5[g]	8	SE/BA/5
Bruce[b]	M/7/2	26	15	74	16	6	1	2	30/9	8	SE/MA/4
Diana[b]	F/7/K	8	6	45	11	0	0	2	23/6	3	SE/MA/14
Hal[b]	M/8/2	7	0	40	0	0	0	0	47/14	0	SE/MA/4
Jason[b]	M/8/2	7	1	40	9	4	0	0	46/14	0	Para/HS/2
Clark[b]	M/9/3	23	13	54	14	8	0	1	20/8	6	SE/MA/9
Arthur[b]	M/10/2	4	3	40	12	1	1	2	59/15[g]	8	SE/MA/5
Raven[c]	F/12/7	20	17	44	15	8	0	2	36/10[g]	7	SE/MA/17
Eddie[c]	M/13/7	16	14	40	13	1	0	2	29/10	8	SE/BA/4
Julia[c]	F/15/8	34	11	43	17	3	0	0	27/10[g]	8	SE/BA/1
Nina[c,d]	F/17/8	12	12	40	12	7	1	2	37/12[g]	8	SA/BA/2

Note. Ps = Participants; CX = Composite; LID = Letter identification; WID = Word identification; WATK = Word attack; Comp = comprehension; Sess = intervention sessions; Wks = Weekly probes during intervention; Mast = Lessons mastered; SE = Special educator; Para = Paraprofessional; Edu = Education; Exp = Years of experience. Italicized names received services in Pennsylvania; others received services in Tennessee. Age reported in years.

[a]Mastery data appears in Lemons et al., 2017.

[b]Mastery data appears in Lemons et al., 2018.

[c]Mastery data unpublished.

[d]Student attended private school.

[e]Intelligence scores derived from The Kaufman Brief Intelligence Test, 2nd Ed.

[f]Reading ability scores derived from Woodcock Johnson Test of Achievement, 3rd Ed.

[g]Student received additional 15-min oral reading of intervention not required for all students.

(Lemons et al., 2017, 2018) was determined through the administration of intervention-aligned measures corresponding with staggered elements of the scope and sequence. That is, progress on all content was measured over the course of the study, but instruction related to a new lesson was introduced only after the mastery of an earlier lesson. However, we also staggered implementation of intervention across participants in order to evaluate the global impact of the intervention on CBM. Multiple-probe across tasks designs featured in previous research were embedded within the current delayed multiple-probe across participants design. The embedded design concurrently arranges separate experimental conditions to examine the impact of an intervention on different variables resistant to simultaneous examination in a single or combined design (e.g., King et al., 2020).

The delayed multiple-probe across participants design in which (a) the intervention is delayed across participants and (b) the administration of baseline assessment generally overlaps, but does not begin at the same time for all participants (Cooper et al., 2020), addresses threats to internal validity in the same fashion as traditional multiple-probe designs (Ledford & Gast, 2018). Although not randomized, participants' entry into the intervention and movement through each lesson occurred independently of responding on CBM. That is, participants began intervention based on the availability of cooperating instructors, and instructional decisions and outcome evaluations were based on intervention-aligned measures rather than CBM. This provides a measure of protection against threats to internal validity associated with traditional response-guided designs (e.g., experimenter bias; Hwang et al., 2018) as there was no relation between the measures and experimental procedures. The dataset associated with the current study describes changes in participants' CBM and is available online (King et al., 2021).

Measures

Participant characteristics and intervention progress. In addition to the demographic variables of participants (i.e., age, gender), we obtained the education level of cooperating instructors (i.e., high school, BA, MA). We also recorded information regarding the participants' engagement with the intervention, including the number of sessions completed and exposure to optional lesson components. The number of lessons mastered over the course of the intervention provides an indication of participants' success on intervention-aligned measures. Specifically, we determined mastery using Lesson Mastery Probes administered during each intervention session. Participants who vocally identified seven of eight lesson targets for three consecutive sessions proceeded to the next lesson (Lemons et al., 2017, 2018).

Reading ability. We administered a pre-assessment literacy battery prior to intervention. Project staff assessed the reading ability of participants using subtests of the *Woodcock Johnson Reading Mastery Test* ([WJMT] 3rd Ed.) and obtained raw scores for participants in Word Identification (WID; $M = 3.06$; $SD = 2.82$; range = 0–8), Word Attack (WAT; $M = 0.29$; $SD = 0.59$; range = 0–2), Letter Identification (LID; $M = 11.82$; $SD = 5.02$; range = 0–17), and Passage Comprehension (PCM; $M = 1.53$; $SD = 1$; range = 0–3). Average internal consistency of the WRMT-3 is 0.91 (range = 0.68–0.98), and split-half reliability is 0.95 (range = 0.87–0.98).

Cognitive ability. We administered the Verbal Knowledge, Riddles, and Matrices subtests of the *Kaufman Brief Intelligence Test* (2nd Ed. [KBIT-2]; Kaufman & Kaufman, 2004), an individually administered IQ assessment. Assessments yielded verbal ($M = 14.47$; $SD = 8.87$; range = 4–34), nonverbal ($M = 7.88$; $SD = 5.92$; range = 0–17), and full-scale IQ scores (FSIQ; $M = 47.24$; $SD = 10.26$; range = 40–74). The FSIQ from the KBIT-2 is correlated with FSIQ from the *Wechsler Intelligence Scale for Children* (3rd Ed.; 0.76; Wechsler, 1991).

Dependent Variables

Immediately following enrollment, research staff assessed participants once per week during baseline and intervention with CBM. Staff administered the first baseline CBM a minimum of 3 weeks prior to the intervention. The final baseline probe occurred prior to the administration of three consecutive Lesson Mastery Probes and consisted of the children identifying lesson targets presented on flash-cards (Mastery Probe results appear in Lemons et al., 2017, 2018; previously unpublished mastery results appear in Supplemental Figure S1). Intervention did not occur on CBM administration days.

Each CBM was aligned with K-1 grade content and assessed the participants' performance of a specific skill within a 1-min period. For each variable, we recorded the number of correct and incorrect responses participants exhibited, for a total of eight outcomes. Specific procedures, including discontinuation rules, were derived from DIBELS (Good et al., 2013). Adapted from the letter naming fluency assessment, the LSF assessment required participants to provide sounds corresponding with a random assortment of letters. The WIF assessment required participants to read unconnected words from forms encompassing equivalent proportions of words targeted in Lessons 1–4 and Lessons 5–8, as well as untaught decodable words featuring targeted letter sounds. The ORF test assessed the ability to read connected text. For the FSF assessment, children produced the sounds of words presented by the administrator and received full credit (2 points) or partial credit (1 point) for identified first sounds.

Validity and reliability. A panel of experts in reading and language ($n = 3$), as well as the instruction of individuals with developmental disabilities ($n = 2$), verified the content validity of all CBM measures prior to the beginning of the study. Extensive detail regarding the technical adequacy of DIBELS measures for the general population, as well as measures similar to the LSF and WIF, is available (e.g., Anderson et al., 2012; Good et al., 2013; Zumeta et al., 2012). Recent scholarship attests to the suitability of CBM for older individuals with ID (e.g., Hosp et al., 2014). An analysis involving a sample of children with intellectual disabilities ($n = 74$) likewise revealed strong correlations (see Good et al., 2013 for correlation interpretations) between CBM (e.g., WIF, ORF, LSF, phoneme segmentation fluency) and standardized literacy tests (e.g., Rodgers et al., 2021). For the current sample, we correlated average correct responses on baseline CBM scores with subtests from the WJMT. Results should be interpreted with caution due to the sample size; nonetheless, correlations support the validity of CBM. The LID subtest exhibited moderate-to-strong correlations with LSF (0.52), WIF (0.45), and ORF (0.38). We observed similar correlations between the WID subtest and WIF (0.74), ORF (0.60), and LSF (0.37). WAT was moderately correlated with FSF (0.34). We assessed reliability by correlating the first two data points of each CBM. Reliability for LSF (0.95), FSF (0.81), WIF (0.84), and ORF (0.81) was acceptable.

Baseline

Participants received their typical reading instruction throughout the study. Teacher reports, instructional documentation, and researcher observations regarding typical instructional practices prior to the study suggested participants received an average of 110 min of daily reading instruction outside of the intervention ($SD = 60$; range $= 30$–270). This included an average of 65 min of instruction in special education ($SD = 50$; range $=$ 0–180) and 45 min in a general education setting ($SD = 60$; range $= 0$–90). Special education typically involved small group or individualized instruction. General education instruction primarily consisted of whole-group instruction or small group instruction. With the exceptions of Eddie, Bruce, and Clark, all participants received some form of standardized reading curriculum (e.g., *Harcourt Storytown*). In terms of content, instruction for older participants (e.g., Eddie) focused on text reading and comprehension, and instruction for younger children focused on phonics, alphabetic knowledge, and word study.

Intervention

The intervention consisted of four primary instructional components requiring 20 min of instructional time and an optional two steps instructors could choose to deliver over an additional 15 min. The number of intervention sessions and participant involvement in the optional instruction appears in Table 1. The number of intervention sessions received by participants differed due to student absences, the point of the year in which students entered the intervention, and the speed at which students met mastery criterion for intervention-aligned measures (i.e., a score ≥ 87.5% for three consecutive sessions). Lemons et al. (2017, 2018; see, too, King et al., 2020) provide details regarding the instructional protocol (Supplemental Table S1) and the scope and sequence (Supplemental Table S2). The scope and sequence were divided into eight lessons, which participants moved through based on intervention-aligned measures (i.e., Lesson Mastery Probes).

Primary lesson activities addressed decodable word and letter sound acquisition, phonics, and the recognition of high-frequency words. In addition to words targeted for mastery, each lesson incorporated partner words (i.e., shared target sounds) and vocabulary words (e.g., prepositions) that were not directly assessed during the intervention. Optional activities involved comprehension of adjectives and prepositions as well as reading connected text and writing. From the outset of the study, cooperating instructors were permitted to omit optional activities due to scheduling conflicts, the lack of instructionally aligned measures related to the activities, and instructor views regarding the relevance of the activities to each learner's academic objectives. Instructional procedures consisted of presenting material (e.g., picture card, word card) and a model prior to an instructional directive. Instructors praised correct responses. Incorrect responses and non-responses were followed by a model. All activities included lesson targets and randomly selected mastered items to provide opportunities for success.

Fidelity

Project staff evaluated fidelity of instructors using a checklist corresponding with the administration of instruction, materials, error correction, and other lesson components (see Lemons et al., 2017, 2018). Observations encompassed the first three intervention sessions and continued until fidelity exceeded 90% for three consecutive days. After instructors met criterion, project staff obtained additional fidelity data for one session each week and provided additional training in the event that fidelity fell below 90%. Fidelity was obtained for an average of 32% of sessions with a range of 54%–100% across sessions. Low scores on fidelity partially reflect instructor performance during the initial sessions prior to when cooperating educational personnel met criterion. In addition, personnel with limited experience or background in education (e.g., paraprofessionals) frequently required additional training sessions. Aggregate fidelity for individual instructors ranged from 84%–99%, with an average score of 92.5% across instructors. Fidelity scores for individual instructors appear in Supplemental Table S3.

Interobserver Agreement

Two MA-level students in special education obtained interobserver agreement (IOA) from video recordings of 60% of baseline and 21% of intervention CBM. Agreement was defined as an item for which scorers observed the same response and was calculated by dividing the number of agreements by the number of items. Average aggregate IOA for each participant was 93.8% ($SD = 3.7$; range = 87–99%). IOA for each participant appears in Supplemental Table S3.

Data Analysis

Visual analysis. We determined the responsiveness of participants to instruction for each dependent measure through visual analysis. Procedures followed the steps outlined by Lane and Gast (2014). Within condition analyses assessed stability and changes in level and trend via the split-middle method. We defined stability criterion as 80% of data points within ± 25% of the median. Between condition analyses involved evaluating (a) changes in trend, and (b) immediate and sustained changes in level. We determined the non-overlap in data-points for each contrast as well as a weighted mean for each measure using Tau_{bc}, an estimate of overlap ranging between −1 and 1 (Tarlow, 2017). Tau_{bc} values were interpreted as very large (> 0.8), large (0.61–0.8), moderate (0.21–0.6), and negligible (0–0.2). We defined responders as participants who exhibited a change in level immediately following the intervention or whose data assumed a positive trend inconsistent with baseline within 3 weeks of the intervention (Bruhn et al., 2020).

Multilevel modeling. For the primary analysis, we opted to aggregate and statistically analyze the data due to (a) the limited sensitivity of CBM to short-term gains, and (b) the use of instruction consisting of discrete, sequentially administered lessons. We initially considered analyzing the data using the gradual effects model for single-case designs (Swan & Pustejovsky, 2018). However, this approach presents multiple issues related to accuracy when applied to data featuring low levels of baseline responding. Data were therefore analyzed using multilevel modeling, which accounts for dependency and autocorrelation in nested data and is therefore well suited for single-case design (Ferron et al., 2009).

Multilevel modeling has recently been applied to single-case design (see Shadish et al., 2013) because these designs include multiple measurements across time nested within individuals (Ferron et al., 2009; Shadish et al., 2013). In addition to accounting for nested, dependent data, multilevel modeling allows researchers to quantify critical components of traditional visual analysis (e.g., trends in baseline, changes in level and slope) and examine the influence of moderators (Shadish et al., 2013). In the present study, we applied a series of two-level models wherein measurement sessions (Level 1) were nested within individuals (Level 2):

$$Y_{ij} = \beta_{0j} + \beta_{1j}Phase_i + \beta_{2j}Phase_i * Time_{ij} + \beta_{3j}Covariates + e_{ij}$$

In this model, Y is a child-level outcome (e.g., ORF corrects) for the ith child across j measurement periods. The intercept is represented by β_{0j}. The next coefficient (β_{2j}) is a dichotomous variable called *phase*, which indicates whether a participant's data comes from baseline or intervention phases. The third coefficient (β_{2j}) is an interaction term between the previously mentioned *phase* variable and a continuous *time* variable centered on when an individual began intervention (i.e., *time* = 0 on the first instructional session, *time* = 1 on the second instructional session). The *phase* and *phase x time* variables yield values that correspond to significant changes in level and trend, respectively. The final coefficient (β_{3j}) represents key covariates introduced into the models, described in more detail below.

To capture variability between cases, we allowed the phase variable to vary at the second level of the model, meaning participants were allowed to vary with respect to their immediate treatment effect (Rodadaugh & Moeyaert, 2017). The random effects can be presented as:

$$\beta_{1j} = \theta_{10} + \mu_{1j}$$

Where overall performance (β_{1j}) are comprised of individual-level performances θ_{10} plus random variation of means (μ_{1j}).

The use of multilevel modeling does not necessitate the determination of specific outcome-covariate combinations a priori; the adequacy of various models can be compared and the best model identified. We therefore compared a series of models for each of the eight outcomes and selected the best model for each measure. All comparison models included the two key variables phase and phase x time, described above, and we report the resulting coefficients for each model regardless of statistical significance given their importance to single-case design data. In addition to phase and phase x time, comparison models incorporated some combination of the following covariates: age; sex; verbal IQ; nonverbal IQ; FSIQ; LID; WAT; PCM; the number of intervention sessions and lessons mastered; whether the participant received an optional intervention; and instructors' level of education and years of experience. As a measure of phonological awareness, we also included the baseline phase median FSF for all models except when FSF corrects and incorrects were the primary outcomes. These combinations included two- and three-way interactions between variables. When evaluating significance of the covariates, we used the Kenward-Roger (1997) method for estimating the degrees of freedom, which adjusts for small sample bias and is a more accurate method of estimating treatment relative to traditional approaches (Ferron et al., 2009).

The use of multilevel modeling does not necessitate the determination of specific outcome-covariate combinations a priori; the adequacy of various models can be compared and the best model identified.

We used the corrected Akaike Information Criterion which adjusts for small sample bias (AICc; Hurvich & Tsai, 1989) to identify the best model for each measure. AICc is a relative comparison index used to rank competing models, maximize fit, and minimize information loss (Symonds & Moussalli, 2010). We also report the corresponding AICc cumulative weight, which presents the probability that the selected model is more appropriate than the preceding model.

We analyzed data using analysis RStudio 3.6.0 (R Core Team, 2019), with the *lme4* (Bates et al., 2015), *AICcmodavg* (Mazerolle, 2019), and *sjPlot* (Lüdecke, 2020) packages. Prior to analysis, continuous data were evaluated for multicollinearity. All correlations were below 0.70 (Supplemental Table S4). There were 34 instances of missing outcome data across four participants. The *lme4* package, by default, removes observations with missing data which is acceptable given the limited instances of missing data and the lack of support for methods of multiple imputation for single-case designs (Lüdtke et al., 2017). Once the best model was chosen for each outcome, we evaluated model linearity and residuals. Overall, there were no concerns with linearity. The residuals for several models (e.g., FSF Incorrects, WIF Corrects) indicated that the models struggled to predict outcomes at extreme values. This may be due to the low scores of several participants on most assessments, as such variance can impact model residuals (Baek & Ferron, 2013). All data and code associated with the final model appear online (King et al., 2021).

Social Validity

Cooperating instructors completed a survey comprised of 11 items for the purposes of providing a subjective measure of the skill importance (one item), intervention acceptability (i.e., whether interventionists were comfortable with the intervention; six items), and effectiveness of the intervention (four items) using a 6-point Likert-type scale (Supplemental Table S5). To provide an objective measure of social validity, we compared correct responses on all measures to goal lines calculated using a procedure adapted from the National Center on Response to Intervention (2012) and the National Center for Intensive Intervention (Bailey & Weingarten, 2019). When individualizing instruction, an intra-individual goal-setting framework is advised for children who require intensive academic supports. This process entailed (1) determining the baseline rate of improvement (ROI) for each child by finding the midpoint for each baseline, subtracting the median of the first half of baseline from the median of the second-half, and dividing by the number of datapoints; (2) multiplying ROI by a desirable rate of change (1.5) and the number of weeks the child received intervention; and (3) adding the product to the mean of the last three baseline datapoints. For children who exhibited no growth during baseline, we used an arbitrary ROI that assumed an increase in performance by one item per week. Figure 1 presents the goals for each child as lines extending from baseline to the final treatment data point. Goals could not exceed the typical first-grade benchmark of 60 correct items per min.

Results

On average, participants received 35.41 intervention sessions ($SD = 12.32$; range = 20–64). Intervention sessions were delivered

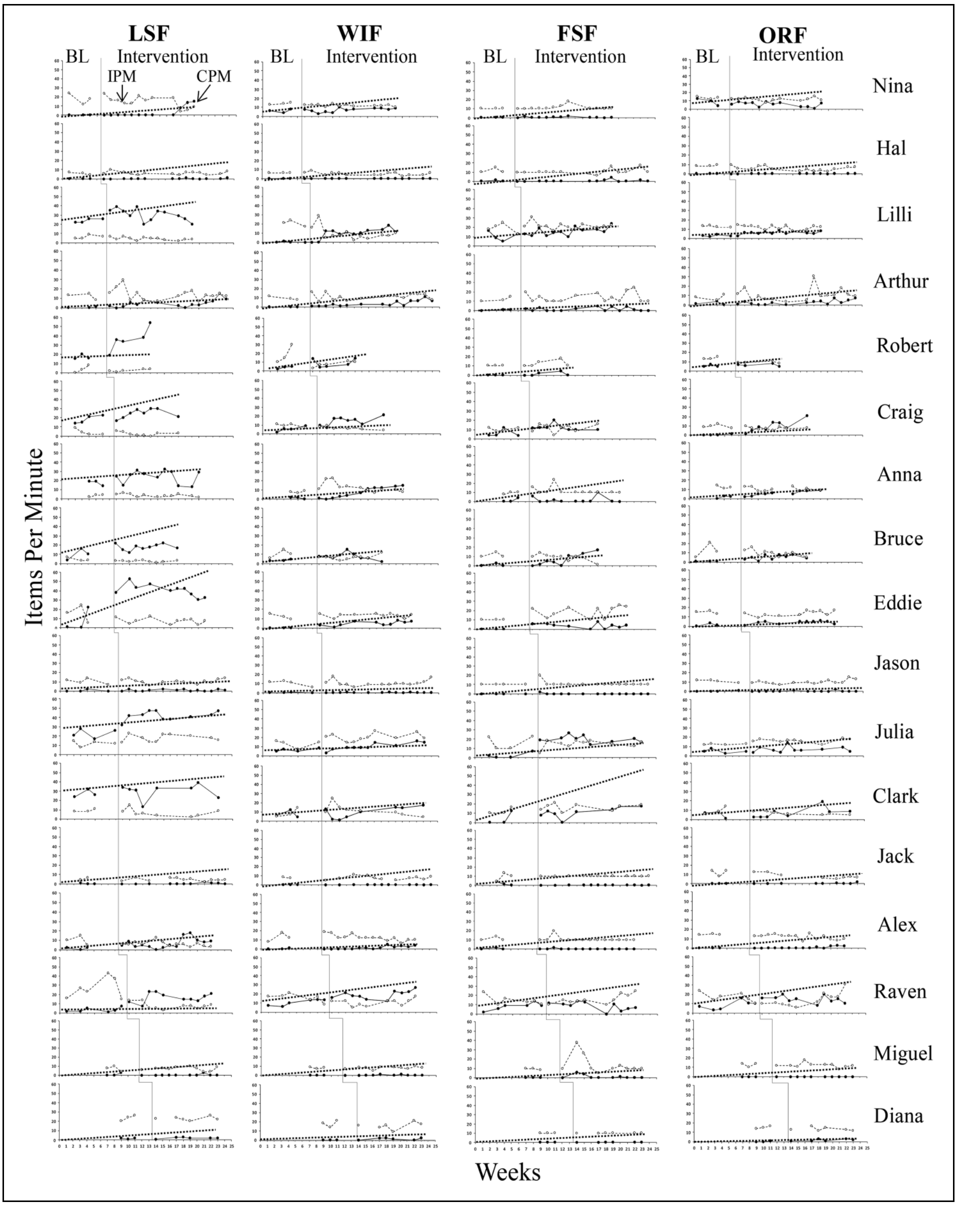

Figure 1. Participant CBM Performance.
Note. Participants arranged based on start of intervention. Y-axis values range from 0–60; x-axis values range from 1–25. Closed circles = correct per min (CPM); Open circles = incorrect per min (IPM). LSF = Letter sound fluency; WIF = Word identification fluency; FSF = first sound fluency; ORF = Oral reading fluency. Dashed lines represent post-hoc individualized goals for correct responses based on performance in baseline.

over an average of 10.59 weeks (SD = 2.85; range = 5–15). Participants mastered an average of 5.71 lessons (SD = 3.37; range = 0–8). Descriptive statistics for dependent measures appear in Table 2. Figure 1 graphically depicts study outcomes.

Visual analysis. For measures of correct items per min, 35% of participants (n = 6) exhibited data patterns consistent with response to instruction for LSF and ORF. Responsiveness was observed in approximately 29% of participants (n = 5) on WIF and FSF measures. Forty-one

Table 2. Outcomes for Multilevel Models.

Variable	Mean BL (SD/R)	Mean Int (SD/R)	Est	SE	*p* value	AICc	Cum Wt
LSF Corrects	9.22 (10.28/0–32)	14.67 (15.34/0–54)				1430.75	0.19
Intercept			14.46	9.15	.126		
Phase*			7.06	2.35	.007		
Phase x Time			0.19	0.12	.110		
Verbal IQ*			0.74	.026	.009		
LSF Incorrects	11.25 (8.72/0–43)	8.39 (6.03/0–29)				1287.76	0.75
Intercept			119.07	55.00	.058		
Phase			−0.87	1.52	.572		
Phase x Time			−0.33	0.31	.284		
Letter ID*			2.38	0.93	.030		
Sex (Male)*			−12.67	3.96	.010		
Verbal IQ*			1.61	0.98	.162		
Nonverbal IQ*			−3.84	1.61	.041		
Passage Comp*			−11.92	4.37	.023		
Instructor Education (MA)*			16.21	5.02	.010		
FSF Baseline Median*			−4.26	1.87	.048		
Nonverbal IQ x Passage Comp*			2.64	0.93	.019		
WIF Corrects	3.10 (3.97/0–14)	5.66 (6.23/0–27)				1086.67	0.35
Intercept*			−32.58	8.15	.012		
Phase			1.15	0.74	.162		
Phase x Time			0.40	0.21	.059		
Lsns Mastered*			1.08	0.17	< .001		
Age*			1.47	0.37	.014		
Full Scale IQ*			0.43	0.11	.014		
Nonverbal IQ*			−0.73	0.19	.014		
Word ID*			0.56	0.19	.034		
Word Attack**			−6.49	1.09	.003		
FSF Baseline Median**			0.85	0.12	.004		
WIF Incorrects	12.14 (5.23/5–30)	10.60 (4.96/2–29)				1255.35	0.15
Intercept			19.66	14.41	.246		
Phase			−0.45	1.35	.747		
Time x Phase			−0.40	0.31	.203		
FSF Corrects	2.65 (4.27/0–16)	4.59 (6.67/0–26)				1181.11	0.13
Intercept			−0.37	1.17	.755		
Phase*			2.86	1.25	.033		
Time x Phase			0.05	0.07	.491		
FSF Incorrects	11.93 (4.15/4–25)	13.09 (5.34/1–38)				1299.85	0.25
Intercept*			37.31	7.87	.001		
Phase			1.63	0.99	.113		
Phase x Time			−0.03	0.10	.738		
Full Scale IQ*			−0.52	0.14	.004		
Verbal IQ*			0.38	0.16	.031		
Word Attack*			4.30	1.22	.004		
Age*			−1.18	0.39	.012		

(continued)

Table 2. (continued)

Variable	Mean BL (SD/R)	Mean Int (SD/R)	Est	SE	p value	AICc	Cum Wt
ORF Corrects	2.85 (3.72/0–16)	4.14 (4.60/0–21)				1072.42	0.48
Intercept			−22.66	12.34	.076		
Phase			1.09	0.78	.403		
Time x Phase			0.08	0.20	.668		
Lsns Mastered*			0.55	0.16	.002		
Word Attack*			−3.58	1.68	.044		
FSF Baseline Median*			0.48	0.22	.040		
ORF Incorrects	12.53 (3.60/5–24)	10.88 (4.32/3–31)				1197.34	0.25
Intercept*			23.37	6.90	.002		
Phase			−1.36	0.70	.061		
Time x Phase			0.01	0.08	.920		
Full Scale IQ*			−0.26	0.12	.044		
Nonverbal IQ*			0.45	0.22	.043		
Letter ID*			0.43	0.21	.046		

Note. BL = baseline; Int = intervention; R = range; SE = standard error; Est = estimate; AICc = Akaike Information Criterion for small sample size; Cum Wt = cumulative weight for AICc; Lsns = lessons; Sess = sessions; Comp = comprehension; LSF = Letter Sound Fluency; WIF = Word Identification Fluency; FSF = First Sound Fluency; ORF = Oral Reading Fluency; Phase = change in level after intervention; Time x Phase = intervention trend; Letter ID = subtest from WRMT-3; Word ID = subtest from WRMT-3.
*significant at or below 0.05.

percent of students (n = 7) did not exhibit progress consistent with response to intervention on any measure of correct response (see Supplemental Table S7 for details). Weighted Tau_{bc} for LSF (M = 0.23; SD = 0.23; range = −0.23–0.66), WIF (M = 0.33; SD = 0.23; range = −0.03–0.60), and ORF (M = 0.27; SD = 0.27; range = −0.29–0.64) was indicative of moderate nonoverlap. Tau_{bc} for FSF (M = 0.16; SD = 0.32; range = −0.72–0.62) was consistent with negligible nonoverlap between conditions. Fewer participants exhibited reduced errors on measures of LSF (29%; n = 5), WIF (18%; n = 3), and ORF (24%; n = 4). No effect was observed for decreasing FSF errors. Sixty-five percent of students (n = 11) did not exhibit progress consistent with response to intervention on any measure of incorrect response (Supplemental Table S7). Tau_{bc} was consistent with a moderate nonoverlap for ORF (M = −0.22; SD = 0.24; range = −0.71–0.54). Negligible nonoverlap was observed for LSF (M = −0.15; SD = 0.23; range = −0.68–0.24), WIF (M = −0.09; SD = 0.31; range = −0.51–0.48), and FSF (M = 0.11; SD = 0.22; range = −0.25–0.48).

Multilevel modeling. The best model for each outcome, the AICc values, and the AICc cumulative weights are reported in a supplemental file (Supplemental Table S6). Full models for each of the eight outcomes showed considerable variability with respect to included covariates, with some models (e.g., FSF and LSF corrects) including few predictors beyond the phase and phase x time variables and other models (e.g., ORF corrects) including up to 16 predictors. AICc values ranged from 1072.42 to 1430.75, with cumulative weights ranging from 0.13–0.48.

Table 2 is a truncated version of the eight full models, reporting the AICc and cumulative weights for each model, the coefficients for phase (i.e., level) and phase x time (i.e., trend), and any significant covariates. The models revealed significant immediate changes in level for LSF and FSF corrects. Participants earned, on average, 7.06 more LSF corrects and 2.86 more FSF corrects immediately following baseline. The results for trend were not significant.

Several variables appeared as significant covariates in at least three models. All IQ variables were significant predictors in a number of models. FSIQ was a significant predictor in WIF corrects, FSF incorrects, and ORF incorrects. Higher FSIQ scores were associated with more corrects (0.43 for WIF) or fewer incorrects (−0.52 for FSF and −0.26 for ORF). The number of mastered lessons appeared in WIF and ORF corrects and was statistically significant. WAT was a significant predictor in the WIF corrects, FSF incorrects, and ORF corrects models. Higher WAT scores were associated with lower numbers of corrects (−6.49 for WIF and −3.58 for ORF) and higher number of incorrects on outcome measures (4.30 on FSF). Participants' median baseline FSF performance was a significant predictor in LSF incorrects, WIF corrects, and ORF corrects. Higher FSF performance was associated with more corrects (0.85 for WIF and 0.48 for ORF) and fewer incorrects (−4.26 on LSF). Two variables (intervention sessions, the optional intervention) did not appear as a significant predictor in the top eight models.

Social Validity

Instructors indicated the intervention generally targeted important skills ($M = 5.71$; range = 3–6), resulted in positive changes in the participants' reading ($M = 5.24$; range = 2–6), and was acceptable ($M = 5.14$; range = 1–6). Scores regarding the time required for implementation suggest that some instructors encountered difficulty in scheduling the intervention ($M = 3.88$; range = 1–6). Participant performance relative to data-derived objectives varied across measures (Figure 1). For LSF, 29% of participants ($n = 5$) met or exceeded their individualized goals. Fewer participants met targets in WIF (18%; $n = 3$), FSF or ORF (18%). Thirty-five percent of participants ($n = 6$) did not meet data-derived goals on a single measure (Supplemental Table S7).

Discussion

This study examined the effect of a literacy intervention on CBM administered to children with DS, many of whom exhibited FSIQ scores consistent with moderate ID, and who improved performance on intervention-aligned measures associated with the same intervention. Results revealed inconsistent gains across domains, with small immediate improvements relegated to correct LSF and FSF responses. Findings further suggest the trend of correct responses following intervention, though approaching significance on WIF, did not improve. Covariates associated with improved performance varied across measures, with FSF, cognitive ability and lessons mastered appearing across multiple measures. Visual analysis suggested a minority of participants exhibited performance consistent with response to intervention. Although consumers reported high levels of satisfaction, few participants met goals derived through the intra-individual goal-setting framework (Bailey & Weingarten, 2019).

Changes in LSF and FSF were uncharacteristically immediate relative to the more gradual gains observed in other studies (e.g., Lemons and Fuchs, 2010). Notwithstanding the performance of individual students, most models did not show a significant change at the onset of intervention, and only WIF correct responses approached significance in terms of trend. This result may be due to the relatively brief duration of instruction, limited growth for children with moderate ID or DS on CBM (e.g., Allor et al., 2010; Lemons & Fuchs, 2010; Lemons et al., 2013), or the minimal opportunities for engagement with full text. Additionally, the scope and sequence intentionally focused on foundational skills including common sounds for single letter and decoding for VC and CVC words. Consequently, the absence of significant changes in ORF—a skill not targeted by the primary intervention components—could be anticipated. The limited performance relative to recent intervention studies (e.g., Lim et al., 2019) may also be due to sample composition, as the current study intentionally selected children with limited literacy skills, most of whom exhibited moderate ID. Challenges posed to multilevel models by extremely low measurement values may also have influenced outcomes. That performance across CBM did not accord with intervention-aligned measures suggesting most participants (76%, $n = 13$) mastered content (Table 1) provides further evidence of the divergence between intervention-aligned and distal measures (e.g., Dessemontet et al., 2019).

While certainly not dispositive, the current study contributes to the discussion surrounding the relevance of predictors to the development of reading in DS. Findings regarding the significant association between FSF and improved performance in LSF, WIF, and ORF affirm earlier research supporting the role of phonological awareness as an instructional target for children with DS (e.g., Dessemontet and de Chambrier, 2015). The restriction of beneficial associations between participant age to correct responses (WIF) and incorrect responses (FSF) provides qualified support regarding the limited influence of age on reading skills relative to children without disabilities (Roch et al., 2019). Results linking IQ to improved performance in LSF, FSF, and ORF are also consistent with previous research (e.g., Roch et al., 2019). Combinations of verbal, nonverbal and FSIQ were also significant predictors for multiple variables (e.g., WIF corrects, ORF incorrects). As verbal IQ reflects prior experience, this suggests the conception of FSIQ as an inherent, immutable trait should not determine whether reading instruction is included in an academic program (Otero, 2017).

Sample-specific idiosyncrasies can result in the identification of significant variables with no bearing on the larger population, as when small samples with high autocorrelation yield narrower confidence intervals than larger samples (Baek & Ferron, 2013). Support for the relationship between educator variables was not consistently observed across measures and strongly contributed to an increase in the number of incorrect responses on the LSF measure. Burroughs et al. (2019) noted the wide range of findings in this regard, and our results support the need for further research. In addition, the apparent relationship between (a) higher IQ and standard reading scores (e.g., WAT, LID), and (b) lower scores on measures of reading ability (e.g., WIF, ORF) may have implications for instructional design and evaluation. The finding regarding WAT, which was significantly associated with poor performance in FSF, WIF, and ORF contributes to contradictory findings regarding the relation of WAT to reading performance (Saunders & Defulio, 2007). At the clinical level, we often see a child's performance plateau or decrease as they begin to read connected text. Children who have specific prerequisite skills may attempt more items unsuccessfully or exhibit just enough correct responses to circumvent discontinuation rules (e.g., King et al., 2020). This might result in more errors relative to children with lower ability. Patterns of correct and incorrect responses on measures potentially provide greater insight into the nature of a child's performance issues. We therefore recommend assessing correct and incorrect responses during instruction.

Support for the relationship between educator variables was not consistently observed across measures and strongly contributed to an increase in the number of incorrect responses on the LSF measure.

Limitations

This study has several notable limitations. WIF measures created specifically for this study included much of the material featured in the scope and sequence. Although not uncommon for CBM, this measure may have been more responsive to intervention as a result. WIF outcomes should therefore be interpreted with caution. Measures included in the model captured a limited dimension of phonological awareness (i.e., FSF). Tasks related to other elements of phonological awareness (e.g., phoneme blending) may have resulted in different findings. Our use of the intra-individual framework was not entirely consistent with prescribed DBI procedures (e.g., number of baseline datapoints; Bailey & Weingarten, 2019). Alterations were consistent with changes that may occur in practical settings, where interventionists must draw conclusions from limited data or develop benchmarks based on near-zero levels of performance. Our analysis is defensible due to its supplementary nature, variability in application likely to occur in practice, and the lack of advice relevant to students with extremely low baseline outcomes.

Implications for Practice

Success on measures aligned with an intervention does not necessarily predict CBM performance, at least across the relatively brief duration of the study. However, the performance of students on intervention-aligned measures (i.e., the number of lessons mastered) significantly predicted WIF and ORF scores. The latter is significant given that ORF was not targeted for most participants. These findings support assertions by Fuchs et al. (2018) regarding the value of intervention-aligned measures. Specifically, measures aligned with an intervention can assist in the development of instructional programs and the determination of correspondence between student learning and instructional objectives. Additionally, teachers are more likely to use intervention-aligned measures to assess student progress. Tasks which provide students with a higher likelihood of success may also decrease motivation challenges for students with DS who struggle on CBM (Grieco et al., 2015).

This study did not evaluate the use of DBI; therefore, comments regarding the application of the procedure in practice should be interpreted cautiously. Regardless, findings do provide some insight into DBI, as typically presented (e.g., Fuchs et al., 2014), for this population. We agree DBI holds promise for students who exhibit limited response to conventional interventions (Fuchs et al., 2020). Nonetheless, we need more research to better understand goal setting and progress evaluation for children with DS and moderate ID (Lemons et al., 2019). Grade level benchmarks would not have been appropriate for most of our participants. Yet, calculated tertiary goals often did not result in usable objectives due to negative or zero-level baseline trends. This limitation of DBI is observed in other methods of analyzing data, which are incompatible with students who exhibit exceptionally low levels of performance (e.g., Swan and Pustejovsky, 2018). Our intervention integrated many components associated with intervention intensity, yet most participants' responding did not match data-based intervention goals.

While previous research suggests the timeline of our intervention may have been insufficient to yield significant progress for many children with DS and moderate ID (e.g., Burgoyne et al., 2012), standard DBI guidelines suggest teachers use CBM to evaluate and potentially alter instruction within a shorter timespan (Fuchs et al., 2020). However, CBM do not appear to capture growth quickly enough with this population for the purposes of rapid instructional modification. Consequently, asking teachers to repeatedly modify interventions after four datapoints below the goal line may not be feasible due to their limited resources. Cooperating instructors identified time as an impediment to using the intervention. Scholars (e.g., Brownell et al., 2010) have also suggested typical preparation programs do not equip instructors with the skills needed to continually modify instruction based on CBM.

Given the dearth of research in this area, we encourage practitioners to flexibly apply DBI and incorporate broader sets of data (e.g., student motivation) into evaluations. There is much research needed in the area of monitoring adequate response to reading interventions for students with ID. Nonetheless, we believe this study supports educators' use of early reading CBM. Reliable, valid indicators of academic progress can provide a useful perspective on the effectiveness of instruction and guide long-term planning. Due to the limited focus on CBM for children with ID, we encourage educators to also monitor progress with intervention-aligned assessments such as mastery of an intervention's scope-and-sequence. Collecting data on student engagement and motivation during reading intervention can also increase the effectiveness of instruction.

There is much research needed in the area of monitoring adequate response to reading interventions for students with ID. Nonetheless, we believe this study supports educators' use of early reading CBM.

Directions for Future Research

We have much to learn about how children with DS respond to intervention. Researchers should perform more assessment-oriented work with this population, using a broad array of distal and proximal measures. Given that special educators are increasingly asked to use DBI and other methods predicated on CBM, use of these measures in research appears to be appropriate. The routine use of CBM would provide some indication of outcomes to be expected in practice and contribute to the social validity of this line of research (Callender et al., 2020). As CBM may not be sufficiently sensitive to improvements or deficits in specific skills (Van Norman et al., 2018), a place remains for intervention-aligned mastery measures in research and practice.

Future research should also consider incorporating alternatives to traditional single-case designs (e.g., Hwang et al., 2018) capable of simultaneously evaluating short- and long-term outcomes. The use of CBM and more rigorous designs may produce less robust findings, however (e.g., Hua et al., 2020). Although new tools capable of estimating minute or gradual changes in single-case design may eliminate some of the ambiguity associated with CBM (Swan & Pustejovsky, 2018), the field may need to become more tolerant of mixed or negative findings derived from otherwise sound studies in the short-term (Kittelman et al., 2018).

It is clear that a great deal of variability exists among individuals with DS. More specific instructional modifications may be more suitable for children with specific characteristics. Future studies that permit moderator analyses would help provide insight into this process and thus help interventionists match instruction and needed adaptations to individual students. As outcomes of high-achieving children with DS may not be representative of the entire population, researchers should provide detailed descriptions of the participants to allow sufficient opportunities to estimate the likelihood of replicating research results among specific populations. This type of focused research has the potential to enhance educators' abilities to individualize and intensify intervention for a broader group of learners and improve outcomes for a greater number of students with disabilities.

References

Ahlgrim-Delzell, L., & Rivera, C. (2015). A content comparison of literacy lessons from 2004 and 2010 for students with moderate and severe intellectual disability. *Exceptionality*, *23*(4), 258–269. https://doi.org/10.1080/09362835.2015.1064417

Allor, J. H., Gifford, D. B., Al Otaiba, S., Miller, S. J., & Cheatham, J. P. (2014). Teaching students with intellectual disability to integrate reading skills: Effects of text and text-based lessons. *Remedial and Special Education*, *34*(6), 346–356. https://doi.org/10.1177/0741932513494020

Allor, J. H., Mathes, P. G., Roberts, J. K., Jones, F. G., & Champlin, T. M. (2010). Teaching students with moderate intellectual disabilities to read: An experimental examination of a comprehensive reading intervention. *Education and Training in Autism and Developmental Disabilities*, *45*(1), 3–22. https://www.jstor.org/stable/23880147

Anderson, D., Lai, C. F., Park, B. J., Alonzo, J., & Tindal, G. (2012). An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 1. *Behavioral Research and Teaching*. https://www.brtprojects.org/

Baek, E. K., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across participant variation in autocorrelation and residual variance. *Behavior Research Methods*, *45*(1), 65–74. http://dx.doi.org/10.3758/s13428-012-0231-z

Bailey, T. R., & Weingarten, Z. (2019). *Strategies for setting high-quality academic individualized education program goals*. Office of Special Education Programs, U.S. Department of Education. https://intensiveintervention.org/resource/high-quality-academic-IEP-goals

Bates, D., Maechler, 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

Bird, E. K. R., Cleave, P. L., & McConnell, L. (2000). Reading and phonological awareness in children with down syndrome: A longitudinal study. *American Journal of Speech-Language Pathology*, *9*(4), 319–330. https://doi-org.proxy.lib.uiowa.edu/10.1044/1058-0360.0904.319

Boat, T.F., & Wu, J. T. (2015). *Mental disorders and disabilities among low-income children*. National Academies Press.

Brownell, M. T., Sindelar, P. T., Kiely, M. T., & Danielson, L. C. (2010). Special education teacher quality and preparation: Exposing foundations, constructing a new model. *Exceptional Children*, *76*(3), 357–377. https://doi.org/10.1177/001440291007600307

Bruhn, A. L., Rila, A., Mahatmya, D., Estrapala, S., & Hendrix, N. (2020). The effects of data-based, individualized interventions for behavior. *Journal of Emotional and Behavioral Disorders*, *28*(1), 3–16. https://doi-org.proxy.lib.uiowa.edu/10.1177/1063426618806279

Burgoyne, K., Duff, F. J., Clarke, P. J., Buckley, S., Snowling, M. J., & Hulme, C. (2012). Efficacy of a reading and language intervention for children with down syndrome: A randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *53*(10), 1044–1053.

Burgoyne, K., Duff, F., Snowling, M., Buckley, S., & Hulme, C. (2013). Training phoneme blending skills in children with down syndrome. *Child Language Teaching and Therapy*, *29*(3), 273–290 https://doi.org/10.1111/j.1469-7610.2012.02557.x

Burroughs, N., Gardner, J., Lee, Y., Gou, S., Touitou, I., Jansen, K., & Schmidt, W. (2019). *Teaching for excellence and equity: Analyzing Teacher Characteristics, Behaviors and Student Outcomes with TIMSS* (Vol. *6*; pp.) Springer.

Callender, A., King, S. A., & Howard, M. (2020). Social validity assessment: Examples from a functional skills training intervention. In J. A. Jaworski (Ed.), *Advances in sociology research* (Vol. 31). Nova Science Publishers.

Clemens, N. H., Lee, K., Henri, M., Simmons, L. E., Kwok, O. M., & Al Otaiba, S. (2020). Growth on sublexical fluency progress monitoring measures in early kindergarten and relations to word reading acquisition. *Journal of School Psychology*, *79*, 43–62 https://doi.org/10.1016/j.jsp.2020.01.003

Cologon, K., Cupples, L., & Wyver, S. (2011). Effects of targeted reading instruction on phonological awareness and phonic decoding in children with Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *116*(2), 111–129 https://doi.org/10.1016/j.jsp.2020.01.003

Cooper, J. O., Heron, W.L., & Heward, W.L. (2020). *Applied behavior analysis* (3rd ed.). Pearson.

Dessemontet, R. S., & de Chambrier, A. F. (2015). The role of phonological awareness and letter-sound knowledge in the reading development of children with intellectual disabilities. *Research in Developmental Disabilities*, *41*, 1–12 https://doi.org/10.1016/j.ridd.2015.04.001

Dessemontet, R. S., Martinet, C., de Chambrier, A. F., Martini-Willemin, B. M., & Audrin, C. (2019). A meta-analysis on the effectiveness of phonics instruction for teaching decoding skills to students with intellectual disabilities. *Educational Research Review*, *26*, 52–70. https://doi.org/10.1016/j.edurev.2019.01.001

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Gobioff, G., & Hibbard, S. T. (2009). Making treatment effect inferences from multiple-baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods*, *41*(2), 372–384. https://doi.org/10.3758/BRM.41.2.372

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

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (2020). Bringing data-based individualization to scale: A call for the next-generation technology of teacher supports. *Journal of Learning Disabilities*. https://doi.org/10.1177/0022219420950654

Fuchs, D., Hendricks, E., Walsh, M. E., Fuchs, L. S., Gilbert, J. K., Zhang Tracy, W., Patton III, S., Davis-Perkins, N., Kim, W., Elleman, A. M., & Peng, P. (2018). Evaluating a multidimensional reading comprehension program and reconsidering the lowly reputation of tests of near-transfer. *Learning Disabilities Research & Practice*, *33*(1), 11–23. https://doi.org/10.1111/ldrp.12162

Grieco, J., Pulsifer, M., Seligsohn, K., Skotko, B., & Schwartz, A. (2015). Down syndrome: Cognitive and behavioral functioning across the lifespan. *American Journal of Medical Genetics-C*, *169*(2), 135–149 https://doi.org/10.1002/ajmg.c.31439

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2013). *DIBELS Next technical manual*. Dynamic Measurement Group.

Hosp, J. L., Hensley, K., Huddle, S. M., & Ford, J. W. (2014). Using curriculum-based measures with postsecondary students with intellectual and developmental disabilities. *Remedial and Special Education*, *35*(4), 247–257. https://doi-org.proxy.lib.uiowa.edu/10.1177/0741932514530572

Hua, Y., Hinzman, M., Yuan, C., & Balint Langel, K. (2020). Comparing the effects of Two reading interventions using a randomized alternating treatment design. *Exceptional Children*, *86*(4), 355–373. https://doi-org.proxy.lib.uiowa.edu/10.1177/0014402919881357

Hurtado-Parrado, C., & López-López, W. (2015). Single-case research methods: history and suitability for a psychological science in need of alternatives. *Integrative Psychological and Behavioral Science*, *49*(3), 323–349.

Hurvich, C. M., & Tsai, C-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307 https://doi.org/10.1093/BIOMET/76.2.297

Hwang, Y., Levin, J. R., & Johnson, E. W. (2018). Pictorial mnemonic-strategy interventions for children with special needs: Illustration of a multiply randomized single-case crossover design. *Developmental Neurorehabilitation*, *21*(4), 223–237. https://doi.org/10.3109/17518423.2015.1100689

Jung, P. G., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis. *Learning Disabilities Research & Practice*, *33*(3), 144–155. https://doi-org.proxy.lib.uiowa.edu/10.1111/ldrp.12172

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test* (2nd ed.). NCS Pearson.

Kenward, M. G., & Roger, J. H. (2017). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. https://doi.org/10.2307/2533558

King, S. A., Lemons, C. J., Davidson, K. A., Fulmer, D., & Mrachko, A. A. (2020). Reading instruction for children with down syndrome: Extending research on behavioral phenotype aligned interventions. *Exceptionality*, 1–17. https://doi.org/10.1080/09362835.2020.1749631

King, S.A., Lemons, C.J., & Rodgers, D. (2021). Curriculum based measures and code for study involving children with down syndrome. [Data set]. https://osf.io/5vbzy/?view_only=2773239290974e98abbbf700c5e0b614

Kittelman, A., Gion, C., Horner, R. H., Levin, J. R., & Kratochwill, T. R. (2018). Establishing journalistic standards for the publication of negative results. *Remedial and Special Education*, *39*(3), 171–176. https://doi.org/10.1177/0741932517745491

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, *24*(3–4), 445–463 https://doi.org/10.1080/09602011.2013.815636

Laws, G., & Gunn, D. (2002). Relationships between reading, phonological skills and language development in individuals with Down syndrome: A five year follow-up study. *Reading and Writing, 15*(5), 527–548. https://doi.org/10.1023/A:1016364126817.

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology: Applications in special education and behavioral sciences*. (3rd ed.) Routledge.

Lemons, C. J., & Fuchs, D. (2010). Modeling response to reading intervention in children with down syndrome: An examination of predictors of differential growth. *Reading Research Quarterly*, *45*(2), 134–168. https://doi.org/10.1598/RRQ.45.2.1

Lemons, C.J., Gesel, S.A., & LeJune, L.M. (2019). Intensive intervention for students with intellectual and developmental disabilities. In R. Z. Edmonds, A. G. Gandhi, & L. Danielson (Eds.), *Essentials of intensive intervention* (pp. 71–96). Guilford Press.

Lemons, C. J., King, S. A., Davidson, K. A., Puranik, C. S., Al Otaiba, S., Mrachko, A. A., Partanen, J., & Fidler, D. J. (2017). Developing an early reading intervention aligned with the down syndrome behavioral phenotype. *Focus on Autism and Other Developmental Disabilities*, *32*(3), 176–187. https://doi.org/10.1177/1088357615618941

Lemons, C. J., King, S. A., Davidson, K. A., Puranik, C. S., Al Otaiba, S., & Fidler, D. J. (2018). Personalized reading intervention for children with Down syndrome. *Journal of School Psychology*, *66*, 67–84. https://doi.org/10.1016/j.jsp.2017.07.006

Lemons, C. J., Zigmond, N., Kloo, A. M., Hill, D. R., Mrachko, A. A., Paterra, M. F., Bost, T. J., & Davis, S. M. (2013). Performance of students with significant cognitive disabilities on early-grade curriculum-based measures of word and passage reading fluency. *Exceptional Children*, *79*(4), 408–426. https://doi-org.proxy.lib.uiowa.edu/10.1177/001440291307900402

Lim, L., Arciuli, J., Munro, N., & Cupples, L. (2019). Using the MULTILIT literacy instruction program with children who have Down syndrome. *Reading and Writing*, *32*(9), 2179–2200 https://doi.org/10.1007/s11145-019-09945-8

Lüdecke, D. (2020). Sjplot: Data visualization for statistics in social science. R package Version 2.8.3. https://CRAN.R-project.org/package=sjPlot.

Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple imputation of missing data in multilevel designs: A comparison of different strategies. *Psychological Methods*, *22*(1), 141–165 https://doi.org/10.1037/met0000096.

Mazerolle, M. J. (2019). AICcmodavg: Model selection and multimodel inference based on AIC. R package Version 2.2–2. https://CRAN.r-project.org/package=AICcmodavg.

Melby-Lervåg, M., Lyster, S. A., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, *138*(2), 322–352. https://doi.org/10.1037/a0026744.

National Center on Response to Intervention. (2012). RTI Implementer series Module 2: Progress monitoring [PowerPoint presentation]. http://www.rti4success.org/resourcetype/rti-implementer-series-module-2-progress monitoring.

Otero, T. M. (2017). Brief review of fluid reasoning: conceptualization, neurobasis, and applications. *Applied Neuropsychology: Child*, *6*(3), 204–211. https://doi.org/10.1080/21622965.2017.1317484

Psyridou, M., Eklund, K., Poikkeus, A. M., & Torppa, M. (2018). Reading outcomes of children with delayed early vocabulary: A follow-up from age 2–16. *Research in Developmental Disabilities*, *78*, 114–124. https://doi.org/10.1016/j.ridd.2018.05.004

R Core Team. (2019). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/

Roch, M., Mattera, L., & Simion, E. (2019). Reading for comprehension: The contribution of decoding, linguistic and cognitive skills. In S. Lanfranchi (Ed.), *International review of research in developmental disabilities* (Vol. 56, pp. 213–255). Academic Press.

Rodadaugh, E., & Moeyaert, M. (2017). *Multilevel modeling of single-case data: An introduction and tutorial for the applied researcher*. Northeastern Educational Research Association Conference, Mansfield, CT, United States. https://opencommons.uconn.edu/nera-2017/8

Rodgers, D., King, S. A., & Lemons, C. J. (2021). *Validity of curriculum-based measures in reading for children with intellectual disabilities*. Manuscript in preparation.

Saunders, K. J., & DeFulio, A. (2007). Phonological awareness and rapid naming predict word attack and word identification in adults with mild mental retardation. *American Journal on Mental Retardation*, *112*(3), 155–166. https://doi.org/10.1352/0895-8017(2007)112[155:PAARNP]2.0.CO;2

Shadish, W. R., Kylse, E. N., & Rindskopf, D. M. (2013). Analyzing data from single-case designs using multilevel models: New applications and some agenda items for future research. *Psychological Methods*, *18*(3), 385–405. https://doi.org/10.1037/a0032964

Steele, A., Scerif, G., Cornish, K., & Karmiloff-Smith, A. (2013). Learning to read in Williams syndrome and Down syndrome: Syndrome-specific precursors and developmental trajectories. *Journal of Child Psychology and Psychiatry*, *54*(7), 754–762. https://doi.org/10.1111/jcpp.12070

Swan, D. M., & Pustejovsky, J. E. (2018). A gradual effects model for single-case designs. *Multivariate Behavioral Research*, *53*(4), 574–593.

Symonds, M. R. E., & Moussalli, A. (2010). A brief guide to model selection, multimodel inference and model averaging in behavioral ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13–21.

Tarlow, K. R. (2017). An improved rank correlation effect size statistic for single-case designs: Baseline corrected Tau. *Behavior Modification*, *41*(4), 427–467. https://doi.org/10.1080/00273171.2018.1466681

Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. *ISRNEducation*, 1–29.

VanMeveren, K., Hulac, D., & Wollersheim-Shervey, S. (2020). Universal screening methods and models: Diagnostic accuracy of reading assessments. *Assessment for Effective Intervention*, *45*(4), 255–265. https://doi.org/10.1177/1534508418819797

van Bysterveldt, A. K., Gillon, G., & Foster-Cohen, S. (2010). Integrated speech and phonological awareness intervention for pre-school children with Down syndrome. *International Journal of Language & Communication Disorders*, *45*, 320–335. https://doi.org/10.3109/13682820903003514

Van Norman, E. R., Maki, K. E., Burns, M. K., McComas, J. J., & Helman, L. (2018). Comparison of progress monitoring data from general outcome measures and specific subskill mastery measures for reading. *Journal of School Psychology*, *67*, 179–189. https://doi.org/10.1016/j.jsp.2018.02.002

What Works Clearinghouse. (2020). *What works clearinghouse standards handbook, Version 4.1*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/handbooks.

Yoder, P. J., Bottema-Beutel, K., Woynaroski, T., Chandrasekhar, R., & Sandbank, M. (2013). Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders. *Evidence-Based Communication Assessment and Intervention*, *7*(4), 150–174. https://doi.org/10.1080/17489539.2014.917780

Zumeta, R. O., Compton, D. L., & Fuchs, L. S. (2012). Using word identification fluency to monitor first-grade reading development. *Exceptional Children*, *78*(2), 201–220. https://doi-org.proxy.lib.uiowa.edu/10.1177/001440291207800204

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Institute of Education Sciences, (grant number R324A110162).

ORCID iD

Seth King https://orcid.org/0000-0001-7142-8694

Supplemental Material

Supplemental material for this article is available online.

Manuscript received June 2021; accepted January 2022.