Original Research

Implementing Comprehensive Literacy Instruction for Students With Severe Disabilities in General Education Classrooms

Exceptional Children
1–18

DOI: 10.1177/0014402919880156
journals.sagepub.com/home/ecx
SAGE

Pam Hunt[1], Elizabeth Kozleski[2], Jaehoon Lee[3], Kathleen Mortier[1], Danielle Fleming[1], Tyler Hicks[2], Lakshmi Balasubramanian[1], Grace Leu[2], Leslie Ann Bross[2], Vidya Munandar[2], Kristin Dunlap[2], Inna Stepaniuk[2], Corrine Aramburo[1], and Youngha Oh[3]

Abstract

The purpose of this conceptual replication study was to investigate the efficacy of an early literacy intervention when it was implemented by special educators in general education classrooms with students in the class participating in the lessons. The study was conducted in 16 schools in three states. Eighty students with severe disabilities participated in the study. Students in the intervention group received Early Literacy Skills Builder (ELSB) instruction, and students in the "business-as-usual" control group received literacy instruction planned by special education teachers to address the students' individualized education program literacy goals. Literacy assessments were conducted in five waves scheduled across the school year. Results showed that students receiving ELSB instruction made greater gains in assessed literacy skills than students in the control group. These findings provide evidence that students with severe disabilities can benefit from comprehensive emergent literacy instruction when it is implemented in general education settings.

For students severely affected by intellectual disabilities or autism, the ability to gain meaning from text increases access and independence, provides a vehicle for cultural and social engagement, expands unique interests and talents, and fosters lifelong learning. Yet, students with severe disabilities have had little access to comprehensive, research-based literacy instruction (Browder, Wakeman, Spooner, Ahlgrim-Delzell, & Algozzine, 2006). Historically, their curriculum focused on "functional" vocabulary instruction associated with the development of functional skills to increase the students' independence and participation in home, school, and community living (Ayres, Lowrey, Douglas, & Sievers, 2011; Browder et al., 2006). Factors contributing to this historical lack of access may include inadequate teacher preparation to implement comprehensive literacy instruction as well as low expectations of the students' ability to benefit from such instruction (Petersen, 2016; Ruppar, Gaffney, & Dymond, 2014; Timberlake, 2014).

During the past two decades, there has been a notable emphasis on research-based literacy instruction to increase students' reading

[1]San Francisco State University
[2]University of Kansas
[3]Texas Tech University

Corresponding Author:
Pam Hunt, Department of Special Education, San Francisco State University, San Francisco, CA 94132, USA.
E-mail: hunt@sfsu.edu

success (e.g., Every Student Succeeds Act [ESSA], 2015; National Reading Panel [NRP], 2000; National Early Literacy Panel [NELP], 2008; No Child Left Behind [NCLB], 2002). In addition, federal policy has mandated that all students have access to and make progress in the general curriculum (e.g., ESSA, 2015; Individuals With Disabilities Education Improvement Act, 2004; NCLB, 2002). This increased emphasis on scientifically based literacy instruction has been associated with an emerging body of research documenting the effectiveness of literacy instruction for students mildly to severely affected by intellectual disabilities and autism (e.g., Allor, Mathes, Roberts, Jones, & Champlin, 2010; Bradford, Shippen, Alberto, Houchins, & Flores, 2006; Browder, Ahlgrim-Delzell, Courtade, Gibbs, & Flowers, 2008; Browder, Ahlgrim-Delzell, Flowers, & Baker, 2012; Lemons, Mrachko, Kostewicz, & Paterra, 2012; Whalon, Otaiba, & Delano, 2009).

Previous studies have demonstrated improvement in students' achievement in areas that include phonological awareness and phonics, word recognition and vocabulary, and comprehension. For example, Bradford et al. (2006) provided systematic instruction on decoding skills to middle school students with moderate intellectual disabilities using the Corrective Reading Program (Engelmann, Becker, Hanner, & Johnson, 1980), which addresses decoding, comprehension, and fluency skills. Three participants received 65 lessons provided 3 days a week over a 6-month period. At the completion of the study, students were able to identify letter-sound correspondences, sound out words, and read short passages written at a second-grade level. Lemons et al. (2012) compared the effectiveness of two research-based early reading interventions when they were implemented with 15 children with Down syndrome between the ages of 5 and 13 years. The phonological awareness and decoding interventions were taught by educators during an average of 25 sessions across approximately 12 weeks. The study results showed improvements in reading phonetically regular and high-frequency words associated with the decoding intervention. The addition of the phonological awareness intervention did not result in consistent gains across participants. Finally, Allor et al. (2010) examined the efficacy of a comprehensive early reading intervention to teach 28 elementary students with moderate intellectual disabilities to read. Students in the treatment group received daily instruction in small groups for 40- to 50-min sessions taught by trained special education teachers. Students in the control group received "typical special education." The comprehensive early reading intervention addressed concepts of print, phonological and phonemic awareness, oral language, letter knowledge, word recognition, vocabulary, fluency, and comprehension. The results indicated that students who received comprehensive and explicit instruction made greater gains than students participating in "typical special education" on measures of phonemic awareness, phonics, word recognition, and comprehension.

The results of these studies are promising, suggesting that students with mild to moderate intellectual disabilities can acquire early reading skills given consistent, explicit literacy instruction. However, a shared characteristic of these and other early literacy studies is that the interventions were not designed for nonverbal responders or students who require systematic prompting procedures and opportunities for repeated practice to develop new skills. An exception to this is the research by Browder and her colleagues at the University of North Carolina at Charlotte (UNCC; Browder et al., 2008, 2012) investigating the effectiveness of a comprehensive early literacy intervention for students with severe disabilities. The literacy program, Early Literacy Skills Builder (ELSB), draws on the NRP (2000) report to include both code and meaning-focused skills while using instructional methods and systematic prompting procedures found to be effective with students with severe disabilities (e.g., Allor et al., 2014; Bradford et al., 2006; Browder, Ahlgrim-Delzell, Spooner, Mims, & Baker, 2009). In addition, ELSB was designed to address the needs of nonverbal responders. For example, students learn to clap out syllables in words, tap out phonemes, and point to words to complete sentences. Shared reading of adapted, grade-level literature provides the context for teaching reading conventions and comprehension and vocabulary skills (Bock & Erickson, 2015). Students learn, for example, to point to the title and author of the book, open the book to get the story started, and turn pages to keep the story going. They point to pictures arrayed on a communication board to predict what the story is about and respond to listening comprehension questions.

Two randomized controlled trials demonstrated the efficacy of ELSB (Browder et al., 2008, 2012). The 2008 study was conducted with 23 students with severe disabilities. Students randomly assigned to the treatment group received ELSB instruction addressing phonemic awareness, phonics, comprehension, and vocabulary skills. Students in the control group received sight word–picture instruction using the Edmark Reading Program. Teachers delivered daily instruction either to individual students or in a small group of two to four students. Results indicated that students in the treatment group significantly outperformed students in the control group on a research team–designed measure of early literacy and two standardized measures. The 2012 study was conducted with three cohorts of students (N= 93 total) who received early literacy instruction for 1, 2, or 3 years. As with the 2008 study, students randomly assigned to the intervention group received the ELSB curriculum, and students in the control group received the Edmark Reading Program. Results showed that there were statistically significant main effects for treatment and year of the study. In addition, there were small to moderate effect sizes (.30–.49) for the two early literacy measures.

Although the results of the studies conducted by Browder and her colleagues provide evidence of the effectiveness of ELSB instruction on the emergent reading performance of students with severe disabilities, the instructional settings for those and other studies of multicomponent early literacy interventions for students with moderate to severe disabilities were limited to special education classrooms (e.g., Allor et al., 2010; Bradford et al., 2006; Lemons et al., 2012). Hudson and her colleagues (Hudson, Browder, & Wood, 2013) identified 17 studies on academic learning for students with severe disabilities in general education settings that met their inclusion criteria. Of those studies, none examined the efficacy of comprehensive literacy interventions implemented in general education classrooms. Instead, the majority of the studies of literacy and other academic instruction evaluated the effectiveness of embedded trial instruction of discrete academic skills distributed across activities in the general education classroom. While there is evidence that this approach is effective for teaching academic content in general education settings (Hudson et al., 2013; McDonnell et al., 2006), it does not allow for the comprehensive research-based approach to reading instruction provided by a literacy intervention like ELSB. By conducting efficacy trials for comprehensive literacy interventions in separate settings only, practitioners and policy makers may conclude that systematic and comprehensive literacy instruction for students with severe disabilities can be implemented only in settings outside general education classrooms (Hunt, 2019; Toews & Kurth, 2019).

The primary purpose of this study was to address this concern by conducting a conceptual replication (Bonett, 2012; Coyne, Cook, & Therrien, 2016; Doabler et al., 2016) that tested the efficacy of the ELSB intervention when it was implemented in general education classrooms in small instructional groups with peers participating in the lessons as reading buddies. Additional differences between this study and previous ELSB efficacy studies (Browder et al., 2008, 2012) were the different geographical locations of the participating schools (one Pacific coast state and two midwestern states), the statistical analyses employed (hierarchical linear modeling), the literacy instruction that served as the control condition ("business-as-usual" [BAU] instruction), and the administration of a standardized early literacy measure not included in previous studies (Gates-McGinitie Reading Test, Pre-Reading Level). Change in the setting for this ELSB efficacy evaluation to general education classrooms allowed us to examine whether findings from the original research would hold up with this difference in instructional setting as well as with the other methodological variations described earlier (Coyne et al., 2016; Doabler et al., 2016).

This conceptual replication was conducted via a longitudinal randomized controlled trial (RCT) with matched-pair randomization. We asked three specific questions:

1. Does the ELSB emergent literacy curriculum increase the literacy skills of students with severe intellectual disabilities or autism in phonemic awareness and phonics, listening comprehension, word study and vocabulary, and reading conventions when special educators implement it in general education classrooms in integrated, small-group contexts?
2. Does the ELSB curriculum do so over and above standard (BAU) instruction implemented in general education classrooms?
3. How much of the variance in students' emergent literacy skills, and of the variance associated with change over time, is accounted for by the following: (a) student characteristics, including disability diagnosis, use of verbal or nonverbal communication, and grade level, and (b) fidelity of intervention implementation?

Method

Participants

Schools. The study was conducted during the 2016–2017 school year in 16 schools and 11 school districts across a Pacific coast state and two midwestern states. The nine elementary schools in the Pacific coast state had an average of 481 students (range: 290–681). On average, 53% of the students at each site (range: 17%–74%) were eligible for free or reduced-cost lunch. Forty-three percent of the students were Latinx American (range: 6%–88% across schools), 29% were Asian or Pacific Islander American (range: 5%–61%), 14% were European American (range: 2%–33%), 7% were two or more races (range: 0.9%–17%), and 6% were African American (range: 0.5%–19%) (California Department of Education, 2018). Three of the districts were urban and three were large suburban; however, six of the nine schools were in urban districts (National Center for Education Statistics [NCES], 2018). The seven elementary schools in the midwestern states had an average of 432 students (range: 286–525). Fifty-five percent of the students at each site (range: 20%–94%) were eligible for free or reduced-cost lunch. On average, 55% of the students in the participating schools were European American (range: 13%–79%), 22% were Latinx American (range: 6%–80%), 13% were African American (range: 3%–33%), 8% were two or more races (range: 3%–13%), and 2% were Asian or Pacific Islander American (range: 0.6%–6%). All of the midwestern districts were identified as urban (NCES, 2018).

Students. Eighty students moderately to severely affected by intellectual disabilities and autism (focal students) participated in the study (four to eight students per school site). All the students were eligible to take their state's alternate assessment. In addition, they met all of the inclusion criteria for participation in the Browder et al. (2008, 2012) studies, including that they (a) were affected by moderate to severe intellectual disabilities as reflected by developmental screening completed by school district psychologists, (b) were enrolled at the time of the study in Grades K to 4, (c) read below the first-grade level as determined by a review of school records, (d) had adequate hearing and vision to respond to curricular materials and instruction, and (e) responded to instruction in English. After discussion with Browder and her research colleagues and based on their recommendation, we added the following inclusion criteria: that they (f) demonstrated picture discrimination skills as determined by assessments conducted by the special education teacher or speech language pathologist and (g) were able to sit for short periods of time for instruction based on special education teacher observation and student records.

The eligible students at each school site were matched into pairs independently for each teacher based on the disability (intellectual disability or autism) listed on their individualized education programs (IEPs) and their verbal status (verbal or nonverbal), gender (male or female—matched to the extent possible), and grade (Grades K–1 or 2–4). If more than the targeted number (i.e., four to eight) of students at each school site were eligible, pairs of the most comparable students were selected from the pool of eligible students. Within each matched pair, one student was randomly assigned to the intervention group and the other to the control group.

Table 1 (available in the online supplemental materials) presents student demographic characteristics and baseline literacy scores for both groups. The general education classroom was the primary placement and setting for receiving special education services for only four of the students. The remaining participants were mainstreamed into general education classrooms for literacy instruction and, for many of the students, other activities during the day.

General education students in each participating classroom were recruited to participate in the lessons as "reading buddies" by the special education teachers during a whole-class presentation at the start of the school year. The teachers described the "fun" reading lessons that they would be teaching in the general education classroom each day and then asked students if they would like to participate in the lessons as reading buddies. It was explained that they would have to take turns, but anyone who wanted to be a reading buddy could. The majority of the students in each class indicated that they wanted to be a reading buddy; however, only students who brought signed permission forms to school could do so. On average, 66% of the students in each class (range: 16%–96%) served as reading buddies. Each day, the opportunity to be a reading buddy was rotated through the list of volunteers. Reading buddies participated in the literacy activities with the focal students. They did not take an instructional role; however, they served as proficient models of targeted emergent reading behaviors, thereby providing repeated opportunities for observational learning by the focal students (Bandura, 1977, 1986).

Educators. The 20 special education teachers (nine in the Pacific coast state; 11 in the midwestern states) had a graduate-level Education Specialist Credential in Moderate/Severe Disabilities (Pacific coast state) or an elementary teacher education license at the baccalaureate level with a graduate-level endorsement in special education for those with moderate or severe disabilities (midwestern states). They had served as special education teachers for an average of 9.7 years (range: 2–27 years). The 25 participating paraprofessionals (14 in the Pacific state; 11 in the midwestern states), who implemented components of the literacy instruction under the supervision of the special education teachers, had served in that position an average of 7.2 years (range: 1–17 years). The 46 general education teachers were recruited to participate in the study through informational presentations by research team members at their schools and follow-up conversations with the special education teachers. Their role was to participate in recruiting reading buddies, managing the calendar identifying the reading buddy for each day, and collaborating with the special education teacher to determine the ways in which ELSB and BAU instruction would be physically integrated into the classrooms' literacy activities.

Setting

Early literacy instruction for students in the intervention and control groups was delivered in general education classrooms during the classroom's literacy block. Literacy instruction for students in the intervention group was delivered in small-group contexts by special educators (i.e., special education teachers alternating with paraprofessionals).

The spaces designated by the general education teachers for the small-group lessons were in proximity to the other small- or large-group instructional activities. Sixty percent of BAU control group instruction was also delivered in small-group contexts by special education teachers and paraprofessionals. However, other configurations were also used, including whole-class instruction delivered by the general education teachers with paraprofessional support (20% of classrooms) and whole-class instruction delivered by the classroom teacher, followed by small-group instruction led by the special education teacher or paraprofessional (20% of the classrooms).

Emergent Literacy Instruction

Intervention group. Students in the intervention group received ELSB instruction (Browder et al., 2008, 2012). ELSB is grounded in principles and practices from research on early literacy (NELP, 2008; NRP, 2000). It includes the NRP components of phonological awareness, phonics, vocabulary, and comprehension and focuses on early literacy skills that build each of those components. In addition, ELSB incorporates direct, systematic instruction based on the principles of applied behavior analysis and is designed for students who are verbal or nonverbal, ages 5 to 10 years.

ELSB has two components: Building With Sounds and Symbols, and Building With Stories. Building With Sounds and Symbols comprises seven levels with five lessons at each level. Lessons focus on stories from a book, *All About Moe*. Moe is a frog, and a Moe puppet is used during the lessons to motivate students to attend and participate. Objectives include the following: (a) reading vocabulary words, (b) pointing to vocabulary words to complete sentences, (c) pointing to words in a story as they are read by a teacher, (d) saying or pointing to a word to fill in the missing word in a repeated story line, (e) responding to literal questions about a story by selecting a picture or answering verbally, (f) demonstrating understanding of syllable segmentation by clapping out syllables in words, (g) demonstrating understanding of phoneme segmentation by tapping out sounds in consonant-vowel-consonant (CVC) words, (h) identifying letter-sound correspondences, (i) identifying first and last sounds in vowel-consonant (VC) and CVC words, (j) pointing to pictures that begin or end with a given consonant sound, (k) pointing to the letters in words when the word is segmented slowly, (l) blending sounds to form a word and then identifying the picture that represents the word, and (m) pointing to pictures or words representing new vocabulary.

Building With Stories implements shared story-reading instruction using literature from the students' grade level to teach students to interact with books and listening comprehension and vocabulary skills. For students in the third and fourth grades, the selected books were adapted to use common words, simple sentence structures, and less text. Instructional objectives included (a) book orientation, (b) identifying the title and author of the book, (c) opening the book and turning pages, (b) pointing to text as it is read, (d) completing repeated story lines, (e) reading new vocabulary, and (f) answering prediction, literal, and inferential comprehension questions.

Control group. We used a BAU approach for the control condition rather than using the Edmark Reading Program employed in previous ELSB efficacy studies. This allowed us to compare the relative effectiveness of reading instruction implemented by the participating special education teachers to meet the literacy objectives of their students in the control group with their implementation of ELSB with their students in the intervention group. Thirteen percent of the special education teachers implemented a published reading program designed to address beginning reading skills (e.g., Starfall; Elliott, Ferguson, & Riess, 2017), and 9% selected pieces of the published reading program used by the classroom teachers (e.g., Reading Street; Scott Foresman, 2011) as the basis for instructional activities for the focal students. The majority of special education teachers (78%) used a variety of teacher-made materials and lesson plans to address focal students' emergent literacy objectives (e.g., letter names and sounds, blending letter sounds, vocabulary word reading, answering comprehension questions).

Procedural fidelity. Procedures were put into place to promote high levels of implementation fidelity, including the following: (a) research team members were certified in ELSB implementation by members of the UNCC research team who developed and evaluated the curriculum; (b) participating special education teachers participated in a 2-day ELSB training by research team members during the summer, with a "booster" session scheduled close to the start of the school year; and (c) school site coordinators observed literacy lessons implemented each week across school sites and provided corrective feedback if needed. During the 2-day (6 hr per day) training, members of the research team provided an overview of the ELSB objectives and instructional methods. This was followed by repeated cycles (taking two learning objectives at a time) of (a) verbal descriptions of the objectives and the procedures for teaching each one, (b) video demonstrations of those instructional procedures, and (c) role-playing to practice implementing the instruction.

Procedural fidelity checklists (Browder et al. 2008, 2012) were used by research team members to document the fidelity of ELSB implementation. The fidelity checklist for Building With Sounds and Symbols was a list of educator behaviors for implementing the scripted lessons addressing the 13 objectives. The checklist for Building With Stories was a task analysis of educator behaviors for teaching story-based lessons (e.g., "Reads title and gives students an opportunity to point to/say the title."). Each item on the checklists was rated with a "+" if present and performed correctly and a "–" if absent or performed incorrectly, to yield a percentage of instructional steps implemented correctly during each session.

Because special education teachers providing BAU instruction used a variety of procedures and materials to address a range of literacy objectives, the procedural fidelity checklist addressed general instructional behaviors relevant to early literacy instruction. The checklist included the following items: (a) Is a reading buddy participating with each focal student in his or her literacy activities in the general education classroom? (b) Is the focal student receiving instruction that addresses early literacy development? (c) Are materials adapted for the focal student (when needed) to increase accessibility (e.g., adding pictures to increase comprehension)? (d) Is physical, verbal, visual, or gestural support from an adult or peer used to assist the focal student to engage in the targeted literacy activities? In addition, the checklist included two items documenting the absence of ELSB materials from the BAU classrooms.

Research team members used procedural fidelity checklists to observe reading lessons once a week at each school site (rotating between ELSB and BAU instruction) to document the extent to which lessons were implemented with fidelity. A second member of the research team joined the primary data collector to collect interrater reliability (IRR) data during 22.4% of the fidelity sessions for the Pacific coast state and 20% of the fidelity sessions for the midwestern states.

Procedures

Thirty- to 40-min ELSB and BAU lessons were delivered daily in general education classrooms during scheduled literacy blocks from September (following the first assessment period) to the end of the school year. Occasional disruptions to scheduled instruction were due to staff and student absences, the scheduling of schoolwide activities (e.g., assemblies and field trips), and emergency situations that required the attention of the special education teacher.

Building With Sounds and Symbols and Building With Stories instruction alternated each day. Building With Sounds and Symbols instruction for all students began with the first lesson because all focal students were reading below a first-grade level. No other literacy instruction was provided for any of the participating students at other times during the day. Typically, two focal students and two general education classmates participated in the lessons. The classmates were designated reading buddies for the day and engaged in the reading group on a rotating daily schedule. To help manage the rotating partner schedule, a calendar identifying the reading buddy for the day was posted in many of the classrooms. As described earlier, all ELSB, and the majority of BAU instruction, was implemented in small-group contexts by special educators. In some classrooms, BAU instruction was delivered through whole-class lessons with adapted materials and support from special educators and peers or a combination of whole-class and small-group lessons.

Research team members observed ELSB or BAU lessons each week at each school site not only to collect fidelity of implementation data but also to provide feedback when intervention procedures were not implemented as described by instructional behaviors listed on the implementation fidelity checklists. For ELSB instruction, feedback consisted of reminders of missed instructional steps or discussion of instructional steps performed inaccurately. For BAU instruction, reminders were given if instruction did not address literacy skills; reading buddies were not present; instruction was not adequately adapted for accessibility by the focal student; the focal student was not assisted, when needed, to engage in the literacy activities; or ELSB materials were present during instruction.

Measures

There were five testing points during the school year: September (baseline), November, January, March, and May and June. The Nonverbal Literacy Assessment (NVLA; Baker, Spooner, Flowers, Ahlgrim-Delzell, & Browder, 2010; Browder et al., 2008, 2012) and the Gates-MacGinitie Reading Test, Pre-Reading Level (4th ed.; GMRT; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) served as the early literacy measures. They were chosen because they (a) addressed key literacy concepts identified by the NRP (2000) report and the NELP (2008) report, (b) had strong psychometric properties, and (c) accommodated students who were nonverbal. Two days were scheduled for assessing each student. The length of each session and the number of breaks from assessment activities varied to maintain student comfort and interest.

NVLA. The NVLA was one of the measures used in previous ELSB efficacy studies. It has a receptive response format with answers provided in four-choice arrays. Answers can be communicated with speech and pointing, pulling cards held in place with Velcro, and eye gaze. The original version had 218 items and was divided into two subtests: Conventions of Reading (CVR) and Phonological Awareness/Phonics Skills (PhonSk). The CVR subtest measures interactions with books and comprehension in the context of shared story reading. The PhonSk subtest measures skills of word study (matching words, picture-word matching) and vocabulary, alphabetic principle, and beginning phonics; phonological awareness (breaking words into syllables); blending; and phonemic awareness (e.g., identifying first and last sounds in words).

Test-retest reliability for the NVLA was .97 ($p < .001$). Internal consistency reliability estimates for the NVLA and the CVR and PhonSk subtests were .98, .80, and .97, respectively (Browder et al., 2008, 2012). Fidelity of administration of the NVLA and IRR were tested by having a second observer score administration procedures and student responses. Mean fidelity of administration was 95.5% (range: 93.1%–98.5%). IRR was 96% (Browder et al., 2008, 2012).

Content validity was established by a national panel of six experts in emergent literacy, severe disabilities, and assessment who reviewed the assessment items. After reviewing the NVLA, they reported that the items reflected the range of early literacy skills (Baker et al., 2010). A confirmatory factor analysis of the NVLA indicated that the

NVLA measures a unitary construct of literacy (Baker et al., 2010).

Four-dimensional version of the NVLA. Fleming, Wilson, and Ahlgrim-Delzell (2018) used item response theory to investigate the NVLA to create a four-dimensional Rasch model using 133 items from the original NVLA. The reduced number of items resulted in high levels of test reliability despite decreasing the number of questions: That is, person separation reliability for CVR was .80; for comprehension, .64; for word study and vocabulary, .83; and for PhonSk, .85. The posteriori–plausible value separation reliability for CVR was .88; for listening comprehension, .91; for word study and vocabulary, .96; and for PhonSk, .96. Therefore, this version of the NVLA provided the same information about student abilities but with fewer implementation hours required. In addition, a four-dimensional version of the NVLA (CVR, comprehension, word study and vocabulary, and PhonSk) provided more detailed information about student abilities. The four-dimensional version of the NVLA was used for this study and included the subtests (a) CVR (book orientation, identifying the title and author of the book, opening the book and turning pages, pointing to text as it is read, and completing repeated story lines), Listening Comprehension (recall of facts, classifying and categorizing, making inferences and predictions), Word Study/Vocabulary (word matching, picture-word identification, vocabulary comprehension and reading), and PhonSk (e.g., letter identification, letter sounds, blending, identifying first and last letters or sounds in words, syllabication).

GMRT. The GMRT was not one of the standardized measures used in previous ELSB efficacy studies. It was selected for this study because it could be accessed by students who were nonverbal and students who had physical disabilities and because it measured a range of early literacy skills. The GMRT was designed as a "paper-and-pencil test" (students fill in the circle next to the correct response from a choice of four). Administration procedures were adapted to increase accessibility by (a) enlarging the symbols and presenting each four-symbol array on a separate response plate and (b) allowing students to respond by pointing to or touching a symbol, pulling a symbol card held in place with Velcro, or gazing to a symbol. Two of the four GMRT subtests were used—Phonological Awareness and Letters/Letter-Sound Correspondences. The Phonological Awareness subtest measured phoneme matching and rhyme. The Letters/Letter-Sound Correspondences subtest measured ability in the areas of matching letters and words, letter name recognition, letter-sound correspondences, and sound-to-letter correspondences.

The development of the GMRT included pilot studies, field testing with thousands of students, and gathering of expert input from test users, measurement specialists, curriculum specialists, and cultural specialists (Johnson, 2005; MacGinitie, MacGinitie, Maria, & Dreyer, 2002). Reliability estimates included total test and subtest internal consistency levels with coefficient values at or above .90 for the total tests and subtests at all levels except adult reading.

Content validity was documented through a test development process to design the scope of the subtests and to identify effective items within subtests. Item bias studies were used to eliminate problematic items. Construct validity was suggested by the strong intercorrelations between the test and subtest scores (Johnson, 2005; MacGinitie et al., 2002).

Data collection. To minimize task demands, we used a three-form planned missing data collection protocol. Three test forms were created with (a) a common block, which contained demographic questions and the CVR subtest; (b) three blocks, each of which covered one-third of the items from the other three NVLA subtests; and (c) the GMRT. Consequently, each form had three-fourths of the total items. Students were randomly assigned to complete a single form at each time point.

IRR. A second member of the research team joined the primary data collector to score student responses during 24% of the assessment sessions across testing points for the Pacific coast state and 20% of the sessions for the midwestern states. The calculation of point-by-point scoring agreement (i.e., correlation) between the primary data collector and independent observer revealed a high level of IRR: that is, 98.3% for the intervention group (range: 98%–98.5%) and 98.6% for the control group (range: 98%–99%).

Statistical Analysis

Normality was checked for the GMRT and NVLA scores (percentage correct) at each measurement time point (September, November, January, March, and May and June). Three NVLA subtest scores (CVR, Listening Comprehension, and Word Study/Vocabulary) did not follow a normal distribution, with a standardized skewness score of $z \geq 2.83$ and a Q-Q plot deviating from normalcy. Thus, they were transformed to satisfy the normality assumption for data analysis—square transformation for CVR and Word Study/Vocabulary and square root transformation for Listening Comprehension—commonly for each time point.

Both the GMRT and the NVLA demonstrated adequate reliability. Internal consistency was acceptable at each measurement, with Cronbach's α ranging from .87 to .90 ($M = .89$, $SD = .01$) for the GMRT and ranging from .69 to .94 ($M = .85$, $SD = .08$) for the NVLA subtests. Test-retest reliability of those scores were also confirmed by moderate or higher correlations between different time points, $r = .63$ to .83 ($M = .75$, $SD = .06$) for GMRT and $r = .39$ to .93 ($M = .73$, $SD = .13$) for NVLA.

Missing data handling. In this study, data were missing by design (i.e., three-form planned missing data collection) or due to participant attrition. A Monte Carlo Markov chain multiple imputation procedure was employed to handle both types of the missing data (Enders, 2010). This procedure generated 200 complete data sets by imputing the original incomplete data and then combined the analysis results from each imputed data set—that is, parameter estimates were averaged over the 200 analyses with imputed data sets, and standard errors were computed using the average of the standard errors over the 200 analyses and the between-analysis variation in the parameter estimates (Rubin, 1987; Schafer, 1997). All measured variables and design factors (e.g., group, wave) were incorporated into the imputation process as auxiliary variables, thereby achieving greater recovery of the missing data (Schafer & Graham, 2002). The missing data mechanisms, assumed as at least missing at random, were confirmed by inspecting distributions of the original and imputed data.

Data analysis. Sample demographics were compared between students who received ELSB instruction and those who received BAU instruction to determine success or failure of the group randomization executed, using an independent-samples *t* test (with Satterthwaite approximation if necessary) and chi-square or Fisher's exact test (as appropriate). An effect size, Cohen's *d* or Cramer's *V*, was calculated for each comparison.

To evaluate the efficacy of ELSB (Research Questions 1 and 2), multilevel modeling, also known as hierarchical linear modeling (HLM), was conducted separately for GMRT score and NVLA (and subtest) scores (percentage correct) (Raudenbush & Bryk, 2002; Singer & Willett, 2003). We selected HLM for this study, rather than the analyses used in previous ELSB efficacy studies, because an important analytic consideration for our study was nonindependence of data. That is, data were collected at multiple time points (Level 1) from students (Level 2) who were nested within schools (Level 3). Thus, we selected statistical models and techniques (i.e., HLM) that could properly handle the nested data (Dorman, 2009; Hedges, 2007).

Models estimated overall group difference across time (i.e., group effect), linear or nonlinear (quadratic, cubic, etc.) change over time (i.e., time effect), and group difference in this change (i.e., group-by-time interaction) while accounting for the clustering of measurements (Level 1) repeated for students (Level 2) within schools (Level 3). Models also accounted for differences in scores that are attributable to locations (Pacific coast and Midwest states), thereby providing an unbiased estimate of the intervention effect. For example, models can be written as

$$y_{tij} = \gamma_{000} + \gamma_{001}\left(\text{location}_j\right) + \gamma_{010}\left(\text{group}_{ij}\right) + \gamma_{100}\left(\text{time}_{tij}\right) + \gamma_{110}\left(\text{group}_{ij} \times \text{time}_{tij}\right) + u_{00j} + u_{0ij} + e_{tij},$$

where y_{tij} is student i's literacy score measured at time t, γ_{000} is the overall mean of the score across students and schools at pretest (t= 0), γ_{001} indicates location of the student's school j (1= Pacific coast, 0= Midwest), γ_{010} represents the group effect, γ_{100} represents the time effect (e.g., linear change), γ_{110} represents the group-by-time interaction, and e_{tij}, u_{0ij}, and u_{00j} are random errors assumed to be independently and identically distributed at the time, student, and school levels, respectively. A proper covariance structure for repeated measurements was determined by evaluating relative model fit (e.g., Akaike information criterion [AIC], Bayesian information criterion [BIC]).

Once the efficacy of ELSB was confirmed by a significant group effect or group-by-time interaction, analysis further examined whether the effect of the intervention was comparable between student subgroups of disability (intellectual disabilities or autism), verbal status (verbal or nonverbal), or grade (Grades K–1 or 2–4) and whether the impact was moderated by intervention fidelity (moderating variables analyses; Research Question 3). Examination of these student characteristics as potential moderating variables were exploratory in nature because previous studies of literacy interventions for students with severe disabilities did not do so; therefore, we had no results from previous studies to confirm.

The literacy scores were compared between the two instructional conditions similarly as described earlier but separately for student subgroups—for example, score differences among students in Grades K and 1 and then among students in Grades 2 through 4. HLM models also further estimated all possible interactions with the moderator being tested. For example, models of testing for moderation can be written as

$$y_{tij} = \gamma_{000} + \gamma_{001}\left(\text{location}_j\right) + \gamma_{010}\left(\text{group}_{ij}\right) + \gamma_{020}\left(\text{grade}_{ij}\right) + \gamma_{100}\left(\text{time}_{tij}\right) + \gamma_{021}\left(\text{location}_j \times \text{grade}_{ij}\right) + \gamma_{030}\left(\text{group}_{ij} \times \text{grade}_{ij}\right) + \gamma_{110}\left(\text{group}_{ij} \times \text{time}_{tij}\right) + \gamma_{120}\left(\text{time}_{tij} \times \text{grade}_{ij}\right) + \gamma_{130}\left(\text{group}_{ij} \times \text{time}_{tij} \times \text{grade}_{ij}\right) + u_{00j} + u_{0ij} + e_{tij},$$

where γ_{020} is overall grade difference across time (i.e., grade effect), γ_{030} represents overall group-by-grade difference across time (i.e., group-by-grade interaction), γ_{120} represents grade difference in linear change over time (i.e., time-by-grade interaction), and γ_{130} represents group-by-grade difference in this change (group-by-time-by-grade interaction).

In a subsequent analysis, independent-samples t test (with Satterthwaite approximation if necessary) was performed to find the specific time points where the group (and subgroup) differences are significant and the differences become enlarged or diminished (i.e., effect sizes). All analyses were conducted using SAS 9.4 (SAS Institute, 2002–2012).

Results

Sample Characteristics

Table 1 (available online) presents the students' demographic characteristics and pretest scores on the literacy measures (GMRT and NVLA). The students who received ELSB instruction (intervention group; n= 40) and those who received BAU instruction (control group; n= 40) did not differ in terms of disability status (p= .82, V= 0.03), school grade ($p = .37$, $V = 0.23$), gender ($p = .23$, $V = 0.13$), verbal status ($p = 1.00$, $V = 0.00$), or pretest literacy scores ($p = .35$–$.98$, $d = 0.06$–0.21; before transformation). This finding indicated that the matched-pair randomization used in this study yielded comparable treatment groups of equal size, which may increase the precision and power of our statistical inference. In addition, attrition was the same in the two treatment groups ($n = 3$, 7.5% in each group).

Efficacy of ELSB

The intraclass correlations in the GMRT score and NVLA (and subtest) scores showed that 41.9% to 57% of the score variation occurred at the student level (Level 2) and 24.2% to 37.7% at the school level (Level 3). The compound-symmetric covariance structure produced the smallest AIC and BIC values and thus were chosen for the HLM models examined.

The efficacy of ELSB (Research Questions 1 and 2) was confirmed by the HLM results, in which a significant group-by-time interaction (i.e., a significant group difference in change) indicates a significant effect of the intervention. Indeed, the interaction was significant for all literacy scores (all $p < .01$), except for NVLA Word Study/Vocabulary, suggesting that after controlling for score differences related to geographic locations, improvement in literacy over the school year was significantly greater for the intervention group than for the control group. When the total percentage correct on the NVLA was analyzed, the group-by-time interaction was also significant ($p < .001$). In regard to the NVLA Word Study/Vocabulary subtest, the intervention and control groups as a whole made a significant increase over time (time effect, $p < .001$), and their improvement did not differ from each other (group-by-time interaction, $p = .14$). Overall, the analysis results supported the efficacy of ELSB when implemented in the context of general education using an integrated, small-group instructional model.

Figures 1 and 2 show the literacy scores at five different time points over the 9-month school year (September, November, January, March, and June). The results of follow-up independent-samples t test showed that the two treatment groups had equivalent scores at pretest (September; $p = .39$–$.88$, $d = 0.04$–0.19), but the intervention group achieved significantly higher scores at posttests: GMRT at 9 months ($p < .05$, $d = 0.49$; see Figure 1) and NVLA CVR at 9 months ($p < .05$, $d = 0.48$); Listening Comprehension at 4 months ($p < .05$, $d = 0.47$), 6 months ($p < .01$, $d = 0.62$), and 9 months ($p < .01$, $d = 0.69$); and PhonSk at 9 months ($p < .05$, $d = 0.54$; see Figure 2). When the total percentage correct on NVLA was analyzed, similar results were found—that is, the score was comparable at pretest ($p = .76$, $d = 0.07$) but significantly higher at posttests ($p < .05$, $d = 0.51$, at 6 months; $p < .05$, $d = 0.55$, at 9 months) in the intervention group compared to the control group. However, the group difference in the Word Study/Vocabulary score was not significant at any posttest ($p = .09$–$.97$, $d = 0.01$–0.40).

Moderation of the ELSB Effect

As this study found that ELSB had a significant effect on literacy learning in students with intellectual disabilities or autism, the data were further examined by student subgroups to determine if the intervention was more (or less) effective for students based on their disability, verbal status, or grade (Research Question 3). The results described next are presented in Table 2 (available online).

Although the group-by-time-by-disability interactions were significant in HLM ($p < .05$ in GMRT; $p < .01$ for NVLA), the posttest differences between ELSB and BAU instruction were similar for students with autism ($d = 0.01$–0.71 in GMRT; $d = 0.40$–0.74 in NVLA) and those with intellectual disabilities ($d = 0.06$–0.44 in GMRT; $d = 0.23$–0.52 in NVLA).

The HLM results confirmed that the intervention effect was moderated by verbal status, with significant group-by-time-by-verbal status interactions (both $p < .01$ in GMRT and NVLA). The verbal students

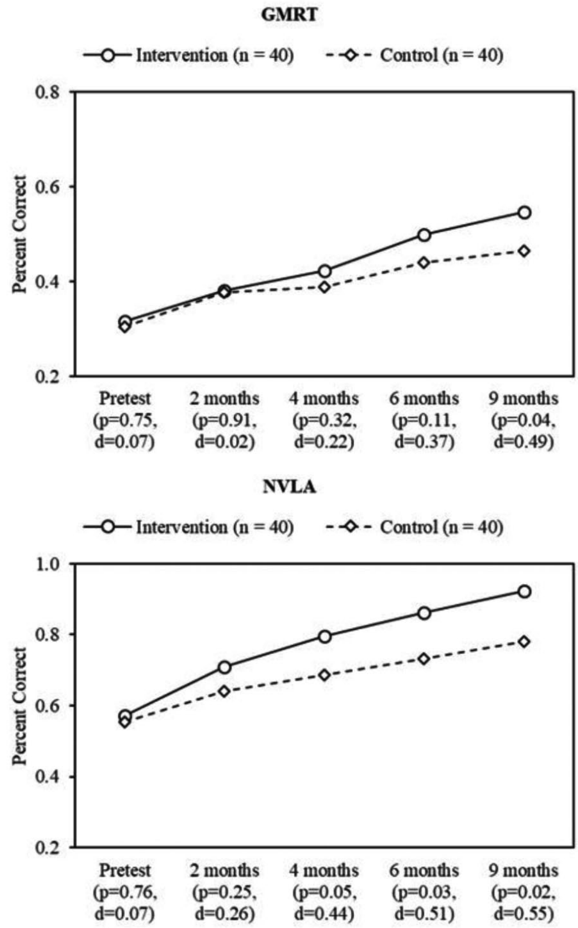

Figure 1. Gates-MacGinitie Reading Test and Nonverbal Literacy Assessment total scores.

who received ELSB instruction had greater gains in literacy scores than did those who received BAU instruction, which led to medium to large differences at posttests (d = 0.18–0.61 in GMRT; d = 0.44–0.75 in NVLA) and statistically significant differences at 4, 6, and 9 months (all $p < .05$) in this subgroup. The 12 nonverbal students randomly assigned to the ELSB condition had considerably lower pretest scores than the 12 nonverbal students assigned to the control condition (d = −0.59 in GMRT; d = −0.58 in NVLA). Nevertheless, the nonverbal students receiving ELSB instruction achieved significantly greater gains in the literacy scores ($M\Delta$ = 0.35, $SD\Delta$ = 0.11, in GMRT; $M\Delta$ = 0.30, $SD\Delta$ = 0.13, in NVLA) by the end of the study (9 months) as compared to those receiving BAU instruction ($M\Delta$ = 0.15, $SD\Delta$ = 0.19, in GMRT; $M\Delta$ = 0.16, $SD\Delta$ = 0.09, in NVLA; $p < .01$, d = 1.31 in both GMRT and NVLA). Given the small number of students in the study who were nonverbal, caution should be taken when interpreting these results.

Significant group-by-time-by-grade interactions in HLM ($p < .01$ in GMRT; $p < .001$ for NVLA) suggested that the effect of ELSB instruction differed between younger and older students. The follow-up independent-

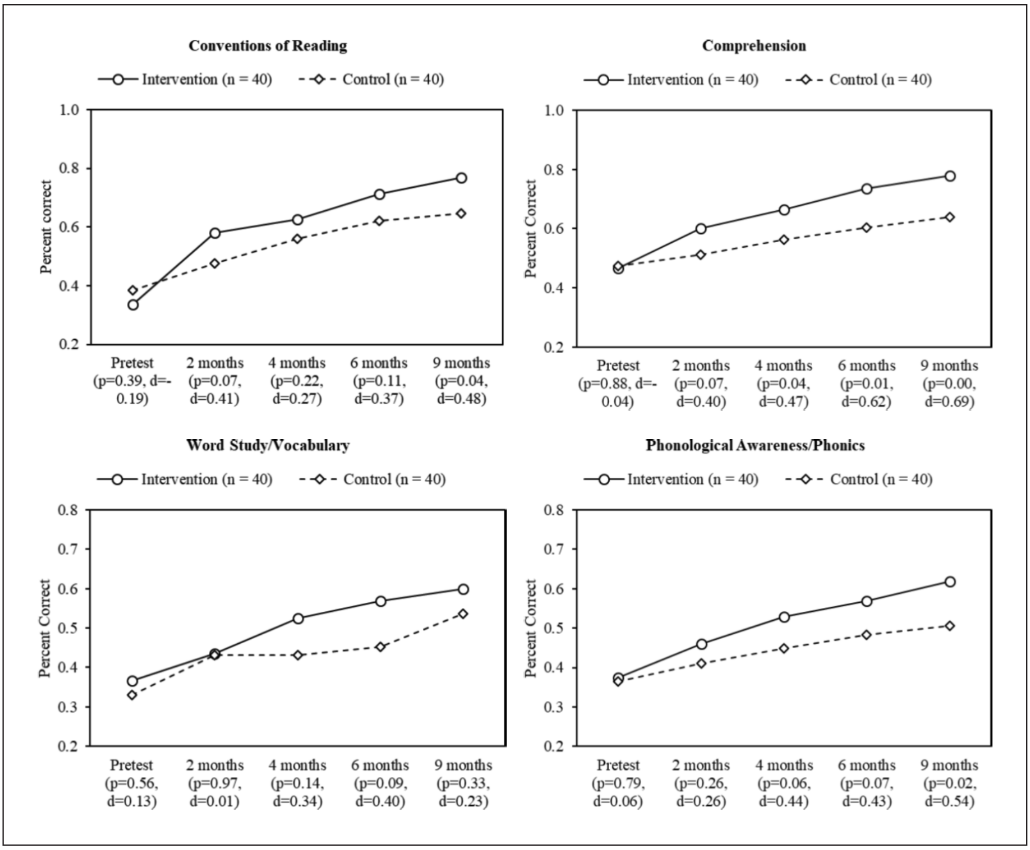

Figure 2. Nonverbal Literacy Assessment dimension scores (after transformation).
Note. The Conventions of Reading and Word Study/Vocabulary scores (percentage correct) were square transformed, and the Listening Comprehension score was square root transformed to meet the normality assumption.

samples *t* test results also showed that for both younger and older students, ELSB instruction produced greater gains in the posttest literacy scores than did BAU instruction. However, these gains were greater among students in Grades K and 1 (d= 0.07–0.91 in GMRT; d= 0.26–0.79 in NVLA) than among students in Grades 2 through 4 (d= 0.081–0.24 in GMRT; d= 0.41–0.50 in NVLA).

Procedural fidelity. The procedural fidelity checklists for ELSB and BAU instruction are described on page 12. Measures of procedural fidelity were high for both ELSB and BAU instruction (95.1% and 98.1%, respectively), and the fidelity of implementation did not moderate the impact of ELSB instruction (p = .40 in GMRT; p =.20–.98 for NVLA). IRR of the procedural fidelity measures was 96.2%.

Discussion

In this conceptual replication study, we investigated the efficacy of ELSB, a comprehensive early literacy intervention for students with severe disabilities, when it was implemented in small-group contexts in general education classrooms with peers participating in the lessons. Participants randomly assigned to the intervention group (ELSB instruction) and the control group (BAU instruction) received literacy instruction each day implemented by special educators during the classroom's literacy period. Two early literacy measures were administered in five waves scheduled across the school year. Results indicated that while scores for students in both the ELSB group and the BAU group showed a significant increase for each of the literacy measures (i.e., the GMRT and NVLA) over the 9-month school year, scores for students receiving ELSB instruction made greater gains than students in the control group. These gains were made in foundational early literacy skill areas, including early phonics, phonological awareness, comprehension, and the conventions of reading and print awareness. The one exception to this was the finding for the NVLA subtest Vocabulary/Word Study. For this subtest, scores for both groups showed a significant increase over time (time effect; $p < .001$); however, there was no difference between the groups.

The moderate effect sizes for our study were promising, particularly given the targeted student population. Effect sizes at 9 months were .49 for the total GMRT score and .55 for the total NVLA score. Effect sizes for the NVLA subtests ranged from .48 to .69, with the largest effect size found for measures of comprehension. These effect sizes were larger than those reported by the Browder et al. (2012) study (.30–.49). However, it is difficult to make comparisons between the studies because of the difference in control group interventions as well as differences in statistical analyses employed, literacy measures implemented, and geographical location.

The results of our study suggest that the effects of ELSB instruction may be generalized to integrated, small-group instructional contexts in general education classrooms in which students with and without disabilities participate in lessons together. This is a noteworthy finding given the opportunities provided by this setting for students with disabilities to benefit from peer modeling of targeted literacy skills. In addition, this setting provides extended opportunities for supported, positive interactions between the students with and without disabilities that may produce positive changes in the peers' perceptions of the focal students' characteristics and abilities.

Exploratory analyses of moderation associated with student characteristics revealed three interesting findings. First is the finding that the nonverbal students receiving ELSB instruction ($n = 12$) made significantly greater gains in the literacy scores by the end of the school year than the nonverbal students in the control group ($n = 12$). This provocative finding warrants further research to determine its generalizability beyond the small number of nonverbal students in our study. Second, although ELSB instruction produced greater gains for both younger students (Grades K–1) and older students (Grades 2–4) than BAU instruction, it produced the greatest gains for younger students. If additional research reveals similar patterns, as is the case for research on early literacy instruction for students with high-incidence disabilities (e.g., Lovett et al., 2017), then a case can be made for the importance of early literacy intervention for students with severe disabilities. Finally, the finding of no difference in the gains in literacy scores for participants with an autism diagnosis and participants with an intellectual disability diagnosis suggests that ELSB instruction might be equally effective for these different disability subgroups.

Study Limitations

The purposive sampling of school sites and the small sample size ($N = 80$) are potential threats to the generalizability of our study findings. In addition, special educators at each school site implemented both ELSB and BAU instruction. Although this controlled for differences between the intervention and control groups in educator characteristics, it also created the possibility that BAU instruction was enriched by the educators' development of skills implementing the systematic and comprehensive instruction associated with ELSB lessons, thereby lessening differences between the groups in intervention effectiveness. Also, the current study did not address the degree to which the focus on fidelity of implementation (weekly observation and feedback by researchers) helped teachers to maintain teaching precision across the year. However, the consistently high level of procedural fidelity (i.e., 95.1% for ELSB instruction) suggests that extensive feedback was often not needed. Finally, the lack of specific dosage data is a limitation of the study. Educators implemented 30 to 40 min of literacy instruction in general education classrooms to students in the intervention and control groups across the school year. However, there is no record of the days when instruction did not take place due to factors such as school holidays, special events, and educator or student absences.

Practical Implications

The results of the current study suggest that ELSB can be implemented effectively and with high procedural fidelity across a wide range of general education classrooms with diverse educator characteristics and varying classroom cultures, resources, and practices. In addition, the preliminary finding that nonverbal students with significant disabilities receiving ELSB instruction made greater gains in the development of early literacy skills than nonverbal students receiving BAU instruction is promising. This outcome suggests that ELSB objectives, instructional procedures, materials, and response formats may be a valuable resource in the design of early literacy instruction for this population of students.

Future Research

The conditions under which the students with severe disabilities learned merit additional research. We need more information about how peers mediated learning for the focal students in both the experimental and control groups. How did their presence inform or influence the focal students' performances or teachers' instruction? In addition to questions related to peer mediation, we need to explore the organizational, structural, and professional barriers that need to be identified and addressed so that ELSB or its equivalent can be implemented without the resources of a research team. Finally, as more services are delivered in general education with interventions that produce strong learning results, research is needed to address how the educator workforce can be prepared and supported to deliver high-quality, evidence-based results. The research and practice communities need to work together to ensure that educators have the knowledge base and organizational supports to design and deliver effective instruction for all their students.

References

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*, 3–22.

Ayres, K. M., Lowrey, K. A., Douglas, K. H., & Sievers, C. (2011). I can identify Saturn but I can't brush my teeth: What happens when the curriculum focus for students with severe disabilities shifts. *Education and Training in Autism and Developmental Disabilities*, *46*, 11–21.

Baker, J. N., Spooner, F., Flowers, C., Ahlgrim-Delzell, L., & Browder, D. M. (2010). A measure of emergent literacy skills. *Psychology in the Schools*, *47*, 501–513. doi:10.1002/pits.20486

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bock, A. K., & Erickson, K. A. (2015). The influence of teacher epistemology and practice on student engagement in literacy learning. *Research and Practice for Persons With Severe Disabilities*, *40*, 138–153. doi:10.1177/1540796915591987

Bonett, D. G. (2012). Replication-extension studies. *Current Directions in Psychological Science*, *21*, 409–412. doi:10.1177/0963721412459512

Bradford, S., Shippen, M. E., Alberto, P., Houchins, D. E., & Flores, M. (2006). Using systematic instruction to teach decoding skills to middle school students with moderate intellectual disabilities. *Education and Training in Developmental Disabilities*, *41*, 333–343.

Browder, D., Ahlgrim-Delzell, L., Courtade, G., Gibbs, S., & Flowers, C. (2008). Evaluation of the effectiveness of an early literacy program for students with significant developmental disabilities. *Exceptional Children*, *75*, 33–52. doi:10.1177/001440290807500102

Browder, D. M., Ahlgrim-Delzell, L., Flowers, C., & Baker, J. N. (2012). An evaluation of a multicomponent early literacy program for students with severe developmental disabilities. *Remedial and Special Education*, *33*, 237–246. doi:10.1177/074193251038730

Browder, D., Ahlgrim-Delzell, L., Spooner, F., Mims, P. J., & Baker, J. N. (2009). Using time delay to teach literacy to students with severe developmental disabilities. *Exceptional Children*, *75*, 343–364. doi:10.1177/001440290907500305

Browder, D. M., Wakeman, S., Spooner, F., Ahlgrim-Delzell, L., & Algozzine, B. (2006). Research on reading for students with significant cognitive disabilities. *Exceptional Children*, *72*, 392–408. doi:10.1177/001440290607200401

California Department of Education (2018). *DataQuest*. Sacramento, CA: Author. Retrieved from https://data1.cde.ca.gov/dataquest/

Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework of systematic, conceptual replications. *Remedial and Special Education*, *37*, 244–253. doi:10.1177/0741932516648463

Doabler, C. T., Clarke, B., Krosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a Tier 2 mathematics intervention: A conceptual replication study. *Exceptional Children*, *83*, 92–110. doi:10.1177/0014402916660084

Dorman, J. P. (2009). Partitioning the variance in scores on classroom environment instruments. *Australian Journal of Educational & Developmental Psychology*, *9*, 18–31.

Elliott, J., Ferguson, P., & Riess, S. (2017). Starfall ELA Teacher's Guide, (2nd ed.). Boulder, CO: Starfall Education Foundation.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Engelmann, S., Becker, W. C., Hanner, S., & Johnson, G. (1980). *Corrective reading program*. Chicago, IL: Science Research Associates.

Every Student Succeeds Act of 2015, Pub. L. No. 114-95 §114 Stat. 117 (2015).

Fleming, D., Wilson, M., & Ahlgrim-Delzell, L. (2018). Using item response theory to describe the Nonverbal Literacy Assessment (NVLA). *Psychology in the Schools*, *55*, 341–349. doi:10.1002/pits.22110

Hedges, L. V. (2007). Correcting a significance test for clustering. *Journal of Educational and Behavioral Statistics*, *32*, 151–179. doi:10.3102/1076998606298040

Hudson, M. E., Browder, D. M. & Wood, L. A. (2013). Review of experimental research on academic learning by students with moderate and severe intellectual disability in general education. *Research & Practice for Persons with Severe Disabilities*, *38*(1), 17–29.

Hunt, P. (2019). Implementing comprehensive early literacy instruction in general education classrooms: A response to Toews' and Kurth's call to action. *Research and Practice in Severe Disabilities*, *44*, 147–152. doi:10.1177/1540796919862874

Individuals With Disabilities Education Improvement Act of 2004, 20 U.S.C. §§1400 et seq. (2004).

Lemons, C. J., Mrachko, A. A., Kostewicz, D. E., & Paterra, M. F. (2012). Effectiveness of decoding and phonological awareness interventions for children with Down syndrome. *Exceptional Children*, *79*, 67–90. doi:10.1177/001440291207900104

Lovett, M., Frijters, J., Wolf, M., Steinbach, K., Sevcik, R., & Morris, R. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. *Journal of Educational Psychology*, *109*, 889–914. doi:10.1037/edu0000181

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie Reading Tests, Fourth Edition, Forms S and T*. Itasca, IL: Riverside Publishing.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates-MacGinitie Reading Tests technical report*. Itasca, IL: Riverside Publishing.

McDonnell, J., Johnson, J. W., Polychronis, S., Riesen, T., Jameson, J. M., & Kercher, K. (2006). A comparison of one-to-one embedded instruction in general education classes with small group instruction in special education classes. *Education and Training in Developmental Disabilities*, *41*, 125–138.

National Center for Education Statistics. (2018). *Common Core of Data*. Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/ccd/schoolsearch/

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy. Retrieved from http://www.nifl.gov/earlychildhood/NELP/NELPreport.html

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Pub. No. 00-4769). Washington, DC: Government Printing Office.

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2002).

Petersen, A. (2016). Perspectives of special education teachers on general education curriculum access: Preliminary results. *Research and Practice for Persons With Severe Disabilities*, *4*, 19–35. doi:10. 1177/1540796915604835

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley. doi:10.1002/9780470316696

Ruppar, A. L., Gaffney, J. S., & Dymond, S. K. (2014). Influences on teachers' decisions about literacy for secondary students with severe disabilities. *Exceptional Children*, *81*, 209–226. doi:10.1177/0014402914551739

SAS Institute. (2002–2012). *SAS/STAT 9.4 user's guide*. Cary, NC: Author.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London, UK: Chapman and Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037/1082-989X.7.2.147

Scott Foresman (2011). *Reading Street*. Upper Saddle River, NJ: Pearson Education.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.

Timberlake, M. T. (2014). Weighing costs and benefits: Teacher interpretation and implementation of access to the general education curriculum. *Research and Practice for Persons With Severe Disabilities*, *39*, 83–99. doi:10.1177/1540796914544547

Toews, S., & Kurth, J. (2019). Inclusive literacy instruction: A call to action. *Research and Practice in Severe Disabilities*, *44*, 135–142. doi:10.1177/1540796919855373

Whalon, K. J., Otaiba, S. A., & Delano, M. E. (2009). Evidence-based reading instruction for individuals with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *24*, 3–16. doi:10.1177/1088357608328515

Authors' Note

This research was supported by Grant R324A150021 from the Institute of Education Sciences, National Center for Special Education Research. The article does not necessarily reflect the positions or policies of this funding agency, and no official endorsement should be inferred.

We thank the teachers and students who participated in this research. We wish to acknowledge the contributions of research team members Isabella Brown, Renee Starowicz, and Dorcas Yap.

Supplemental Material

Supplemental material for this article is available online.

Manuscript received December 2018; accepted September 2019.