

An Examination of Treatment Intensity with an Oral Reading Fluency Intervention: Do Intervention Duration and Student–Teacher Instructional Ratios Impact Intervention Effectiveness?

Sarah G. Ross · John C. Begeny

Published online: 3 September 2014
© Springer Science+Business Media New York 2014

Abstract With an increasing percentage of schools moving toward approaches to data-based instructional problem-solving and early remediation of learning difficulties, the development and execution of intervention plans often warrants the pragmatic question: How intensive should an intervention be so that it is effective, while also feasible and time efficient to implement? In other words, educators must prudently balance treatment intensity, anticipated effectiveness, and available implementation resources. This study examined the effects of an evidence-based reading fluency intervention that included the same instructional components but was implemented with varying treatment durations and student–teacher ratios. Using an alternating-treatments design, four second-grade struggling readers received four treatment conditions (small group with ~14 min of intervention; small group with ~7 min of intervention; one on one with ~14 min of intervention; and one on one with ~7 min of intervention) and a no-intervention control condition. Using three data-analytic strategies and two dependent measures, overall findings suggested that all intervention conditions led to reading improvements but that (a) the longer intervention duration appeared more effective than the shorter duration, and (b) there was little difference in intervention effectiveness between the small-group and one-on-one conditions.

Keywords Treatment intensity · Reading fluency · Reading intervention · Small group

S. G. Ross (✉) · J. C. Begeny
North Carolina State University, Raleigh, NC, USA
e-mail: Rossg1@gmail.com

Introduction

Far too many students in the USA do not meet proficiency standards in key academic areas, such as reading, math, and science. Educational data indicate that 60 % or more of US students scored below a Proficient level in these key academic areas when measured at fourth and eighth grade (National Center for Education Statistics 2011). In reading, 33 % of fourth-grade students and 24 % of eighth-grade students scored below even a basic level (National Center for Education Statistics). These statistics underscore the numerous challenges and complexities related to effective teaching. Teachers are expected to provide quality instruction in core academic areas; however, given that a large percentage of students struggle in at least one academic area, teaching strategies that engage an entire class at once are often not feasible or effective, as students often need more concentrated instruction to support their learning challenges (U.S. Department of Education 2010).

It is often the job of the regular education teacher to provide more intensive instruction for students with academic difficulties (Burns and Gibbons 2008), therefore putting a strain on already sparse resources. Sharp budget cuts in state departments of education mostly began in 2008–2009 as a result of the US recession (Johnson et al. 2010). For example, North Carolina public schools lost over 17,000 positions and laid off over 6,000 people between the 2008–2009 and 2011–2012 school years. Thirty-five percent of these losses were teacher positions, and 33 % were teaching assistant positions (NC Department of Public Instruction 2012). Yet, teachers are asked to respond to heightened, local, state, and national expectations for student achievement with fewer resources to support increased demands (Block 2010). In addition, loss of teaching positions often results in larger class sizes, which can further amplify the diversity of students' skill levels and learning needs within a given classroom, as well as require increased time and resources devoted to behavior management. As noted above, teachers are increasingly left to manage these multiple challenges with little to no assistance from teaching assistants.

Unsurprisingly, poor student achievement has also brought about reform efforts and new models of instructional problem-solving over the past couple decades, many of which are intended to ensure that struggling learners receive intervention as soon as a learning difficulty is identified. Examples of these models include (a) prereferral intervention teams designed to identify and provide support to at-risk students before they are referred to larger school-wide support teams for more complex intervention (Buck et al. 2003), (b) professional learning communities (DuFour et al. 2004), which place teachers into working groups to promote collegial learning and problem-solving of particular student needs, and (c) response to intervention (RTI), which includes providing intervention to students within the regular education setting, monitoring relative student progress, and later determining the need for additional specialized instruction (Burns and Gibbons 2008).

Consistent among the aforementioned models is the notion of providing students with targeted intervention services at the initial stages of identifying the learning difficulty and well before a student is considered for special education. But this process is not as straightforward as it may seem (Begeny et al. 2012b; Burns and Gibbons 2008) because educators must develop and execute a presumably effective

intervention plan that includes numerous features and components that require staff time, skills, and material resources. For example, as one piece of a more integrated approach to maximizing time and effectiveness when supporting struggling learners, Begeny, Schulte, and Johnson describe a system for intervention planning and delivery, and in developing *only* the intervention plan, educators need to consider at least 16 different features of delivering an intervention for a student or small group of students (e.g., what will the intervention be and who will implement it?; how many minutes of intervention and/or lessons will occur each day and for how many days per week?; how will student progress be monitored, with what form of assessment, how often, and by whom?; how will periodic “follow-ups” of intervention success occur, who will initiate them, and approximately how often will they occur?). In short, the development and execution of intervention plans warrants the pragmatic question: How intensive should an intervention be so that it is effective, while also being feasible and time efficient for educators in the school to implement with integrity and according to the documented intervention plan? Put differently, educators must prudently balance treatment intensity, anticipated effectiveness, and available implementation resources.

Treatment intensity can be conceptualized in a variety of ways, including the frequency, duration, and/or dosage (total number or weeks of sessions) of an intervention, the modality within which an intervention is conducted (one on one, small group, classroom based), and the ratio of students to teachers within an intervention setting (Begeny et al. 2011; Fuchs and Fuchs 2006). Alternative ways for conceptualizing treatment intensity may include factors such as opportunities for student response, frequency and immediacy of corrective feedback, the ecology of the classroom setting, and the integration of intervention techniques (Alberto and Troutman 2009; Greenwood 1991; Kamps et al. 2008).

Thus far, the research on treatment intensity, when considering it within the context of intervention resources, is still relatively sparse, with most studies focusing on the effects of treatment dosage as measured by differences in total hours or weeks of intervention on student reading outcomes (e.g., Begeny 2011; Ukrainetz 2009; Wanzek et al. 2013). For example, Wanzek et al. utilized a meta-analysis to examine the effects of intervention dosage among other factors on student reading outcomes. Studies were included that utilized interventions that lasted from 8 weeks to up to a full school year. The meta-analysis incorporated 10 studies that examined the effects of reading interventions targeted to students in fourth through twelfth grades and found that, among the included studies, there were no significant differences in student outcomes based on the dosage of the reading intervention. Yurick et al. (2012) also examined the effect of overall treatment duration among 38 kindergarten students who received varying total durations (between 6.85 and 13.70 h) of an early reading skills intervention program. Similarly, the authors did not find significant differences in student outcome based on the number of intervention hours received. Ukrainetz (2009) examined previously published intervention studies targeting phonemic awareness and tentatively determined that between 10 and 20 h of phonemic awareness instruction (which may include a combination of classroom instruction and targeted intervention) can result in children with language impairment achieving the phonemic awareness skills needed for reading and spelling.

Recent research has begun to examine the differential effects of treatment modalities that require various levels of instructional intensity, including small groups, one-to-one, and peer tutor deliveries (Begeny et al. 2011; Ross and Begeny 2011; Vaughn et al. 2003). For example, Vaughn et al. examined the differential effect of group size on second-grade student reading improvement, comparing one-to-one, one-to-three, and one-to-ten modalities. Results indicated that there was no difference in reading improvement between the one-to-one and one-to-three modalities, but that they both led to significantly greater reading improvements than the one-to-ten modality. Similarly, Begeny et al. (2011) examined the effects of treatment modality on the reading performance of 6 second grade students. The study compared the use of a standard reading fluency intervention package delivered in either a small group (6 students), one-to-one, or peer-dyad format. In general, results showed that there was little difference in student outcomes between the small group and one-to-one intervention modalities and that both of these formats resulted in better student performance than the peer-dyad modality. In another example, Ross and Begeny examined the relative effects of a reading fluency intervention delivered in either a small group (5 students) or one-to-one context when utilized with five second-grade, English language learners. Findings showed that all students made reading fluency gains when the intervention was delivered in the more intensive, one-to-one modality, but two students made meaningful gains when the small group modality was used.

Another way to conceptualize and evaluate treatment intensity is through the duration of the intervention session when the same instructional components are included within a specified instructional protocol, but one protocol integrates those components more or less frequently or for longer or shorter durations of time (e.g., four versus two rounds of repeated readings, or 1 vs. 3 min of modeling procedures). Length of intervention session therefore examines the number of total minutes an intervention is provided during each session while otherwise controlling for the same instructional components or strategies. In practice, a systematic evaluation of session duration (as an element of treatment intensity) has pragmatic value because it would be useful for practitioners to know, for example, if 10 min of using specific evidence-based instructional strategies would be equally effective for most students compared to 20 or 30 min of using those same strategies. To date, we are not aware of any studies that have examined this element of treatment intensity in the context of academic interventions. There have been several studies that involve a “layering” of evidence-based intervention components that are systematically added or removed—done in what is commonly referred to as a brief experimental analysis (e.g., Anderson et al. 2013; Dufrene and Warzak 2007)—and this process of adding or removing intervention components can thereby alter the duration of an intervention session. But brief experimental analysis studies and procedures are not aimed at comparing specific or “standard” intervention protocols, such as the standard intervention protocols that are more commonly across multiple students and used in day-to-day educational practice. Instead, brief experimental analysis is usually used as a means to identify the most effective and efficient intervention protocol for one specific student. Our conceptualization of session duration as a form of treatment intensity applies to standard intervention

protocols that are research based according to the existing literature and can be evaluated as two or more standard protocols that differ only by duration of implementation.

Purpose

This study attempts to add to the treatment intensity research literature by further investigating instructional techniques that could be compared according to relative effectiveness and efficiency. Specifically, the overall purpose was to examine the effects of an evidence-based reading fluency intervention that included the same instructional components but was implemented with varying treatment durations and student–teacher ratios. We examined different student–teacher ratios (small group and one-to-one) because these are commonly used intervention formats in schools, and additional research is needed to examine the relative effectiveness of these instructional contexts. We examined treatment duration (~7 vs. ~14 min per intervention session) because there is no previous research examining this aspect of treatment intensity, so increased research about the effects of intervention duration should have practical value for education researchers and practitioners.

Given the overall purpose to examine the relative effectiveness of the selected interventions, there were two primary research questions. First, which of the interventions will evidence efficacy compared to a control condition? Based on previous research supporting the effects of the intervention components we used in our study (e.g., Morgan and Sideridis 2006; Therrien 2004), we hypothesized that the intervention conditions would be more effective than the control condition. Our second main research question (are any of the intervention conditions relatively more efficacious than the others?) generated two particular questions of interest, both based on our manipulation of treatment intensity according to student–teacher ratios and duration of implementation. Specifically, we wanted to know whether one-on-one interventions were more or less effective than small group interventions and whether the longer (14 min) interventions were more or less effective than the shorter (7 min) interventions. Based on a very small body of research (e.g., Klubnik and Ardoin 2010; Ross and Begeny 2011), we hypothesized that for some students, but not all, the small group intervention would be equally or more effective than the one-on-one intervention. We did not have specific hypotheses regarding the relative effectiveness between longer and shorter interventions because we could not find prior research evaluating this topic in the context of reading fluency.

Method

Participants and Setting

Student participants included four second-grade students from three classrooms in one rural school in the Southeast. Students (Lisa, D’Andra, Michael, and Ben) were randomly selected for this study among a list of 121 low-performing second-grade readers. Other low-performing readers were randomly assigned to separate

intervention research studies taking place in the participating school, and all parents of the students consented to their students' participation. Of the 4 participants in the present study, 2 were female, 2 were African-American, 2 were Caucasian, and their average age was 8.0 years. None of the students received English as a second language services or special education services, and none received remedial support outside of what was provided for the purposes of this study. Within the participating school, 33 % of students received free or reduced-price lunch. Core reading instruction across all second-grade teachers in the participating school was similar. All teachers integrated language arts into their daily curriculum for approximately 90 min, each utilized a *Houghton Mifflin* basal reading series, and each included daily reading groups, independent reading, phonics and vocabulary lessons, and writing activities. However, at the time of this study, reading fluency was not targeted in students' core curriculum.

In addition to the students being identified by their teacher as being approximately one-grade level behind in reading and needing additional assistance with reading fluency, student participants met screening criteria during a winter benchmark assessment that occurred just prior to the start of the project. As indicated by their second-grade oral reading fluency scores and the oral reading fluency norms reported by AIMSweb (2013), all participants read between the 15th and 35th percentile (range 51–74 words read correct per minute).

Throughout the course of the study, all study conditions and procedures were implemented by one of four members of the research team (henceforth referred to as *interventionists*). All aspects of this study occurred in an empty classroom within the school that was free from noise and distractions. Each interventionist learned the instructional protocol for each condition, and they shared implementation responsibilities equally throughout the duration of the project, which was done to minimize the potential influence of interventionist characteristics on student performance. Before implementation of any condition, all interventionists were trained on all procedural roles. This involved didactic instruction by the project lead, monitored practice with peers, and finally, demonstration of mastery criterion. Mastery was observed by the interventionists role-playing each condition individually while the project lead marked each required step of the procedure as it was completed. Interventionists were required to complete 100 % correct implementation of each condition in two consecutive practice trials in order to demonstrate mastery.

Materials

Instructional Passages

Throughout the project, a separate reading passage was used during each session. Reading passages were developed from the Silver, Burdett, and Ginn (Pearson et al. 1989) reading series and represented second- and third-grade reading material. Each passage contained approximately 150 words and was prepared in the standard format for oral reading fluency passages (see Shinn 1989). The passages chosen were at the second- to third-grade reading level, according to the Spache (1953) readability formula. Readability levels ranged from 2.85 to 3.99, with an average

readability of 3.34 ($SD = .35$). This level of passage difficulty was purposefully selected so that materials would be at and slightly above students' reading level, which is a common procedure for reading fluency intervention studies (e.g., Begeny et al. 2009; Begeny and Silber 2006).

Word-Overlap Passages

Word-overlap passages were developed by creating stories with similar words to those words found in the “matched” instructional passage. Specifically, each instructional passage was “dissected” into an alphabetical list of all unique words from the passage. Then, using approximately 80 % of the words from that passage, a new story (i.e., the matched word-overlap passage) was written. Consequently, there was one matched word-overlap passage for each instructional passage used in the study. By creating word-overlap passages in this way, each passage had a high proportion of the same words and an almost identical readability level compared to its matched instructional passage, but the content of the word-overlap passage (e.g., theme, plot) differed from the matched instructional passage. The average word overlap of the word-overlap passages was 80.7 % ($SD = .40$; range 80.0–81.8 %). Readability levels of the word-overlap passages ranged from 2.88 to 3.98, with an average readability of 3.31 ($SD = .33$). As intended, the readability levels between each word-overlap passage and its matched instructional passage were strongly correlated ($r = .95$). Also, like the instructional passages, the word-overlap passages contained approximately 150 words and were prepared in the standard format for oral reading fluency passages.

Experimental Conditions and Design

This study included four intervention conditions that were designed to improve students' reading fluency, as well as a no-intervention control condition (Control). Intervention conditions included: (a) a small-group “longer” intervention (SG-L) that lasted 12–15 min; (b) a small-group “shorter” intervention (SG-S) that lasted 6–8 min; (c) a one-on-one (i.e., interventionist-student) longer intervention (1:1-L) that lasted 12–15 min; and (d) a one-on-one shorter intervention (1:1-S) that lasted 6–8 min. As will be described later, all intervention conditions integrated the same basic instructional strategies (e.g., repeated readings, systematic error correction, modeling of the passage by an adult) and motivational components (e.g., rewards for student on-task behavior). Differences between conditions only varied according to the duration (longer vs. shorter) and instructional context (SG vs. 1:1), therefore allowing us to examine treatment/intervention intensity according to intervention duration and student–teacher instructional ratios. The durations of the interventions were selected because they are consistent with other reading fluency interventions published in the literature (e.g., Begeny 2011; Klubnik and Ardoin 2010).

An alternating-treatments experimental design was used to assess the relative effectiveness of the four intervention conditions and the control condition. To adhere to the basic principles of this design, the order of conditions was counterbalanced and done so in such a way that (a) each student received the

same condition on the same day and (b) each condition occurred at least once in the beginning, middle, and end of the study. The study was designed so that each student would receive five sessions per intervention condition and four control condition sessions. Due to absences, Michael missed one session of SG-L and one session of SG-S. Lisa missed one session of SG-S. Ben and D'Andra did not miss any sessions.

Procedures

All condition sessions occurred during the morning on Mondays, Wednesdays, and Fridays across approximately 8 weeks during the winter months. Procedures across all intervention conditions (SG-L, SG-S, 1:1-L, and 1:1-S) were developed so that each condition integrated the same research-supported instructional and motivational components. Specifically, each condition included (a) the interventionist reading the day's instructional passage aloud to the student (i.e., Modeling); (b) repeated readings (RR), where the student reads the day's selected passage aloud multiple times during the session; (c) the student orally retelling the day's passage aloud after having read it (i.e., Retell); (d) the student practicing words that were read incorrectly during RR by practicing the incorrect word within the 3–7 word phrase that included the incorrect word (i.e., phrase-drill error correction; PD), and practicing each respective phrase three consecutive times; and (e) a reward procedure that functioned as a token economy to reinforce students' on-task behavior and effort. Several previous studies (e.g., Begeny 2011; Begeny et al. 2012a; Klubnik and Ardoin 2010) and meta-analyses (Morgan and Sideridis 2006; Therrien 2004) have evidenced the effectiveness of these procedures in a one-on-one and/or small group instructional context.

What follows is an elaborated description of the instructional procedures and how they were sequenced in each of the intervention conditions. When RR is specified, this means the student (or group of students) orally read the day's passage aloud, but "RR" is noted because the multiple oral readings of the same passage, within the same session, reflects the repeated reading procedure.

SG-L and SG-S Conditions

The sequence of procedures during the SG-L condition was as follows: Modeling, RR, Retell, PD, RR, Modeling, RR, PD, RR, enthusiastic praise and feedback (described later), immediate assessment (described later), and reward (described later). The sequence of procedures for the SG-S condition was as follows: Modeling, RR, Retell, PD, RR, enthusiastic praise and feedback, immediate assessment, and reward. With respect to how these procedures were implemented in small groups, the modeling procedure involved the interventionist reading the passage aloud at a pace only slightly faster than students' oral reading speed and randomly calling on one of the four students in the group to read the next word in the passage, and then continuing with the model reading. This "pause" aspect of modeling occurred 5–7 times during the reading and has been used and specifically evaluated in previous research (e.g., Begeny et al. 2009, 2014). During the oral

reading (i.e., RR) procedure, students read the story aloud chorally at the same pace and in “quiet voices,” with a randomly selected student leader guiding the reading by reading slightly louder than the other three students in the group. During this procedure, which lasted 1.75 min, the interventionist made sure all students were reading aloud and at the same pace. During Retell, the interventionist randomly called on each student once and asked the student to say one thing he/she remembered about the passage that was just read, while also encouraging students to say what they remembered in the correct sequence of what occurred in the story. Finally, during PD, students chorally read the respective phrase three times, after the interventionist specified the phrase to be read and the incorrectly read word from that phrase.

1:1-L and 1:1-S Conditions

The sequence of procedures during the 1:1-L condition was as follows: Modeling, RR, Retell, PD, RR, Modeling, RR, PD, RR, enthusiastic praise and feedback, immediate assessment, and reward. The sequence of procedures for the 1:1-S condition was as follows: Modeling, RR, Retell, PD, RR, enthusiastic praise and feedback, immediate assessment, and reward. The execution of each of these procedures emulated the descriptions above, except that all procedures only involved the one student and interventionist. For example, during modeling, the interventionist paused 5–7 times while reading aloud to ask that one student to read the next word in the passage.

No-Treatment Control Condition (Control)

During the control condition, reading instruction procedures were not used and students only completed the session assessment procedures (described below). To account for possible intervention effects not specifically related to the instructional procedures (e.g., the student’s time with the interventionist, the student’s time out of class), the student was given math fluency worksheet to complete for 10 min. Also, interventionist praise was provided for student effort during the control condition, and the student could earn up to two stars on his or her star chart (described in the next section). We included these motivational components during the control condition because we were most interested in examining the relative effects of the instructional procedures within each condition and therefore held the motivational components consistent across all five conditions.

Motivation Components (Praise, Feedback, and Rewards)

As stated previously, motivational strategies were integrated into each of the intervention conditions because a critical element to improving fluency of any skill includes motivation and (usually) planned motivational strategies (e.g., Haring et al. 1978). One aspect of motivation included the interventionist providing the student(s) with enthusiastic praise and specific feedback about students’ reading skills and/or effort (e.g., “You did such a great job reading aloud with speed,

accuracy, and good expression” or “I really appreciate that as I read out loud you were all following along and knew the correct place in the story when I paused at a word”). Based on students’ effort during the session (e.g., visible signs each student, or the group of students, was following instructions and putting forth their best effort), the students could earn 0, 1, or 2 stars on their “star chart” (Begeny 2009). During SG-L and SG-S sessions, stars were awarded based on the entire group’s effort (i.e., all students in the group would earn the same number of stars on his or her individual star chart, based on the performance of the group). When a student received 15 stars, he or she was able to select a small, inexpensive reward (e.g., pencil, sticker). In all sessions (including control condition sessions), students demonstrated good effort and on-task behavior, and therefore, each session resulted in students receiving two stars.

Immediate and Retention Assessment Procedures

All assessment procedures were conducted one on one (student with interventionist) in a location free from noise and distractions. Both the immediate and retention assessments involved the standard assessment procedures (e.g., Shinn 1989) for curriculum-based measures of reading (CBM-R). Specifically, the student read aloud for 1 min while the interventionist scored words read incorrectly (i.e., mispronunciations, hesitation, substitutions, omissions, and transpositions of words). If the student came to a word that he or she did not know, the interventionist told the word to the student after 3 s and marked that word as an error. The primary score obtained from this assessment procedure was the number of words read correctly per minute (WCPM). Concurrent validity for the CBM-R technique has been reported in the medium to high range (.70–.90; Marston 1989).

For all five conditions, assessment procedures occurred during each session and on two specific occasions: (a) immediately prior to the instructional procedures (and for the control condition, just before students were given the math worksheet and instructions), and (b) at the end of the day’s instructional procedures (or at the end of the math worksheet time for the control condition) and just before students were awarded stars on their star chart. The purpose of the assessments *prior to instructional procedures* was to obtain the student’s “cold read” WCPM score (i.e., the number of WCPM on the student’s very first reading of the passage that would be practiced in that day’s session) and a retention score. The student began this assessment period by reading the passage that was used in the most previous session (approximately 2 days earlier), and this allowed our research team to establish a retention score (i.e., the WCPM score on this reading minus the WCPM score from the cold read of that passage approximately 2 days earlier). The retention score therefore measured how much oral reading fluency growth (as measured in WCPM) occurred from the condition procedures from the previous session. After the reading to obtain the retention score, the student would do the cold read of the passage selected for that’s day’s session.

On the second assessment occasion of each session (i.e., just before students received stars on the star chart), students would read the word-overlap passage associated with the instructional passage selected for that’s day’s session. Both the

retention assessment and word-overlap passage assessment measured aspects of skill generalization. With the former, we measured retention of potential oral reading fluency gains over time (i.e., over approximately 2 days since sessions occurred Mondays, Wednesdays, and Fridays). With the latter, we measured the immediate effects of the condition procedures on similar, but different, reading material.

Implementation Integrity and Inter-Scorer Reliability

Implementation integrity of the conditions was observed on 35 % of the 1:1-L condition, 35 % of the control condition, 55 % of the 1:1-S condition, and 100 % of both the SG-L and SG-S conditions. Implementation integrity was completed by having a second member of the research team either directly observe the session or listen to an audio-recorded version of the session. All members of the research team were observed for implementation integrity approximately the same number of times. The implementation protocols/checklists that were developed for training and implementation purposes were also used to evaluate implementation integrity. Average implementation integrity (across all interventionists) for each of the conditions was high (i.e., Control = 100 %; 1:1-L = 100 %; 1:1-S = 97.0 %; SG-L = 97.1 %; and SG-S = 98 %). Inter-scorer agreement of the timed readings that occurred during assessment procedures was examined through audio recordings. Across all assessment passages administered (i.e., cold reads, the retention assessment, and the word-overlap assessment), inter-scorer agreement was calculated for 28.7 % of the readings. Average agreement was 99.7 %.

Data Analysis

Retention and word-overlap scores served as our two dependent variables and were determined by calculating WCPM gain scores. As described above, assessment procedures required a cold read of the target passage for each session, which resulted in the precondition WCPM score. At the end of each session, students were asked to read a word-overlap passage, which included ~80 % of the words from the target passage, and this resulted in the word-overlap WCPM score. Finally, at the beginning of the following session, procedures required students to again read the target passage from the previous session, resulting in the retention WCPM score. Retention assessment gain scores were calculated by subtracting the precondition WCPM score from the retention WCPM score (henceforth, this will simply be referred to as the *retention gain score*). Word-overlap gain scores were calculated by subtracting the precondition WCPM score from the word-overlap WCPM score (henceforth, this will simply be referred to as the *word-overlap gain score*).

To best compare the differential effects of the five conditions using students' retention and word-overlap gain scores, we used three separate types of single-case design data analysis strategies. Using visual inspection, each student's gain scores were examined at the individual session level. Mean gain scores per condition were used for the standard error of measurement (SEM) analysis and randomization tests.

Visual Inspection

Visual inspection is the most common method of data analysis with single-case designs. Using specific criteria developed by Franklin et al. (1997), each graph was analyzed using three principles of visual inspection: the central location of the data, the variability among the data, and trend location within each group of data (when appropriate). When clear differentiation between conditions emerged, the percentage of non-overlapping data (PND) was calculated. This was calculated by determining the highest score in one condition (e.g., Control) and calculating the percentage of scores in a comparison condition (e.g., 1:1-L) that fall above the identified score (procedure described in Scruggs et al. 1986). This describes the percentage of scores in one condition (e.g., 1:1-L) that do not overlap with the second condition (e.g., Control).

Standard Error of Measurement (SEM) Analysis

To analyze retention and word-overlap gain scores at a mean level across conditions, a conservative approach that utilized the expected standard error score was employed. According to data reported in the Christ and Silberglitt (2007), for a typical second-grade sample with high levels of experimental control, one can assume 6–7 WCPM attributed to SEM on any given CBM-R assessment. In accordance with the *t*-distribution, the SEM and a critical *t* score of 1.65 (for a *p* value of .05) were used to calculate a significant WCPM gain score for a one-tailed test. Therefore, $1.65 \times 2 \times 7$ results in a gain score of 23.1 being considered a significant gain 95 % of the time. Although researchers have differing opinions about the use of *t* tests with data from the same individual, many argue that it is an appropriate way to compare changes within an individual (e.g., Busk and Marascuilo 1992). This level of analysis was the most accurate way to make use of the SEM data reported by Christ and Silberglitt and helped us better understand trends in our data.

Randomization Test Analysis

Randomization tests are statistical tests whose validity is based on the random assignment of units to treatments (Bulté and Onghena 2008). Using permutations of the order of the data, these tests determine whether the same results would have been obtained if the data were randomly assigned to rearranged placements (Busse et al. 1995). Randomization tests are nonparametric and consequently are not based on distributional assumptions, assumptions about the homogeneity of variances, or the independence of residuals (Wilson 2007). To use randomization tests, treatment conditions must be randomly assigned. Thus, we randomly assigned conditions (1:1-L, 1:1-S, SG-L, SG-S, and Control) to the set of possible days we could work with the participants. This makes randomization tests a valid analytic tool without having a random sample of participants (Edgington 1980; Todman and Dugard 1999).

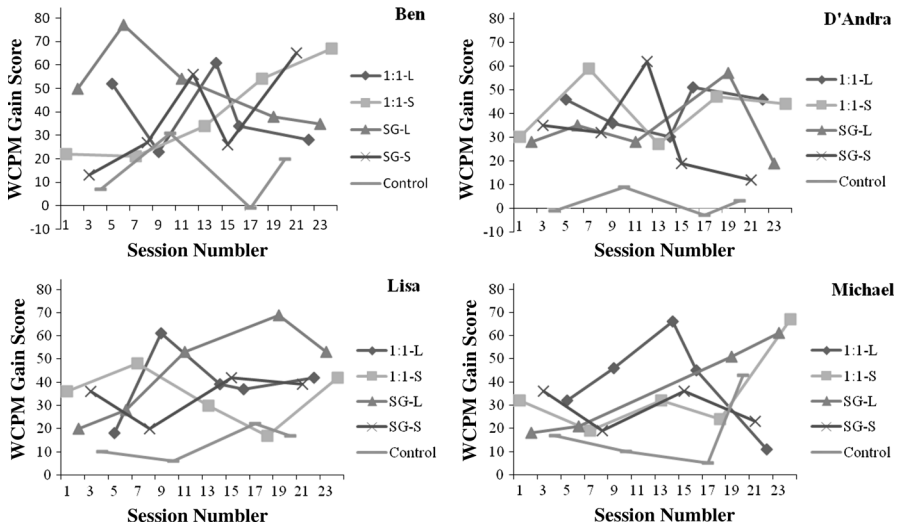


Fig. 1 WCPM gain scores during retention assessment

Results

Visual Inspection

Figure 1 displays retention gain scores for each participant. Graphs were analyzed visually and by using the percentage of non-overlapping data (PND) technique. Based on commonly used considerations and guidelines for interpreting PND scores (e.g., Scruggs and Mastropieri 1997; Scruggs et al. 1987), PND scores of 60 % were interpreted to show moderate differential effectiveness, while PND scores at or above 80 % were interpreted as highly differentially effective. For D’Andra, all treatment conditions were associated with higher gain scores than the control condition, with 100 % PND between treatment conditions and the control condition. For Lisa, the 1:1-L and SG-L conditions resulted in 80 % PND compared to the control condition, while the SG-S and 1:1-S conditions resulted in 75 and 60 % PND (respectively) compared to Control. For Michael, the 1:1-L and SG-L conditions resulted in 60 and 50 % PND when compared to the control condition, but there was substantial overlap between the two shorter conditions and Control. For Ben, the SG-L condition resulted in higher gain scores than the control condition, with 100 % PND present between the two conditions. Ben’s 1:1-L and 1:1-S conditions resulted in 60 % PND with Control, and his SG-S condition overlapped 40 % with the control condition. Across students, the two longer conditions appeared to most consistently outperform the control condition, with the SG-L condition doing this most frequently. Visual inspection of retention gain scores also indicated that for each student, there was extensive overlap between the treatment conditions, with none of the comparisons between treatment conditions exceeding 60 % PND.

Figure 2 shows word-overlap gain scores for each student. Consistent with scores from the retention assessment, the 1:1-L and SG-L conditions resulted in more

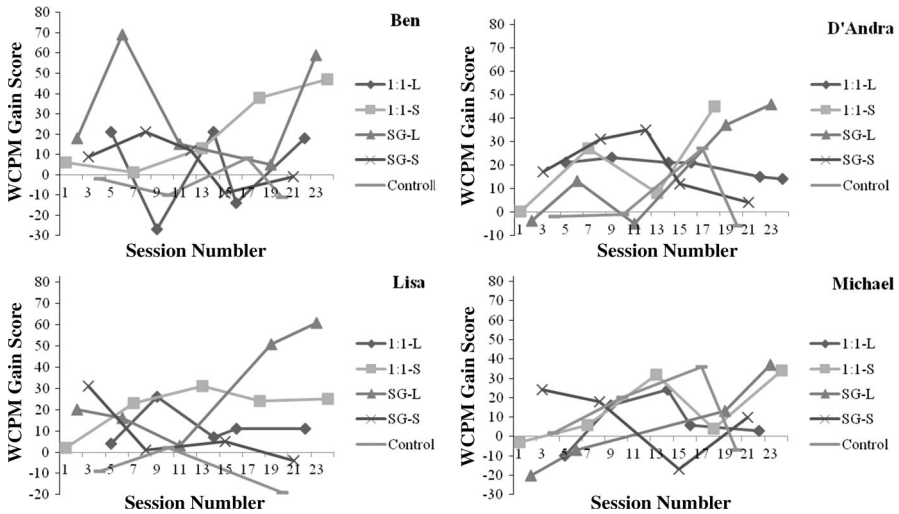


Fig. 2 WCPM gain scores during word-overlap assessment

WCPM gains than the control condition for Lisa (100 % PND between the treatment conditions and Control) with the 1:1-S and SG-S resulting in 80 and 50 % PND with Control, respectively. Similarly for Ben, the SG-L condition resulted in 80 % PND with Control, and the 1:1-L, 1:1-S, and SG-S conditions each resulted in 60 % PND with Control. There was substantial overlap between all conditions for D’andra and Michael.

SEM Analysis

Table 1 reports mean gain scores across each condition and both dependent variables (retention and word-overlap scores). Using the SEM analytic approach, scores greater than 23.1 are indicated with an asterisk. For retention gain scores, each condition resulted in significant mean gains for all students. Only the SG-L condition was associated with significant word-overlap mean gains, and these gains were significant for two students: Ben and Lisa. The control condition did not result in any significant mean gain scores for either retention or word overlap. Collectively, the findings from SEM analysis were generally commensurate with findings from the visual inspection; all treatment conditions were more effective than the control condition according to the retention gain scores, and the SG-L condition appeared relatively more effective for two of the students on the word-overlap measure.

Randomization Test Analysis

Table 2 shows the differences in retention mean gain scores across the five conditions, as well as the findings from the randomization tests. Findings did not indicate significant differences between treatment conditions at retention. When

compared to Control, each treatment condition resulted in significantly larger mean gain scores for D'Andra. The 1:1-L, SG-L, and SG-S mean scores were significantly larger than Control for Lisa, and the 1:1-L and SG-L conditions were significantly larger than Control for Ben. Michael's retention mean gain scores were not significantly larger than his Control scores when analyzed using randomization tests.

Table 3 shows the differences in word-overlap mean gain scores across the five conditions, as well as the findings from the randomization tests. Similar to the retention gain scores, randomization tests of the mean word-overlap scores did not indicate significant differences between treatment conditions. Lisa's mean word-overlap gain scores for the 1:1-L, 1:1-S, and SG-L conditions were all significantly larger than her Control mean word-overlap gain scores. Additionally, for Ben, the SG-L condition resulted in a larger mean word-overlap gain score compared to Control. For D'Andra and Michael, randomization tests did not indicate significant differences between mean word-overlap gain scores between any of the conditions. Overall, the findings from the randomization tests were similar to the findings from the SEM and graphical analyses.

Summary of Findings

Data from this study suggest that for at least one of the two dependent variables, all students benefitted from at least one of the treatment conditions, compared to Control. Although there were no statistically (or visually) significant differences between the four treatment conditions for any of the students, there were clear patterns of differential effectiveness and these patterns differed per student. For Lisa, the SG-L and 1:1-L conditions appeared most effective when considering retention gain scores to the control condition, but for word-overlap gain scores, the SG-L condition was consistently shown to be most effective across each of the analytic methods. For Ben, the SG-L condition appeared most effective when taking into consideration each of the analytic methods and his performance on both the retention and word-overlap gain scores. With D'Andra, her performance during each of the treatment conditions exceeded that of the control condition, but this was only true for the measure of retention gain scores; none of the treatment conditions evidenced word-overlap gain scores that were significantly better than her performance during the control condition. Finally, for Michael, the SEM analysis suggested that his retention gain scores during each of the treatment conditions were superior to his scores during Control, and visual analysis evidenced relatively positive effects for the 1:1-L condition, but the randomization tests did not reveal significant effects for any of the treatment conditions compared to the Control. Also, consistent across each of the analytic methods, Michael's word-overlap gain scores did not meaningfully differ across any of the five conditions.

Discussion

Educators are increasingly asked to respond to local, state, and national expectations for improved student achievement with fewer resources to support these

Table 1 Mean WCPM gain scores across conditions and dependent variables

	Retention		Retention		Retention		Retention		Retention		Retention		Retention		
	1:1-L	1:1-S	1:1-S	SG-L	SG-S	control	1:1-L	1:1-S	SG-L	SG-S	control	1:1-L	1:1-S	SG-L	SG-S
Ben	39.6* (16.2)	36.2* (15.4)	50.8* (16.7)	37.4* (22.0)	14.25 (14.1)	3.8 (22.7)	21 (20.3)	33.2* (28.7)	6.4 (11.7)	-3.75 (8.8)					
D'Andra	41.8* (8.6)	38.2* (14.2)	33.4* (14.4)	32* (19.2)	2 (5.3)	20.2 (3.0)	18.8 (17.7)	17.4 (23.35)	19.8 (13.0)	4.5 (15.2)					
Lisa	39.4* (15.3)	30.2* (12.5)	44.6* (20.1)	34.25* (9.8)	13.75 (7.1)	11.8 (8.5)	16 (14.1)	30.2* (24.6)	8.25 (15.6)	-8.7 (10.5)					
Michael	40* (20.3)	25* (6.8)	37.8* (21.5)	28.5* (8.8)	18.75 (16.9)	7.8 (13.0)	14.6 (17.1)	5.75 (24.9)	8.75 (18.1)	12.75 (19.1)					
Mean across all students	40.2* (1.1)	32.4* (6.0)	41.6* (7.7)	33.0* (3.7)	12.2 (7.2)	5.9 (5.0)	17.6 (2.9)	21.6 (12.6)	10.8 (6.1)	1.2 (9.4)					

SD shown in parentheses

* = mean gain score >23.1. Retention = mean retention gain score; Overlap = mean word-overlap gain score; 1:1-L = one-on-one long condition; 1:1-S = one-on-one short condition; SG-L = small group long condition; SG-S = small group short condition

Table 2 WCPM randomization test results for retention assessments

	1:1-L versus 1:1-S	SG-L versus SG-S	1:1-L versus SG-L	1:1-S versus SG-S	1:1-L versus control	1:1-S versus control	SG-L versus control	SG-S versus control
Ben	3.4	13.4	-11.2	-1.2	25.4*	22.0	36.6**	23.2
D’Andra	3.6	1.4	8.4	6.2	39.8**	36.2**	31.4**	30.0**
Lisa	9.2	10.4	-5.2	-4.05	25.7*	16.3	30.9*	20.5*
Michael	15.0	9.3	2.3	-3.5	21.3	6.3	19.0	9.8

* $p \leq .05$, ** $p \leq .01$

Table 3 WCPM randomization test results for word-overlap assessments

	1:1-L versus 1:1-S	SG-L versus SG-S	1:1-L versus SG-L	1:1-S versus SG-S	1:1-L versus control	1:1-S versus control	SG-L versus control	SG-S versus control
Ben	-17.2	26.8	-29.4	14.6	7.6	24.8	37.0*	10.2
D’Andra	1.4	-2.4	2.8	-1.0	15.7	14.3	12.9	15.3
Lisa	9.2	22.0	-18.4	7.8	20.5*	24.7*	38.9*	16.9
Michael	-6.8	-3.0	2.1	5.9	-5.0	1.85	-7.0	-4.0

* $p \leq .05$, ** $p \leq .01$

expectations. When developing interventions to support student needs, the importance of balancing treatment intensity, anticipated effects, and available implementation resources cannot be overstated. This study examined the effects of an evidence-based reading fluency intervention that included the same instructional components but was implemented with varying resource requirements, including time (measured through intervention session duration) and interventionists (measured through student–teacher ratio). The results of this study indicate that all students responded positively, by making significant WCPM gains to at least one of the treatment conditions (1:1-L; 1:1-S; SG-L; SG-S). This finding is consistent with previous research showing that packaged evidence-based reading intervention strategies (Modeling, RR, Retell, PD, enthusiastic praise and feedback, and reward) are effective in promoting reading fluency gains when used in a one-on-one or a small group context (e.g., Begeny 2011; Malloy et al. 2007; Ross and Begeny 2011).

Although there were no statistically significant differences between treatment conditions for any of the students, the relative performance of some treatment conditions compared to the control condition suggested that some treatments may be more effective. Although patterns of differential treatment effectiveness varied across students, the findings generally suggested that the longer treatment conditions had more of an impact on student performance than the shorter treatment. This was largely true when student performance was measured during retention assessment and true for at least two students during the word-overlap assessment. This indicates that increased exposure (14 vs. 7 min) to evidence-based reading intervention

techniques seems to result in better student outcomes. The briefer intervention periods may have been relatively less effective because they were briefer and allowed fewer opportunities to respond and practice the instructional material, or because more of the intervention component (e.g., more repeated readings) may have been needed.

Although results indicate that longer exposure to evidence-based strategies led to increased student reading gains, in the case of this study, the longer condition was still relatively brief at approximately 14 min. Given that many reading intervention programs require significantly more time for implementation (e.g., approximately 45 min for *Corrective Reading*, Engelmann et al. 2008; approximately 30 min for *REWARDS*, Archer et al. 2005), the 14 min intervention package was relatively efficient in supporting student gains, even for the 1:1 condition. This is not to suggest that the fluency building interventions used in this study are comparable or should be replaced with evidence-based programs such as *Corrective Reading*, *REWARDS*, or other research-supported intervention programs, but it highlights that (a) the longer intervention in this study was still relatively time efficient, and that (b) approximately 12–15 min may be needed to demonstrate meaningful improvements in oral reading fluency. In other words, we discourage educators from trying to treat reading fluency deficits with interventions as short as 7 min because using just a little more time appears to be worth the relative effectiveness.

When examining the effects of the longer treatment conditions, there did not seem to be a clear advantage to the 1:1 treatment compared to the SG instructional context, as one might expect from a treatment intensity perspective, whereby 1:1 intervention is generally considered more intensive. For Michael, there was a little evidence (based on visual analyses) that the 1:1-L condition was slightly more effective than SG-L procedures, but for two students (Ben and Lisa), the SG intervention appeared relatively more effective than the 1:1 condition. The fact that multiple students, but not all students, seem to benefit as much or more from SG interventions that are structurally similar to 1:1 interventions is continuing to be a consistent trend in research that evaluates the effects of SG interventions targeting reading fluency (e.g., Begeny et al. 2011, 2012a; Klubnik and Ardoin 2010; Ross and Begeny 2011). This repeated finding has important implications for educators, as it may be possible to provide sound intervention for reading fluency for four or more students at a time, and for most (but not all) of those students, this SG intervention should be as effective as providing individualized intervention for each student. The SG-L intervention protocol described in this study may therefore be one viable option for targeting reading fluency deficits using SG instruction, but based on other factors (such as access to all implementation materials and guidance during all intervention scenarios), we also encourage educators to consider more manualized SG reading fluency programs. At the present time, we know of only one such program and it is described by Begeny et al. (2012a).

Although this study offers encouraging results to support the use of targeted reading fluency interventions for students with reading difficulties, one limitation is that it only measured student reading outcomes on two generalized measures of reading fluency and did not incorporate additional measures (e.g., assessments of comprehension or vocabulary) that could also help to evaluate the relative

differences in effectiveness between treatment conditions. At this time, there are little to no options of psychometrically sound measures of comprehension or vocabulary that are sensitive enough to change to be used as part of an alternating-treatments design like we used in the present study. However, future research might use group designs (and multiple measures of reading performance) to examine the kinds of treatment intensity variables measured in this study, as group designs have been used previously to examine aspects of treatment intensity such as student–teacher ratios (e.g., Vaughn et al. 2003) and intervention implementation frequency per week (e.g., Begeny 2011).

Additionally, although the treatment condition schedule was initially randomized and passages were randomly assigned to those conditions, each student followed the same schedule and utilized the same order of passages. This was largely due to the need for all students to be in the small group conditions at the same time in order to have enough students to comprise a small group. However, this lack of randomization can lead to potential confounds, such as time and passage effects, and therefore, represents a limitation in the study. Future research should utilize larger pools of students in order to fully randomize students to treatment conditions and passages.

Future research is also needed to further examine the effects of varying levels of treatment intensity, including the effects of increasing one aspect of treatment intensity while decreasing another. For example, would a SG intervention with a student–teacher ratio 10:1 and implemented 5 days per week lead to similar improvements to a SG intervention with a 5:1 ratio administered 3 days per week? Or, would a 10 week intervention administered in a SG format produce similar improvements as a 5 week intervention administered in a 1:1 format? Educators are continuously faced with creatively balancing combinations such as these, and additional research is needed to inform the most effective and resource efficient practice.

Finally, future studies should examine treatment intensity as it relates to students at varying levels of educational need. The students in this study were identified as low-performing readers, but there are many students who need targeted reading support who are not represented in our small sample of struggling readers. Future research should therefore examine treatment intensity with students of varying characteristics (e.g., those with severe and less severe reading difficulties, those with attention difficulties, students of different grade levels, etc.). Furthermore, future research might examine different intervention combinations and/or components that aim to improve reading fluency, perhaps with particular consideration about the time needed to implement that component within the treatment package. For example, including a RR component in a fluency intervention has been well advised for many years based on research examining RR (e.g., Therrien 2004), but recent evidence suggests that the effects of RR may be small for a subpopulation of students. In a recent report from the US Department of Education (What Works Clearinghouse [WWC] 2014), the two studies that met the WWC group design standards showed only small effects for RR when used with grade 5–12 students classified with a learning disability.

Despite its limitations, this study adds to the literature by providing a systematic evaluation of two types of treatment intensity and one in particular (session duration) that has otherwise not previously been evaluated. Furthermore, the findings from this study suggest that SG interventions of approximately 12–15 min may be the most effective and efficient place for educators to begin when trying to improve reading fluency for students with reading challenges that are similar to those of the participants in our study. With this said, there is paramount need for continued research on the numerous and complex aspects of understanding the most optimal balance of intervention effectiveness and efficiency. This is critical because in today's complicated educational environment, there are far too many students performing below even basic levels in core academic areas, and there are far too many educators being asked to support these students without sufficient resources and adequate guidance for achieving instructional effectiveness and efficiency.

References

- AIMSweb. (2013). *Reading-CBM*. Retrieved August 25, 2013 from <https://aimsweb.pearson.com/Report.cfm>.
- Alberto, P. A., & Troutman, A. C. (2009). *Applied behavior analysis for teachers* (8th ed.). Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Anderson, M. N., Daly, E. J., & Young, N. D. (2013). Examination of a one-trial brief experimental analysis to identify reading fluency interventions. *Psychology in the Schools, 50*, 403–414.
- Archer, A., Gleason, M., & Vachon, V. (2005). *REWARDS: Multisyllabic word reading strategies*. Longmont, CO: Sopris West.
- Begeny, J. C. (2009). *Helping early literacy with practice strategies (HELPS): A one-on-one program designed to improve students' reading fluency*. Raleigh, NC: The Helps Education Fund. Retrieved from <http://www.helpsprogram.org>.
- Begeny, J. C. (2011). Effects of the helping early literacy with practice strategies (HELPS) reading fluency program when implemented at different frequencies. *School Psychology Review, 40*, 149–157.
- Begeny, J. C., Braun, L. M., Lynch, H. L., Ramsay, A. C., & Wendt, J. M. (2012a). Initial evidence for using the HELPS reading fluency program with small instructional groups. *School Psychology Forum: Research in Practice, 6*, 50–63.
- Begeny, J. C., Hawkins, A. L., Krouse, H. E., & Laugle, K. M. (2011). Altering instructional delivery options to improve intervention outcomes: Does increased instructional intensity also increase instructional effectiveness? *Psychology in the Schools, 48*, 769–785.
- Begeny, J. C., Krouse, H. E., Ross, S. G., & Mitchell, R. C. (2009). Increasing elementary-aged students' reading fluency with group-based interventions: A comparison of repeated reading, listening passage preview, and listening only strategies. *Journal of Behavioral Education, 18*, 211–228.
- Begeny, J. C., Mitchell, R. C., & Whitehouse, M. (2014). *Integrating a prompting procedure to improve teachers' modeling of reading. An experimental analysis of the 'icing on the cake'* (manuscript in preparation).
- Begeny, J. C., Schulte, A. C., & Johnson, K. (2012b). *Enhancing instructional problem solving: An efficient system for assisting struggling learners*. New York: Guilford Press.
- Begeny, J. C., & Silber, J. M. (2006). An examination of group-based treatment packages for increasing elementary-aged students' reading fluency. *Psychology in the Schools, 43*, 183–195.
- Block, M. (Host). (2010). *N.C. schools official lauds education proposal*. [Radio transcript] Washington, DC: National Public Radio. Retrieved July 15, 2010 from <http://www.npr.org/templates/story/story.php?storyId=124913035&sc=emaf>.
- Buck, G. H., Polloway, E. A., Smith-Thomas, A., & Cook, K. W. (2003). Prereferral intervention processes: A survey of state practices. *Exceptional Children, 69*, 349–360.

- Bulté, I., & Onghena, P. (2008). An R package for single-case randomization tests. *Behavior Research Methods*, *40*, 467–478.
- Burns, M. K., & Gibbons, K. A. (2008). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices*. New York: Routledge.
- Busk, P. L., & Marascuilo, L. A. (1992). Statistical analysis in single-case research: Issues, procedures, and recommendations, with applications to multiple behaviors. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 159–185). Hillsdale, NJ: Lawrence Erlbaum.
- Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology*, *33*, 269–285.
- Christ, T. J., & Silbergitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, *36*, 130–146.
- DuFour, R., Eaker, R., Karhanek, G., & DuFour, R. (2004). *Whatever it takes: How professional learning communities respond when kids don't learn*. Bloomington, IN: Solution Tree.
- Dufrene, B., & Warzak, W. J. (2007). Brief experimental analysis of Spanish reading fluency: An exploratory evaluation. *Journal of Behavioral Education*, *16*, 143–154.
- Edgington, E. S. (1980). Validity of randomization tests for one-subject experiments. *Journal of Educational Statistics*, *5*, 235–251.
- Engelmann, S., Johnson, G., Carnine, L., & Meyer, L. (2008). *Corrective reading: Decoding [curriculum program]*. Columbus, OH: McGraw-Hill Education.
- Franklin, R. D., Gorman, B. S., Beasley, T. M., & Allison, D. B. (1997). Graphical display and visual analysis. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 119–158). Mahwah, NJ: Lawrence Erlbaum.
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93–99.
- Greenwood, C. R. (1991). Longitudinal analysis of time engagement and academic achievement in at-risk and non-risk students. *Exceptional Children*, *57*, 521–535.
- Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). *The fourth R: Research in the classroom*. Columbus, OH: Charles E. Merrill.
- Johnson, N., Oliff, P., & Williams, E. (2010). *An update on state budget cuts: Governors proposing new round of cuts for 2011; At least 45 states have already imposed cuts that hurt vulnerable residents*. Retrieved January 6, 2011 from <http://www.cbpp.org/cms/?fa=view&id=1214>.
- Kamps, D. M., Greenwood, C. R., Arreaga-Mayer, C., Veerkamp, M. B., Utley, C., Tapia, Y., et al. (2008). The efficacy of classwide peer tutoring in middle schools. *Education and Treatment of Children*, *31*, 119–152.
- Klubnik, C., & Ardoin, S. P. (2010). Examining immediate and maintenance effects of a reading intervention package on generalization materials: Individual versus group implementation. *Journal of Behavioral Education*, *19*, 7–29.
- Malloy, K. J., Gilbertson, D., & Maxfield, J. (2007). Using brief experimental analysis for selecting reading interventions for English language learners. *School Psychology Review*, *36*(2), 291–310.
- Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18–78). New York: Guilford Press.
- Morgan, P. L., & Sideridis, P. D. (2006). Contrasting the effectiveness of fluency interventions for students with or at risk for learning disabilities: A multilevel random coefficient modeling meta-analysis. *Learning Disabilities Research & Practice*, *21*, 191–210.
- National Center for Education Statistics (2011). *The Nation's report card: Reading 2011 (NCES 2012-457)*. Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved June, 2012 from <http://nces.ed.gov/nationsreportcard/pdf/main2011/2012457.pdf>.
- NC Department of Public Instruction. (2012). *2012–13 budget frequently asked questions*. Retrieved March 23, 2014 from <http://www.ncpublicschools.org/docs/budget/communication/budgetguide.pdf>.
- Pearson, P. D., Johnson, D. D., Clymer, T., Indrisano, R., Venezky, R. L., Baumann, J. F., Hiebert, E., & Toth, M. (1989). *Silver, Burdett, and Ginn*. Needham, MA: Silver, Burdett, and Ginn, Inc.
- Ross, S. G., & Begeny, J. C. (2011). Improving Latino, English language learners' reading fluency: The effects of small group and one-on-one intervention. *Psychology in the Schools*, *48*, 604–618.
- Scruggs, T. E., & Mastropieri, M. A. (Eds.). (1997). *Advances in learning and behavioral disabilities* (Vol. 11). Greenwich, CT: JAI.

- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8*, 24–33.
- Scruggs, T. E., Mastropieri, M. A., Cook, S. B., & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders, 11*, 260–271.
- Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: The Guilford Press.
- Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal, 53*, 410–413.
- Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*, 252–261.
- Todman, J. B., & Dugard, P. (1999). Accessible randomization tests for single-case and small-n experimental designs in AAC research. *Augmentative & Alternative Communication, 15*, 69–82.
- Ukrainetz, T. (2009). Phonemic awareness: How much is enough in the changing picture of reading instruction. *Topics in Language Disorders, 29*, 344–359.
- U.S. Department of Education. (2010). *Condition of education*. Retrieved from <http://nces.ed.gov/programs/coe>.
- Vaughn, S., Linan-Thompson, S., & Kouzekanani, K. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education, 24*, 301–315.
- Wanzek, J., Vaughn, S., Scammacca, N. K., Metz, K., Murry, C. S., & Roberts, J. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research, 83*, 163–195.
- What Works Clearinghouse. (2014). *WWC intervention report, students with learning disabilities: Repeated reading*. Retrieved July 24, 2014 from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_repeatedreading_051314.pdf.
- Wilson, J. B. (2007). Priorities in statistics, the sensitive feet of elephants, and don't transform data. *Folia Geobotanica, 42*, 161–167.
- Yurick, A., Cartledge, G., Kourea, L., & Keyes, S. (2012). Reducing reading failure for kindergarten urban students: A study of early literacy instruction, treatment quality, and treatment duration. *Remedial and Special Education, 33*, 89–102.