International Journal for Research in Learning Disabilities

volume 4, issue 1, 2019

formerly *Thalamus*

I J R L D

I J R L D

I J R L D

I J R L D

I J R L D

Journal of the
International Academy for Research in Learning Disabilities

Editor: David Scanlon, Boston College, United States

iarld

International Academy for Research in Learning Disabilities

International Journal for Research in Learning Disabilities

formerly *Thalamus*

journal of the *International Academy for Research in Learning Disabilities*

IJRLD
International Journal for Research in Learning Disabilities

Volume 4, No. 1, 2019

Table of Contents

Guest Reviewers

William M. Cruickshank Memorial Lecture Delivered at the 2018 Conference of the International Academy for Research in Learning Disabilities

Solving the Problem of Learning Disabilities

Linda S. Siegel
University of British Columbia

Dyslexia and other learning disabilities are not being properly recognized and treated in our educational system or society at large. Unrecognized and untreated learning disabilities represent a serious social and economic problem, not only to the individual but to society as a whole. For example, antisocial behavior, as seen in prison populations and juvenile detention centers (e.g., Grigorenko, 2006; Quinn, Rutherford, Leone, Osher, & Poirier, 2006; Zhang, Barrett, Katsiyannis, & Yoon, 2011), homelessness (Barwick & Siegel, 1996), substance abuse (McClelland, Elkington, Teplin, & Abram, 2004), suicide (McBride & Siegel, 1997), and emotional difficulties (Livingston, Siegel, & Ribary, 2018) are often a result of dyslexia and other learning disabilities that have not been properly identified and/or treated.

The system has failed on all levels. It is chaotic and unsystematic, the result of flawed logic and ignorance of, or unwillingness to consider, the available research. In this article, I document these issues and propose solutions. The solutions are relatively inexpensive and practical, certainly much cheaper and more humane than the costs of the serious secondary problems created by the widespread lack of attention to learning disabilities.

Improving Identification Procedures

In the current system, learning disabilities are not being properly identified and treated. Identification procedures, in most jurisdictions, are costly and unnecessary. The first step in solving the issues of dyslexia and other learning disabilities is to develop accurate and systematic, but not costly, identification procedures. Currently the procedures for identifying students with learning disabilities are complicated and require extensive testing. A detailed psychoeducational evaluation is usually mandated, typically necessitating an intelligence (IQ) test. However, in most cases, an IQ test is unnecessary, as illustrated in the following.

Often, identification procedures for dyslexia and learning disabilities require that difficulties in achievement, particularly reading and/or mathematics, are unexpected in relation to "general cognitive abilities." This "unexpectedness" is typically interpreted with regard to intelligence, as defined by a score on an IQ test – an archaic and inappropriate strategy referred to as the discrepancy definition of learning disabilities. In order to be considered to have a learning disability, in other words, an individual must demonstrate a significant difference, or discrepancy, between his or her IQ score and reading or mathematics achievement score.

The use of discrepancy in the definition of learning disabilities is incorrect and not supported by empirical research. The inclusion of a measure of intelligence is unnecessary, and has excluded students from being identified as having a learning difference who have, in fact, experienced learning differences (for a review of the evidence, see Fletcher, Francis, Rourke, Shaywitz, & Shaywitz, 1992; Siegel, 1988, 1989a, 1989b, 1992). The use of the discrepancy definition in any form has been and is detrimental to a great many individuals. As the theoretical and empirical shortcomings are being identified, the use

Siegel, L. (2019). Solving the Problem of Learning Disabilities. *International Journal of Research on Learning Disabilities, 4*(1), 3-11.
https://doi.org/10.28987/ijrld.4.1.3

of the discrepancy model to identify learning differences is indefensible, therefore.

Although a discrepancy is expected in terms of reading or mathematics "… in relation to other cognitive abilities" (Lyon, Shaywitz, & Shaywitz, 2003, p. 2), this is a characteristic demonstrated by many individuals who have learning difficulties, but it is by no means an important or relevant aspect of learning disabilities identification and, therefore, should not be part of the identification criteria.

Specifically, the current model requires a predetermined discrepancy or gap between achievement and ability, such that a deficit exists that necessitates special education or related services for a student to benefit from education. In some schools, the gap has to be quite large, as much as 1.5 standard deviations. That is, the difference between the student's IQ score and his or her achievement score would need to be 22 standard score points or more. For example, if the IQ was 100, the standard score in reading would have to be 78 or lower. Typically, individuals with standard scores below 86 or even 91 on tests of reading and/or mathematics are experiencing difficulties, and if the discrepancy definition is used, these individuals would not have access to services unless they had IQ test scores in the above-average range.

Historically, the discrepancy definition has led not only to under-identification but also exclusion from interventions that could have assisted students in becoming competent readers, for example (Siegel, 1989a, 1989b). Research has substantiated these potential difficulties. For example, in various studies, when good, dyslexic readers (poor readers who met the discrepancy definition) and intelligence-commensurate poor readers (poor readers who did not meet the discrepancy definition) were compared, both groups of poor readers performed significantly poorer than the good readers, but not statistically different from each other (e.g., Hurford, Johnston et al., 1994; Hurford, Schauf et al., 1994; Siegel, 1992; Stanovich, 1991; Vellutino, Scanlon, & Lyon, 2000). In another study, 250 students with reading disabilities and 719 nondisabled students were placed into IQ-level groups based on their performance on the *Wechsler Intelligence Scale for Children-Revised*. Results showed that measures of reading, spelling, and the understanding of syntax were better predicted by the presence or absence of a reading disability than by IQ test scores (Siegel, 1988). These studies highlight the reality that poor readers, that is, individuals who had significantly poor reading skills, read similarly regardless of their scores on IQ tests. Thus, the most important finding was that their reading performance could not be differentiated based upon measures of intelligence. In addition, there are no differences between individuals with or without an ability achievement discrepancy on functional MRIs (Tanaka et al., 2011), suggesting a lack of neurological differences between the discrepant and non-discrepant groups.

Under the discrepancy formula, it is assumed that even though poor readers from different IQ group levels read comparably, the poor readers with higher levels of intelligence benefit more from intervention. But the data do not support this assumption. In fact, regardless of their intellectual abilities, poor readers benefit from interventions at statistically identical levels. The major factor is whether or not students receive interventions appropriate for the reading difficulties they are experiencing (e.g., Hurford, Johnston et al., 1994; Pogorzelski & Wheldall, 2002; Stage, Abbott, Jenkins, & Berninger, 2003; Weber, Marx, & Schneider, 2002).

Even if the discrepancy definition is not used, there are conceptual problems with the use of intelligence tests. Intelligence tests are generally very heavily loaded on vocabulary and other language measures, memory, and motor measures, which we now understand are weaknesses for many individuals with learning disabilities (Siegel, 1989a, 1989b). As a result their intellectual functioning is more likely to be underestimated.

To make matters worse, the development of reading and literacy skills fosters the cognitive skills assessed with intelligence and aptitudes tests. Not only do good readers become more competent readers as a function of applying and practicing their reading skills, these readers are further developing their cognitive skills en route to commensurate improvements in scores on cognitive and intelligence tests. The opposite would be true for poor readers. Consequently, according to the discrepancy model, many poor readers would not reach a large enough gap between measures of reading and their IQ score to warrant inclusion in interventions, despite being in desperate need of interventions for their poorly developed reading skills.

Siegel and Himel (1998) directly examined this issue and found an inverse relationship between the age of children with dyslexia and their IQ scores. That is, the older children had lower IQ scores than the younger ones. This finding not only indicates that intelligence tests tap into the weak verbal abilities of the child with dyslexia, which then leads to an underestimate of his or her intelligence, but also highlights the absurdity of the discrepancy definition. That is, the discrepancy is reduced, thus making the student no longer eligible for intervention services. According to Stanovich (1986), "The cognitive consequences of the acquisition of literacy may be profound" (p. 374). Poor readers have less experience reading and gain less from the process of reading due to their inferior reading skills, which then affects the development of knowledge, memory, and other cognitive abilities.

Given these limitations, the discrepancy formula is untenable and unethical. It is inappropriate and unjust to deny a child educational services based on an empirically discredited form of classifying children into those who receive intervention services and those who do not.

The Patterns of Strengths and Weaknesses Model

Currently, an assessment for dyslexia or other learning disabilities often includes a number of tests of cognitive processes. The idea behind this type of assessment is that individuals with learning disabilities show a pattern of high ability in some areas of cognitive functioning and significant weaknesses in others, called the Pattern of Strengths and Weaknesses model (PSW). According to this approach, this unevenness is the defining characteristic of a learning disability and is a requirement for a person to be said to have a learning disability. If a student has a flat profile, without much deviation in terms of strengths and weaknesses, the proponents of this model would not identify this individual as having a learning disability.

The Wechsler Intelligence Scale for Children (WISC-V; Wechsler, 2014), for example, measures verbal comprehension, fluid reasoning (a cognitive ability that requires minimal prior knowledge to solve novel tasks), visual processing, processing speed, and working memory. The *Woodcock-Johnson Tests of Cognitive Abilities* (WJ IV; Schrank, McGrew, & Mather, 2014) measures comprehension-knowledge, long-term retrieval, visual-spatial thinking, auditory processing, fluid reasoning, processing speed, and short-term memory. These tests, along with similar instruments, were designed to examine several aspects of cognitive functioning and are used to provide a pattern of an individual's strengths and weaknesses.

Although the PSW model appears to provide an excellent basis for identification of learning disabilities, including dyslexia, it is essentially the discrepancy model lightly disguised. Thus, it requires that there be irregular patterns among the various cognitive abilities and achievement scores and that the individual's intellectual functioning fall in the average range. The requirement that intelligence be in the average range is analogous to the discrepancy model and, once again, will exclude students who have reading difficulties who do not meet this requirement. In short, the evidence for the PSW approach, as outlined below, suggests that, like the discrepancy model, it is not very useful in identifying learning differences.

There are several forms of the PSW model: the Concordant/Discordant model, the Cattell-Horn-Carroll Operational model, the Discrepancy/Consistency model, and the Hypothesis Testing (HT-CHC) Cattell-Horn-Carroll Operational model.

The Concordant/Discordant model (Hale & Fiorello, 2004) identifies individuals with specific learning disabilities by providing evidence that a weakness in achievement is related to a weakness in cognitive ability. Evidence must also be presented that demonstrates that the cognitive abilities that are not related to the achievement difficulties are not weak. For example, in this formulation, a student who has weaknesses in phonological/auditory processing that lead to weaknesses in the achievement areas of reading and spelling cannot have any other weaknesses in processing or achievement. Of course, this situation is unlikely; most people with learning difficulties have weaknesses in a variety of areas.

The Cattell-Horn-Carroll Operational model (Flanagan, Ortiz, & Alfonso, 2007) uses the Cattell-Horn-Carroll model of cognitive abilities to examine strengths and weaknesses through an examination of the broad cognitive abilities (i.e., Comprehension-Knowledge, Fluid Reasoning, Quantitative Knowledge, Reading and Writing Ability, Short-Term Memory, Long-Term Storage and Retrieval, Visual Processing, Auditory Processing and Processing Speed) and compares the relationships between these areas and corresponding areas of achievement. This model, then, identifies students with specific learning disabilities based on evidence showing that a weakness in achievement is linked to a weakness in one of the broad cognitive abilities outlined above, and that the broad cognitive abilities not related to the achievement weakness are not weak.

The Discrepancy/Consistency model (Naglieri, 1999) identifies specific learning disabilities by examining the variability of an individual's cognitive scores. If a particular score is significantly low compared to the individual's other cognitive scores and meets the threshold of significantly low based on predetermined values (e.g., less than a standard score of 85), then that area is determined to be a weakness. Strengths are identified in the same, but opposite manner (i.e., scores greater than a predetermined value).

The Hypothesis Testing Cattell-Horn-Carroll model (Flanagan, Fiorello, & Ortiz, 2010) combines aspects of the Cattell-Horn-Carroll and concordant/discordant models. Flanagan et al. (2013) define a cognitive or academic weakness as a score lower than a standard score of 85 and a cognitive or academic strength as a standard score greater than 90.

All the variations of the PSW model require a weakness in a cognitive area for identification of specific learning disability to occur. Unfortunately, "… experimental studies showed that cognitive profiles had limited diagnostic accuracy in identifying individuals with learning difficulties" (Beaujean, Benson, McGill, & Dombrowski, 2018, p. 2). (Beaujean and colleagues are not the only researchers critical of the PSW model [e.g., McGill & Busse, 2016; Miciak, Taylor, Denton, & Fletcher, 2015].)

In addition, one of the premises of all of these models is that the individual who is being considered for a specific learning disability must possess "… at least an average level of general cognitive ability or intelligence" (Flanagan et al., 2010, p. 745). As noted above, there is no evidence that this is the case. For individuals with dyslexia/reading difficulties, demarcating groups by intelligence provides an artificial differentiation between those groups – groups that perform nearly identically on measures of reading and its subskills and groups that benefit equally from interventions.

One of the assumptions of the PSW model is that the performance of students with learning disabilities differ from that of their typically achieving peers. But there is great individual variation in using PSW analysis, and this difference between the performance of students with and without learning disabilities is not always found. Therefore, the diagnostic utility and validity of these models is questionable. For example, D'Angiulli and Siegel (2003) found that typically achieving children and children with learning disabilities did not differ in patterns of performance on IQ tests. That is, similar numbers of the two groups who demonstrated discrepancies between their verbal and performance IQ scores and had significantly low scores on Digit Span, a working memory test, and Coding, a memory and motor test. Finally, in a simulation study examining various models of patterns of strengths and weaknesses, Stuebing, Fletcher, Branum-Martin, and Francis (2012) noted that none of them was very useful in identifying students with learning disabilities. Most important, a particular cognitive profile of strengths and weaknesses does not predict who will benefit from remediation or what particular intervention strategy should be employed. This is particularly the case for individuals with reading difficulties.

Opportunity Costs of Current Assessment Models

Requiring these lengthy and detailed assessments may mean that many people who are struggling with reading, writing, and/or mathematics do not get the assessment that is required. For example, there is often an 18- to 24-month wait for assessment within school systems. If an assessment is not conducted by the school system, and often it is not because it is impossible for staff psychologists to test all the students who need to be tested, it is costly (1,500-3,000 US dollars) to seek private testing – a cost that is too much for many families to afford. Further, in most postsecondary institutions, there is no provision for any free or low-cost assessment. Nevertheless, a psychoeducational evaluation is usually necessary for a designation and subsequent intervention and accommodations, meaning that children whose parents cannot afford it, or postsecondary students who cannot afford it, are denied intervention and/or accommodations. This situation is clearly an indication of a lack of social justice.

Appropriate Assessment

All that is necessary to confirm whether or not somebody has a learning disability is to assess their reading, spelling, and mathematics and, perhaps, writing. Standardized tests are available in English and some other languages for this purpose, and teachers and other school personnel can easily be trained to administer them.

What then is necessary? To meet this definition and to understand a student's academic strengths and challenges, we need a variety of achievement tests whose subtests directly measure the potential deficiencies. Tests of reading are obviously essential. It is critical to measure two basic reading skills, word recognition and decoding.

Reading

Word recognition. On word recognition tests (e.g., *Wide Range Achievement Test*, Letter and Word Identification on the *Woodcock-Johnson Test of Achievement,* the *Wechsler Individual Achievement Test*), the person reads words out loud that vary in difficulty, and his or her performance is scored and compared to that of other people of the same age.

Decoding. Decoding skills involve the ability to associate sounds with letters such that letters are translated or decoded into their respective sounds. This ability is assessed by asking a person to read words and pseudowords (nonwords that are legal and pronounceable combinations of letters but not real words in the language in question). The use of pseudowords in these tasks is very useful because some struggling readers use strategies to memorize the visual representations of words. In these cases, it may appear that the individual has appropriately developed decoding skills, even if it is not the case. The use of pseudowords prevents the use of this strategy and, therefore, provides a more accurate representation of the individual's ability to decode as he or she has not encountered these pseudowords previously (e.g., *pum, nafotbil*). In other words, pseudowords force the individual to engage in decoding.

Decoding is essential for reading acquisition and word recognition, which eventually becomes automatic in skilled readers. The efficiency with which the processes occur is also important. If a student takes considerable time to decode and read a word, the time and effort involved in those tasks will consume cognitive effort, which will then interfere with understanding the text. One test that measures the speed of reading words and pseudowords is the *Test of Word Reading Efficiency* (TOWRE), which measures how many words and pseudowords the person can read in 45 seconds.

Accuracy and speed. Tests of accuracy and speed of word recognition and pseudoword reading are absolutely essential for understanding whether a student is experiencing reading difficulties. In fact, it is possible to make a case for dyslexia just based on poor performance on any one of these tasks.

When evaluating the reading skills of struggling readers, it is important to analyze the items on which they make errors and the types of errors that are made. Error and item analyses provide clues as to the types of interventions that would be of most benefit to a given student. For example, if someone is dictated the word take and spells it tak, this is a sign of a problem with understanding that the "final e" indicates that the vowel will likely be long rather than short. Examples of approaches that have used analyses of items and errors and related the results to intervention include Kern and Hosp (2018), Odegard, Cooper, Hirschmann, and Alexander (2017), and Steacy, Elleman, Lovett, and Compton (2016).

Spelling. Other tests are useful for understanding academic performance. For example, learning about a student's spelling of words is important. Spelling is evaluated by dictating a list of words to the person, who then writes them (e.g., Spelling subtest of the *Wechsler Individual Achievement Test* or the Spelling subtest of the *Wide Range Achievement Test*; WRAT).

Additionally, measuring reading comprehension is important in understanding what meaning the person is able to get from the text. However, consideration should be given to the test passages selected, because reading comprehension depends at least in part on the reader's background knowledge that is assumed as part of the reading of the passage. Thus, although decoding, vocabulary, and knowledge very accurately predict reading comprehension, a lack of content knowledge likely will hamper reading comprehension, whereas expert knowledge likely will enhance reading comprehension (e.g., Alfassi, 2004). In addition, good reading skills are related to vocabulary growth (Duff, Tomblin, & Catts, 2015).

Time limits. Most reading comprehension tests require that the test be completed within a time limit. Lesaux, Pearson, and Siegel (2006) examined the reading comprehension ability of dyslexics to perform in two conditions: the typical condition in which students have a limited amount of time to complete the task vs. an unlimited amount of time to complete the task. Although the conditions had little effect on good readers, the dyslexics performed significantly better in the untimed condition.

Even for tests of reading comprehension, we should not just consider the overall score, but ask whether or not the questions were answered correctly within the time limit. For example, if a person answers few questions but all the answers are correct, then he or she understands what has been read, but just reads slowly. In addition, reading comprehension tests can place different demands on the respondent. Some item can be answered using verbatim information provided in the text, while others can only be answered by going beyond the explicitly stated information and making an inference.

Dictation. A spelling dictation test for both words and pseudowords should be included. Analyses of error patterns, as conducted by Lennox and Siegel (1996), would be useful for planning interventions. For example, errors such as *nature* spelled as *nachure* indicate reasonable phonetic skills but problems with orthographic processing. *Circle* spelled as *sicel* may indicate a problem with hearing and/or remembering the sounds in words and suggests phonemic awareness training. Of course, one would not recommend an intervention based on one error; these are just included as examples.

Mathematics

Speed and accuracy. The assessment of mathematical skills should include the speed and accuracy of mathematical calculation and problem solving. An analysis of errors is critical. It is important to note whether or not the individual makes errors that indicate that he or she does not understand place value; for example, adding 29 and 13 and producing 312 as the answer.

Mathematical reasoning. It is important to note that existing tests of mathematical reasoning are not independent of computational skill. There was, however, one good test of mathematical reasoning that did not involve computational skills, the Missing Elements test of the *Key Math Test*. Using this test, the student was presented with a problem such as "There are 36 chairs in a room. How many rows are there?" and was asked to say what additional information would be needed to solve the problem. Unfortunately, this subtest disappeared from later versions of the *Key Math Test*.

Writing

Assessment of writing skills is important but extremely difficult. The writing process is not something that can easily be captured in a brief assessment. In addition, there is the issue of handwriting; many people write directly on the computer. Of course, one could always ask for samples of written work, but there is no information about the conditions under which it was done and, therefore, no standardization. The *Woodcock Johnson* has Writing Fluency and Writing Samples subtests; these can be useful, but there is some subjectivity in the scoring. Another instrument is the *Test of Written Language* (TOWL), but it is also difficult to score objectively. It may be useful to have students tell a story orally based on the pictures on the TOWL. Often, one can see creativity in the oral story that one does not find in a poorly handwritten story. Again, objective scoring of the story is not practical, but the test can yield useful information about a student.

Early Identification

Early identification of children at risk for learning disabilities is critical. Although we know how to detect children at risk for reading difficulties early in their school career, even before they start reading instruction, systems are not in place in most jurisdictions to find these children.

In studies of early identification of and intervention for children at risk for reading difficulties, we found that children at risk for reading difficulties can be detected in kindergarten, even before reading instruction begins. The screening is efficient and easy to administer. Children are administered a letter recognition test and some phonological awareness tasks, such as phoneme deletion, which requires them to say, for example, *pink* without the "p" sound (the sound is said, not the letter). For a more complete description of this technique, see Siegel (2018).

Early identification for potential reading difficulties is cheap, easy to administer, and effective. With brief training, teachers can do it. The advantages of having teachers do the identification is that they become more aware of the child's strengths and difficulties. Unfortunately, research on identifying children at risk for mathematical difficulties and providing an intervention is still in its infancy stage.

Early Intervention

Early identification is meaningless unless an intervention program is in place to help students who are identified as being at risk. We know that if we intervene with children with reading difficulties early, intervention is more effective and cheaper (Fawcett, Lee, & Nicholson, 2014; Griffiths & Stuart, 2013).

With regard to early reading intervention, we have conducted studies of Firm Foundations (summarized in Siegel, 2018), a locally developed (British Columbia, Canada) program to help children understand the sounds of the English language and how letters have these sounds. The program is effective, and helps both children whose first language is English and children who are learning English as a second language. Brief teacher training and the purchase of the Firm Foundations are the only costs involved, and these are minimal.

Improving the Curriculum

In many faculties of higher education, teachers are not properly taught about teaching reading, writing, and mathematics. For example, in a review of studies about the knowledge and understanding of teachers about teaching reading and dyslexia, Washburn, Mulcahy, Joshi, and Banks-Cantrell (2016) found that teachers often lacked knowledge about important concepts related to reading, such as phonemic awareness and morphology. Teacher knowledge of reading-related concepts impacts student performance, and inadequate knowledge on the part of teachers is correlated with poor performance of students. Therefore, improvement of teacher education should be an important priority.

Recognizing Parents and Teachers

One aspect of solving the issues related to learning disabilities involves providing appropriate educational support. Within the educational system, there is a tendency for some school personnel to belittle or even ignore the role of parents in helping their children succeed (Siegel, 2013, 2016), thereby creating a difficult educational environment for the child. It is also sometimes the case that teachers' careful observations about a child's difficulties are discounted by the school system.

Coping Strategies

To help people with learning disabilities, we need to recognize their talents and abilities. Traditionally, we have not paid enough attention to the strengths of people with learning disabilities. Many have excellent abilities in nonacademic areas such as music, art, dance, drama, sports, or carpentry, to name a few. Some have developed coping strategies, but these strategies should be taught on a larger scale. For example, some use humor and good reasoning skills to help them deal with bullies. Madaus (2005) has outlined strategies for dealing with the transition to postsecondary education.

Solutions

In summary, the following are solutions to the problems of learning disabilities.

1. Streamline assessments of learning disabilities and make them available without charge to anyone.
2. Practice early identification and intervention on a large scale in every district, as essential for success.
3. Improve the teaching of reading and mathematics, using evidence-based methods.
4. Respect parents and teachers and their observations.

How can governments and school systems find the money to fund these initiatives? There are several ways. The system of psychoeducational testing currently used is unnecessarily complex and costly. A more streamlined system would save thousands of dollars. For example, abolishing "high-stakes testing," because group multiple-choice tests tell us little about the basic skills of an individual. If governments and school systems put serious efforts into help peopling with learning disabilities using the methods outlined above, we would significantly reduce the incidence of academic, social, and emotional problems, which come at a great cost to both the individual and society.

References

Alfassi, M. (2004). Reading to learn: Effects of combined strategy instruction on high school students. *The Journal of Educational Research*, *97*(4), 171-185. https://doi.org/10.3200/JOER.97.4.171-185

Barwick, M. A., & Siegel, L. S. (1996). Learning difficulties in adolescent clients of a shelter for runaway and homeless street youths. *Journal of Research on Adolescence, 6,* 649-670.

Beaujean, A. A., Benson, N. F., McGill, R. J., & Dombrowski, S. C. (2018). A misuse of IQ scores: Using the dual discrepancy/consistency model for identifying specific learning disabilities. *Journal of Intelligence, 6*(3), 36. https://doi.org/10.3390/jintelligence6030036

D'Angiulli, A., & Siegel, L. S. (2003). Cognitive functioning as measured by the WISC-R: Do children with LD have distinctive patterns of performance? *Journal of Learning Disabilities, 36,* 48-58. https://doi.org/10.1177/00222194030360010601

Duff, D., Tomblin, J. B., & Catts, H. (2015). The influence of reading on vocabulary growth: A case for a Matthew effect. *Journal of Speech, Language, and Hearing Research*, *58*(3), 853-864. https://doi.org/10.1044/2015_JSLHR-L-13-0310

Fawcett, A. J., Lee, R., & Nicolson, R. (2014). Sustained benefits of a multi-skill intervention for preschool children at risk of literacy difficulties. *Asia Pacific Journal of Developmental Differences, 1*(1), 62-77.

Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A "third method" approach to specific learning disability identification. *Psychology in the Schools*, *47*(7), 739-760. https://doi.org/10.1002/pits.20501

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). Use of the cross-battery approach in the assessment of diverse individuals. In D. P. Flanagan, S. O. Ortiz, & V. C. Alfonso (Eds.), *Essentials of cross-battery assessment, second edition* (pp. 146-205). New York, NY: John Wiley and Sons.

Fletcher, J. M., Francis, D. J., Rourke, B. P., Shaywitz, S. E., & Shaywitz, B. A. (1992). The validity of discrepancy-based definitions of reading disabilities. *Journal of Learning Disabilities*, *25*(9), 555-561. https://doi.org/10.1177/002221949202500903

Griffiths, Y., & Stuart, M. (2013). Reviewing evidence-based practice for pupils with dyslexia and literacy difficulties. *Journal of Research in Reading, 36*(1), 96-116. https://doi.org/10.1111/j.1467-9817.2011.01495.x

Grigorenko, E. L. (2006). Learning disabilities in juvenile offenders. *Child and Adolescent Psychiatric Clinics of North America, 15*, 353-371.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook.* New York, NY: Guilford Press.

Hurford, D. P., Johnston, M., Nepote, P., Hampton, S., Moore, S., Neal, J., ... & Huffman, D. (1994). Early identification and remediation of phonological-processing deficits in first-grade children at risk for reading disabilities. *Journal of Learning Disabilities*, *27*(10), 647-659. https://doi.org/10.1177/002221949402701005

Hurford, D. P., Schauf, J. D., Blaich, T., Moore, K., & Bunce, L. (1994). Early identification of children at risk for reading disabilities. *Journal of Learning Disabilities*, *27*(6), 371-382. https://doi.org/10.1177/002221949402700604

Kern, A. M., & Hosp, M. K. (2018). The status of decoding tests in instructional decision-making. *Assessment for Effective Interventions*, *44*(1), 32-44. https://doi.org/10.1177/1534508417749874

Lennox, C., & Siegel, L. S. (1996). The development of phonological rules and visual strategies in average and poor spellers. *Journal of Experimental Child Psychology, 62,* 60-83. https://doi.org/10.1006/jecp.1996.0022

Lesaux, N. K., Pearson, M. R., & Siegel, L. S. (2006). The effects of timed and untimed testing conditions on the reading comprehension performance of adults with reading disabilities. *Reading and Writing*, *19*(1), 21-48. https://doi.org/10.1007/s11145-005-4714-5

Livingston, E. M,,. Siegel, L. S., & Ribary, U. (2018). Developmental dyslexia: emotional impact and consequences. *Australian Journal of Learning Difficulties*, 1-29

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, *53*(1), 1-14.

Madaus, J. W. (2005). Navigating the college transition maze: A guide for students with learning disabilities. *Teaching Exceptional Children, 37*, 32-37. https://doi.org/10.1177/004005990503700305

McBride, H., & Siegel, L. S. (1997). Learning disabilities and adolescent suicide. *Journal of Learning Disabilities, 30,* 652-659. https://doi.org/10.1177/002221949703000609

McClelland, G. M., Elkington, K. S., Teplin, L. A., & Abram, K. M. (2004). Multiple substance use disorders in juvenile detainees. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(10), 1215-1224.

McGill, R. J., & Busse, R. T. (2017). When theory trumps science: a critique of the PSW model for SLD identification. *Contemporary School Psychology*, *21*(1), 10-18. https://doi.org/10.1007/s40688-016-0094-x

Miciak, J., Taylor, W. P., Denton, C. A., & Fletcher, J. M. (2015). The effect of achievement test selection on identification of learning disabilities within a patterns of strengths and weaknesses framework. *School Psychology Quarterly*, *30*(3), 321-334. https://doi.org/10.1037/spq0000091

Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York, NY: John Wiley and Sons.

Odegard, T. N., Cooper, J., Hirschmann, M., & Alexander, E. (2017). *Dyslexia within RTI.* Murfreesboro, TN: Middle Tennessee State University.

Pogorzelski, S., & Wheldall, K. (2002). Do differences in phonological processing performance predict gains made by older low-progress readers following intensive literacy intervention? *Educational Psychology*, *22*(4), 413-427.

Quinn, M. M., Rutherford, R. B., Leone, P. E., Osher, D. M., & Poirier, J. M. (2005). Youth with disabilities in juvenile corrections: a national survey. *Exceptional Children, 71*, 339-345. https://doi.org/10.1177/001440290507100308

Schrank, F. A., McGrew, K.S., & Mather, N. (2014). *Woodcock-Johnson IV Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside

Siegel, L. S. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *42*(2), 201-215.

Siegel, L. S. (1989a). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22,* 469-478, 486.

Siegel, L. S. (1989b). Why we do not need intelligence test scores in the definition and analyses of learning disabilities. *Journal of Learning Disabilities*, *22*(8), 514-518.

Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, *25*(10), 618-629.

Siegel, L. S. (2013). *Understanding dyslexia and other learning disabilities*. Vancouver, BC: Pacific Educational Press.

Siegel, L. S. (2016). *Not stupid, not lazy: Understanding dyslexia and other learning disabilities.* Baltimore, MD: International Dyslexia Association.

Siegel, L. S. (2018). A case study of successful early screening and intervention. *Perspectives on Language and Literacy, 44,* 29-32.

Siegel, L. S., & Himel, N. (1998). Socioeconomic status, age and the classification of dyslexics and poor readers: The dangers of using IQ scores in the definition of reading disability. *Dyslexia*, *4*(2), 90-104.

Snowling, M. J., Adams, J. W., Bowyer-Crane, C., & Tobin, V. (2000). Levels of literacy among juvenile offenders: The incidence of specific reading difficulties. *Criminal Behaviour and Mental Health, 10*, 229-241. https://doi.org/10.1002/cbm.362

Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ – word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities*, *36*(1), 24-33. https://doi.org/10.1177/00222194030360010401

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407. https://doi.org/10.1598/RRQ.21.4.1

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray? *Reading Research Quarterly*, *26*(1), 7-29. https://doi.org/10.2307/747729

Steacy, L. M., Elleman, A. M., Lovett, M. W., & Compton, D. L. (2016). Exploring differential effects across two decoding treatments on item-level transfer in children with significant word reading difficulties: A new approach for testing intervention elements, *Scientific Studies of Reading*, *20*(4), 283-295. https://doi.org/10.1080/10888438.2016.1178267

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review*, *41*(1), 3-22.

Tanaka, H., Black, J. M., Hulme, C., Stanley, L. M., Kesler, S. R., Whitfield-Gabrieli, S., ... Hoeft, F. (2011). The brain basis of the phonological deficit in dyslexia is independent of IQ. *Psychological Science, 22*(11), 1442-1451. https://doi.org/10.1177/0956797611419521

Vellutino, F. R., Scanlon, D. M., & Lyon, R. G. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities*, *33*(3), 223-238. https://doi.org/10.1177/002221940003300302

Washburn COMPLETE REFERENCE TO BE ADDED

Weber, J-M., Marx, P., & Schneider, W. (2002). Are there different remedial potentials for dyslexics and garden-variety poor readers? *Psychologie in Erziehung und Unterricht*, *49*, 56-70.

Wechsler, D. (2014). *Wechsler intelligence scale for children-fifth edition*. Bloomington, MN: Pearson.

Zhang, D., Barrett, D. E., Katsiyannis, A., & Yoon, A. (2011). Juvenile offenders with and without disabilities: Risks and patterns of recidivism. *Learning and Individual Differences, 21*, 12-18. https://doi.org/10.1016/j.lindif.2010.09.006

Note

Parts of this article appeared in Perspectives on Language and Literacy. This material is used by permission of the International Dyslexia Association (DyslexiaIDA.org).

Peer-Assisted Assessment in Reading: Two Exploratory Studies

Paul J. Riccomini,[1] Sheri Berkeley,[2] Allison Neally,[3] Christina Stagliano,[3] Leigh Ann Kurz,[2] and Frederick Brigham[2]

Abstract

Oral reading fluency (ORF) assessments measure how quickly and accurately students read within one minute. They are widely used at the elementary level; however, due to the typical structure and class sizes in middle schools, such individualized assessments are less feasible and, therefore, less frequently used. Two exploratory studies investigated potential methods for efficiently administering ORF measures at the middle school level by utilization of peers: Peer-Assisted Assessment in Reading (PAAR). Findings from both studies showed that after a short training, students were highly accurate in identifying words read in one minute regardless of grade or instructional level, but they were less consistent in correctly identifying miscues. Comparisons between student and student teacher accuracy were also made. Implications for practice are discussed.

Keywords: peer assisted, oral reading fluency, middle school, preservice teachers

Reading and assessment are two of the more widely discussed aspects of current American educational systems. For example, assessments are embedded in major federal regulations both in general (i.e., Every Student Succeeds Act, 2015) and special education (i.e., Individuals With Disabilities Education Improvement Act, 2004). Further, the American Recovery and Reinvestment Act (2009) included additional emphasis on the use of assessment data to improve educational outcomes by offering funds to states that close achievement gaps between the general population and students with disabilities, students from minority backgrounds, low-income students, and students with limited proficiency in the English language. The emphasis on verifying academic performance through assessments is also clear in the Common Core State Standards that have been adopted by a wide majority of states (Calkins, Ehrenworth, & Lehman, 2012; Kendall, Ryan, Alpert, Richarson, & Schwols, 2012).

The initial assessments created in the wake of regulations were most often static end-of-course assessments to validate progress and educational accountability. Such assessments are summative in nature and are of little value in adjusting instructional procedures for individuals who are enrolled in a course (Brigham, Tochterman, & Brigham, 2000). However, a wider array of sensitive measures that can track progress across the school year are available now to educators and are becoming prominent in professional practice in schools (Brigham, Berkeley, & Walker, 2012). For example, progress monitoring is increasingly employed by schools to detect and address educational problems at an early point (Berkeley & Riccomini, 2017; Deno, 2003; Fuchs, Fuchs, & Compton, 2012; Petscher, Cummings, Biancarosa, & Fien, 2013; Santi & Vaughn, 2007).

Among the common elements of progress monitoring procedures is the use of materials that are drawn directly from the instructional materials to be used with students, or at least are very similar to those that students will encounter in their studies (Venn, 2013). Using the same or similar materials for both instruction and assessment has the benefit of reduc-

[1]The Pennsylvania State University; [2]George Mason University; [3]The University of Georgia

Riccomini,P. J., Berkeley, S., Neally, A., Stagliano, C., Kurz, L., & Brigham F. (2019). Peer-Assisted Assessment in Reading: Two Exploratory Studies. *International Journal for Research in Learning Disabilities, 4*(1), 12-26. https://doi.org/10.28987/ijrld.4.1.12

ing the amount of inference required to interpret the assessment data, thereby enhancing their predictive validity (Lembke, Hampton, & Hendricker, 2013) as well as their validity in guiding interventions, also known as treatment utility (Hayes, Nelson, & Jarrett, 1987). Hayes et al. (1987) defined treatment utility as simply "the degree to which assessment is shown to contribute to beneficial treatment outcome" (p. 963). Roach (2008) described the amount of inference required to employ assessment data in useful interventions as the "X factor," in which the connections between assessments and instructional actions are often tenuous and poorly understood. Thus, reducing the conceptual distance between measures and the actions to be based upon them is likely to minimize error and response time as well as make the process of assessment more transparent.

Several variables have been identified as useful targets for screening students for risk of failure and for monitoring general development within a performance domain. Most of these are fluency variables where speed and accuracy of task completion are important (Schwanenflugel & Ruston, 2008).

Reading Assessment and Middle School

The goal of reading is comprehension. Reading involves a union of decoding, fluency, and comprehension that is integrated and automatized in effective readers so that they are able to devote their full attention to the meaning of the text (Soriano, Miranda, Soriano, Nievas, & Felix, 2011). In contrast, readers who are not fluent in reading understand less of what they read, are less motivated to practice, and struggle more in learning academic content (Kuhn & Stahl, 2003; Meisinger, Bloom, & Hynd, 2010). While adequate fluency does not ensure good comprehension of text, poor fluency virtually ensures that the reader will have a great deal of difficulty in grasping the meaning of what he or she reads.

It is well established that oral word reading fluency is a strong predictor of overall reading competence (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001; Harniss, Caros, & Gersten, 2007; Jenkins, 2009; Stecker, Fuchs, & Fuchs, 2005) and that a key relationship exists between fluency and comprehension (Hosp & Fuchs, 2005; Neddenriep, Fritz, & Carrier, 2011; Petscher et al., 2013). It is important to note that the relationship between ORF and comprehension decreases with reading development (e.g., Denton et al., 2011; Jenkins & Jewell, 1993; Silberglitt, Burns, Madyun, & Lail, 2006). However, although the relationship is lower than among elementary children, ORF is still widely used as a screening and progress monitoring measure with students in middle school.

One simple and effective method of assessing fluency is to compute the correct number of words read aloud per minute (CWPM). To conduct this assessment, teachers listen to an individual student read for one minute, note the number of words read and the number of errors, and compute the student's reading rate (Hasbrouck & Tindal, 2006). The structure of elementary schools makes individual assessment more practical than at the secondary levels. At the elementary level, teachers have 20 to 30 students in the same class for much of the school day. In addition, schools often have extended language arts blocks where teachers and/or reading specialists work with students in small groups or individually on a regular basis. Within this context, monitoring of student progress in reading, including fluency development, is a part of typical practice that emphasizes instruction in basic reading.

Middle and high schools, on the other hand, are typically organized by specific disciplinary classes (e.g., science, social studies, English, math) that become increasingly specialized by high school (e.g., biology, chemistry, physics) (Siskin, 1994; Troia, 2006). This organizational structure creates time constraints as teachers are responsible for a larger number of students (e.g., 30 students in up to 8 classes) with whom they have contact for a shorter amount of instructional time (e.g., 45-60 minute instructional periods), which, in turn, limits individualized assessment opportunities.

Another limiting factor at the middle school level is the amount of content that teachers are required to cover within the curriculum, including English language arts classes, where basic reading instruction is no longer a focus. The amount of time that teachers need in order to cover the required curriculum limits the class time available for individualized assessment or instruction. Therefore, individually administered types of assessments at the middle school level are generally not feasible (Berkeley & Riccomini, 2017). For this reason, teachers often opt for group-administered alternate assessments of comprehension (e.g., MAZE; Hosp & Hosp, 2003) rather than individual assessments of oral reading ability (Barth, Stuebing, Fletcher, Denton, Vaughn, & Francis, 2014; Espin, Wallace, Lembke, Campbell, & Long, 2010).

This is highly unfortunate, because as students progress in school, reading fluency becomes more critical in content-area classes. That is, students are expected to read greater amounts of more difficult material in less time. Because the focus of instruction is on subject content, rather than basic reading skills, adolescents with reading problems are often overlooked and struggle with these reading tasks (Mercer, Campbell, Miller, Mercer, & Lane, 2000). Therefore, teachers would be well advised to consider the reading fluency levels of their students when selecting instructional texts and contemplating the provision of instructional supports. Doing so means that teachers need to find alternative ways to assess students' reading fluency.

Alternative Administration

While ORF measures may not be practical for ongoing progress monitoring, they may be feasible as a screening measure that can help inform instruction both classwide and for individual students. The information gained from a classwide screening of reading fluency can assist a teacher in selecting instructional materials that match the skill levels of the class (Hosp, Hosp, & Howell, 2016) as well as selecting instructional methods/approaches to build background knowledge and vocabulary that may help students compensate for weaknesses in basic reading skills (Espin & Deno, 1995; Espin, Shin, & Busch, 2005). The information gained can also assist a teacher in quickly identifying students who do not have satisfactory basic reading skills, making modifications or accommodations based on this information, targeting additional supports for the student, and monitoring the progress of the student more closely throughout the year (Hosp et al., 2016; Wayman, Wallace, Wiley, Ticha, & Espin, 2007).

A possible solution to the logistical dilemmas and time constraints associated with individually administered assessments at the secondary level may lie in findings from studies on classwide peer tutoring (CWPT). CWPT is an instructional approach where students are taught by their peers (Greenwood, Maheady, & Delquadri, 2002). Unlike other approaches that utilize peers, such as cooperative learning, CWPT employs a large amount of structure, with both tutor and tutee being trained in tutoring procedures and supervised by classroom teachers (Maheady, Harper, & Mallette, 2003).

The benefits of peer tutoring include the ability of students to (a) learn more in less time, (b) improve oral reading rates, and (c) increase academic responding (Morano & Riccomini, 2017). It has been successfully implemented at the secondary level in the areas of mathematics (e.g., Calhoon & Fuchs, 2003), social studies (e.g., Marshak, Mastropieri, & Scruggs, 2011), and English language arts (Mastropieri et al., 2001). Further, students have been successfully paired to work together to improve reading (e.g., Fuchs, Fuchs, & Kazdan, 1999; Harris, Marchand-Martella, & Martella, 2000; Spencer, Scruggs, & Mastropieri, 2003). Because students can be taught to work together to teach each other content and reading skills, it seems probable that students could also be taught to recognize and note errors when listening to each other read.

The Current Study

Peer-Assisted Assessment in Reading (PAAR) has the potential to help middle school teachers screen entire classes of students in order to quickly and efficiently identify students who are not fluent readers and are likely to struggle. Using PAAR, students' ability to read text is generally described using three categories with the following criteria: independent (≥97% accuracy), instructional (90-96% accuracy), and frustrational (<90% accuracy) (see Gillett, Temple, Temple, & Crawford, 2017).

In order to investigate whether PAAR is viable for middle school classrooms with a wide range of reading abilities, two exploratory studies investigated whether students who read below grade level could accurately score a wide range of readers, including more proficient readers (i.e., levels above their instructional level). Specifically, the following research questions were investigated:

1. Can struggling readers accurately determine the number of words read and miscues made during a one-minute ORF assessment by (a) an average reader (instructional level), (b) an above-average reader (independent level), and (c) a below-average reader (frustrational level)?
2. Does student performance differ when reading text that is (a) on grade level, (b) one year below, (c) two years below, or (d) three years below?

In addition to the above questions, Study 2 also investigated how accurate students were compared to experienced and preservice teachers; specifically, Is student scoring of words read and miscues made comparable to that of (a) master's-level student teachers, and (b) undergraduate student teachers?

General Method

Overview

To explore the plausibility of using PAAR as a fluency screening administration procedure, researchers developed audio passages read by a range of readers (independent, instructional, and frustrational) with a series of passages (on grade level, and one, two, and three years below). Then participants were trained in ORF scoring. Finally, students listened to and scored the pre-recorded audio passages.

Audio passages. The audio stories used for the practice activities were developed by the two lead researchers. For each grade level (3 through 8), three comparable passages from the Reading Fluency Monitor (RFM; Read Naturally, 2008a) were selected and designated as *frustration, instructional,* or *independent*. All passages were expository in nature (see Table 1 for more information about the passages selected). Then a corresponding rate (number of words read) and errors (number of miscues) were determined for each passage. The planned errors included mispronunciations, omissions, and hesitations, consistent with scoring conventions of oral reading fluency (see Gillet et al., 2017).

Once the errors were included, the lead researcher read aloud the story, making the predetermined errors and stopping at the predetermined number of words read while being recorded by a standard digital microphone. Each digital audio clip included the stating of the directions, the reading aloud of the story with predetermined miscues, and a beeping sound to indicate the beginning and ending of the one-minute time frame. Links to these audio files were embedded into a PowerPoint slide for easy access.

Measure

Test of scoring proficiency. The Test of Scoring Proficiency captured participants' accuracy in scoring the prerecorded audio passages from the RFM. For each passage, participants notated the last word read in one minute and reading miscues, and then calculated correct words per minute (CWPM). Two independent researchers counted the number of words read and miscues that were notated by participants, computed CWPM, and reconciled any differences from 95% (mostly due to illegible writing) to 100% agreement. These scores were used in the analysis.

Procedures

Participant training. One of the researchers provided training to all participants. At the time of the study, the trainer had been providing professional development on oral reading fluency to practicing teachers for seven years. The trainer held a doctoral degree in special education and had previously taught in middle school. The trainer also developed the training materials and activities for the project based on previous work (Riccomini & Stecker, 2005).

After a brief overview of oral reading fluency, the trainer provided a model for each scoring convention with an example that was displayed to participants

Table 1
Passage Technical Information

Grade-Level Passages	Length of Passage Range	Benchmark Correlation Range[a]	Range of Difficulty[b]
3	152 to 202	.92 to .97	.4 to .7
4	171 to 210	.95 to .96	-1.2 to 1.2
5	237 to 254	.92 to .95	-4.1 to 3.9
6	217 to 249	.89 to .95	-5.4 to -.5
7	231 to 247	.93 to .95	-1.3 to -3.9
8	228 to 311	.91 to .95	-4.7 to -3.07

[a]Benchmark correlations are a measure of validity with correlations of .8 or higher.

[b]Range of difficulty indicates median words correct per minute of norming sample compared to medium benchmark. Negative scores are more difficult than the benchmark and positive scores are less difficult than the benchmark. Information available from: https://www.readnaturally.com/knowledgebase/documents-and-resources/25/296

using PowerPoint. Following the model, the trainer reviewed each scoring rule, and then the participants were given an opportunity to practice scoring single sentences that contained the target error. The participants' performance was closely monitored, and any errors were re-explained until corrected. Students had to demonstrate a 90% scoring accuracy before proceeding to online practice activities.

Scoring. Participants were given pencils and folders containing sets of three passages (frustrational, instructional, independent) for each grade level: on the student's grade level as well as one, two, and three years below that grade level ($N = 12$). Passages were randomly ordered by reading level and organized in increasing difficulty from three years below to grade level. When scoring passages, participants removed the first passage from the folder, placed the headphones over their ears, clicked on the audio file that matched the passage, and completed the scoring. This process was repeated until all passages were scored.

Data Analysis

Two procedures were used to compare participant scores of fluency passages to the accurate scoring value. First, a series of one-sample *t*-tests were used to help determine whether the population mean (participant scoring) was equal to our hypothesized valued (accurate scoring). When employing multiple *t*-tests, an adjustment of alpha (e.g., Bonferroni) is typically employed to correct for a familywise error rate. However, for the purpose of this study, an uncorrected alpha of .05 is a more rigorous standard, and thus was employed. In addition, because scores on several passages did not follow a normal distribution, a Wilcoxon Signed-Rank Test was used to verify the results from the *t*-tests. Finally, 95% confidence intervals (CI) were determined and visually analyzed to evaluate whether the upper and lower ranges were consistent with acceptable error ranges when administering and scoring ORF assessments.

Experiment 1

Method

The general methods described above were used for this study. A description of the participants and specific training procedures specific to Experiment 1 follows.

Participants. The participating middle school was located in a small rural district in the southeastern United States; it had an enrollment of 1,000 students in grades six, seven, and eight. The student population was 49% male, and student ethnicities were: 86% Caucasian, 10% African-American, 3% Hispanic, and less than 1% Asian/Pacific Islander. The school was categorized as Title I, where more than 50% of students received free or reduced-price lunch. In the 2005-2006 and 2006-2007 school years, the school did not make "Adequate Yearly Progress" and had a "Needs Improvement" status. Students were selected by the school for a remedial reading class based on the results of the prior year's high-stakes tests in reading and language arts. The required reading class replaced one of the students' elective classes.

Student assent and parental consent were obtained for 88 students enrolled in the reading class; however, nine students were absent during one or more of the testing sessions and, therefore, were not included in the final data analysis. The resulting number of subjects was 79 students (sixth grade = 31, seventh grade = 22, eighth grade = 26) – 32 boys and 47 girls. Subjects were primarily Caucasian (75.9%), followed by African American (12.7%), Hispanic (7.6%), and mixed ethnicities (3.8%). No students with identified disabilities in reading participated in the study.

Prior to the study, a series of assessments was given to gain a comprehensive picture of participants' reading levels. All students scored below the 1st percentile in total word reading (phonetic decoding and sight word recognition) on the Test of Word Reading Efficiency (TOWRE; Torgesen, Rashotte, & Wagner, 2012). In addition, students at all grade levels read fewer correct words per minute (CWPM) on grade-level passages of the Reading Fluency Benchmark Assessor (RFBA; Read Naturally, 2008b) than expected for their respective grade when compared to national norms (see Hasbrouck & Tindal, 2006). Specifically, sixth graders read an average of 105.53 CWPM vs. the expected rate of 127 CWPM; seventh graders read an average of 83.27 CWPM vs. the expected rate of 128 CWPM; and eighth graders read an average of 66.83 CWPM vs. the expected rate of 133 CWPM. To provide a more completed description of the participants, additional achievement data for the Gates-MacGinitie Reading Tests (GMRT; MacGinitie, MacGinitie, Maria, & Dreyer, 2000) and student lexiles are presented in Table 2.

Training procedures. All participants received a 45- to 60-minute training session on the parame-

Table 2
Study 1 & 2 Student Achievement Data by Grade Level

	RFBA CWPM *M* (range)	GMRT %tile *M* (range)			Lexile *M* (range)
		Vocabulary Subtest	Comprehension Subtest	Total Score	
Study 1					
6th Grade (*N* = 28)	94.51 (13 to 145)	21.00 (2 to 66)	26.58 (4 to 62)	25.48 (7 to 57)	735.00 (525 to 935)
7th Grade (*N* = 12)	87.67 (49 to 146)	21.25 (1 to 72)	16.42 (1 to 54)	18.71 (1 to 50)	702.92 (525 to 935)
8th Grade (*N* = 11)	73.24 (44 to 118)	10.27 (1 to 34)	13.45 (1 to 31)	10.64 (2 to 32)	741.36 (560 to 875)
Study 2					
6th Grade (*N* = 16)	98.25 (48 to 136)	26.68 (1 to 37)	35.37 (15 to 60)	32.11 (15 to 55)	763.42 (650 to 885)
7th Grade (*N* = 13)	84.78 (23 to 130)	11.92 (1 to 23)	17.83 (6 to 34)	13.25 (1 to 27)	728.33 (630 to 820)
8th Grade (*N* = 21)	67.45 (35 to 105)	17.18 (1 to 50)	16.45 (3 to 42)	15.73 (2 to 48)	757.73 (630 to 910)

Note. RFBA CWPM: Reading Fluency Benchmark Assessor Correct Words Per Minute; GMRT: Gates-MacGinitie Reading Tests.

ters and scoring conventions for oral reading fluency (described under General Procedures). Participants were trained using single sentences and not complete paragraphs similar in length and frequency of miscues to those used to measure their scoring. The scoring conventions included in the training were mispronounced words, omissions, word reversals, hesitations, and self-corrections.

Results

Student scoring. As shown in Table 3, findings from analysis of student accuracy in scoring *words read* in one minute produced insufficient evidence to conclude that student scoring errors were significantly different from zero (i.e., accurate scoring). Further, statistical analysis could not be conducted on 12 of 36 passages (33%) because all students scored the passage accurately. Visual inspection of confidence intervals showed that upper and lower ranges were within acceptable ranges on all passages.

Statistically significant differences between student scoring of reading *miscues* and zero (i.e., accurate scoring) for each of the passages were found. However, visual inspection of confidence intervals showed that upper and lower ranges were only marginally outside acceptable ranges on all passages (see Table 4).

Counting and computation accuracy. Counting and subtraction errors were tallied to determine accuracy (see Table 5). Overall, students had high levels of accuracy counting notated words read (*M* = 88%), counting notated reading miscues (*M* = 99%), and computing CWPM (*M* = 88%).

Experiment 2

Method

The general methods mentioned above were also utilized for Study 2; however, this study extended the work of Study 1, as follows. First, the training was streamlined so that the critical content could be addressed within 20 minutes. In addition, students were required to pass a "certification test" with 90% accuracy before proceeding to scoring the prerecorded audio passages. Finally, in addition to comparing student performance to accurate scoring, student scoring performance was compared to the performance of master's-level and undergraduate student teachers who were trained to score an ORF as part of their coursework. A description of participants and revised training procedures follows.

Participants. Students were recruited from the same school as in Study 1 during the following school year. No students participated in both Study 1 and Study 2, so participants in Study 2 represented an independent sample. Student teachers were recruited from a university teacher preparation program.

Students. Participating students consisted of 21 sixth graders, 13 seventh graders, and 16 eighth

Table 3
Study 1: Student Identification of Words Read

	8th Grade ($N = 11$)		7th Grade ($N = 12$)		6th Grade ($N = 28$)	
	M (SD)	95% CI	*M (SD)*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	0.00 (0.00)	[0.00, 0.00]	0.17 (0.58)	[-0.20, 0.53]	2.14 (11.14)	[-2.18, 6.46]
Instructional	3.09 (10.25)	[-3.80, 9.98]	-1.25 (4.33)	[-4.00, 1.50]	0.00 (0.00)	[0.00, 0.00]
Frustrational	0.00 (0.00)	[0.00, 0.00]	-1.00 (3.46)	[-3.20, 1.20]	-1.11 (6.26)	[-3.53, 1.32]
One Level Below						
Independent	0.55 (1.29)	[-0.32, 1.41]	0.00 (0.00)	[0.00, 0.00]	-1.43 (7.56)	[-4.36, 1.50]
Instructional	0.00 (0.00)	[0.00, 0.00]	0.00 (0.00)	[0.00, 0.00]	0.35 (1.34)	[-0.16, 0.88]
Frustrational	0.09 (0.30)	[-0.11, 0.29]	0.00 (0.00)	[0.00, 0.00]	0.00 (0.00)	[0.00, 0.00]
Two Levels Below						
Independent	0.00 (0.00)	[0.00, 0.00]	-0.08 (0.29)	[-0.27, 0.10]	0.96 (4.15)	[-0.64, 2.57]
Instructional	0.00 (0.00)	[0.00, 0.00]	0.00 (0.00)	[0.00, 0.00]	0.07 (0.26)	[-0.03, 0.17]
Frustrational	0.36 (0.81)	[-0.18, 0.91]	0.00 (0.00)	[0.00, 0.00]	0.04 (0.19)	[-0.04, 0.11]
Three Levels Below						
Independent	0.09 (0.30)	[-0.11, 0.29]	-0.17 (0.58)	[-.053, 0.20]	2.18 (11.33)	[-2.22, 6.57]
Instructional	0.27 (0.91)	[-0.33,0.88]	0.17 (0.39)	[-0.08, 0.41]	0.00 (0.00)	[0.00, 0.00]
Frustrational	1.36 (3.91)	[-1.26, 3.99]	0.00 (0.43)	[-0.27, 0.27]	-0.04 (0.19)	[-0.11, 0.04]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero). No p-values were < .05 from one-sample t-test analyses, indicating that student scoring was not significantly different than accurate scoring of number of words read. (No p-values were < .05 from Wilcoxon Signed Rank Test Analyses.) 0.00 (0) indicates t-test could not be run because standard deviation was zero (all students scored this passage correctly).

Table 4
Study 1: Student Identification of Reading Miscues

	8th Grade ($N = 11$)		7th Grade ($N = 12$)		6th Grade ($N = 28$)	
	M (SD)	95% CI	*M (SD)*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	4.09 (1.51)*	[3.07, 5.11]	3.08 (1.00)*	[2.45, 3.72]	3.04 (1.29)*	[2.54, 3.54]
Instructional	7.09 (1.22)*	[6.27, 7.91]	4.00 (1.71)*	[2.92, 5.08]	5.11 (2.54)*	[4.12, 6.09]
Frustrational	5.00 (1.18)*	[4.21, 5.79]	4.67 (1.88)*	[3.48, 5.86]	3.00 (1.72)*	[2.33, 3.67]
One Level Below						
Independent	2.64 (1.43)*	[1.67, 3.60]	3.42 (1.38)*	[2.54, 4.29]	2.61 (1.34)*	[2.09, 3.13]
Instructional	4.18 (1.25)*	[3.34, 5.02]	5.50 (2.88)*	[3.67, 7.33]	4.18 (2.00)*	[3.40, 4.95]
Frustrational	3.09 (1.58)*	[2.03, 4.15]	2.67 (1.07)*	[1.98, 3.35]	2.64 (1.59)*	[2.03, 3.26]
Two Levels Below						
Independent	3.64 (1.12)*	[2.88, 4.39]	3.17 (1.64)*	[2.12, 4.21]	1.93 (1.25)*	[1.45, 2.41]
Instructional	6.00 (1.10)*	[5.26, 6.74]	4.08 (1.78)*	[2.95, 5.22]	3.79 (1.77)*	[3.10, 4.47]
Frustrational	4.18 (1.25)*	[3.34, 5.02]	2.42 (1.00)*	[1.78, 3.05]	1.75 (1.27)*	[1.26, 2.24]
Three Levels Below						
Independent	2.18 (0.98)*	[1.52, 2.84]	1.67 (1.44)*	[0.75, 2.58]	1.86 (1.65)*	[1.22, 2.50]
Instructional	4.73 (1.49)*	[3.73, 5.73]	4.00 (2.05)*	[2.70, 5.30]	1.82 (1.09)*	[1.40, 2.24]
Frustrational	3.09 (1.14)*	[2.33, 3.85]	1.67 (1.37)*	[0.80, 2.54]	0.50 (0.92)*	[0.14, 0.86]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero).
*p values < .05 from one-sample t-test analyses, indicating that student identification of reading miscues was significantly different than accurate scoring. (All passages had p-values < .05 on Wilcoxon Signed Rank Tests.)

Table 5
Percentage of Passages With Accurate Counting or Computation

	Accurate Counting of Words Read (mistakes on 2 or fewer passages)	Accurate Counting of Marked Miscues (mistakes on 2 or fewer passages)	Accurate Computation of CWPM (mistakes on 2 or fewer passages)
Study 1			
Students			
6th Grade (*N* = 28)	89%	96%	89%
7th Grade (*N* = 12)	83%	100%	100%
8th Grade (*N* = 11)	91%	100%	82%
Study 2			
Students			
6th Grade (*N* = 21)	81%	90%	95%
7th Grade (*N* = 13)	69%	100%	69%
8th Grade (*N* = 16)	50%	75%	94%
Student Teachers			
Undergrad (*N* = 26)	46%	100%	100%
Graduate (*N* = 26)	100%	96%	100%

Note. CWPM: Correct Words Per Minute. Passages were counted as "accurate" if there were 2 or fewer errors in counting (words read and/or counting miscues) or computation of CWPM (i.e., subtraction) based on student notations on scoring protocols.

graders (*N* = 50). The majority of participants were male (58%); student ethnicities were Caucasian (78%), African American (13%), Hispanic (4%), and mixed ethnicities (4%). Table 2 summarizes achievement data for these students by grade level.

Student teachers. Fifty-two undergraduate (*n* = 26) and graduate (*n* = 26) students enrolled in two special education teacher preparation courses volunteered to participate in the study. A majority were female (94%) and Caucasian (98%). None of the undergraduate students had any years of teaching experience; graduate students' teaching experience ranged from 2 to 27 years. Only 1 of the 52 participants reported having learned about CBM in a previous course.

Training procedures. Two adjustments were made to the training materials in Study 2 based on our experience from the first experiment. First, the materials were revised to shorten the duration required to deliver the training and better focus the participants on the scoring conventions within a shorter amount of class time. Specifically, all participants (students and student teachers) received a more focused and streamlined 20-minute training session on the scoring conventions for ORF (described under General Procedures and streamlined based on the experiences in Study 1). However, to streamline the training, only the most frequently observed scoring conventions were included in the training: mispronounced words, omissions, and hesitations.

Second, a certification test was developed and used to determine proficiency in (a) notating scoring errors within passages, (b) counting words read and reading miscues, and (c) calculating CWPM. The certification test was added based on observed errors made in these areas by participants in Experiment 1. After all students had demonstrated proficiency in the individual errors in isolation, students were provided an opportunity to score a short paragraph containing 3-4 sentences. Finally, participants were required to demonstrate 100% accuracy on the certification test prior to moving forward; students who did not reach 100% were provided one-on-one reteaching until achieving 100% accuracy on the certification test.

Student teachers received the same training as the students but in an online format. This presentation format was selected because past research (Riccomini & Stecker, 2005) demonstrated its effectiveness for training preservice teachers to perform reading fluency assessments. Student teacher participants did not have to take the certification test because it was assumed that adults would not have problems with counting or basic calculation. After training, all 52 student teachers rated the online training as either useful or very useful.

Results

Student scoring. Findings from analysis of student accuracy in scoring words read in one minute produced insufficient evidence to conclude that student scoring errors were significantly different from zero (i.e., accurate scoring). Further, statistical analysis could not be conducted on 15 of 36 passages (42%) because all students scored the passages accurately. Inspection of confidence intervals showed that upper and lower ranges were within acceptable ranges on all passages (see Table 6).

Statistically significant differences between student scoring of reading miscues and zero for each of the passages were found, except for one. However, inspection of confidence intervals showed that upper and lower ranges were only marginally outside acceptable ranges on all passages (see Table 7).

Student teacher scoring. Findings from analysis of student teacher accuracy in scoring words read in one minute produced insufficient evidence to conclude that errors made by graduate-level student teachers were significantly different from zero, and only one passage was significantly different for undergraduate student teachers. Further, statistical analysis could not be conducted on 12 of 18 passages (67%) for graduate student teachers and 13 of 18 passages (72%) for undergraduate student teachers, because all student teachers scored the passage accurately. Inspection of confidence intervals showed that upper and lower ranges were within acceptable ranges on all passages (see Table 8).

Statistically significant differences between student teacher scoring of reading miscues and zero (accurate scoring) were obtained for both 44% of passages scored by graduate student teachers and 78% of passages scored by undergraduate student teachers, indicating that graduate teachers were more accurate than undergraduate teachers. However, inspection of confidence intervals showed that upper and lower ranges were only marginally

Table 6
Study 2: Student Identification of Words Read

	8th Grade (N = 16)		7th Grade (N = 13)		6th Grade (N = 21)	
	M (SD)	95% CI	*M (SD)*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	-0.69 (2.75)	[-2.15, 0.78]	0.62 (1.50)	[-0.29, 1.52]	1.33 (6.11)	[-1.45, 4.11]
Instructional	0.00 (0)	[0.00, 0.00]	-1.77 (4.90)	[-4.73, 1.19]	0.00 (0)	[0.00, 0.00]
Frustrational	-0.53 (2.07)	[-1.68, 0.61]	0.00 (0)	[0.00, 0.00]	0.14 (0.48)	[-0.07, 0.36]
One Level Below						
Independent	0.80 (1.82)	[-0.21, 1.81]	0.00 (0)	[0.00, 0.00]	-0.14 (0.66)	[-0.44, 0.16]
Instructional	0.44 (1.03)	[-0.11, 0.99]	0.00 (0)	[0.00, 0.00]	0.48 (1.63)	[-0.27, 1.22]
Frustrational	-0.50 (2.28)	[-1.72, 0.72]	-0.15 (0.56)	[-0.49, 0.18]	0.00 (0)	[0.00, 0.00]
Two Levels Below						
Independent	0.00 (0)	[0.00, 0.00]	-0.38 (1.39)	[-1.22, 0.45]	0.19 (0.75)	[-0.15, 0.53]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]	0.05 (0.22)	[-0.05, 0.15]
Frustrational	0.13 (0.50)	[-0.14, 0.39]	0.00 (0)	[0.00, 0.00]	0.05 (0.22)	[-0.05, 0.15]
Three Levels Below						
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.19 (0.75)	[-0.21, 0.59]	0.08 (0.28)	[-0.09, 0.24]	1.00 (2.65)	[-0.20, 2.20]
Frustrational	0.00 (0)	[0.00, 0.00]	-0.08 (0.28)	[-0.24, 0.09]	0.00 (0)	[0.00, 0.00]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero). No p-values were < .05 from one-sample t-test analyses, indicating that student scoring was not significantly different than accurate scoring of number of words read. (No p-values were < .05 from Wilcoxon Signed Rank Test Analyses except for the 6th-grade instructional passage, p = .024.) 0.00 (0) indicates t-test could not be run because standard deviation was zero (all students scored this passage correctly).

Table 7
Study 2: Student Identification of Reading Miscues

	8th Grade (*N* = 16)		7th Grade (*N* = 13)		6th Grade (*N* = 21)	
	M (SD)	95% CI	*M (SD))*	95% CI	*M (SD)*	95% CI
On Grade Level						
Independent	3.69 (1.62)*	[2.82, 4.55]	2.38 (2.10)*	[1.11, 3.66]	3.81 (1.44)*	[3.16, 4.46]
Instructional	6.31 (1.74)*	[5.39, 7.24]	3.54 (1.98)*	[2.34, 4.74]	5.90 (3.10)*	[4.50, 7.31]
Frustrational	4.00 (2.16)*	[2.85, 5.15]	2.85 (1.82)*	[1.75, 3.95]	3.29 (1.90)*	[2.42, 4.15]
One Level Below						
Independent	3.25 (1.34)*	[2.54, 3.96]	2.46 (1.71)*	[1.43, 3.50]	3.62 (1.20)*	[3.07, 4.17]
Instructional	3.75 (1.61)*	[2.89, 4.61]	4.85 (1.77)*	[3.78, 5.92]	4.90 (1.81)*	[4.08, 5.73]
Frustrational	3.06 (1.98)*	[2.01, 4.12]	1.92 (1.66)*	[0.92, 2.92]	2.62 (1.50)*	[1.94, 3.30]
Two Levels Below						
Independent	3.13 (1.50)*	[2.33, 3.92]	2.08 (1.71)*	[1.05, 3.11]	2.43 (1.29)*	[1.84, 3.01]
Instructional	4.94 (2.21)*	[3.76, 6.11]	2.69 (1.70)*	[1.66, 3.72]	5.33 (1.43)*	[4.68, 5.98]
Frustrational	2.50 (1.75)*	[1.57, 3.43]	2.46 (1.39)*	[1.62, 3.30]	2.10 (0.94)*	[1.67, 2.52]
Three Levels Below						
Independent	2.69 (1.49)*	[1.89, 3.48]	0.46 (2.40)	[-0.99, 1.91]	2.33 (1.56)*	[1.62, 3.04]
Instructional	3.69 (1.54)*	[2.87, 4.51]	2.46 (1.76)*	[1.40, 3.53]	2.33 (1.24)*	[1.77, 2.90]
Frustrational	2.63 (0.81)*	[2.20, 3.05]	1.62 (1.12)*	[0.94, 2.29]	0.76 (1.00)*	[0.31, 1.21]

Note. Numbers in table indicate deviation from correct scoring (i.e., zero).
*p values < .05 from one-sample t-test analyses, indicating that student identification of reading miscues was significantly different from accurate scoring. (All passages had p-values < .05 on Wilcoxon Signed Rank Tests.)

outside acceptable ranges on all passages for both groups (see Table 9).

Counting and computation accuracy. Counting and subtraction errors made by participants were tallied to determine accuracy (see Table 5). Overall, student teachers were more accurate than students. When counting the number of *words read*, students (in both studies) were not as accurate as the graduate program student teachers in Study 2, but they were more accurate that the undergraduate student teachers. Almost half of the undergraduate student teachers miscounted by one word (suggesting they did not understand how to use the preprinted numbers indicating cumulative numbers of words for each line). When counting the number of marked *reading miscues*, students in Study 1 were as accurate as both graduate and undergraduate student teachers, but sixth and eighth graders in Study 2 were not as accurate as the participating student teachers. When subtracting to compute CWPM, only seventh graders in Study 1 were as accurate as the student teachers.

General Discussion

In both studies, findings showed that after a focused training in ORF assessment scoring, students were highly accurate in identifying *words read* in one minute regardless of grade level (on grade or one, two, and three grades below) or instructional level (independent, instructional, frustration). This finding was the same for both graduate and undergraduate student teachers in Study 2. Because determining words read in an ORF assessment requires counting skills, we documented accuracy here and found that average student performance varied (Study 1 = 88%; Study 2 = 67%), as did student teacher performance (undergraduate teachers = 46%; graduate teachers = 100%).

In both studies, students were less consistent in correctly identifying *reading miscues* regardless of grade level (on grade, or one, two, and three grades below) or instructional level (independent, instructional, frustration). However, inspection of the findings showed that student teachers were not consistently accurate in scoring miscues either. Determining errors made in an ORF assessment also requires counting skills; students were highly accurate in both studies (Study 1 = 96%; Study 2 = 88%) as were the student teachers (undergraduate teachers = 100%; graduate teachers = 96%). Further, determining CWPM in an ORF assessment requires basic computation. Findings here showed that, on average, students were highly accurate (Study 1 = 90%; Study 2 = 86%), as were the student teachers (graduate and undergraduate = 100%).

Implications for Practice

These combined findings provide initial support for the use of PAAR to conduct ORF measures as a screening for struggling readers at the middle school level, where this type of measure is typically not feasible when administered one-on-one by the classroom teacher. In addition to having the potential to serve as a screening tool to identify individual students who may be struggling, PAAR also has the potential to help teachers in selecting instructional materials and designing classwide instruction.

Although the use of PAAR in middle school may be more feasible than assessing individual students, it is not intended as a comprehensive measure to be used for educational decisions such as placement or formal evaluation. The overarching goal for the design and implementation of this research was to determine if peers could assess ORF with acceptable levels of reliability that would provide teachers with useful information for immediate classroom instructional

Table 8
Study 2: Student Teacher Identification of Words Read

	Graduate Level		Undergraduate Level	
	M (SD)	95% CI	*M (SD)*	95% CI
8th Grade				
Independent	0.00 (0)	[0.00, 0.00]	5.38 (15.20)	[-7.33, 18.08]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Frustrational	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
7th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Frustrational	0.11 (0.32)	[-0.05, 0.27]	0.13 (0.34)	[-0.06, 0.31]
6th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.04 (0.20)	[-0.04, 0.12]
Instructional	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Frustrational	0.00 (0.30)	[-0.12, 0.12]	0.00 (0)	[0.00, 0.00]
5th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.23 (0.82)	[-0.10, 0.56]	0.12 (0.59)	[-0.12, 0.35]
Frustrational	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
4th Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	0.18 (0.39)	[-0.03, 0.38]	0.22 (0.43)*	[0.01, 0.43]
Frustrational	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
3rd Grade				
Independent	0.00 (0)	[0.00, 0.00]	0.00 (0)	[0.00, 0.00]
Instructional	-0.13 (0.35)	[-0.42, 0.17]	0.00 (0)	[0.00, 0.00]
Frustrational	-0.13 (0.35)	[-0.42, 0.17]	0.00 (0)	[0.00, 0.00]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero). 0.00 (0) indicates *t*-test could not be run because standard deviation was zero (all teachers scored this passage correctly).
**p* values < .05 from one-sample *t*-test analyses, indicating that teacher identification of words read was significantly different from accurate scoring. (Identical results were found in analysis using Wilcoxon Signed Rank Tests.)

Table 9
Study 2: Student Teacher Identification of Reading Miscues

	Graduate Level		Undergraduate Level	
	M (SD)	95% CI	*M (SD)*	95% CI
8th Grade				
Independent	0.78 (1.09)	[-0.06, 1.62]	1.88 (1.73)*	[0.43, 3.32]
Instructional	2.11 (0.78)*	[1.51, 2.71]	2.13 (1.25)*	[1.08, 3.17]
Frustrational	1.89 (0.60)*	[1.43, 2.35]	2.88 (0.64)*	[2.34, 3.41]
7th Grade				
Independent	0.11 (1.45)*	[0.83, 0.61]	0.06 (1.48)*	[-0.73, 0.85]
Instructional	0.39 (1.04)	[-0.34, 0.45]	0.31 (0.79)*	[0.11, 0.74]
Frustrational	0.33 (1.24)	[-0.28, 0.95]	0.75 (0.93)*	[0.25, 1.25]
6th Grade				
Independent	0.08 (0.56)	[-0.15, 0.30]	0.50 (0.91)*	[0.13, 0.87]
Instructional	0.00 (1.47)	[-0.59, 0.59]	1.54 (2.02)*	[0.72, 2.36]
Frustrational	0.46 (0.72)*	[0.15, 0.76]	0.29 (0.49)	[0.17, 0.74]
5th Grade				
Independent	0.23 (0.95)	[-0.15, 0.61]	0.85 (1.35)*	[0.30, 1.39]
Instructional	-0.23 (0.99)	[-0.63, 0.17]	0.42 (1.17)	[-0.05, 0.90]
Frustrational	0.46 (0.91)*	[0.10, 0.83]	0.81 (1.02)*	[0.40, 1.22]
4th Grade				
Independent	0.06 (0.43)	[-0.16, 0.28]	0.00 (0.59)	[-0.30, 0.30]
Instructional	1.41 (1.23)*	[0.78, 2.04]	1.83 (1.42)*	[1.12, 2.54]
Frustrational	1.53 (0.51)*	[1.26, 1.79]	1.72 (0.58)*	[1.44, 2.01]
3rd Grade				
Independent	1.88 (1.25)*	[0.83, 2.92]	2.56 (0.53)*	[2.15, 2.96]
Instructional	0.63 (1.41)	[-0.55, 1.80]	0.78 (0.67)*	[0.27, 1.29]
Frustrational	0.25 (0.46)	[-0.14, 0.64]	0.11 (0.33)	[-0.15, 0.37]

Note. Numbers in the table indicate deviation from correct scoring (i.e., zero).
*p values < .05 from one-sample t-test analyses, indicating that teacher identification of words read was significantly different from accurate scoring. (Identical results were found in analysis using Wilcoxon Signed Rank Tests.)

decisions (e.g., use of assistive technology, alternative reading sources, small-group instruction). We do not recommend using peers for formal assessments, although our results indicated student peers were more accurate than preservice teachers.

Teachers will have to provide training to students prior to conducting peer assessment, which may be time consuming. As middle school teachers consider the use of peers for assessing ORF, they should carefully consider and weigh the cost-to-benefit ratio; our results indicate the potential benefits of using peers as a screening measure in middle school grades.

Another important and surprising finding was that middle school students were more accurate than the preservice student teacher participants. This could be explained through anecdotal observations made of the preservice teachers. These participants were taking an entire course on teaching reading where significant time was allocated to assessing reading beyond correct and incorrect and recording the type of miscue (e.g., mispronunciations, omissions, and hesitations), consistent with scoring conventions of oral reading fluency (see Gillet et al., 2017).

It is possible that the preservice teachers were trying to record too much information in the early stages of practicing measuring ORF, which negatively impacted their accuracy. The middle school participants, on the other hand, were just recording correct or incorrect, and this may have resulted in the higher accuracy. Either way, the training necessary to achieve high levels of accuracy in measuring ORF along with specific miscues used in this research project was likely not sufficiently intensive.

When screening to identify struggling readers, accuracy of *words read* in a minute has practical importance and may be more important than identification of specific miscues. For example, through peer-assisted assessment in a seventh-grade social studies class, Suzie may identify that Johnny reads 86 CWPM. Because the peer is more likely not to identify all miscues made by the reader, Johnny may actually read 82 CWPM (i.e., Suzie failed to identify 4 errors). However, considering that a typical seventh grader reads approximately 150 CWPM, the educational impact is the same – the student is reading far fewer CWPM with grade-level material than expected for his or her grade. In this example, regardless of the peer's failure to identify all of the reader's miscues, a classroom teacher would immediately see that this is a student who is potentially at risk and should be monitored more closely by a teacher. As with any screening measure, the result should be considered as a single data point that can be used to make decisions about a student (Hasbrouck & Tindal, 2006).

Clearly, data obtained from PAAR is not appropriate for formal identification or progress monitoring purposes if obtained from student scoring. However, these findings have implications for content-area classrooms at the middle school level. If utilized classwide, content-area teachers could potentially teach students how to score ORF assessments and obtain data on students in all of their classes in an extremely short period of time. This information could then be used both to inform instruction and to identify students whose progress should be closely monitored.

Conclusions

Nationally, reading performance continues to be a major concern. By the time students reach middle school, they are expected to read purposefully and comprehend, but many fail to do so (Ivey & Broaddus, 2000; Troia, 2006). In 2017, 37% of fourth-grade students performed at or above the *Proficient* level in reading skills, and only 36% of eighth-grade students performed at or above the *Proficient* level (National Center for Educational Statistics, 2018). Because today's U.S. classrooms are more diverse than ever before, students' reading levels vary widely in a single classroom, and the higher the grade level, the wider the range of reading abilities. To deliver quality instruction to meet the individual needs of all students in the classroom, teachers need to be aware of their students' current levels of performance.

Because middle school teachers often do not hear students read aloud, they may be unaware of which students in their classes are struggling with the assigned reading material. Fluency assessments are efficient, reliable, and valid indicators of reading proficiency when used as screening measures (Fuchs et al., 2001; Hasbrouck & Tindal, 2006). Knowing which students have the most difficulty in reading a textbook fluently can help teachers select appropriate texts for homework and instruction, match students to appropriate modifications (e.g., use of a text-to-speech version of the content), tailor instructional approaches to meet the needs of entire classes, and/or intensify instruction to address the reading capabilities of all students in the class (Berkeley & Lindstrom, 2011; Riccomini, Morano, & Hughes, 2017). Findings from this study suggest that using PAAR as a screening assessment has potential to help teachers make these important instructional decisions.

References

American Recovery and Reinvestment Act (ARRA) of 2009, Pub. L. No. 111-5, 123 Stat. 115, 516 (2009).

Barth, A. E., Stuebing, K. K., Fletcher, J. M., Denton, C. A., Vaughn, S., & Francis, D. (2014). The effect of reading duration on the reliability and validity of middle school students' ORF performance. *Assessment for Effective Intervention : official journal of the Council for Educational Diagnostic Services*, *40*(1), 53-64. https://doi.org/10.1177/1534508414545643

Berkeley, S., & Lindstrom, J. H. (2011). Technology for the struggling reader: Free and easily accessible resources. *Teaching Exceptional Children*, *43*, 48-55.

Berkeley, S., & Riccomini, P. J. (2017). Academic progress monitoring. In J. M. Kauffman & D. P. Hallahan (Eds.), *Handbook of special education* (2nd ed., pp. 218-232). New York, NY: Routledge. https://doi.org/10.1177/004005991104300405

Brigham, F. J., Berkeley, S., & Walker, E. (2012). Navigating the measurement and monitoring maze. *Exceptionality, 20*, 133-142. doi:10.1080/09362835.2012.669298

Brigham, F. J., Tochterman, S., & Brigham, M.S.P. (2000). Students with emotional and behavioral disorders and their teachers in test-linked systems of accountability. *Assessment for Effective Intervention, 26*(1), 19-27. https://doi.org/10.1177/073724770002600104

Calhoon, M. B., & Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial & Special Education, 24*, 235-245. https://doi.org/10.1177/07419325030240040601

Calkins, L., Ehrenworth, M., & Lehman, C. (2012). *Pathways to the common core: Accelerating achievement*. Portsmouth, NH: Heinemann.

Deno, S. L. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Intervention, 28*(3-4), 3-12. https://doi.org/10.1177/073724770302800302

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., … Francis, D. J. (2011). The relations among oral reading and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading, 15,* 109-135. https://doi.org/10.1080/10888431003623546

Espin, C. A., & Deno, S. L. (1995). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 20*, 121-142.

Espin, C. A., Shin, J., & Busch, T. W. (2005). Curriculum-based measurement in the content areas: Vocabulary-matching as an indicator of social studies learning. *Journal of Learning Disabilities, 38,* 353-363. https://doi.org/10.1177/00222194050380041301

Espin, C. A., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice,* 25, 60-75. https://doi.org/10.1111/j.1540-5826.2010.00304.x

Every Child Succeeds Act. (2015). Public Law No. 114-95. Available from https://www.congress.gov/bill/114th-congress/senate-bill/1177.

Fuchs, D., Fuchs, L., & Compton, D. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78*, 263-279. https://doi.org/10.1177/001440291207800301

Fuchs, L. S., Fuchs, D., Hosp, M., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239-256. https://doi.org/10.1207/S1532799XSSR0503_3

Fuchs, L., Fuchs, D., & Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial & Special Education, 20*, 309-318. https://doi.org/10.1177/074193259902000507

Gillet, J. W., Temple, C. A., Temple, C. N., & Crawford, A. N. (2017). *Understanding reading problems* (9th ed.). Boston, MA: Pearson.

Greenwood, C. R., Maheady, L., & Delquadri, J. C. (2002). Class-wide peer tutoring. In G. Stoner, M. R. Shinn, & H. Walker (Eds.), *Interventions for achievement and behavior problems* (2nd ed., pp. 611-649). Washington, DC: National Association of School Psychologists.

Harniss, M. K., Caros, J., & Gersten, R. (2007). Impact of the design of U.S. history textbooks on content acquisition and academic engagement of special education students: An experimental investigation. *Journal of Learning Disabilities, 40*, 100-110. https://doi.org/10.1177/00222194070400020101

Harris, R., Marchand-Martella, N., & Martella, R. (2000). Effects of a peer-delivered corrective reading program. *Journal of Behavioral Education, 10*, 21-36. https://doi.org/10.1023/A:1016691705754

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*, 636-644. https://doi.org/10.1598/RT.59.7.3

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42*, 963-974. doi:10.1037/0003-066x.42.11.963

Hiebert, E. H. (2009). The (mis)match between texts and students who depend on schools to become literate. In E. H. Hiebert & M. Sailors (Eds.), *Finding the right texts: What works for beginning and struggling readers* (pp. 1-20). New York, NY: Guilford Press.

Hosp, M. K., & Hosp, J. L. (2003). Curriculm-based measurement for reading, spelling, and math: How to do it and why. *Preventing School Failure, 48,* 10-17. https://doi.org/10.1080/1045988X.2003.10871074

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34*(1), 9-26.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement* (2nd ed.). New York, NY: Guilford Press.

Individuals With Disabilities Improvement Act of 2004, 20 U.S.C. § 1400 *et seq.* (2005).

Ivey, G., & Broaddus, K. (2000). Tailoring the fit: Reading instruction and middle school readers. *The Reading Teacher, 54,* 68-78.

Jenkins, J. R. (2009). Measuring reading growth: New findings on progress monitoring (PM). *New Times for DLD, 27*(1), 1-2.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formatove teaching: Reading Aloud and Maze. *Exceptional Chirldren, 59*(5), 421-432. https://doi.org/10.1177/001440299305900505

Kendall, J., Ryan, S., Alpert, A., Richardson, A., & Schwols, A. (2012). *State adoption of the comon core state standards: The 15 percent rule.* Denver, CO: Mid-Continent Research for Education and Learning.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95,* 3-21. https://doi.org/10.1037/0022-0663.95.1.3

Lembke, E. S., Hampton, D., & Hendricker, E. (2013). Data-based decision making in academics using curriculum-based measurment. In J. W. Lloyd, T. J. Landrum, B. G. Cook, & M. Tankersley (Eds.), *Research-based approaches for assessment* (pp. 18-31). Boston, MA: Pearson.

MacGinitie, W. H., MacGinitie, R., Maria, R., & Dreyer, L. (2000). *Gates-MacGinitie Reading Tests.* Rolling Meadows, IL: Riverside Publishing.

National Center for Educational Statistics. (2018). *NAEP reading report card.* Retrieved from https://www.nationsreportcard.gov/.

Maheady, L., Harper, G. F., & Mallette, B. (2003). A focus on class wide peer tutoring. *Current Practice Alerts, 8.*

Marshak, L., Mastropieri, M. A., & Scruggs, T. E. (2011). Curriculum enhancements in inclusive secondary social studies classrooms. *Exceptionality, 19*, 61-74. https://doi.org/10.1080/09362835.2011.562092

Mastropieri, M. A., Scruggs, Mohler, L., Beranek, M., Boon, R., Spencer, V., & Talbott, E. (2001). Can middle school students with serious reading difficulties help each other and learn anything? *Learning Disabilities Research & Practice, 16*, 18-27. https://doi.org/10.1111/0938-8982.00003

Meisinger, E. B., Bloom, J. S., & Hynd, G. W. (2010). Reading fluency: Implications for the assessment of children with reading disabilities. *Annals of Dyslexia, 60(1)*, 1-17. doi:10.1007/s11881-009-0031-z

Mercer, C. D., Campbell, K. U., Miller, M., Mercer, K. D., & Lane, H. B. (2000). Effects of a reading fluency intervention for middle schoolers with specific learning disabilities. *Learning Disabilities Research & Practice*, *15,* 179-189. https://doi.org/10.1207/SLDRP1504_2

Morano, S., & Riccomini, P. J. (2017). Re-examining the literature: The impact of peer tutoring on higher order learning. *Preventing School Failure: Alternative Education for Children and Youth, 61*(2), 104-115. doi: 10.1080/1045988X.2016.1204593

Neddenriep, C. E., Fritz, A. M., & Carrier, M. E. (2011). Assessing for generalized improvements in reading comprehension by intervening to improve reading fluency. *Psychology in the Schools, 48*(1), 14-27. doi:10.1002/pits.20542

Petscher, Y., Cummings, K. D., Biancarosa, G., & Fien, H. (2013). Advanced (measurement) applications of curriculum-based measurement in reading. *Assessment for Effective Intervention, 38*(2), 71-75. doi:10.1177/1534508412461434

Read Naturally. (2008a). *Reading fluency monitor*. Saint Paul, MN: Read Naturally

Read Naturally. (2008b). *Reading fluency benchmark assessor*. Saint Paul, MN: Read Naturally.

Riccomini, P. J., Morano, S., & Hughes, C. A. (2017). Big ideas in special education: Specially designed instruction, high-leverage practices, explicit instruction, and intensive instruction. *Teaching Exceptional Children, 50*(6), 1-8. https://doi.org/10.1177/0040059917724412

Riccomini, P. J., & Stecker, P. M. (2005). Effects of technology-enhanced practice on scoring accuracy of oral reading fluency. *Journal of Special Education Technology, 20*(3), 5-12. https://doi.org/10.1177/016264340502000301

Roach, A. T. (2008). Intervention validity of assessments used in special education: Introduction to the special series. *Assessment for Effective Intervention, 34*(1), 3-5. doi:10.1177/1534508408314109

Santi, K., & Vaughn, S. (2007). Progress monitoring: an integral part of instruction. *Reading & Writing, 20*, 535-537. https://doi.org/10.1007/s11145-007-9053-2

Schwanenflugel, P. J., & Ruston, H. P. (2008). Becoming a fluent reader: From theory to practice. In M. R. Kuhn & P. J. Schwanenflugel (Eds.), *Fluency in the classroom* (pp. 1-16). New York, NY: Guilford Press.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationships of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*(5), 527-535. doi:10.1002/pits.20175

Siskin, L. (1994). *Realms of knowledge: Academic departments in secondary schools*. Bristol, PA: Falmer Press.

Soriano, M., Miranda, A., Soriano, E., Nievas, F., & Felix, V. (2011). Examining the efficacy of an intervention to improve fluency and reading comprehension in Spanish children with reading disabilities. *International Journal of Disability, Development and Education, 58*(1), 47-59. doi:10.1080/1034912x.2011.547349

Spencer, V. G., Scruggs, T. E., & Mastropieri, M. A. (2003). Content area learning in middle school social studies classrooms and students with emotional or behavioral disorders: A comparison of studies. *Behavioral Disorders, 28*, 77-93. https://doi.org/10.1177/019874290302800208

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools*, *42*, 795-819. https://doi.org/10.1002/pits.20113

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *TOWRE: Test of Word Reading Efficiency.* Austin, TX: Pro-Ed.

Troia, G. A. (2006). Meaningful assessment of content-area literacy for youth with and without disabilities. *Assessment for Effective Intervention, 31*, 69-80. https://doi.org/10.1177/073724770603100206

Venn, J. (2013). Curriculum-based assessment. In J. W. Lloyd, T. J. Landrum, B. G. Cook, & M. Tankersley (Eds.), *Research-based approaches for assessment* (pp. 84-93). Boston, MA: Pearson.

Wayman, M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41,* 85-120. https://doi.org/10.1177/00224669070410020401

A Comparison of the Effectiveness of Using CRA-SIM vs. Direct Instruction to Teach Multiplication With Regrouping

Margaret M. Flores,[1] Bradley J. Kaffar,[2] and Vanessa Hinton[1]

Abstract

Mathematics learning standards across the United States include conceptual knowledge of numbers and use of modeling and explanation within computation. The concrete-representational-abstract (CRA) sequence and the Strategic Instruction Model (SIM) have been shown to be effective in increasing students' fluency in computation, including multiplication with regrouping, with a focus on conceptual knowledge. To date, the CRA-SIM multiplication research has not included a comparison to another method. The purpose of this study was to fill this gap by comparing CRA-SIM to Direct Instruction (DI). Twenty-nine elementary- and middle-level students with and without disabilities received either CRA-SIM or DI. While each program resulted in increased fluency in multiplication with regrouping, CRA-SIM led to slight increases in fluency compared to DI. Implications and program components that influenced these results will be discussed.

Keywords: multiplication, mathematics difficulties, CRA-SIM, Direct Instruction

Mathematics standards adopted by most states across the United States include mathematical practices intended to lead students toward effective mathematical thinking and application within real-world situations (Common Core State Standards Initiative [CCSSI], 2010). Mathematical practices, in turn, include identification of meaning within problems so that reasonable solutions can be developed. Evaluation of possible solutions requires deep knowledge and understanding of numbers and operations such that multiple pathways toward a solution can be generated and then judged for reasonableness (Barlow & Harmon, 2012).

In order for students to engage in these practices, instruction must emphasize conceptual understanding, perseverance in problem solving, as well as proficiency in procedural knowledge and fluency in operations (National Mathematics Advisory Panel, 2008). It is important for students to have conceptual knowledge because poor conceptual understanding often leads to confusion or error patterns as mathematics problem solving becomes more complex (Kroesbergen, van't Noordende, & Kolkman, 2014). Therefore, development of conceptual understanding should begin in the early grades and continue as more complex mathematical processes are introduced (Witzel, Ferguson, & Mink, 2012). Further, mathematics interventions for students who struggle should include modeling and emphasis on conceptual understanding as fluency develops (Dacey & Drew, 2012; Kiuhara & Witzel, 2014).

As operations become more complex, algorithms are included as one way to solve problems. Within the standards adopted by most states across the nation, students are required to show fluency in standard algorithms, including multiple-step processes associated with multiplication with regrouping and long division. When designing interventions that lead to fluency in using traditional algorithms, conceptual understanding as to how and why these algorithms work may be included.

[1]Auburn University; [2]St. Cloud State University

Flores, M. M., Kaffar, B. J., & Hinton, V. (2019). A Comparison of the Effectiveness of Using CRA-SIM vs. Direct Instruction to Teach Multiplication With Regrouping. *International Journal for Research in Learning Disabilities, 4*(1), 27-40. https://doi.org/10.28987/ijrld.4.1. 27

Teaching Conceptual Understanding of Numbers and Operations

The use of the concrete-representational-abstract (CRA) instructional sequence is one way to build or reinforce conceptual understanding of numbers and operations. To assist students in learning and remembering procedural steps, CRA has been combined with the Strategic Instruction Model™ (SIM), a model in which students learn to systematically complete complex academic tasks through explicit instruction and programmed generalization with procedural knowledge acquired through a streamlined set of steps, usually in the form of a mnemonic device (Deshler & Schumaker, 1986).

The combination of CRA and SIM (CRA-SIM) works as follows. First, students learn to compute problems using concrete objects through explicit instruction (i.e., teacher demonstration, guided practice, and independent practice). Next, students solve problems by drawing representations of numbers, fading their dependence on physical objects. After having mastered computation using drawings, students learn a strategy for following algorithmic procedures in the form of a device that aids memory. This mnemonic strategy provides students with a scaffold from modeling problems with drawings to solving problems using numbers only. The goal of CRA-SIM is to develop computational fluency based on conceptual understanding of numbers and operations.

CRA-SIM. Mercer and Miller (1992) began the line of CRA-SIM research with the development of a curriculum called the *Strategic Math Series*, in which the researchers taught basic addition, subtraction, multiplication, and division using CRA-SIM. Within separate curriculum materials, students learned to perform basic operations using concrete objects, drawings, and a procedural mnemonic strategy called "DRAW" for solving problems using numbers only. The procedural steps were as follows: (a) discover the sign; (b) read the problem; (c) answer or draw and check; and (d) write the answer. Mercer and Miller (1992) found that CRA-SIM was more effective than traditional basal mathematics curriculum materials.

In a later study, Miller and Mercer (1993) investigated the effects of CRA-SIM for elementary students with specific learning disabilities (SLD), and obtained similar results with regard to effectiveness. Additional studies have demonstrated positive effects using CRA-SIM to teach basic operations as well as more advanced skills such as fractions and algebraic equations (Harris, Miller, & Mercer, 1995; Morin & Miller, 1998), and integers, fractions, and algebra (Maccini & Hughes, 2000; Maccini & Ruhl, 2000; Witzel, Mercer, & Miller, 2003).

The most recent CRA-SIM research focused on computation involving regrouping, first with addition and subtraction and then with multiplication. For example, Flores (2009, 2010) taught subtraction with regrouping to elementary students experiencing failure in mathematics using explicit instruction combining CRA methods and the DRAW mnemonic strategy. Results of this multiple-probe-across-students study showed a functional relation between CRA-SIM and subtraction with regrouping across all students.

In another study, Miller and Kaffar (2011) compared CRA-SIM to traditional basal materials, teaching addition with regrouping to students with learning difficulties in mathematics. Whereas the comparison group received 16 lessons from a traditional second-grade basal series, the CRA-SIM group received 16 lessons, including explicit instruction using the CRA sequence and a procedural mnemonic strategy (Read the problem, Examine the ones column, Note ones in the ones column, Address the tens column, Mark tens in the tens column, Examine and note hundreds and exit with a quick check; RENAME). The researchers extended the regrouping research by having students use place-value mats to organize base-ten blocks and drawings during the concrete and representational stages of instruction. After solving addition with regrouping problems with objects and drawings, students learned the mnemonic strategy, RENAME, as they transitioned to problem solving using numbers only in the abstract phase. Compared to the basal group, students in the CRA-SIM group demonstrated significant gains in addition with regrouping.

Using a multiple-probe-across-students design, Mancl, Miller, and Kennedy (2012) extended the line of CRA-SIM subtraction with regrouping research to include students with SLD. The students solved problems using base-ten blocks and a place-value mat at the concrete level of instruction. During the representational phase, they learned to solve problems using pictures and a place-value mat. The researchers taught the RENAME strategy between the representational and abstract phases of instruction; finally, students progressed to problem solving using numbers only. The results indicated a functional relation between CRA-SIM and accuracy in subtraction.

Flores, Hinton, and Strozier (2014) taught subtraction with regrouping and multiplication with regrouping using a multiple-probe-across-behaviors design. Students receiving tier three intervention within a response to intervention prevention framework (Vaughn & Fuchs, 2003) learned subtraction with regrouping in the tens place, subtraction with regrouping in the tens and hundreds places, and multiplication with regrouping using one-digit and two-digit multipliers (bottom number). The algorithms were taught using base-ten blocks during the concrete phase and drawings during the representational phase. The students learned the RENAME strategy and applied it when solving problems during the abstract phase. A functional relationship was found between CRA-SIM and students' regrouping performance across subtraction and multiplication.

In an extension of this line of CRA multiplication research, Flores, Schweck, and Hinton (2014) included elementary students with SLD. Focusing solely on problems with two-digit multipliers, their study added place-value mats, which students used to organize base-ten blocks and drawings during the concrete and representational phases. Using a multiple-probe-across-students design, results showed a functional relation between CRA-SIM and multiplication with regrouping performance.

Finally, Flores and Franklin (2014) explored CRA research on methods implemented by practitioners instead of a researcher or graduate assistant, with findings that indicated CRA improves student performance for multiplication of two-digit numbers.

Direct Instruction. As indicated, the research supporting CRA-SIM for teaching multiplication with regrouping has demonstrated promising results for small groups of students. However, to date, no study has compared CRA-SIM to another method as a way of showing its effectiveness compared to other evidence-based practices with a focus on multiplication with regrouping.

One form of instruction classified as an evidence-validated instructional method for students who struggle with mathematic concepts is Direct Instruction (DI) (Gersten et al., 2009). DI has been found to be effective in teaching mathematics for at-risk students and students with disabilities (Przychodzin, Marchand-Martella, Martella, & Azim, 2004). DI programs are comprehensive and establish long-term goals that address strands, or skill sets embedded in larger concepts. For example, *Corrective Mathematics* addresses skill sets of solving for facts, place value, operations, and word problems within the larger idea of what addition, subtraction, multiplication, and division imply (Przychodzin et al., 2004).

When applying DI, the following procedures are used (Watkins & Slocum, 2004). First instruction organizes central concepts and strategies to ensure efficient student learning. Second, clear and systematic methods of teacher communication are implemented to decrease misunderstanding. Third, instruction includes structured verbal exchanges between teachers and students, which increase student engagement, progress monitoring, and repeated practice. Fourth, instruction strategically integrates skills to ensure efficient learning and understanding. Fifth, instructional concepts are arranged in strands, in which learning develops throughout the program while continually reviewing and generalizing information.

Similar to CRA-SIM, researchers have demonstrated positive effects of using DI to teach operations as well as more advanced mathematics such as fractions (Adams & Engelmann, 1996). Studies examining fraction instruction have included students with and without SLD in grades seven and higher (Flores & Kaylor, 2007; Kelly, Gersten, & Carnine, 1990; Scarlato & Burr, 2002). Since several skill sets are addressed in DI research, it is difficult to hone in on specific operations; nevertheless, studies have addressed embedded skills such as multiplication. Most DI research makes comparisons to basal programs used within the school systems and involves solving word problems using mathematic operations.

Darch, Carnine, and Gersten (1984) taught word problems with addition, subtraction, multiplication, and division operations to students in the fourth grade who demonstrated a deficit in addition, subtraction, multiplication, or division, and in solving for word problems. Seventy-three participants were randomly assigned to one of four groups: (a) DI method with fixed amount of practice, (b) basal instruction with fixed amount of practice, (c) DI method with extended practice, and (d) basal instruction with extended practice. The students who received DI outperformed peers who received basal instruction in solving word problems. The DI method with extended practice group scored the highest of all four groups.

In a later study, Wilson and Sindelar (1991) compared the effectiveness of using a DI and a basal program to teach addition and subtraction word problems. The study included 62 students with SLD in mathematics within nine elementary schools. Participants were able to solve addition and subtraction facts with at least 80% accuracy and read at least at a 1.5 grade level. Students were randomly assigned to three groups: (a) DI with word problem types taught individually, (b) DI with word problem types integrated, and a (c) basal program with word problems taught individually. Students who received DI made substantial increases in solving mathematical word problems involving addition and subtraction compared to students who received the basal program instruction.

Other researchers (Woodward & Brown, 2006) compared instructional approaches using the lens of standards-based curriculum. Specifically, Woodward and Brown (2006) examined a DI curriculum and a curriculum that was not explicit, but followed the National Teachers of Mathematics Standards 2000 guidelines. Participants, who were in the sixth grade, included students with and without disabilities, and were recommended by teachers for intensive, remedial instruction. The students had difficulty solving whole-number problems using the four basic operations (i.e., addition, subtraction, multiplication, and division) and place value. Results of the study showed that students who received DI instruction significantly outperformed the comparison group on posttest measures of the four basic operations and place value.

The purpose of the current study was to extend the CRA-SIM literature by comparing its effects to DI, an established, evidence-validated, and explicit instructional method.

Method

Participants

Participants were 29 students in the United States in grades four through seven enrolled in remedial summer programs based on their lack of response to instruction and their performance on the annual state assessment, indicating that they had not made adequate progress toward mastering state standards. Fifteen of the students had identified disabilities and were eligible for special education services under the categories of specific learning disabilities (7), emotional behavioral disorders (3), autism spectrum disorders (1), and other health impairments (4).

The criteria for student participation were as follows: (a) parental permission to participate in research, and (b) meeting placement guidelines in the DI *Corrective Mathematics Multiplication* (Engelmann, & Carnine, 2005) program, which meant students had to demonstrate proficiency in basic multiplication, addition with regrouping, and multiplication involving one-digit numbers. The students were matched as pairs by assessment performance,

Table 1
Student Demographic Information by Group

Demographic Information		DI Group	CRA-SIM Group
Gender	Male	10	9
	Female	5	5
Grade	Grade 4	6	8
	Grade 5	5	5
	Grade 6	3	1
	Grade 7	0	1
Ethnicity	White	11	13
	African American	2	1
	Latino/a	1	1
Disability	Specific Learning Disability	4	3
	Emotional Behavioral Disorder	2	1
	Autism Spectrum Disorder	1	0
	Other Health Impairment	1	3
	None	6	8
Computation Achievement[a]		Standard Score = 82 Range = 72-98	Standard Score = 81 Range = 68-99

[a]Standard score from Operations Subtest, *KeyMath Diagnostic Assessment* (Connolly, 2007).

disability category when possible, ethnicity, and grade. The members of each pair were randomly assigned to either the CRA-SIM (15 students) or DI (14 students) groups through a coin toss. Statistical analysis showed there were no differences between the groups. Student characteristics by group are presented in Table 1.

Setting

The study was conducted in remedial summer intervention programs at a combined elementary and middle school. The annual programs are a collaborative effort between a university and a local school district in a small Midwestern city. Students attended the program for four and one half hours per day, four days per week for six weeks. District teachers provided general instruction in reading, written expression, and mathematics. In addition, university graduate students pulled children out for up to two periods per day to provide more intensive mathematics interventions. The graduate students were certified teachers enrolled in a special education licensure program for specific learning disabilities; five teachers had recently completed a bachelor's degree and initial certification with less than one year of teaching experience, and one teacher had 17 years of general education teaching experience.

Instruction occurred in separate classroom settings for 50 minutes each day. For both DI and CRA-SIM, the students received small-group instruction with four to six students in each group. Across groups, students were seated at tables or desks surrounding the teacher, who provided instruction using a white board and materials appropriate for each intervention.

Materials

Direct Instruction materials. The teacher materials consisted of the *Corrective Mathematics Multiplication* presentation book (Engelmann & Carnine, 2005). The teachers used the script within the book and the white board to implement the program. The student materials consisted of the *Corrective Mathematics Multiplication* student workbook. The student lesson materials involved a page with multiple sections in which students completed short activities related to multiplication and multiplication with regrouping procedures; the students practiced multiplication facts, mental computation, application of place value, addition with regrouping, word problems, and multiplication with regrouping within each lesson. The materials also included a section in which students recorded points that could be earned for completion and accuracy within each lesson section.

Although it was not part of the curriculum, the students in the DI group also completed a learning contract prior to instruction, whereby they agreed to work hard to learn multiplication, and, in turn, their teachers agreed to work hard to provide instruction. The contract was added to the program because CRA-SIM includes a contract, and the researchers wanted equivalent student motivation within each group.

CRA-SIM materials. The teacher materials included (a) an instructional manual describing procedures and teacher behaviors for each lesson, including suggested scripts and answer keys; (b) a set of oversized magnetic base-ten blocks to be used for teacher demonstration; and (c) a reproduction of the place-value mat projected on the white board and Smart Board™ to organize magnetic blocks or draw representations.

Each student used the following materials: (a) learning contract signifying a commitment to learn a new way of multiplying; (b) a set of base-ten bocks made of foam; (c) a laminated multiplication mat that was used to organize base-ten blocks, learning sheets for each lesson, and a progress chart that was used to monitor daily progress during guided and independent practice. The learning sheets were divided into three sections: demonstration, guided practice, and independent practice. Within each section, problems were written with both words and mathematical symbols (e.g., There are 32 classrooms and each has 24 students. How many students in all? __ groups of __, 32 x 24). The problems within learning sheets 1 through 12 consisted of only multiplication problems. Beginning with lesson 13, instruction included discrimination between multiplication problems and addition or subtraction problems. Learning sheets for lessons 14-16 consisted of words without written problems requiring that students determine the appropriate operation as well as computation. Maintenance lessons began with lesson 17, which did not have demonstration, just guided and independent practice. Finally, lessons 18-20 involved only independent practice. In addition to learning sheets, student materials included progress charts, which were used to record accuracy of computation for each lesson.

Assessment materials. Assessment materials were two timed assessments developed by the first author. Probes consisted of sheets of paper with 20 problems requiring multiplication with regrouping. Both probes included similar problems, but the problems were not presented in the same order to ensure students would not memorize answer patterns from pretest to posttest.

Reliability of the researcher-developed probes was assessed prior to the study; results from the internal consistency test revealed Cronbach's alpha coefficient of r = .83 for all probe items. Timed assessments were used since previous CRA research used fluency as a dependent measure.

Assessment Procedures

Students completed timed probes before and after the DI or CRA-SIM interventions, as follows. The teachers gave each student a probe and told them that they would be taking a timed test and were to begin when a timer started. Students were also told to answer as many problems as possible until they were told to stop and the timer sounded. After these instructions, the teachers told the students to begin, and started a timer set for two minutes. When the timer sounded, the teachers told students to put their pencils down and collected the completed probes. The probes were analyzed by counting the number of correct digits written below the answer line, meaning the numbers used in calculating the problem as well as the final answer.

Instructional Procedures

Difference between DI and CRA-SIM procedures. Both DI and CRA-SIM are explicit instructional methods, but differ in several ways. DI requires more frequent choral responding than CRA-SIM, which affords students more opportunities to practice correct responses when solving operations with multiplication with regrouping. DI also offers more opportunities for students to practice with repetition solving problems using the procedural components of the standard algorithm. CRA-SIM, on the other hand, places heavy emphasis on conceptual understanding by making models and drawings of quantities, and demonstrating the operation as students solve problems. Once students have solving problems in a manner that requires multiple representations of the numbers and operations (e.g., base-ten blocks, drawings, and numbers), they are allowed to solve problems using numbers only, with a focus on the procedures of the standard algorithm.

DI. The procedures for the DI group were followed, as prescribed by the program (Engelmann & Carnine, 2005). Each instructional period lasted 50 minutes; the size of the instructional group influenced the lesson length since individual turns and responses required more time. The students sat at tables or desks around the teacher. The teachers used the program script to present instruction. Instruction began with lesson 28, the entry point into the program for students who have mastered basic multiplication and need instruction in regrouping. Seventeen lessons that involved multiplication with regrouping (lessons 28-44) were implemented. Lessons included multiple skills that involved review of multiplication facts, reading and writing numbers, multiplication with single- and multi-digit multipliers, and word problems involving addition, subtraction, and multiplication.

A typical lesson went as follows. The first part was rehearsal of basic multiplication facts. Students completed addition with regrouping. Another section provided instruction in place value and reading numbers (e.g., 2,006 read as "two thousand six"). Here students completed word problems that involved addition, subtraction, and multiplication.

The section of the lesson devoted to multiplication with regrouping procedures provided students with tasks that were progressively more complex. Multiplication problems were written using grids that separated numbers by place value, and small boxes were written above the problem to be used for regrouping. The students were presented with a partially completed multiplication problem that had two-digit multipliers (bottom number).

Instruction began with the teacher's verbal description of how to multiply the numbers in the multiplicand (top number) by the number in the ones place of the multiplier (bottom number). Next, the program presented students with a problem written with grids, regrouping boxes, and completed computation using the number in the ones place of the multiplier. The students computed the rest of the problem using the number in the tens place of the multiplier and added the results to arrive at the final answer. Next, the program presented students with problems written with grids and regrouping boxes, but students computed the whole problem. Finally, the grids were removed, and students computed problems that included regrouping boxes. A visual representation of problem presentation for each program is presented in Figure 1.

Instruction involved activities requiring both oral and written responses using the student workbooks. Throughout each lesson, the program includes a point system whereby students earn points for completing lesson components correctly. The teachers implemented this reinforcement system across all lessons. Students were asked to respond to questions or engage in tasks as a group and individually. The teachers used signals to prompt group responses. The teachers followed the prescribed error-correction procedures for the following errors: responses prior to the teacher's signal, late responses, failure to complete tasks, and incorrect responses. Correction procedures included modeling the correct response,

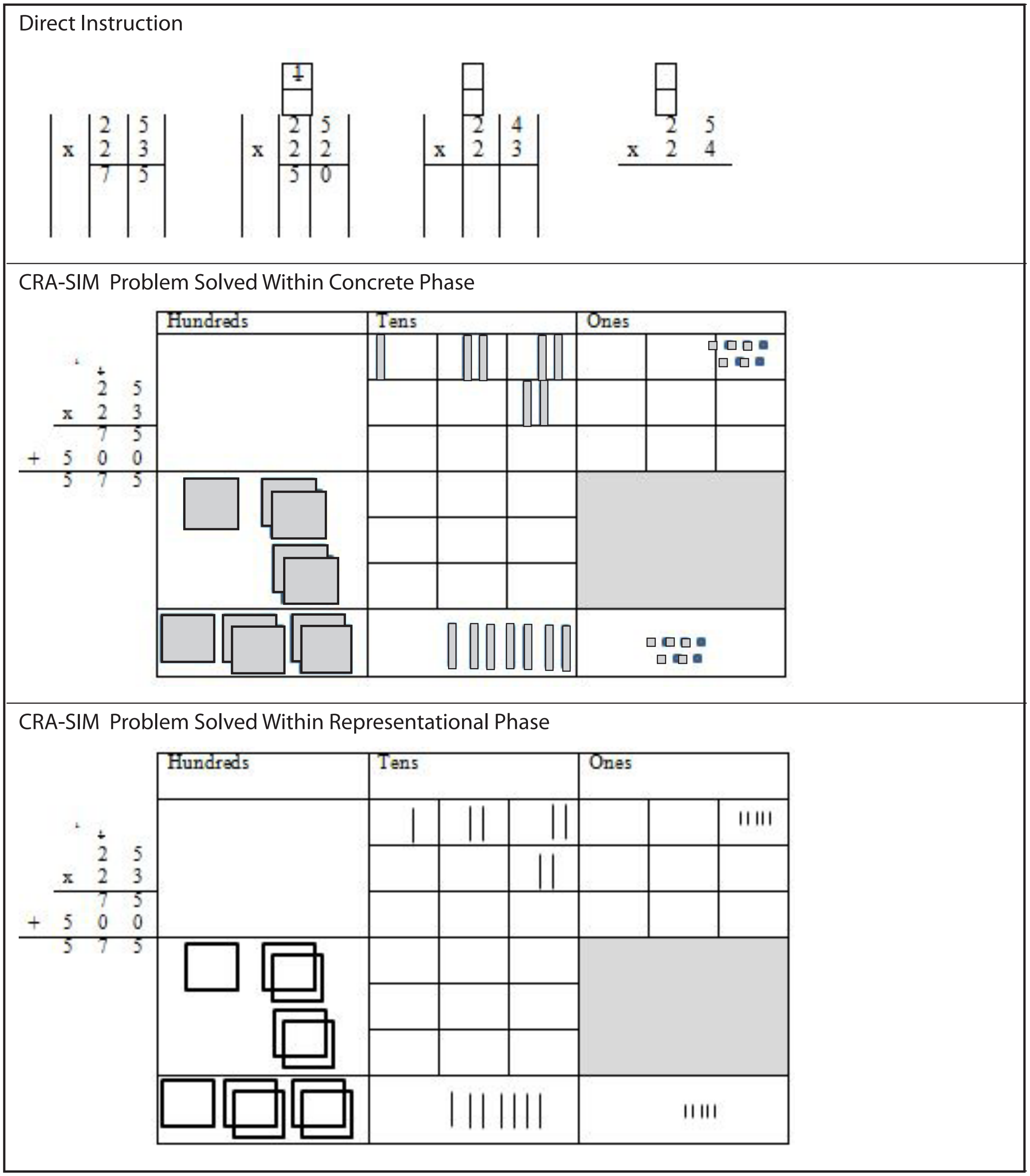


Figure 1. Visual presentation of instructional aids within each program.

leading the students in the correct response, and asking students to respond independently.

CRA-SIM instruction. For CRA-SIM, the teachers implemented instruction according to the intervention manual. The manual includes suggested scripts for each lesson intended to provide guidance but not to be read word for word. Each lesson consists of five parts: the advance organizer, demonstration, guided practice, independent practice, and post-organizer. During the advance organizer, the teachers provided the students with an overview of the lesson activities and stated expectations for student behavior. During the demonstration, the teachers showed the students how to solve problems through physical as well as mental processes by thinking aloud. Students were asked to participate by responding to questions related to prior knowledge or repeating words or phrases used by the teacher.

Guided practice involved problem solving by both the teacher and the students. The teachers completed problems with the students' assistance. That is, the students told the teachers how to complete each step of the problem and, when necessary, the teachers provided verbal prompts. Independent practice involved student completion of problems without the teachers' assistance. During this stage, the teachers monitored students' work and provided feedback when errors occurred, but did not complete portions of problems or prompt students during their work. The lesson ended with a post-organizer, in which the teachers briefly reviewed the lesson activities and provided group feedback.

The teachers measured students' progress using the number of correct digits written within the last problem completed during guided practice as well as the independent practice problems. There was no timing component to these portions of the lesson, but number of digits correct was used rather than the number of problems correct because it is a more sensitive measure of improvement. Lesson mastery was defined as 80% of digits correct. The students tracked their own progress after each lesson by marking their percent of digits correct on a chart.

The intervention was divided across 20 lessons, 17 of which were completed within this study. Prior to the first lesson, students took a pretest, discussed their performance with the teacher, and made a commitment to learning a new way of solving multiplication problems. The first five lessons involved solving multiplication problems with regrouping using base-ten blocks and a multiplication mat; instruction began with one-digit multipliers in lesson 1, and became more complicated with each lesson until problems included two-digit multipliers in lesson 5. After the advance organizer, the teachers demonstrated problems using large base-ten block magnets and the multiplication mat on a white board, large enough for students to see. The teachers read the written problem, emphasizing that there were groups of items/people and each containing the same amount and that this could be solved through multiplication. The teachers set out the multiplicand (top number) on the mat using base-ten blocks and used the digit in the ones place of the multiplier (bottom number) to make groups. They told students about a rule, "if there are ten or more, go next door." If the groups of ones combined to make 10 or more, the ones blocks were exchanged for tens blocks and placed in the tens' place of the mat. The teachers wrote a small number (crutch number) over the digit in the tens' place of the written problem to indicate regrouping. The teachers marked the remaining ones on the written problem and moved to the digit in the tens' place of the multiplicand.

The teachers made groups of tens using the digit in the ones' place of the multiplier. When combining the groups, they added the crutch number previously added to the tens' place during regrouping. Upon adding the number, the teachers crossed out the crutch number, telling the students that they did not want that number to confuse them later. The teachers applied the "ten or more" rule by exchanging 10 tens blocks for hundreds, if necessary. They marked the written problem. The teachers told the students that they were finished multiplying by the bottom number in the ones' place. They crossed out the digit in the ones' place of the multiplier and wrote a zero in the ones' place underneath the first row of answers, emphasizing that they were multiplying by tens. Then they multiplied the digit in the ones' place of the multiplicand by the digit in the tens' place of the multiplier. They made groups of tens and put them on the mat. They applied the "ten or more" rule if necessary and marked the problem. Then the teachers told the students that they were working the hundreds' place. The teachers showed the students the problem (e.g., 30x20, 30 groups of 20) and demonstrated that this resulted in six hundred using the blocks (30 groups of 20 made 60 tens blocks, which were combined into 6 hundreds blocks). The teacher told the students that 30 groups of 20 is the same as 3 groups of 2 hundreds.

After the teachers multiplied the multiplicand by both numbers in the multiplier using blocks and completed the written problem, they combined the blocks within each row of answers, beginning with the ones column. After combining blocks in each column, the teachers marked the written problem. Finally, the teachers checked the students' work by comparing the blocks with the written answers. The teachers demonstrated three problems. Guided practice involved completion of three problems through cooperation with students and assistance with verbal prompts. Finally, the students completed three problems independently.

Lessons 6 through 10 involved the use of drawings to represent the problem on the multiplication mat rather than base-ten blocks. The procedures described above were followed again. Ones were drawn using horizontal lines (groups) with short vertical tallies (items in each group). Tens were drawn using long vertical lines. Hundreds were drawn using squares. Visual representations of the instructional aids used in each program were presented in Figure 1; CRA-SIM representations show a completed problem rather than the entire process of manipulating objects and drawings.

During lesson 11, the students learned the strategy for solving regrouping problems. The RENAME strategy was taught and practiced until the students could recite the steps upon seeing the mnemonic device. This lesson served as a bridge between problem solving using representations and problem solving using abstract symbols or numbers only. With lesson 12, students solved problems using the RENAME strategy without concrete or representational tools.

Lessons 13 through 16 involved solving problems at the abstract level as well as discriminating between multiplication and other operations. It also included learning an additional strategy: Find what you are solving for, Ask yourself, "What are the parts of the problem?," Set up the numbers, Tie down the sign (FAST). The students practiced FAST RENAME until they could recite each the steps of FAST RENAME upon presentation of the mnemonic device. Lessons 14 through 16 involved solving problems using numbers only. These lessons also included discrimination between word problems (addition, subtraction, or multiplication), identifying the appropriate operation as well as computation of the problem.

When demonstrating problem solving, the teacher emphasized analysis of the words within problems and thinking about the whole problem. For example, if groups of items were being joined together, were there multiple groups with the same amount in each (multiplication) or were groups of different sizes combined (addition)? The inclusion of different types of problems was similar to the DI lessons and ensured that students understood real-world application of multiplication rather than simply assuming that word problems within a multiplication lesson would require multiplication.

Lessons 17 through 20 were maintenance lessons. Lesson 17 involved guided practice with two problems, and the remainder of the lesson was independent practice. Lessons 18 through 20 involved only independent practice; however, the study ended with Lesson 17.

Treatment Integrity

Treatment integrity was measured using checklists completed after the researchers watched both live and videos of instruction. Twenty-five percent of the lessons were observed, with one observation and simultaneous video recording each week so that observations were spread throughout the intervention. Using a checklist based on checklists used in previous DI and CRA-SIM research (Flores , Hinton, & Stonier, 2014a; Marchand-Martell et al., 1995), the observer indicated whether 15 teacher behaviors that were required throughout the lesson were present or not. Teacher behaviors included (a) organization; (b) accurate presentation of lesson using scripts (DI) or suggested scripts (CRA-SIM); (c) accurate use of program procedures; (d) accurate correction and feedback procedures; (e) accurate use of program materials; (f) appropriate pacing and affect such as smooth phrasing, eye contact with students, and engaging tone of voice; and (g) maintenance of progress monitoring (CRA-SIM) or point system (DI). One of the authors observed a lesson live, another observed the recordings and calculated interobserver agreement. Interobserver agreement for treatment integrity scoring was 98%.

The treatment integrity for the DI group was 96%, with scores ranging from 80% to 100%. The behaviors lacking were related to organization, correction procedures, and management of choral responses. The treatment integrity for the CRA-SIM group was 93%. The behaviors lacking here were related to organization, pacing, procedures, and teacher affect.

The researchers met with the teachers after poor ratings (i.e., poor ratings were defined as scores below 80% for not having materials ready, not modeling or solving problems with students; and displaying disinterest to students' reactions and cues) to provide feedback and practice; to ensure remediation, the researchers observed the subsequent lesson (this follow-up observation was not included in treatment integrity).

Interscorer Reliability and Social Validity

The teachers and a researcher scored the assessment probes, calculating the number of correct digits written. A researcher compared scores and calculated interscorer reliability by dividing the number of agreements and the sum of agreements and disagreements. Interscorer reliability was 97%.

Social validity was measured using both teacher and student surveys. After the study, teachers and students completed a survey regarding their satisfaction with and the effectiveness of the multiplication interventions. Both surveys included questions about liking instruction, increasing their multiplication skills, and whether they would recommend the instruction type for other students. Specific questions were related to each type of instruction, such as the type of instructional materials (multiple skills in one lesson for DI or use of base-ten blocks and drawings for CRA-SIM), procedures such as frequent choral responding, and use of motivational systems such as earning points and using a progress chart.

Students. Survey questions were read aloud to the students, who completed them as a group without providing their names. The DI student survey results indicated that 93% of students increased their multiplication skills and liked the materials. Most (71%) students liked the fast pace of instruction, the point system, as well as the lesson format, which involved different types of activities in each lesson. The students' responses were more divided regarding choral responses; 57% reported that this was difficult.

The CRA-SIM student survey results indicated that all of the students in this group perceived that their multiplication skills increased. Most (93%) students reported that they liked the lesson format and would recommend it for other students. More than half (60%) of the students did not like using blocks and reported that manipulation of blocks with the place-value mat was difficult. However, 100% of the students liked solving problems using drawings on the place-value mat, and 93% reported that they would continue to use the RENAME strategy to solve problems. Most (93%) students liked the progress chart.

Teachers. The teachers all reported that the students were in need of a multiplication with regrouping intervention and that the programs were beneficial. The DI teachers agreed that the lesson design with multiple skills was effective and also liked the scripted lessons, but they were split in their responses about other components of the program. Two teachers liked the pacing, frequent questioning, and choral response, and felt DI was easy to implement. The other teacher did not like these features of the program, and reported that she would not use it again.

The three CRA teachers' responses were similar to each other. That is, the teachers reported that the three phases of instruction were beneficial for students, the program was easy to implement, and they would use it again. Both CRA-SIM teachers reported that their least favorite part of the program was management of materials at the concrete level.

Results

Prior to the intervention, the researchers established that the DI and CRA-SIM groups were equivalent with regard to pretest performance and mathematics achievement, defined as standard scores on the Operations subtest of the *KeyMath-3* (Connolly, 2007). A one-way analysis of variance (ANOVA) was conducted, with the between-subjects factor being group (DI or CRA-SIM) and the within-subjects factors being the number of correct digits written on the pretest and the operations standard score on the *KeyMath-3*.

Results of the ANOVA indicated that there were no significant differences between groups with regard to pretest performance, $F(1, 27) = 0.10$, $p = 0.75$. Similarly, there were no differences between groups with regard to computation achievement, $F(1, 27) = 0.05$, $p = 0.82$. Therefore, both groups were equivalent. The means and standard deviations are presented in Table 2.

The researchers analyzed data for all students present for the pre- and posttest using a repeated-measures ANOVA. A two-way ANOVA was conducted with the between-subjects factor being group (DI or CRA-SIM) and the within-subjects factor or dependent variable being number of digits correct as written on timed probes before and after the intervention. The means and standard deviations for number of digits correct are presented in Table 2.

The results of the ANOVA indicated a significant change in student performance across groups, Wilks' $\Lambda = 0.495$, $F(1, 27) = 27.54$, $p < 0.000$, multivariate $\eta^2 = 0.51$. The results also revealed a significant effect for group, meaning that there was a difference between groups, Wilks' $\Lambda = 0.78$, $F(1, 27) = 7.61$, $p < 0.01$, multivariate $\eta^2 = 0.22$.

Table 2
Means and Standard Deviations for the Two Groups

Analysis of Group Equivalence		Mean	Standard Deviation
Pretest (correct digits written on 2-minute probe)			
	DI	15.29	15.36
	CRA-SIM	17	13.83
Standard Score on Operations Subtest *KeyMath-3*			
	DI	82.14	6.92
	CRA-SIM	81.47	9.1

Number of Digits Written Correctly on Timed Test		Mean	Standard Deviation	Effect Size
DI Group				
	Pretest	15.29	3.9	
	Posttest	22.0	4.52	
				$\eta^2 = 0.22$
CRA-SIM Group				
	Pretest	17	3.77	
	Posttest	38.6	4.52	

Discussion

The purpose of the study was to compare CRA-SIM to another explicit evidence-validated practice, DI, in the form of *Corrective Mathematics Multiplication* (Engelmann & Carnine, 2005). Both instructional programs were implemented with fidelity by teachers in natural settings within a summer intervention program. The average fluency for both groups increased, and participation in either intervention resulted in moderate gains ($\eta^2 = 0.51$). While there was a larger increase for students within the CRA-SIM group, the effect size was minimal ($\eta^2 = 0.22$). For 12 of the students in the CRA-SIM group, their fluency score increased, and of those, 9 students demonstrated increases of more than 20 correct digits (range of 21-58). In the DI group, the fluency of 9 students increased, and of those, 2 demonstrated increases of 20 and 23 correct digits, respectively.

Three students in the CRA-SIM group and six students in the DI group did not demonstrate progress. The researchers informally compared the characteristics of these students to students whose fluency increased in an attempt to find a reason for this finding. The presence of disability did not differ since students with the same disabilities within their instructional groups showed increases from 15 to 39 digits. In addition, *KeyMath* scores for these students were similar; the lowest standard score within the group of non-responders was 69, but some students with *KeyMath* scores of 72 and 73 increased their fluency by 12 to 44 digits. Students who did not demonstrate progress were in grades four and five across both groups, but their same-age peers showed increases. Finally, three non-responding students from each group had five or more absences whereas students who demonstrated increases in fluency had one or fewer absences. As a result, exposure to instruction through attendance likely influenced student progress.

In addition, results indicated that (a) more than half (60%) of the students did not like using blocks and reported that manipulation of blocks with the place-value mat was difficult; (b) 100% of the students liked solving problems using drawings on the place-value mat; and (c) 93% reported that they would continue to use the RENAME strategy to solve problems. Base ten-blocks are used at the beginning of the intervention to represent quantities and operations in a concrete manner as students solve problems and write the numbers that correspond to the concrete representations. Perhaps it was difficult and cumbersome for students to manipulate and trade out blocks to regroup at the beginning of the intervention. Using base-ten blocks may also be viewed negatively by students, as a sign that they need help to solve the operation. When drawing representations or using RENAME to solve problems, students do not trade out quantities to regroup, which can make it obvious to others what they may know or not know. Instead, they draw out how the quantities were regrouped on their personal mat, or solve problems using RENAME and numbers only on their paper.

The students within the CRA-SIM group demonstrated learning, consistent with previous CRA-SIM research related to multiplication with regrouping (Flores & Franklin, 2014 ; Flores, Hinton, & Strozier, 2014; Flores, Schweck, & Hinton, 2014). The students' performance on probes was also consistent with previous research. Students' fluency did not show increasing trends until the end of the abstract phase. In the current study, the abstract phase was presented, but there was not time for further practice.

The results of this study are significant because CRA-SIM was compared to a research-validated explicit practice, whereas previous studies in this line of research have only documented student progress and response to CRA-SIM. Findings showed that CRA-SIM was as effective as DI and resulted in slight gains in fluency, or students' ability to solve problems accurately and quickly. This finding is significant because CRA-SIM was compared to a research-validated intervention program with an emphasis on explicit, teacher-directed instruction in very specific task-oriented skills systematically leading to fluent procedural knowledge. The DI program also included frequent choral responding, so students had more opportunities to practice correct responses. This is a different approach than CRA-SIM, which does not include these features; however, students in the CRA-SIM group made at least as much or more progress in fluency. The CRA-SIM intervention emphasizes conceptual understanding and mathematical thinking as students make models and representations while computing problems, and this approach resulted in equivalent gains in procedural fluency.

Both programs were explicit; the DI program was likely more explicit and provided more practice and repetition with regard to the procedural components of the algorithm using numbers only. However, the DI program lacked conceptual instruction using base-ten blocks, place-value mats, drawings, and a mnemonic for remembering the algorithmic procedures – components of CRA-SIM that may have increased students' conceptual understanding of the algorithmic procedures, influencing overall progress and progress toward mastery of the procedures.

Throughout instruction, video evidence from fidelity measures showed students in the CRA-SIM group answering questions and verbally describing the conceptual underpinnings of computation procedures. In addition, researchers engaged in informal interactions with students in CRA-SIM and DI groups regarding their approach to computation. The researchers asked students to talk aloud while solving a problem using numbers only. Students in the CRA-SIM group consistently referred to numbers based on their value (e.g., 20 x 40 with the answer being 8 hundred) whereas students in the DI group referred to these as the written numerals in the algorithm (e.g., 2 x 4). Although students in both groups wrote the number eight correctly in the hundreds place of the problem, the CRA-SIM group appeared to be more aware of the algorithm's meaning.

Limitations

The current study is limited in terms of the length of implementation, approximately four and one half weeks of instruction with 17 lessons. Both programs include additional lessons to which students were not exposed. Additional time may have allowed for greater gains in students' multiplication performance. Thus, future research should be implemented without the time limits imposed by the summer programs within this study. Implementation during a traditional school year would allow for a natural implementation over the course of a grading period.

The study is also limited in terms of the number and representativeness of the subjects. As a result, future research should include larger groups of students receiving both interventions as well as students from different regions of the country.

Conclusions and Implications

The current research provides additional evidence that CRA-SIM is an effective intervention with regard to multiplication with regrouping. This is the first study of multiplication with regrouping that involved groups of this size and comparison with an evidence-validated explicit practice. It is difficult to conclude that previous CRA-SIM multiplication-with-regrouping results would generalize due to the use of single-case methods and the small number of studies conducted. The current study strengthens this line of research and has greater implications for generalization since a larger number of children demonstrated progress, as much progress as an evidence-validated practice. CRA-SIM was implemented within a class period, four days per week – that is, within instructional limits similar to those within typical elementary and middle schools, especially when specific intervention periods are designated. In addition, the CRA-SIM and DI interventions were easily implemented based on teachers' reports.

This study provides data for informed decision making with regard to curricular choices for intervention. Both interventions were effective, although differing in the way conceptual knowledge is included. That is, the DI program addresses conceptual understanding through verbal description whereas the CRA-SIM program addresses conceptual understanding through student manipulation of base-ten blocks and drawings. CRA-SIM may be preferable for students who need extra scaffolding in their abstract thinking, provided through the concrete and representational phases of instruction. CRA-SIM also allows for student demonstration of mathematical practices (CCSSI, 2010) that call for modeling and explanation of problem solving. DI does not preclude this, but the physical process of modeling is not included. Overall, then, for the purposes of using interventions consistent with standards for mathematical practices, CRA-SIM may be preferable, but additional research is necessary to draw those conclusions.

References

Adams, G., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR*. Seattle, WA: Educational Achievement Systems.

Barlow, A. T., & Harmon, S. (2012). CCSSM: Teaching in grades 3 and 4. *Teaching Children Mathematics*, *18*, 498-507. doi:10.5951/teacchilmath.18.8.0498

Common Core State Standards Initiative (CCSSI). (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Connolly, A. J. (2007). *Key Math Diagnostic Assessment*. Minneapolis, MN: Pearson.

Dacey, L., & Drew, P. (2012). Common core state standards for mathematics: The big picture. *Teaching Children Mathematics, 18,* 378-383. doi:10.5951.18.6.0378

Darch, C., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research, 77*, 351-359. https://doi.org/10.1080/00220671.1984.10885555

Deshler, D. D., & Schumaker, J. B. (1986). Learning strategies: An instructional alternative for low achieving adolescents. *Exceptional Children, 52,* 583-590. https://doi.org/10.1177/001440298605200611

Engelmann, S., & Carnine, D. (2005). *Corrective mathematics: Multiplication.* Columbus, OH: SRA/McGraw-Hill.

Flores, M. M. (2009). Teaching subtraction with regrouping to students experiencing difficulty in mathematics. *Preventing School Failure, 53*(3), 145-152. https://doi.org/10.3200/PSFL.53.3.145-152

Flores, M. M. (2010). Using the concrete-representational-abstract sequence to teach subtraction with regrouping to students at risk for failure. *Remedial and Special Education, 31*(3), 195-207. doi:10.1177/0741932508327467

Flores, M. M., & Franklin, T. M. (2014). Teaching multiplication with regrouping using the concrete-representational-abstract sequence and the Strategic Instruction Model. *Journal of the American Academy of Special Education Professionals*, *6*, 133-148. https://doi.org/10.1111/ldrp.12032

Flores, M. M., Hinton, V. M., & Strozier, S. D. (2014). Teaching subtraction and multiplication with regrouping using the concrete-representational-abstract sequence and strategic instruction model. *Learning Disabilities Research & Practice, 29*(2), 75-88. https://doi.org/10.1111/ldrp.12032

Flores, M., M., & Kaylor, M. (2007). The effects of a direct instruction program on the fraction performance of middle school students at-risk for failure in mathematics. *Journal of Instructional Psychology, 34*, 84-94.

Flores, M. M., & Kaffar, B. J. (2018). *Multiplication with regrouping: Standard algorithm*. Lawrence, KS: University of Kansas Center for Research on Learning.

Flores, M., M., Schweck, K. B., & Hinton, V. M. (2014b). Teaching multiplication with regrouping to students with specific learning disabilities. *Learning Disabilities Research & Practice, 29*(4), 171-183. https://doi.org/10.1111/ldrp.12043

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Harris, C. A., Miller, S. P., & Mercer, C. D. (1995). Teaching initial multiplication skills to students with disabilities in general education classrooms. *Learning Disabilities Research & Practice*, *10*(3), 180-195.

Kiuhara, S. A., & Witzel, B. S. (2014). Focus on inclusive education: Math literacy strategies for students with learning difficulties. *Childhood Education, 90*, 234-238. doi:10.1080/00094056.2014.912067

Kelly, B., Gersten, R., & Carnine, D. (1990). Student error patterns as a function of curriculum design: Teaching fractions to remedial high school students and high school students with learning disabilities. *Journal of Learning Disabilities, 23*, 23-29. https://doi.org/10.1177/002221949002300108

Kroesbergen, E. H., van't Noordende, J. E., & Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology, 20,* 23-37. doi:10.1080/09297049

Maccini, P., & Hughes, C. A. (2000). Effects of a problem-solving strategy on the introductory algebra performance of secondary students with learning disabilities. *Learning Disabilities Research & Practice*, *15,* 10-21. https://doi.org/10.1207/SLDRP1501_2

Maccini, P., & Ruhl, K. L. (2000). Effects of graduated instructional sequence on the algebraic subtraction of integers by secondary students with learning disabilities. *Education and Treatment of Children*, *23,* 465-489.

Mancl, D. B., Miller, S. P., & Kennedy, M. (2012). Using the concrete-representational-abstract sequence with integrated strategy instruction to teach subtraction with regrouping to students with learning disabilities. *Learning Disabilities Research and Practice, 27*(4), 152-166. doi:10.1111/j.1540-5826.2012.00363.x

Marchand-Martella, N., Lingnugaris-Kraft, B., Pettigrew, T., & Leishman, R. (1995). *Direct instruction supervision system.* Logan UT: Utah State University.

Mercer, C. D., & Miller, S. P. (1992). Teaching students with learning problems in math to acquire, understand, and apply basic math facts. *Remedial and Special Education, 13*(3), 19-35. doi:10.1177/074193259201300303

Miller, S. P., & Kaffar, B. J. (2011). Developing addition with regrouping competence among second grade students with mathematics difficulties. *Investigations in Mathematics Learning, 4*(1), 24-49. https://doi.org/10.1080/24727466.2011.11790308

Miller, S. P., & Mercer, C. D. (1993). Using data to learn about concrete-semiconcrete-abstract instruction for students with math disabilities. *Learning Disabilities Research & Practice, 8*(2), 89-96.

Morin, V. A., & Miller, S. P. (1998). Teaching multiplication to middle school students with mental retardation. *Treatment of Children*, *21*(1), 22-36.

Przychodzin, A., Marchand-Martella, N., Martella, R., & Azim, D. (2004). Direct Instruction mathematics programs: An overview and research summary. *Journal of Direct Instruction, 4*(1), 53-84. Retrieved from http://www.adihome.org/articles/JDI_04_01_05.pdf

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington, DC: U.S. Department of Education.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

Scarlato, M. C., & Burr, W. A. (2002). Teaching fractions to middle-school students. *Journal of Direct Instruction, 2(1)*, 23-38. Retrieved from https://www.nifdi.org/research/journal-of-di/volume-2-no-1-winter-2002/434-teaching-fractions-to-middle-school-students/file

Vaughn, S., & Fuchs, L. (2003). Redefining learning disabilities as inadequate response to instruction. The promise and potential problems. *Learning Disabilities Research and Practice , 18,* 137-146. https://doi.org/10.1111/1540-5826.00070

Watkins, C. L., & Slocum, T. A. (2004). The components of direct instruction. In N. E. Marchand-Martella, T. A. Slocum, & R. C. Martella (Eds.), *Introduction to Direct Instruction* (pp. 29-65). Boston, MA: Pearson.

Wilson C. L., & Sindelar, P. T. (1991). Direct instruction in math word problems: Students with learning disabilities. *Exceptional Children, 57*, 512-519. https://doi.org/10.1177/001440299105700605

Woodward, J., & Brown, C. (2006). Meeting the curricular needs of academically low-achieving students in middle grade mathematics. *The Journal of Special Education, 40*(3), 151-159.

Witzel, B. S., Ferguson, C. J., & Mink, D.V. (2012). Number sense: Strategies for helping preschool through grade 3 children develop math skills. *Young Children, 67*(3), 89-94. Retrieved from http://www.naeyc.org/yc/pastissues/2012/may

Witzel, B. S., Mercer, C. D., & Miller, M. D. (2003). Teaching algebra to students with learning difficulties in an inclusive setting. *Learning Disabilities: A Contemporary Journal*, *3*(2), 49-60. doi:10.1111/1540-5826.00068

The Effects of a Peer-Tutoring Intervention on the Text Productivity and Completeness of Narratives Written by Eighth Graders With Learning Disabilities

Matthias Grünke,[1] *Bruce Saddler*
Kristie Asaro-Saddler, and
Mariola Moeyaert[2]

Abstract

Writing becomes increasingly more imperative across all content areas as students progress through secondary school. However, many of them struggle with the complex process of putting thoughts and ideas onto paper or into a keyboard. Adolescents with learning disabilities (LDs), in particular, are usually challenged by writing activities. One major cause for their difficulties is their lack of planning skills. This single-case study evaluated a peer-tutoring approach designed to teach adolescents with LDs to better plan narratives by using a simple strategy (story mapping). A multiple-baseline design (AB) across subjects was employed to assess the effects of the intervention. The results indicate large and significant effects of the intervention on the length and completeness of the narratives the students produced. This suggests that writing skills of adolescents with LDs can be improved by way of peer tutoring with limited resources required.

Keywords: writing difficulties, learning disabilities, peer tutoring, writing planning skills

Significance of Writing and Causes for Problems in Text Production

Written language is a critical element in the lives of millions of students enrolled in K–12 schools (National Center for Education Statistics [NCES], 2012). Proficiency in text production is an important skill for elementary-aged children to learn, but writing effectively becomes even more essential for youth in middle and high school as they prepare to make the critical transition to higher education and/or the workforce. In school, writing is often used as an instrument to assess students' understanding of content (Graham, 2008). Text production is also an important means of sharing and communicating in everyday life through social media outlets, emailing, and texting (Graham & Hall, 2016), with an estimated 171 billion e-mail messages sent daily (NCES, 2012). In the workplace, writing is essential to acquiring employment and advancing in a profession, with more workers than ever before required to create reports, PowerPoint presentations, e-mails, or other types of written documentation (National Commission on Writing, 2004).

However, despite the centrality of written expression in learning, social, and work environments, a great percentage of school-aged children struggle with this skill. In fact, according to the most recent National Assessment of Educational Progress (NCES, 2012), of the three possible skill levels (basic, proficient, and advanced), only about a quarter (27%) of grade 8 and 12 students performed at or above a "proficient" level, with only 5% of students with disabilities performing at or above that standard. The

[1]University of Cologne, Germany; [2]University at Albany, USA

Grünke, M., Saddler, B., Asaro-Saddler, K., & Moeyaert, M. (2019). The Effects of a Peer-Tutoring Intervention on the Text Productivity and Completeness of Narratives Written by Eighth Graders With Learning Disabilities. *International Journal for Research in Learning Disabilities,* 4(1), 41-58. https://doi.org/10.28987/ijrld.4.1.41

present study was conducted in Germany. Unfortunately, no comparable surveys have been undertaken in this country on the expressive writing competencies of elementary and/or secondary students.

Perhaps one reason for the difficulties that many learners experience in writing is that they do not spend sufficient time developing writing skills and engaging in writing tasks. For example, secondary school students spend little time writing or being taught how to write (Graham, Capizzi, Harris, Herbert, & Morphy, 2014), despite a large majority of teachers believing that writing is very important beyond high school (Kiuhara, Graham, & Hawken, 2009). Apparently, what time is devoted to writing is not well spent. In one study, Rietdijk, van Weijen, Janssen, van den Bergh, and Rijlaarsdam (2018) documented that communicative writing, process writing, and writing strategy instruction were insufficiently implemented in primary education classrooms. Another survey, by Applebee and Langer (2011), revealed that secondary school students invested the majority (over 80%) of their writing time in engaging in tasks such as completing fill-in-the-blank assignments, short-answer exercises, and copying information from teacher presentations, rather than composing extended writing pieces, such as stories or essays. Such low-level writing tasks often require little analysis and interpretation (Kiuhara et al., 2009) and do not encourage students to think critically, a skill required for effective writing that is highly valued by employers (Graham & Hall, 2016; National Association of Colleges and Employers, 2012; National Commission on Writing, 2004).

Students' lack of proficiency in writing may also be related to teachers' lack of knowledge regarding high-quality writing instruction. Writing is a difficult skill to learn and to teach, and quality pre- or inservice writing instruction must first be experienced by teachers before they can deliver it to students. However, educators may not be receiving the training they need (Grünke & Leonard-Zabel, 2015). For example, Graham et al. (2014), in a survey, found that nearly two-thirds (64%) of middle school teachers reported having received minimal or no preservice training on how to teach writing. A similar lack of training may also have contributed to secondary school teachers reporting only infrequently using evidence-based writing practices and adaptations for their struggling writers (Graham et al., 2014; Kiuhara et al., 2009), an outcome that is especially problematic for students with learning disabilities (LDs).

Students With Learning Disabilities

Many students struggle as they try to put their thoughts onto paper or into a keyboard. However, written communication is often especially challenging for children and youth with LDs. Writing involves the adroit utilization and coordination of multiple processes: cognitive, linguistic, affective, behavioral, and physical (Santangelo, 2014), which can each present barriers to young people with LDs who "… fail to develop the knowledge, skill, will, and self-regulation necessary to succeed in key subject areas" (Grünke & Morrison-Cavendish, 2016, p. 1). For example, these students may have a less mature and developed understanding of the importance of writing. They may lack topic and genre knowledge, and may not know how to effectively engage in the various stages of the writing process. Thus, they often exhibit little to no structured or systematic planning before they commence the writing task (Gillespie & Graham, 2014). Instead, they engage in what Scardamalia and Bereiter (1986) termed "knowledge-telling" behavior, in which a person generates content by writing down whatever information he or she can recall about a topic without regard for the purpose or goal of the assignment. When revising, students with LDs tend to focus only on surface-level changes, such as punctuation and spelling, rather than evaluating their text to make meaningful modifications (Saddler & Asaro, 2007). Additionally, they may have difficulties with handwriting, spelling, and vocabulary, which may limit the amount of writing they produce and the words they choose to use. Finally, they may exhibit low motivation to write and low self-efficacy toward their writing ability (Graham, Collins, & Rigby-Willis, 2017).

Combined, these characteristics have a profound effect on the ability to produce quality written products. In a recent meta-analysis of the writing characteristics of students with LDs, Graham et al. (2017) found that, when compared with their typically achieving classmates, students with LDs produced writing samples that were of lower overall quality and less legible, more disorganized, and with more spelling and grammar errors and less diverse vocabulary and ideas.

Of all these areas of need, effectively organizing their writing may be one of the greatest challenges for students with LDs. Organization is a critical component of planning. During the planning process, writers compile relevant information and develop a blueprint that will assist them in accomplishing the goals of the composition. Along the way, they coordinate three subprocesses: generating (retrieving relevant information), organizing (structuring the information), and goal setting (developing goals and establishing a writing plan to achieve the goals). This component of the overall writing process is so important that skilled writers spend considerable time planning and organizing. Further, it is the basis for the other two major text production activities – translating and reviewing (Flower & Hayes, 1981).

Unfortunately, children and youth with LDs minimize the role of planning and organizing (Scardamalia & Bereiter, 1986). They may have ideas about what they want to write, but they struggle to put their thoughts on paper and visually organize them (Sundeen, 2014). In fact, Koutsoftas (2016) found that many students with LDs may be able to produce as many as or more ideas than students without disabilities; however, they are less efficient at structuring and sequencing their thoughts. As a result, an organizational tool, such as a story map, may be beneficial for such students.

Story Mapping

A story map is a graphic organization tool that outlines the important components of a story. As such, story maps can support writers by providing visual reminders of these important elements while offering a space in which they can brainstorm and write down their notes for the respective element (Grünke & Leonard Zabel, 2015). Story maps also allow visualization of how the elements of a story – settings, characters, and events – are connected (Li, 2007). Such a tool is beneficial for writers with LDs, who tend to have difficulty generating stories (Barenbaum, Newcomer, & Nodine, 1987), and who write relatively short texts consisting of elements that are not logically related (MacArthur & Graham, 1987).

Story mapping has been found to be an effective intervention for students with LDs in both reading comprehension (Boon, Paal, Hintz, & Cornelius-Freyre, 2015) and writing (Hennes, Büyüknarci, Rietz, & Grünke, 2015; Li, 2007). Li (2007), for example, found that three of four fourth- and fifth-graders with LDs increased their writing fluency after learning a story-mapping strategy. This finding is supported by Hennes and colleagues (2015), who noted that when they received a story-mapping intervention, students between 8 and 14 years of age who had LDs significantly increased their number of words written, and, perhaps more important, the quality of their stories.

In two recent studies, peers were used to teach elementary school children with LDs how to apply story mapping to help them plan their narratives. The results indicate that the students with LDs significantly increased the number of words they wrote (Grünke, Janning, & Sperling, 2016; Grünke, Wilbert, Tsiriotakis, & Lopez Agirregoikoa, 2017). Both studies provide insight into the potential of using peer-assisted writing strategies to support students with LDs in story writing.

Peer-Assisted Writing

Peer-assisted writing strategies involve one student systematically assisting, or teaching, another in writing. Typically, this includes pairing a higher-achieving with a lower-achieving writer (Grünke et al., 2016). Peer-assisted writing offers several benefits to learners. First, peer support aids students by providing them with prompting, modeling, and immediate access to assistance that is individualized (Yarrow & Topping, 2001). In addition, peer tutors tend to limit processing overload that is often experienced by writers with LDs, allowing them to focus on higher-level processing rather than low-level skills such as spelling (Yarrow & Topping, 2001). Students who participate in peer writing structures also tend to demonstrate increased time on task and increased levels of engagement and experimentation (Englert, Berry, & Dunsmore, 2001; Yarrow & Topping, 2001). In a meta-analysis of writing interventions, peer-assisted writing was shown to have a positive impact on written products, with an overall effect size of 0.75, indicating a moderate to large effect (Graham & Perin, 2007).

A wide variety of students can benefit from peer assistance in writing, including students as young as kindergarten (Puranik, Patchan, Lemons, & Al Otaiba, 2017) to secondary students (Rensing, Vierbuchen, Hillenbrand, & Grünke, 2016), both with and without disabilities (e.g., Saddler & Asaro, 2008; Saddler & Graham, 2005), as well as English-as-a-foreign-language (EFL) students (Kurihara, 2017).

Research Question

The purpose of the current study was to contribute to the scarce body of empirical literature on the benefits of teaching story mapping to enhance the writing performance of students with LDs through peer tutoring. By testing an older population (three secondary school students with LDs), the study extends the two previously mentioned experiments by Grünke et al. (2016, 2017), which provide evidence that a peer-tutoring story-mapping strategy can help children at risk for failure to catch up with their classmates at a rather early stage of their writing development.

Method

Setting and Participants

The inclusive secondary modern school ("Hauptschule") chosen for this study is located in a metropolitan town in Northrhine-Westfalia (Germany). It enrolls approximately 240 students in grades 5 to 9. A "Hauptschule" offers lower secondary education (primarily for students with average grades and below) according to the International Standard Classification of Education (UNESCO, 2011). Any student who went to a four-year German elementary school can attend a "Hauptschule" afterwards, whereas attendance at a grammar school ("Gymnasium") – the other option for secondary education – requires high grades.

Three tutees and three tutors were chosen from one of the eighth-grade classes by the main teacher and the first author. Five of the students in this class had been diagnosed with LDs by a multidisciplinary team at the end of their elementary education. The process of identification had involved standardized assessment of language, reading, math, and nonverbal cognitive abilities. All five students had acquired basic skills in the aforementioned areas. However, they demonstrated profound difficulty applying them in higher-level schoolwork. Despite having an average IQ level, they had severe trouble applying problem-solving steps and transferring academic skills to other tasks, due to challenges in finding solutions to complex assignments and comprehending the logic behind them. There was a discrepancy of at least 1.5 standard deviations between aptitude (intelligence) and performance in higher-order skills (mathematical problem solving and reading comprehension).

We considered as potential participants of our study students who, according to their school records, did not score below the 30th percentile in a standardized spelling test. Thus, we wanted to make sure that participants were ready to attend to expressive writing without being held back by having to think too much about the correct order of letters in a word, for example. In addition, potential students had to be deemed by their main teacher as being socially capable of independently working in pairs without needing constant attention from an adult. Twelve students in the class met these criteria.

Prior to the start of the intervention, we presented these learners with a writing prompt consisting of a sequence of pictures in the form of a cartoon strip. We asked the students to imagine what was happening in those pictures and to write a story about it. No time limits were given. We allocated the three students who produced the shortest stories to function as tutees and assigned the three students with the longest texts as tutors. Pairs were built on the basis of the teacher's judgment of who got along best with whom. Not surprisingly, none of the five students with LDs were among the allocated tutors; however, all three tutees met the criteria for LDs.

The first team consisted of Aaron (tutee) and Adrian (tutor) (the participants' names have been changed to preserve their anonymity). Aaron was a 15-year-old male. According to his main teacher, he was comparatively difficult to motivate and demonstrated extremely weak analytical skills. Even though he possessed ample spelling abilities, he was a very reluctant writer. Adrian was 14 years old, and a student with average math and language skills. He enjoyed writing, and was viewed by his teacher as extraordinarily cooperative. Both boys were born in Germany and did not have an immigrant background.

The second team was made up of Baci (tutee) and Babak (tutor). Baci was 15 years old, and the daughter of Turkish migrant workers. Even though she was born in Germany, she spoke mostly Turkish at home with her family. Her teacher described her as having very low self-efficacy in writing and as not being performance-oriented. Baci's grades were generally in the D range. Babak was 14 years old, and the son of Iranian parents who migrated to Germany when he was a toddler. According to his teacher, he was very hardworking and organized. He demonstrated excellent written and verbal communication skills, and his grades were in the top quarter of his class.

The third and last team included two girls: Channa (tutee) and Cora (tutor). A daughter of Polish parents, 14-year-old Channa had moved with her family to Germany as an infant. Even though her spelling abilities were acceptable, her vocabulary was regarded as rather limited by her teacher. Her grades were in the range of Cs and Ds. She presented a notably low level of motivation and very little enthusiasm for learning. Cora was 15 years old, did not have an immigrant background, and was deemed by her teacher to be a very skilled reader and writer. She generally received grades in the A-B range, and was well known in her class for her prosocial behavior.

Experimental Design

We applied an AB multiple-baseline design across subjects (Gast, Lloyd, & Ledford, 2018). The data were collected over 13 consecutive school days. Following the single-case reporting guidelines by Tate et al. (2016), a randomization procedure was used to increase the internal validity of the study, utilizing the randomize function in Microsoft® Excel. We stipulated that each phase in our design had to consist of at least three measurement points. Therefore, the start of the intervention for each of the three cases was chosen at random (with the restriction that there had to be at least three probes per participant in every phase). Hence, the treatment could have started any time between the 4th and the 11th probe. A random drawing of all six possible options for each participant resulted in an arrangement whereby the training for Aaron started after the sixth measurement point, and for Baci as well as for Channa after the eighth. Thus, Aaron received seven training sessions, and Baci and Channa each participated in five.

Materials

Baseline and intervention writing prompts consisted of fifty 5 x 8-inch index cards for each team that had story starters printed on one side (e.g., "One day, I discovered that I had a super power"). The prompts were based on the ideas in a book by Kinder (2014), which contains a high number of story starters of an equal difficulty level. Before printing them on the index cards, we simplified them

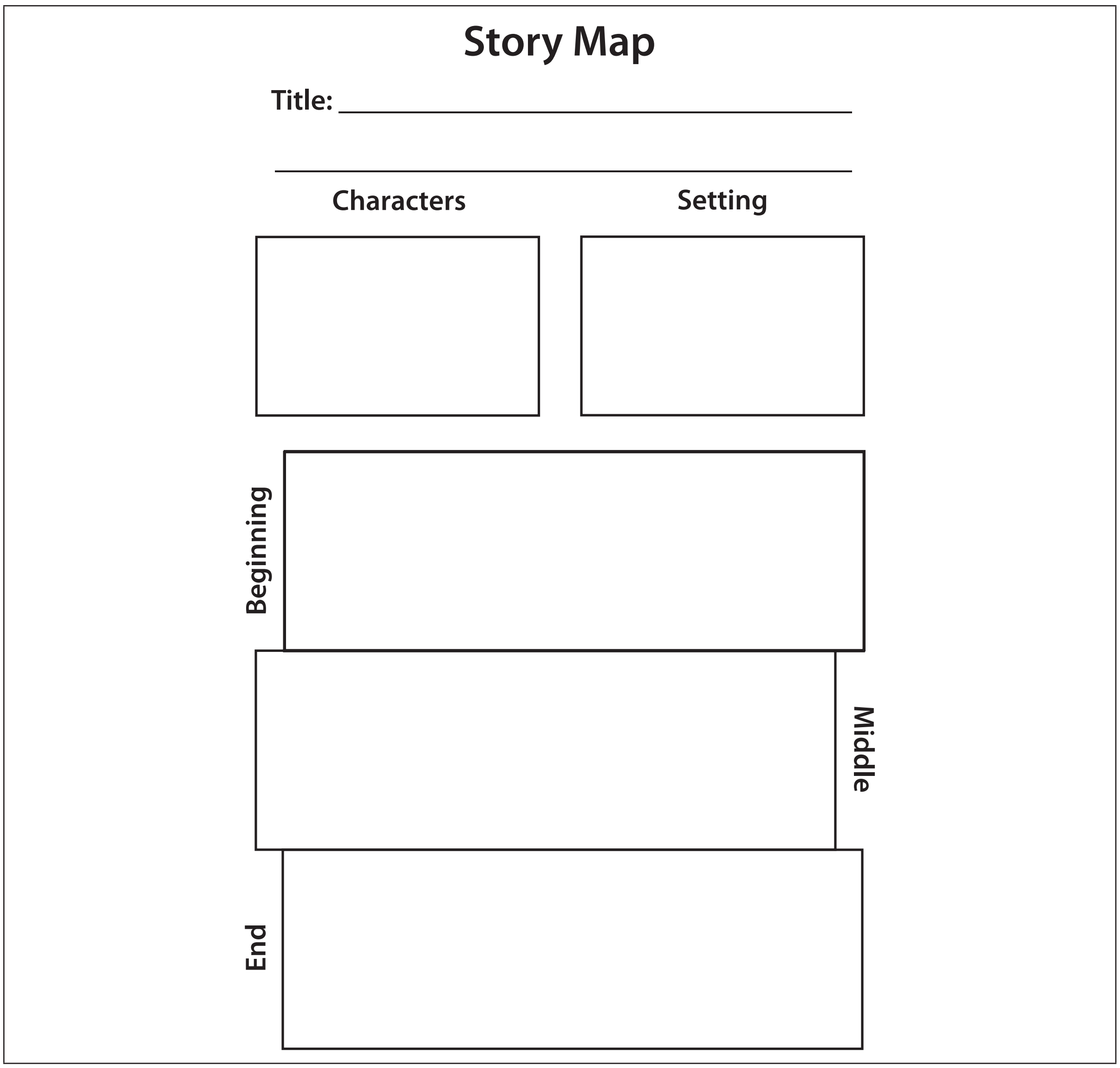


Figure 1. A story map template.

and translated them into German. At each of the 13 measuring points, the tutees were given a pen and four sheets of 8.5 x 11-inch notepaper.

For the intervention, we used a two-page list of 50 narrative writing prompts for sixth and seventh graders that were taken from the website https://k12.thoughtfullearning.com/resources/writingtopics (e.g., "If only I would have listened!," "Summer in a cabin by a lake," or "We couldn't stop laughing!"). Before presenting them to the participants, we translated the prompts into German. We employed simple story map templates as depicted in Figure 1. In addition, we provided each team with an 11.5 x 16.5-inch poster that visualized the steps of the mapping strategy: (a) contemplate the story heading, (b) think about what could happen in the story, (c) review the fields of the story map, and (d) take notes on your ideas about the story while using the appropriate fields. During the intervention, we made sure that the tutors and tutees always had enough pens and notepaper at their disposal. In order to guide the tutors through the lessons, and to provide them with reminders of what to do during the process, we prepared a simple four-page script containing brief instructions and mnemonics in large print (the scripts are available from the first author upon request).

Definition and Measurement of Dependent Variables

At each measurement point, a female research assistant asked the tutees to randomly draw two index cards with story starters from the pile. Aaron, Baci, and Channa were always given a choice to decide which of the two prompts they wanted to use as an initial point for their narratives. We never gave them a pile containing cards that they had already seen. The tutees were given a pen and notepaper. No time limits were set. The research assistant told the tutees that they should use the paper to produce their stories, but that they could also utilize some of it to outline their ideas if they felt a need to do so.

Two dependent variables were used to capture the writing performance of Aaron, Baci, and Channa: the number of total words written (TWW; Hosp, Hosp, & Howell, 2016) and a writing rubric developed by Harris and Graham (1996). The TWW is a widely used production-dependent fluency measure (Furey, Marcotte, Hintze, & Shackett, 2016) defined as the number of recognizable words written, regardless of spelling or context (excluding digits). Any letter or group of letters that has a space before and after it (even if it must be viewed as a nonsense word) is considered a word (Hosp et al., 2016). The writing rubric was included to determine the completeness of the texts. It consists of eight categories of descriptors: main character, locale, time, starter event, goal, action, ending, and reaction (see Figure 2). The rubric includes specifications for how certain criteria must be met in order to earn a specific number of points for one of the eight criteria. In total, a text could be awarded between 0 and 19 points.

Two female graduate students, who were blind to the purpose of the study, served as scorers. Both of them had received extensive training by the first author on how to use the instruments. First, they determined the total number of words the tutees had written (TWW). They randomly chose one text after another and evaluated them. Subsequently, they counted TWW independently and compared their results. Inter-rater agreement was calculated for each text by dividing the smaller number by the larger and multiplying by 100. The scores varied between 95.15 and 100.00%, with an average of 97.05%. In case of discrepancies, the raters discussed them until they reached consensus.

Second, one of the graduate students scored all the stories using the writing rubric whereas the other independently scored a random selection of 20% of the texts. Inter-rater agreement was determined using the same procedure as with TWW. Agreement ranged between 73.33 and 100.00%, with an average of 88.48%. Such a value meets the standards established by Hartmann, Barrios, and Wood (2004) for percentage agreements in single-case designs. We used the scores obtained by the first graduate student for the data analysis. The second rater was involved to gain an indication of the reliability of the scores.

Procedures

Baseline. Each day of the study, at the beginning of the third class period, the research assistant picked up the tutor-tutee teams from their class and brought them to a resource room in the school. The rest of the students in the class remained with their main teacher and engaged in independent reading or writing activities.

In the resource room, the teams were seated at tables as far apart as possible. For about 30 minutes, they participated in different educational games (such as math racetracks; see Skarr et al., 2014). (None of these activities involved story writing or similar tasks.) Then the research assistant gave Adrian, Babak, and Cora some math problems to solve, whereas Aaron,

A Scale for Scoring the Inclusion and Quality of the Parts of a Story

1. Main Character

0	No main character is established.
1	A main character is presented; however, he/she is just a name on a page. Very little information or detail about the main character is provided.
2	A main character is presented and described in such detail that he/she is always "real" for you.

2. Locale

0	No locale or place is mentioned.
1	Locale given, but little description offered.
2	Locale given, with more complete description offered, or unusual locale is chosen.

3. Time

0	No time given.
1	Time given, but traditional in reference.
2	Time given, but unusual in reference or more complete description.

4. Starter Event

0	The precipitating event that causes the main character to establish a goal is not presented.
1	The precipitating event that causes the main character to establish a goal is presented. The precipitating event can be a natural occurrence, an internal response, or an external action.
2	The precipitating event is complex, unusual, or well described.

5. Goal

0	The goal or purpose of the main character is not established.
1	The goal or purpose of the main character is established but not clearly articulated.
2	The goal or purpose of the main character is clearly articulated.
3	Two or more goals are clearly articulated.

6. Action

0	The actions that the main character initiates in order to achieve the goal are not presented.
1	What the main character does in order to achieve the main goals is presented.
0-4	Add one point for each of the following: A. Actions or events happen in a logical order (i.e., they are not inconsistent). B. Ingenuity or originality are used to solve situations or predicaments. C. There is more than one well-defined episode. For example, the main character tries one action, and if it is unsuccessful then tries another action. D. The main character goes to one place during his travels and then another.

7. Ending

0	No real ending, lack of conclusion, or story seems unfinished. In other words, the long-range consequences of the main character's actions are not resolved.
1	Long-range consequences of main character's actions are resolved, but the ending or conclusion is fairly common.
2	Long-range consequences of main character's actions are resolved. In addition, the conclusion or ending is unusual, or the ending contains a moral.

8. Reaction (expressed anywhere in the story)

0	The emotional reactions of the main character are not presented.
1	Some emotional feelings expressed by the main character.
2	Emotions feelings of the main character expressed with depth.

Figure 2. The writing rubric by Harris and Graham (1996) (the original version provides examples for the different categories).

Baci, and Channa were asked to write a story in accordance with the description in the above section on measurement of dependent variables.

Tutor preparation. Before the teams started with the story-mapping intervention, the research assistant provided the tutors with a two-hour training on the components of the instructional framework, including modeling and strategy instruction. She also familiarized them with the script and encouraged them to refer to it as a memory aid.

Intervention. In the treatment condition, the measuring procedures were identical to those of baseline. However, the teams no longer engaged in educational games, but practiced outlining acceptable narratives using the story-mapping strategy. Each lesson lasted 30 minutes. As in the baseline condition, the research assistant was always present to provide support if needed. Training sessions followed the common scaffolding sequence for direct instruction of "I do it, we do it, you do it" (Archer & Hughes, 2010). The tutees often had trouble concentrating during regular school lessons, but being in a situation where they received individual instruction helped them to stay on task. Even the fact that there were three teams in the same room, working on different tasks, hardly seemed to bother or distract anyone. However, if a team got sidetracked or needed additional support, the research assistant intervened.

Lesson 1. In the first lesson, the tutors put the poster with the strategy steps on the table. They read each step out loud and assured their partners that using this technique would help them to write better stories. The tutors then reverted to the two-page list of narrative writing prompts. After choosing one of the headings, they crossed it out and modeled the process by performing the activities listed on the poster using completed story-map templates while thinking aloud. The tutors then repeated the steps of the strategy and asked the tutees follow-up questions to make sure that they understood the purpose of the procedure.

In the remaining time, Aaron, Baci, and Channa filled out their own story maps, while Adrian, Babak, and Cora provided scaffolded feedback and encouragement. Before the tutees' performance was measured, the tutors took the posters away in order to create a situation that conformed with the baseline conditions.

Lesson 2. At the beginning of the second lesson, the tutors again placed the posters on the table and went over the steps of the strategy once more to help the tutees to internalize them. Afterwards, tutors and tutees together selected a story starter from the list, crossed it out, and took turns filling out the fields of the template while executing the four activities on the poster. In the remaining minutes, they repeated the process at least one more time. Finally, the posters were removed, and the tutees wrote a story to measure their performance.

Lesson 3 and remaining lessons. From the third session onward, support from the tutors was gradually reduced to provide the tutees with opportunities for independent practice. The posters were again placed on the table, but this time they were not explicitly mentioned; they just served as a memory aid.

Aaron, Baci, and Channa were now encouraged to choose a writing prompt from the list and fill out a template by themselves. If necessary, Adrian, Babak, and Cora gave corrective feedback, and if the process came to a halt, they offered suggestions on what to fill into the fields. At the end of the lesson, the tutees were prompted to create their own story maps without using one of the pre-prepared templates. After about half an hour, the tutees' performance was measured.

During the remaining sessions, the tutees created their own story maps and were constantly encouraged by their tutors as they devised ideas to enter into the fields of the template. Aaron, Baci, and Channa did not take notes during performance measurements under baseline conditions, but they started doing so after the second treatment session. Moreover, they quickly began to create their own story maps in order to organize their ideas prior to the writing process.

Treatment Fidelity

The research assistant used a checklist to ensure fidelity of implementation. It contained all relevant aspects of the procedure as described above. If the tutors deviated from the script, the research assistant interceded and made sure that the treatment was executed the way in which it was initially planned.

Social Validity

After the completion of the last training session, the research assistant interviewed Aaron, Baci, and Channa individually, asking their opinion about the peer-tutoring intervention. Specifically, she asked (a) if they liked the format of the treatment, (b) if they viewed it as helpful, (c) if they thought it had an impact on their writing performance, and (d) if they would recommend it to other students. The research assistant took notes on the participants' feedback.

Results

Descriptive Analysis

Tables 1 and 2 present the descriptive statistics for the TWW and the writing rubric scores. As illustrated, the two dependent variables are highly correlated across cases ($r_{\text{Spearman's Rank}} = .85$; $p < .001$; one-tailed). In examining the raw scores, it is noteworthy that Channa wrote an exceptionally long story of 359 words after the fourth treatment session. The rest of the narratives that she produced during Phase B were, on average, not even half as long. She may have had an especially creative day and, therefore, been able to demonstrate her capabilities. A comparison between the mean TWW values during baseline and the intervention revealed an increase in performance for Aaron of 112.24%; Baci, 183.17%; and Channa, 49.17%. The rubric rating scores grew by 75.70% for Aaron; Baci, 115.38%; and Channa, 121.09%. In sum, these treatment gains are remarkable.

Visual Analysis

A visual display of the data is provided in Figures 3 and 4 (the graphs were produced using the MultiSCED web application by Cools et al., 2018).

Table 1
The TWW for Each Participant

		Baseline	Intervention
Aaron	*N* (Probes)	6	7
	Raw Scores	81; 76; 70; 88; 47; 79	164; 131; 208; 137; 156; 152; 188
	M	73.50	156.00
	SD	14.27	27.44
	Range	47-88	131-208
Baci	*N* (Probes)	8	5
	Raw Scores	88; 68; 79; 68; 110; 59; 66; 92	176; 223; 192; 318; 307
	M	78.75	223.00
	SD	16.99	65.59
	Range	59-110	176-318
Channa	*N* (Probes)	8	5
	Raw Scores	79; 130; 93; 90; 109; 92; 60; 71	106; 135; 129; 359; 192
	M	90.50	135.00
	SD	21.89	102.71
	Range	60-130	106-359

Table 2
Writing Rubric Scores for Each Participant

		Baseline	Intervention
Aaron	*N* (Probes)	6	7
	Raw Scores	7; 7; 9; 7; 5; 6	12; 11; 12; 12; 13; 12; 12
	M	6.83	12.00
	SD	1.33	0.58
	Range	5-9	11-13
Baci	*N* (Probes)	8	5
	Raw Scores	6; 6; 6, 5, 10; 6; 4; 9	13; 14; 15; 13; 15
	M	6.50	14.00
	SD	2.00	1.00
	Range	4-9	13-15
Channa	*N* (Probes)	8	5
	Raw Scores	5; 4; 6; 8; 7; 5; 4; 8	11; 11; 13; 14; 13
	M	5.88	13.00
	SD	1.64	1.34
	Range	4-8	11-14

For both the TWW and the writing rubric as the outcome score, the visual analysis for the within-condition comparisons in the baseline settings demonstrates a lack of trend, little variability, and shows that the level represented the data points well. In the treatment sessions, the within-condition comparisons suggest an increasing trend in the desired direction. The between-condition comparisons, in turn, reveal a large difference in level, in that higher TWW scores were observed in the sessions preceded by the intervention compared with the no-intervention sessions. An experimental effect can also be determined by the proportion of overlap between the data in different conditions, where fewer overlapping data points indicate a relatively stronger effect (Kratochwill et al., 2010).

We next applied an non-overlap index, the Percentage of Non-Overlapping Data (PND) (Scruggs, Mastropieri, & Casto, 1987), which is defined as the ratio of measurements in Phase B that exceeds the highest measurement from Phase A. For TWW, the PND for Aaron, Baci, and Channa equaled 100%, 100%, and 60%, respectively. For the writing rubric, the PND equaled 100% across the three participants. According to Scruggs and Mastropieri (1998), PND scores above 90 are indications of very effective treatments, scores from 70 to 90 represent effective interventions, scores from 50 to 70 represent

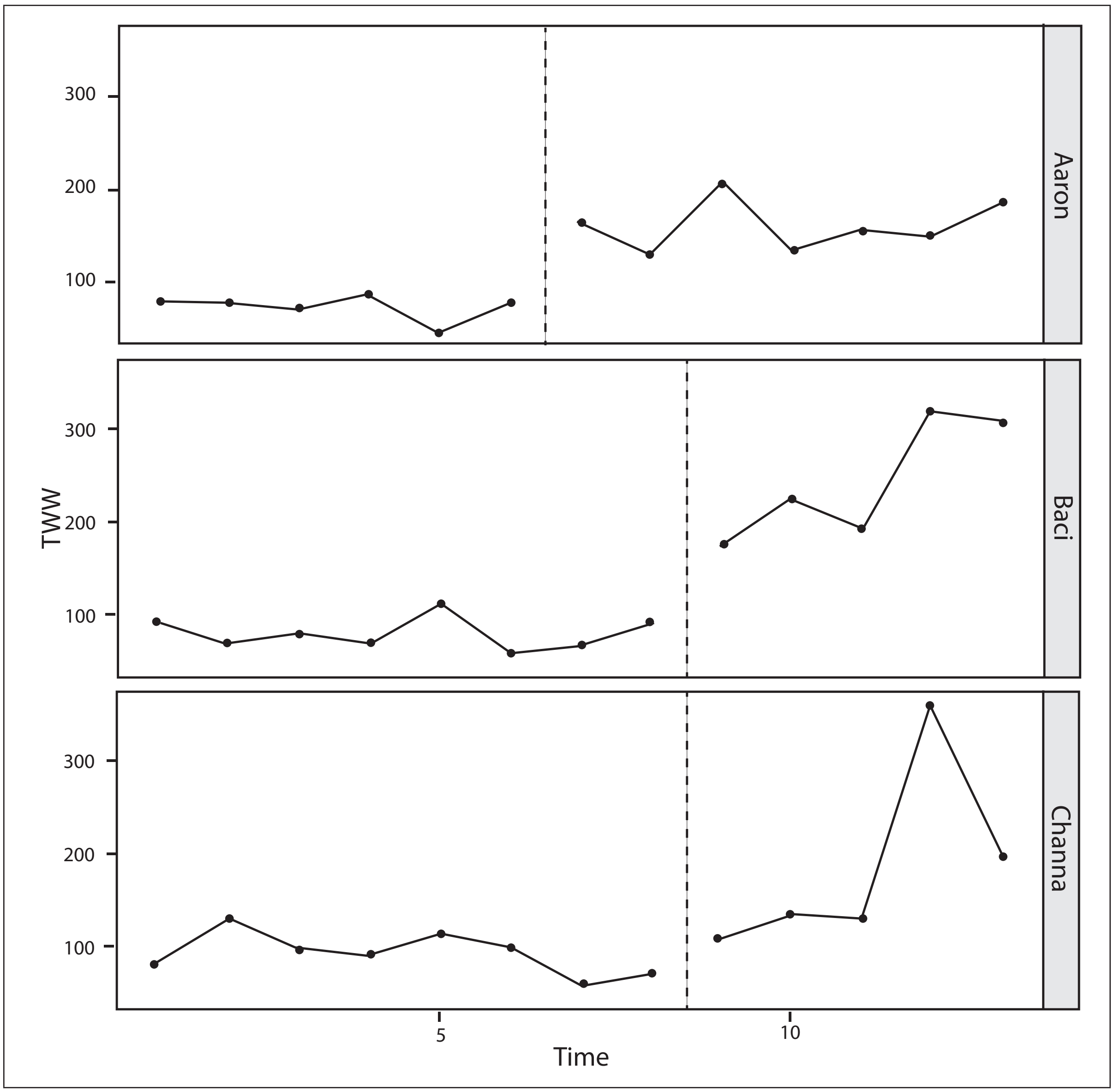


Figure 3. TWW for all three cases.

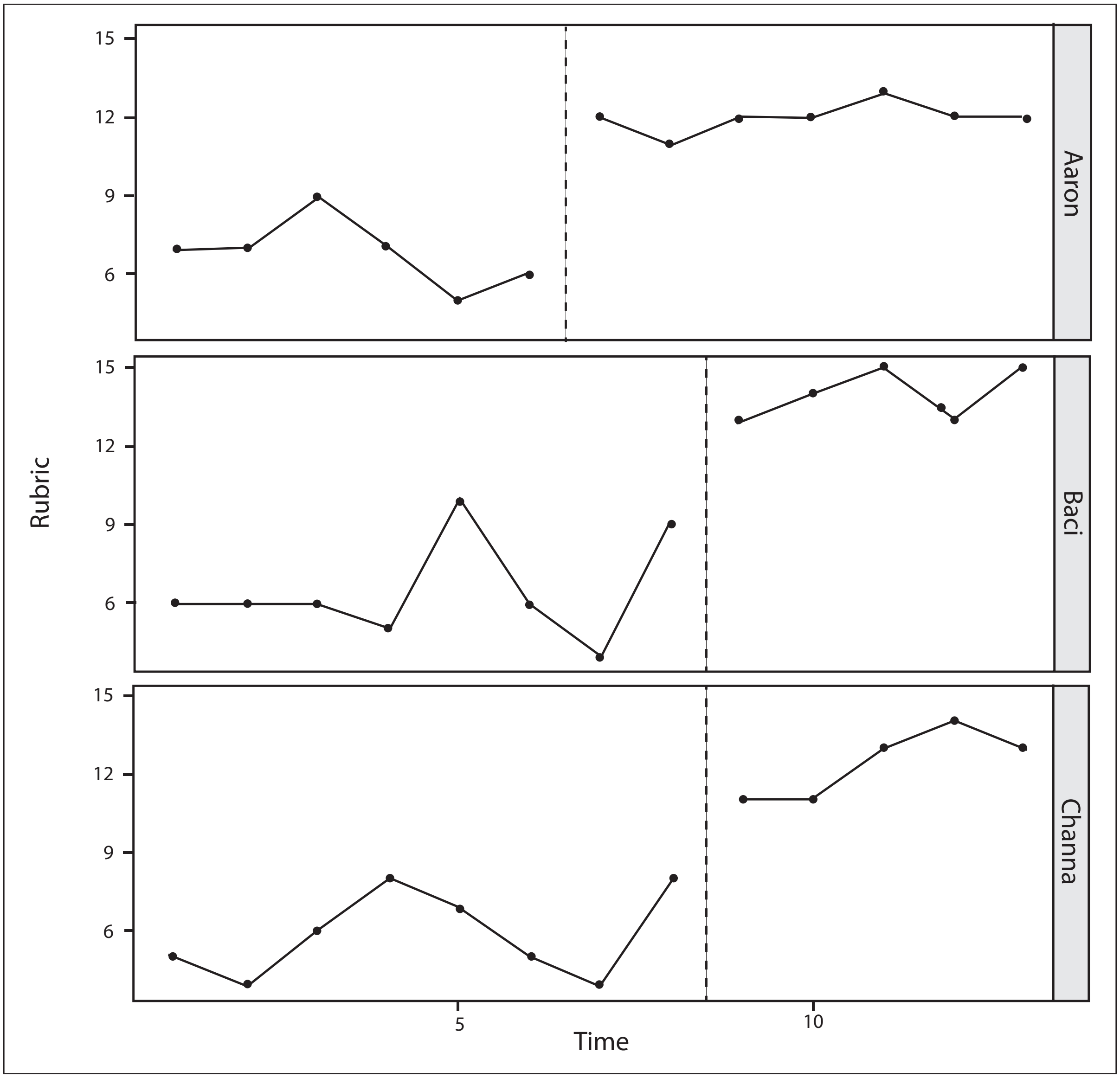


Figure 4. Writing rubric scores for all three cases.

questionable interventions, and scores below 50 are indicative of ineffective interventions. Hence, overall, the strategy instruction may be considered as being very beneficial. Only the PND score of 60% for Channa in the TWW stands out. As mentioned earlier, the number of words that she wrote in her stories increased from Phase A to Phase B by about 50%. However, she produced one relatively short text of just 106 words at the beginning of her relatively brief intervention, which strongly contributed to her low PND score.

Quantitative Analysis

Horner et al. (2005) consider visual inspection the traditional and most common mode of analyzing data from single-case research. However, increasingly more researchers, including the authors of this paper, are calling for a complementary use of statistical techniques to provide a direct test of the null hypothesis and to employ precisely defined criteria for significance (Grünke, Boon, & Burke, 2015; Lobo, Moeyaert, Baraldi, & Babik, 2017; Stone, Friedlander, & Moeyaert, 2018; Tate et al., 2016). As a result, we also performed the following tests.

Randomization test. As previously mentioned, treatment for each of the three cases started at random with the only restriction being that there should be at least three measurements per participant in each phase. Therefore, the total number of unique ways in which the start of the intervention could be delivered equals $8^3 = 343$. The p value assesses the null hypothesis that the experimental manipulation had no impact on the dependent variable by "rearranging the observed scores to all permutations of the possible randomization orders and examining different outcomes" (Barlow, Nock, & Hersen, 2009, p. 284).

In the present study, we conducted the randomization test using the SCDA package within R (Bulté & Onghena, 2013). The difference between the mean of the treatment phase scores and the mean of the baseline phase scores on each dependent variable was used as test statistic (i.e., $\bar{B} - \bar{A}$; Edgington & Onghena, 2007) and compared with a sampling distribution using the set of 343 permutations. For the TWW, the observed test statistic was 115.64 ($p < .01$); for the writing rubric, it was 6.40 ($p < .01$). As a consequence, we can deduce that the story-mapping strategy intervention had a statistically significant and large effect on both text quantity and completeness.

Individual-level analysis. The following data evaluation focused on calculating different effect size measures to determine the quantitative magnitude of the treatment benefits. For our single-level analysis, we included Glass' Δ (Glass, McGaw, & Smith, 1981), Hedges' g (Hedges, Pustejovsky, & Shadish, 2013), the Tau-U (Parker, Vannest, Davis, & Sauber, 2010), and a regression-based approach (Van den Noortgate & Onghena, 2003a, 2003b). The advantage of the regression-based estimate is that, in addition to the changes in level, it is possible to estimate the initial baseline level, the trend during the baseline, the immediate treatment effect, and the change in trend between the baseline and treatment phase (Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2014).

From the single-level analysis, we can deduce that the intervention was consistently statistically significant across a variety of effect sizes for both the TWW and the writing rubric as the outcome (see Table 3). This finding offers evidence that the story-mapping strategy is an effective intervention.

Two-level analysis. In addition to calculating the Tau-U for each student, we generated a weighted, across-case Tau-U, using the Tau-U online calculator (Vannest, Parker, & Gonen, 2011). The analysis yielded a weighted Tau-U of .95 [$SD = 0.20$; $z = 4.86$; $p < .01$) for the TWW and 1.00 [$SD = 0.20$; $z = 5.11$; p < .01) for the writing rubric. Thus, the treatment effect can be considered to be remarkable.

Moreover, after ensuring that the data met the requirements for conducting a regression analysis, we applied multilevel modeling, as outlined by Van den Noortgate and Onghena (2008). The purpose of this last step of the evaluation was to estimate the overall average effect of the treatment across participants, as well as the within- and between-case variability.

Two different two-level models were run. Equations 1 through 3 present "Model 1," in which the overall average baseline level and the overall average change in level is estimated in addition to the between-case variance in the baseline level and between-case variance in

Table 3
Output Effect Sizes for the Single-Level Analysis

		Effect size type			
Variable	Case	Glass' Δ	Hedges, g**	Tau-U	Regression***
TWW	Aaron	5.78	3.68	1.00* [SD = 0.33]	88.79 [SE = 12.48]*
	Baci	8.49	3.65	1.00* [SD = 0.34]	164.45 [SE = 23.84]*
	Channa	2.03	1.35	0.85* [SD = 0.34]	93.70 [SE = 36.69]*
Rubric	Aaron	3.89	4.84	1.00* [SD = 0.33]	5.17 [SE = 0.55]*
	Baci	3.75	4.09	1.00* [SD = 0.34]	7.50 [SE = 0.97]*
	Channa	6.93	3.94	1.00* [SD = 0.34]	6.53 [SE = 0.88]*

* $p < .001$.
**Hedges bias-corrected effect size for small samples.
***Changes in level effect size.

changes in level. In Equation 1, y_{ij} refers to the outcome at measurement occasion i nested within participant j. The outcome is regressed on an intercept and a dummy-coded variable ($Treatment_{ij}$) indicating to which phase the outcome score belongs. If $Treatment_{ij}$ is 0, then the score is part of the baseline phase; otherwise it is part of the treatment phase. Therefore, β_{0j} is the parameter reflecting the baseline level for case j, and β_{1j} represents the change in level, hence the treatment effect for case j.

Level 1:

$$y_{ij} = \beta_{0j} + \beta_{1j} Treatment_{ij} + e_i \text{ with } e_i \sim N(0, \sigma_e^2) \quad (1)$$

It is unlikely that the baseline level and the treatment effect is the same for all participants. Therefore, a second level is added to the model, allowing the subject-specific coefficients (and) to vary across subjects:

Level 2:

$$\begin{cases} \beta_{0j} = \theta_{00} + u_{0j} \\ \beta_{1j} = \theta_{10} + u_{1j} \end{cases} \text{ with } \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \\ \sigma_{u_1u_0} & \sigma_{u_1}^2 \end{bmatrix}\right) \quad (2)$$

Using equations 1 and 2, the overall average baseline level (θ_{00}) and intervention effect (θ_{10}) can be estimated in addition to the between-case variability in baseline level ($\sigma^2_{u_0}$) and treatment effectiveness ($\sigma^2_{u_1}$).

Equations 5 and 6 present "Model 2;" they were used to estimate the overall average baseline level (θ_{00}), trend during baseline (θ_{10}), immediate treatment effect (θ_{20}) and treatment effect on the time trend (θ_{30}), in addition to the between-case variance in these estimates.

Table 4
The Output Effect Sizes for the Two-Level Analysis

		Model 1			*Model 2*		
Variable		Estimate	*SE*	*p*	Estimate	*SE*	*p*
TWW	**Fixed Effects**						
	Baseline level	81.59	9.63	<.001	70.29	17.07	< .001
	Trend baseline	/	/	/	-2.67	3.58	.46
	Immediate treatment effect	114.19	26.01	.028	76.25	25.62	.021
	Treatment effect on trend	/	/	/	26.86	12.74	.126
	Variance Effects						
	Baseline level	0.00	/	/	0.00	/	/
	Trend baseline	/	/	/	0.00	/	/
	Immediate treatment effect	1384.89	1753.36	.21	353.28	1031.25	.366
	Treatment effect on trend	/	/	/	347.12	394.43	.189
	Residual variance	2040.28	487.70	<.001	1359.64	343.83	< .001
Rubric	**Fixed Effects**						
	Baseline level	6.36	0.30	<.001	6.76	0.65	< .001
	Trend baseline	/	/	/	0.08	0.15	.61
	Immediate treatment effect	6.41	0.68	.002	5.40	1.01	< .001
	Treatment effect on trend	/	/	/	0.20	0.25	.43
	Variance Effects						
	Baseline level	0.00	/	/	0.00	/	/
	Trend baseline				0.012	0.03	.32
	Immediate treatment effect	0.74	1.12	.25	0.84	1.25	.25
	Treatment effect on trend				0.00	/	/
	Residual variance	2.07	0.49	<.001	1.93	0.49	< .001

Level 1:

$$y_{ij} = \beta_{0j} + \beta_{1j} Time_{ij} + \beta_{2j} Treatment_{ij} + \beta_{3j} Treatment_{ij} * Time_{ij} + e_i \text{ with } e_i \sim N(0, \sigma_e^2) \quad (4)$$

Level 2:

$$\begin{cases} \beta_{0j} = \theta_{00} + u_{0j} \\ \beta_{1j} = \theta_{10} + u_{1j} \\ \beta_{2j} = \theta_{20} + u_{2j} \\ \beta_{3j} = \theta_{30} + u_{3j} \end{cases} \text{with} \begin{bmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & & & \\ \sigma_{u_1u_0} & \sigma_{u_1}^2 & & \\ \sigma_{u_2u_0} & \sigma_{u_2u_1} & \sigma_{u_2}^2 & \\ \sigma_{u_3u_0} & \sigma_{u_3u_1} & \sigma_{u_3u_2} & \sigma_{u_3}^2 \end{bmatrix} \right) \quad (5)(6)$$

For more details and interpretation of the two-level analysis of single-case experimental designs, we refer the reader to Moeyaert, Ferron, Beretvas, and Van den Noortgate (2014). The output for the present study may be found in Table 4.

From the two-level analysis, we can deduce that, across the three cases and the two models, the treatment appears to be statistically significant at the .05 level (two-tailed). For Model 1, the change in the average TWW between the baseline and treatment phases equaled 114.19 [$t(2.68) = 4.39$, $p = .028$]. For Model 2, the immediate treatment effect (the change in the TWW between the end of the baseline phase and the start of the treatment phase) equaled 25.62 [$t(2.68) = 4.39$, $p = .028$]. For Model 1, the change in the average writing rubric between baseline and treatment phase equaled 6.41 [$t(3.09) = 9.41$, $p = .002$]. For Model 2, the immediate treatment effect (the change in the writing rubric between the end of the baseline phase and the start of the treatment phase) equaled 5.40 [$t(10.4) = 5.32$, $p < .001$].

Qualitative Analysis

Responses from the interviews conducted by the research assistant, which lasted 8-10 minutes each, helped to verify that all three tutees considered the intervention to be very beneficial. However, it needs to be taken into consideration that the survey was conducted in a rather informal way. That is, we did not choose a particular method from the pool of elaborated and well-established approaches in this field of research. Rather, after the intervention, the research assistant took each of the participants aside to ask them if they liked the format of the treatment, if they viewed it as helpful, if they thought it had an impact on their writing performance, and if they would recommend it to other students, to get some general feedback on how the treatment was received by the students (see Snodgrass, Chung, Meadan, & Halle, 2018).

The responses of the three tutees suggest that the peer-tutoring instruction was well received and that they enjoyed participating in the study. Indeed, they all expressed their approval of the treatment very emphatically. Baci stated: "I normally don't like writing; I like reading better. But working together with Babak was fun." All three tutees also stressed that they deemed the intervention as extremely helpful as they tried to structure their ideas before actually writing their stories. Aaron commented: "Finally, someone showed me how to write better. Story maps make everything much easier." The three children appeared to be very proud of their accomplishments and were under the impression that their writing had improved significantly. They attributed the increases in performance to the format of the instruction. Finally, they unanimously stated that they would recommend the treatment to other children who are struggling with text composition. Channa explained: "It feels good to be able to write much longer stories now. I believe that other kids would also find story maps helpful."

Discussion

Main Findings

The purpose of this single-case analysis was to determine the effects of a peer-tutoring graphic organizing strategy on the writing performance of three eighth graders with LDs. Few studies have evaluated the benefits of teaching text-planning skills to struggling learners by using fellow students as instructors. In this experiment, we demonstrated that a rather short intervention of five to seven 30-minute lessons resulted in statistically and practically significant improvements in the length and completeness of narratives written by the tutees.

The performance enhancements reached impressive magnitudes ranging from around 50% to over 180% in the TWW and from about 75% to a little over 120% in the writing rubric. In five of six instances, there was no overlap between the data points during the baseline and intervention. A randomization test revealed a significant overall level change between phases. On an individual basis, different effect sizes indicated remarkable treatment outcomes for all three tutees. Finally, our summary of the data across the three cases via an overall Tau-U and multilevel modeling confirmed the results from all previous analyses and highlight the great benefit that the peer-tutoring intervention had on the performance of the participants.

Thus, the results are in line with our expectations and with the findings from extant studies on teaching students with LDs the story-mapping strategy through peer tutoring (see Grünke et al., 2016, 2017). What is especially gratifying and unique with regard to our experiment is that it shows that, although early intervention is important, it is not too late to teach adolescents with LDs simple text-planning skills using rather simple means. That is, it did not take long for the students to be able to apply the strategy and significantly improve their text-production skills. According to their main teacher, they had been trying unsuccessfully for many years to become more proficient writers. Furthermore, they seemingly enjoyed the intervention and were very happy about the progress that they had made.

Limitations

Notwithstanding the promising results, the study is subject to certain limitations. First and foremost, the small number of participants does not allow for generalization of the data. However, according to the widely accepted standards for single-subject research by Horner et al. (2005), external validity can be achieved through replication. The Council for Exceptional Children (CEC; 2014) stipulates that at least five methodologically sound case reports with positive effects and at least 20 total participants are needed in order to consider a treatment as evidence based. Thus, this experiment is a step toward the goal of establishing the intervention as an approach that meets these standards.

We incorporated a randomization procedure into the design, which increased the internal validity of the single-case analysis. Fortunately, we ended up with at least six baseline measurements, thus exceeding the minimum standard by Horner et al. (2005). The data trend prior to the intervention was quite stable, which also speaks to a high internal validity. However, the final two baseline data points for all three tutees indicate a possible rise in performance rather than level or decreasing performance. To strengthen the assumption of a functional connection between treatment and achievement, additional baseline measurements would have been beneficial. In addition, we did not provide for follow-up data collection. Even though the participants learned quickly how to make efficient use of the strategy, there is no way to know – with a given degree of certainty – whether the intervention effects are lasting. Upcoming school holidays did not allow for the acquisition of additional data directly upon the termination of the treatment. Undoubtedly, information about the stability of the performance gains would have helped to make an even stronger case for applying our intervention in schools.

Furthermore, the process of selecting participants may be considered a limitation. To measure the tutees' learning progress, we asked them to compose texts in response to story starters. However, we predicated the identification of eligible participants on the number of words the students produced while writing about a cartoon strip. Even though the fact that the TWW of the three tutees averaged less than 100 and the number of total points in the rubric averaged less than 7 during baseline speaks to the supposition that our identification process fulfilled its intended purpose, a different option for choosing the participants might have been more suitable. For example, we could have utilized writing prompts with no pictures to better align the selection procedure with the intervention. In addition, we could have considered the quality of the texts while trying to find tutees who would especially benefit from the treatment.

A final limitation pertains to the way the dependent variable was measured. While capturing TWW is the most common way to monitor writing performance, it is not the only one (see, e.g., Dockrell, Connelly, Walter, & Critten, 2015; McMaster & Espin, 2007). The same goes for the writing rubric that we used to measure students' skills in producing narratives. As writing is a complex ability, no one way of measuring could ever capture it completely. Hence, the way in which we performed our assessments is open to criticism. However, although we approached writing skills from two rather different angles (TWW and a writing rubric), the results were highly correlated. Moreover, both procedures were clearly very sensitive to change and were able to reflect the improvements that the participants had undergone.

Implications for Practice and Future Research

Despite any weaknesses that this study might have, the present findings are very encouraging and suggest that despite being such a demanding task, writing can be effectively taught to youth with LDs with very minimal resources. The intervention consisted of only five to seven 30-minute lessons, and instruction for the tutors only took two hours. The treatment was also inexpensive, including the cost of printing the posters, the story-map templates, and the index cards. Finally, ways were found for the teams to work on their assignments, even though the school was not able to provide an individual room for each tutor-tutee pair.

As Johnson and Semmelroth (2014) rightly pointed out, "while arguably no other content area in education has produced more instructional practice research than special education, the profession itself has made little progress in getting these instructional strategies into practice" (p. 71). One major reason for this gap stems from the fact that many teachers are overly burdened with crowded classrooms full of very diverse learners and a wide range of administrational tasks unrelated to teaching (Vannest & Hagan-Burke, 2009). This is particularly true in special education settings (Greenglass, Burke, & Konarski, 1997).

Being able to impart a highly effective strategy such as story mapping by way of peer tutoring opens up opportunities for teachers to provide students with LDs and other struggling learners with much-needed individually tailored support. In this study, the treatment was monitored by a research assistant. Even though the main teachers should generally be capable of overseeing peer-tutoring interventions in their own classrooms, an additional person to provide help whenever necessary is beneficial. To that end, Christle and Schuster (2003) suggested that educators involve interns, administrators, college students, school aides, volunteers, and other temporarily available people into everyday teaching.

Future research is warranted to confirm our findings with larger samples and to investigate the duration of the effects of an intervention such as ours. Such experiments should be conducted using different empirical designs and different ways of measuring text production abilities. In addition, it appears meaningful to not only monitor the tutees' progress, but also to determine the impact on the tutors of a system of instruction in which learners help each other. If tutees also benefited, it would constitute an even stronger case for implementing the strategy into everyday life in school than if only the tutees profited. Finally, the focus in this study was on improving the planning skills of adolescents with LDs in order to help them to write better narratives. Future research should also zero in on supporting these students to produce other text genres (e.g., descriptions, expositions, or poetry) and to foster other writing subskills (e.g., composing, refining, editing, or reviewing).

References

Applebee, A. N., & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal, 100*, 14-27.

Archer, A. L., & Hughes, C. A. (2010). *Explicit instruction: Effective and efficient teaching*. New York, NY: Guilford.

Barenbaum, E., Newcomer, P., & Nodine, B. (1987). Children's ability to write stories as a function of variation in task, age, and developmental level. *Learning Disability Quarterly, 10*, 175-188.

Barlow, D. H., Nock, M., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior for change*. London, UK: Pearson.

Boon, R. T., Paal, M., Hintz, A. M., & Cornelius-Freyre, M. (2015). A review of story mapping instruction for secondary students with LD. *Learning Disabilities: A Contemporary Journal, 13*, 117-140.

Bulté, I., & Onghena, P. (2013). The single-case data analysis package: Analysing single-case experiments with R software. *Journal of Modern Applied Statistical Methods, 12*, 450-478. https://doi.org/10.22237/jmasm/1383280020

Council for Exceptional Children (CEC). (2014). Standards for evidence-based practices in special education. *Teaching Exceptional Children, 46*, 206-212.

Christle, C. A., & Schuster, J. W. (2003). The effects of using response cards on student participation, academic achievement, and on-task behavior during whole-class math instruction. *Journal of Behavioral Intervention, 12*, 147-165.

Cools, W., Declerq, L., Beretvas, S. N., Moeyaert, M., Ferron, J. M., & Van den Noorgate, W. (2018). *MultiSCED user guide*. https://kuleuven.app.box.com/v/MultiSCEDUserGuide

Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children's writing products: The role of curriculum based measures. *British Educational Research Journal, 41*, 575-595.

Edgington, E. S., & Onghena, P. (2007). *Randomization tests*. Boca Raton, FL: Chapman & Hall/CRC.

Englert, C., Berry, R., & Dunsmore, K. (2001). A case study of the apprenticeship process. *Journal of Learning Disabilities, 34*, 152-171. https://doi.org/10.1177/002221940103400205

Flower, L. S., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*, 365-387.

Furey, W. M., Marcotte, A. M., Hintze, J., & Shackett, C. (2016). Concurrent validity and classification accuracy of curriculum-based measurement for written expression. *School Psychology Quarterly, 31*, 369-382. https://doi.org/10.1037/spq0000138

Gast, D. L., Lloyd, B. P., & Ledford, J. R. (2018). Multiple baseline and multiple probe designs. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology* (pp. 239-282). New York, NY: Routledge.

Gillespie, A., & Graham, S. (2014). A meta-analysis of writing interventions for students with learning disabilities. *Exceptional Children, 80*, 454-473. https://doi.org/10.1177/0014402914527238

Glass, G.V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. London, UK; Sage.

Graham, S. (2008). *Effective writing instruction for all students*. Wisconsin Rapids, WI: Renaissance Learning.

Graham, S., & Hall, T. E. (2016). Writing and writing difficulties from primary grades to college: Introduction to the special issue. *Learning Disability Quarterly, 39*, 3-4. https://doi.org/10.1177/0731948715592154

Graham, S., Capizzi, A., Harris, K. R., Herbert, M., & Morphy, P. (2014). Teaching writing to middle school students: A national survey. *Reading and Writing, 27*, 1015-1042. https://doi.org/10.1007/s11145-013-9495-7

Graham, S., Collins, A. A., & Rigby-Wills, H. (2017). Writing characteristics of students with learning disabilities and typically achieving peers: A meta-analysis. *Exceptional Children, 83*, 199-218. https://doi.org/10.1177/0014402916664070

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445-476. https://doi.org/10.1037/0022-0663.99.3.445

Greenglass, E. R., Burke, R. J., & Konarski, R. (1997). The impact of social support on the development of burnout in teachers: Examination of model. *Work and Stress, 11*, 267-278.

Grünke, M., Boon, R. T., & Burke, M. (2015). Use of the randomization test in single-case research. *International Journal for Research in Learning Disabilities, 2*, 44-64.

Grünke, M., Janning, A. M., & Sperling, M. (2016). The effects of a peer-tutoring intervention on the text production of students with learning and speech problems: A case report. *Learning Disabilities: A Contemporary Journal, 14*, 225-235.

Grünke, M., & Leonard Zabel, A. M. (2015). How to support struggling writers. *International Journal of Special Education, 30*, 137-150.

Grünke, M., & Morrison-Cavendish, W. (2016). Learning disabilities around the globe: Making sense of the heterogeneity of the different viewpoints. *Learning Disabilities: A Contemporary Journal, 14*, 1-8.

Grünke, M., Wilbert, J., Tsiriotakis, I. K., & Agirregoikoa, A. L. (2017). Improving the length and quality of texts written by fourth graders with learning disabilities through a peer-tutoring graphic organizing strategy. *Insights into Learning Disabilities, 14*, 167-188.

Harris, K. R., & Graham, S. (1996). *Making the writing process work: Strategies for composition and self-regulation*. Brookline, MA: Brookline Books.

Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In S. N. Haynes & E. M. Heiby (Eds.), *Behavioral assessment* (pp. 108-127). Hoboken, NJ: John Wiley & Sons.

Hedges, L.V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods, 4*, 324-341. https://doi.org/10.1002/jrsm.1086

Hennes, A.-K., Büyüknarci, Ö., Rietz, C., & Grünke, M. (2015). Helping children with specific learning disabilities to improve their narrative writing competence by teaching them the use of the story maps strategy. *Insights into Learning Disabilities, 12*, 35-56.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in in special education. *Exceptional Children, 71*, 165-179. https://doi.org/10.1177/001440290507100203

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2016). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: Guilford.

Johnson, E. S., & Semmelroth, C. L. (2014). Special education teacher evaluation: Why it matters and what makes it challenging. *Assessment for Effective Intervention, 39*, 71-82.

Kinder, R. A. (2014). *1,000 awesome writing prompts*. North Charleston, SC: Create Space.

Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology, 101*, 136-160. https://doi.org/10.1037/a0013097

Koutsoftas, A. D. (2016). Writing process products in intermediate-grade children with and without language-based learning disabilities. *Journal of Speech, Language & Hearing Research, 59*, 1471-1483. https://doi.org/10.1044/2016_JSLHR-L-15-0133

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What Works Clearinghouse*. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kurihara, N. (2017). Do peer reviews help improve student writing abilities in an EFL high school classroom? *TESOL Journal, 8*, 450-470.

Li, D. (2007). Story mapping and its effects on writing fluency and word diversity of students with. learning disabilities. *Learning Disabilities: A Contemporary Journal, 5*, 77-93.

Lobo, M. A., Moeyaert, M., Baraldi, C. A., & Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy, 41*, 187-197. https://doi.org/10.1097/NPT.0000000000000187

MacArthur, C. A., & Graham, S. (1987). Learning disabled students' composing under three methods of text production: Handwriting, word processing, and dictation. *Journal of Special Education, 21*, 22-42.

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing. *The Journal of Special Education, 41*, 68-84. https://doi.org/10.1177/00224669070410020301

Moeyaert, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology, 52*, 191-211. https://doi.org/10.1016/j.jsp.2013.11.003

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2014). The influence of the design matrix on treatment effect estimates in the quantitative analyses of single-subject experimental design research. *Behavior Modification, 38*, 665-704. https://doi.org/10.1177/0145445514535243

National Association of Colleges and Employers. (2012, November). *Job outlook 2011*. Retrieved from http://career.pages.tcnj.edu/files/2011/07/Job_Outlook_2011_Full_Report_PDF1.pdf

National Center for Education Statistics. (2012). *The nation's report card: Writing 2011* (NCES 2012-470). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Commission on Writing. (2004). *Writing: A ticket to work … or a ticket out.* Retrieved from http://www.collegeboard.com/prod_downloads/writingcom/writing-ticket-to-work.pdf.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2010). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy, 42*, 284-299. https://doi.org/10.1016/j.beth.2010.08.006

Puranik, C. S., Patchan, M. M., Lemons, C. J., & Al Otaiba, S. (2017). Using peer assisted strategies to teach early writing: Results of a pilot study to examine feasibility and promise. *Reading and Writing: An Interdisciplinary Journal, 30*, 25-50. https://doi.org/10.1007/s11145-016-9661-9

Rensing, J., Vierbuchen, M.-C., Hillenbrand, C., & Grünke, M. (2016). Implementing peer-assisted writing support in German secondary schools. *Insights into Learning Disabilities, 13*, 151-164.

Rietdijk, S., van Weijen, D., Janssen, T., van den Bergh, H., & Rijlaarsdam, G. (2018). Teaching writing in primary education: Classroom practice, time, teachers' beliefs and skills. *Journal of Educational Psychology, 110*, 640-663. http://dx.doi.org/10.1037/edu0000237

Saddler, B., & Asaro, K. (2007). Increasing story quality through planning and revising: Effects on young writers with learning disabilities. *Learning Disability Quarterly, 30*, 223-234. https://doi.org/10.2307/25474635

Saddler, B., & Asaro, K. (2008). The effects of peer assisted sentence combining practice on four writers with learning disabilities. *Learning Disabilities: A Contemporary Journal, 6*, 17-31.

Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology, 97*, 43-54. https://doi.org/10.1037/0022-0663.97.1.43

Santangelo, T. (2014). Why is writing so difficult for students with learning disabilities? A narrative review to inform the design of effective instruction. *Learning Disabilities: A Contemporary Journal, 12*, 5-20.

Scardamalia, M., & Bereiter, C. (1986). Written composition. In M. Wittrock (Ed.), *Handbook on research on teaching* (pp. 778-803). New York, NY: MacMillan.

Scruggs, T., Mastropieri, M., & Casto, G. (1987). The quantitative synthesis of single subject research. *Remedial and Special Education, 8*, 24-33. https://doi.org/10.1177/074193258700800206

Scruggs, T. E., & Mastropieri, M. A. (1998). Synthesizing single subject studies: Issues and applications. *Behavior Modification, 22*, 221-242. https://doi.org/10.1177/01454455980223001

Skarr, A., Zielinski, K., Ruwe, K., Sharp, H., Williams, R. L., & McLaughlin, T. F. (2014). The effects of direct instruction flashcard and math racetrack procedures on mastery of basic multiplication facts by three elementary school students. *Education and Treatment of Children, 37*, 77-93.

Snodgrass, M. R., Chung, M. Y., Meadan, H., & Halle, J. W. (2018). Social validity in single-case research: A systematic literature review of prevalence and application. *Research in Developmental Disabilities, 74*, 160-173. https://doi.org/10.1016/j.ridd.2018.01.007

Stone, M., Friedlander, M. L., & Moeyaert, M. (2018). Illustrating novel techniques for analyzing single-case experiments: Effects of pre-session mindfulness practice. *Journal of Counseling Psychology, 65*, 690-702. https://doi.org/10.1037/cou0000291

Sundeen, T. H. (2014). Essay development for secondary students with learning disabilities: Graphic organizers for visualizing organizational patterns. *Rural Special Education Quarterly, 33*, 29-36. https://doi.org/10.1177/875687051403300305

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., … Wilson, B. (2016). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016 statement. *Archives of Scientific Psychology, 4*, 1-9.

UNESCO. (2011). *International standard classification of education*. Retrieved from http://uis.unesco.org/

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18*, 325-346. https://doi.org/10.1521/scpq.18.3.325.22577

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers, 35*, 1-10.

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence Based Communication Assessment and Intervention, 2*, 142-151.

Vannest, K. J., & Hagan-Burke, S. (2009). Teacher time use in special education. *Remedial and Special Education, 31*, 126-142. https://doi.org/10.1177/0741932508327459

Vannest, K. J., Parker, R. I., & Gonen, O. (2011). Single case research: Web based calculators for SCR analysis (Version 1.0) [Web-based application]. College Station, TX: Texas A & M University. Available from singlecaseresearch.org.

Yarrow, F., & Topping, K. J. (2001). Collaborative writing: The effects of metacognitive prompting and structured peer interaction. *British Journal of Educational Psychology, 71,* 261-282. https://doi.org/10.1348/000709901158514

Improving Writing: Focus on Elementary-School African American Male Students With a Learning Disability

Svjetlana Curcic and Sara Platt
University of Mississippi

Abstract

This study examined the effectiveness of a writing intervention with three African American male third graders with a learning disability in reading and low writing skills. The participants were instructed in planning, organizing, writing, editing, and revising, supported by dictation and transcription of students' thoughts with the *Dragon Dictation* app. Multiple-baseline data across participants were collected on the following curriculum-based measures: total words written, words spelled correctly, and correct word sequence. Visual data analysis, the percentage of nonoverlapping data, and the post-intervention data indicate large gains on the curriculum-based measures across participants.

Keywords: writing, elementary, learning disability, African American

Writing competence has been recognized as an important, albeit neglected, skill to be mastered during K-12 schooling. For example, in 2003 the National Commission on Writing (2003) drew attention to the neglected "R" (among reading, writing, and arithmetic) and called for a writing revolution even in its report title, *The Neglected "R": The Need for a Writing Revolution.* Almost a decade later, the National Assessment of Educational Progress (NAEP) data showed that one in five 8th- and 12th-grade students in the United States still performed below the basic level in writing (the basic level denotes only partial mastery) (National Center for Education Statistics, 2012).

Disaggregated by race, gender, and ability, data indicate that Black students in eighth grade performed significantly below their White peers: 87% of White students performed at or above the basic level, whereas 65% of Black students performed at or above the basic level (National Center for Education Statistics, 2012). Similarly, 87% of 12th-grade White students performed at or above the basic level, whereas only 61% of Black students performed at or above the basic level. In 2011, the average score for female students in 8th grade was 160 and 140 for male students (on a scale of 0-300), and 157 as opposed to 143 for males in 12th grade. Finally, students with disabilities had an average score of 113 in writing, while students not identified with a disability had an average score of 154 in 8th grade, similar to 12th-grade outcomes: 112 and 153 for students with and without disabilities, respectively (National Center for Education Statistics, 2012).

In short, the NAEP data, known as the nation's report card, support long-standing concerns over African American students' achievement in comparison to their White peers, as well as a discrepancy between male and female student achievement in writing and a discrepancy between the achievement of students with disabilities vs. their nondisabled peers. As a result, a K-12 student who is an African American male with a disability is not likely, at least statistically speaking, to achieve high competency in writing. Also disconcerting is the fact that in spite of an increased focus on evidence-based practices in the field of education, research on writing with African American students, especially those with disabilities, is almost nonexistent (Graham, Harris, & Beard, in press).

The focus of the present study was on improving the writing skills of a small group of African American students with a learning disability (LD) in reading. All three participants were males, with low writing skills and speaking in one of the southern U.S. dialects, often referred to as African American Vernacular English (AAVE).

Writing Intervention Research With African American K-12 Students

Recognizing a dearth of writing intervention research with African American students, Graham and colleagues (in press) re-analyzed data from five studies designed as true experiments to test writing, reading, and math interventions for students with special needs or at risk for learning problems. The majority of the participants were African American elementary-school students who lived in poor neighborhoods in the Washington, DC, area. The specific focus of the re-analysis was on the outcomes of African American male elementary students experiencing difficulties in writing. The intervention strategies included teaching advanced planning, self-regulation, revising, sentence construction, spelling, and handwriting. The effect sizes were corrected for small samples.

Graham and colleagues (in press) taught third graders advanced planning and self-regulation strategies for planning and drafting stories and opinion texts, including 32 African American students (9 control students, 11 students in a self-regulated strategy condition, and 13 students in a self-regulated strategy development plus peer support condition). The effect size (ES) for students in the self-regulated strategy development condition was 0.80 for stories and 1.28 for opinion essays, and for self-regulated strategy development plus peer support for transfer vs. control, ES was 1.24 for stories and 0.98 for opinion essays.

In another study, Graham and Harris (Graham et al., in press) individually instructed six fourth-grade African American male students over four sessions to consider the following revision goals to improve their story writing: (a) set the story in a different time, (b) add a new character, (c) change the location, and (d) change when the story occurred. The revising instruction based on goal setting had a positive impact compared to the control condition with ES of 0.97 for quality of revision and ES of 1.70 for meaning-changing revisions.

Another strategy (Graham et al., in press) that improved African American male fourth graders' writing was based on sentence combining. Students were instructed in developing complex sentences, with seven African American male students paired with more skilled fourth-grade writers, and four African American male students assigned to a control group who were taught grammar skills. The sentence combining comprised five units. In each unit, the teacher provided explanation and modeling in using conjunctions, incorporating adjectives and adverbs, and combining adjectival and adverbial clauses. The instruction was followed by paired students practicing combining the sentences first orally, and then in writing. An effect size of 0.31 was obtained for the sentence-combining measure, and the compositions were qualitatively better than the compositions written by the control group (ES = 0.14).

Instruction in spelling and handwriting have both been found to contribute to increased writing competency. In one study (Graham et al., in press), instruction to second graders assigned to a spelling condition (eight African American males) and math instruction (17 African American males) showed that spelling instruction resulted in ES of 0.64 and 0.47 for the spelling treatment on two spelling tests, and ES of 0.96 for sentence writing. While working with first graders on handwriting in another study, Graham and colleagues (Graham et al., in press) instructed the students in two conditions: half of the students were randomly assigned to the handwriting condition (including seven African American boys), and the other half were randomly assigned to phonological awareness instruction (including 10 African American boys). The boys in the handwriting condition made greater gains than the group in the phonological control condition, with ES = 1.01 in handwriting fluency, ES = 0.54 in writing quality, and ES = 0.75 in sentence construction.

In addition to recent work by Graham and colleagues, Fogel and Ehri (2000) taught third- and fourth-grade students who spoke African American Vernacular English (AAVE) (*N* = 89) to increase the use of Standard English (SE) in their writing. The students were instructed in three conditions: (a) exposure to SE features in stories; (b) story exposure plus explanation of SE rules; and (c) story exposure, SE rule instruction, and guided practices in transforming AAVE to SE. The SE syntactic features taught were possessive "s," past tense "-ed," third-person present-tense singular "s," plural "s," indefinite article, and subject-verb agreement. The students in the third group performed better at posttest on translation tasks and exhibited a higher rate of success (defined as 65% or higher use of SE forms) in their free writing. Also, the students in this group, which included practice as opposed to simply exposure to or explanation of the rules, showed 81% increase in SE features in free writing, in contrast to 33% of the students in the story exposure-only group and 55% of the students in the story-exposure-and-rule-explanation condition. The outcomes were not disaggregated by gender or ability, but all students showed low writing achievement.

A number of writing intervention studies have included African American students as participants. Yet, the outcomes related specifically to African American students with LD tend to be masked within the means of experimental and control groups, limiting our knowledge of effective interventions for this student population (Gillespie & Graham, 2014; Washington, Patton-Terry, & Seidenberg, 2014). In addition, Rogers and Graham (2008) noted in their meta-analysis of single-subject writing intervention studies that many researchers "failed to adequately describe their participants" (p. 900), which limits generalization about the effectiveness of interventions for specific student populations.

To provide a more nuanced understanding of the potential of writing instruction for students with a specific learning disability (LD), we examined the effects of a writing intervention conducted with three African American male third graders with LD in reading and low writing skills using a multiple-baseline research design.

Theoretical Framework

The theoretical framework of the present study was based on the long line of research on explicit and strategy instruction with students with LD (e.g., Englert, 2009; Fletcher, Lyon, Fuchs, & Barnes, 2019; Graham & Harris, 2009; Schumaker & Deshler, 2009). Explicit instruction with modeling and guided practice, frequent opportunities to practice with feedback, along with visual or verbal prompts, has been discussed from different theoretical perspectives (Graham, Harris, & Santangelo, 2015; Hughes, Morris, Therrien, & Benson, 2017). While some researchers suggest that explicit instruction is aligned with applied behavioral analysis, consistent with positive reinforcement (feedback), examples (stimulus control), and modeling (orientation to stimuli), we believe that explicit instruction provides support in reducing the cognitive load for students who lack prerequisite knowledge or automaticity in the application of skills (Hughes et al., 2017).

We focused the instructional framework on an adapted planning, organizing, writing, editing, and revising strategy (captured by Englert and her colleagues [1991] in the acronym POWER) and literacy activities focused on summarization of texts in reading-writing contexts (Englert, Raphael, Anderson, Stevens, & Fear, 1991; Graham & Hebert, 2010; Graham, MacArthur, & Hebert, 2019). Developing literacy skills in reading-writing contexts seems to be neglected in both reading and writing research in spite of increasing evidence that instruction in the summarization of texts read shows large effect sizes in writing quality (Graham & Perin, 2007; Kang, McKenna, Arden, & Ciullo, 2015; Mason, Snyder, Sukhram, & Kedem, 2006). Graham and Hebert (2010) recommended (a) having students write about texts they read; (b) teaching students the writing skills and processes needed to create texts; and (c) increasing the frequency of student writing. Graham and Hebert's (2010) recommendations are reflected in our writing intervention.

To ensure meaningful student participation, we considered the principles of universal design for learning (UDL), which addresses students' potential problems upfront as opposed to retrofitting instruction (Center for Applied Special Education Technology, 2018). Accordingly, a number of the texts used in the present study were made accessible with a text-to-speech accommodation for reading. To provide support for writing development, we used *Dragon Dictation*, a free app. The app served as a bridge between students' mental processes and transcription of their ideas into written products (Hayes & Olinghouse, 2015).

Research Questions

The study attempted to answer the following questions:

1. Do students who participate in writing intervention focused on summarization of fictional and nonfictional texts improve their writing as measured by the following curriculum-based measurements (CBMs): total words written (TWW), words spelled correctly (WSC), and correct word sequence (CWS)?
2. If students improve in their writing skills during the intervention, are they able to generalize and maintain their writing skills?
3. What are the students' perceptions about instruction upon the completion of the intervention?

Method

We employed a single-subject research design across participants to examine whether the proposed writing intervention was effective compared to the baseline of students' writing (Kazdin, 2010). In multiple-baseline-data-across-participants designs, the intervention is staggered across time, and comparisons are made both between and within data series across participants (Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf, & Shadish, 2010). To meet research design standards, a multiple-baseline design must include a minimum of six phases (at least three baselines and three intervention phases) (Kratochwill, Hitchcock, Horner, Levin, Odom, Rindskopf, & Shadish, 2013). If either a baseline or an intervention phase is based on fewer than three data points, an effect cannot be demonstrated, and the study does not meet the single-subject design research standards (Kratochwill et al., 2013).

Visual analysis of level, trend, variability, overlap, immediacy of the effect, and data pattern is used to evaluate "whether there are at least indications of an effect at three different points in time" (Kratochwill et al., 2013, p. 32). While there is no consensus on the preferred method of computing an effect size beyond visual inspection of data, a number of methods have been proposed (Rodgers, Lewis, O'Neill, & Vannest, 2018). To calculate an effect size in the current study, we used the widely applied method of calculating a percentage of nonoverlapping data (Rodgers et al., 2018).

Setting

The study was conducted in a public elementary Title I school situated in a rural area of the southeastern United States (Title I schools enroll large numbers or percentages of students from low-income households who may be at risk academically). At the time of the study, the school had 457 students in elementary grades K-4 who came from two small towns. Four percent of the students had individualized education programs (IEPs), and 4% were English language learners (ELLs). Further, 95% of the students received a free or a reduced-price lunch. Finally, at the time of the study, 10% of the student population was homeless, predominantly defined as not having a permanent address and living with friends or extended families. Demographic school data are presented in Table 1.

Table 1
School Racial Demographics

Race	Percentage
African American	81
Asian/Pacific Islander	0
Hispanic	4
Native American	0
White	15

All instructional sessions took place in a resource room, with three African American students with LD scheduled to work with a special education teacher during the fourth period of the school day. We also occasionally worked with a fourth student whose reading and writing levels were significantly below those of his third-grade special education peers. According to his special education teacher, this student was receiving an extra 30-minute instruction, with a different schedule on some days. Due to inconsistent instruction, we did not include the outcomes of this student in the present study (he did show progress, however).

Participants

The three third-grade participants received services in the special education resource room for approximately 35 minutes per day, but also regularly participated in a general education English language arts class. Demographic information about the students and their achievement levels in reading on the standardized test for the assessment of reading (STAR) is provided in Table 2 (all names are pseudonyms). Informal testing on the Graded Word List indicated that John read independently at the first-grade level; Sam read independently at the kindergarten level, and Kofi read independently at the first-grade level.

Two of the participants (Sam and Kofi) were diagnosed with an LD in reading in the third grade based on the response-to-intervention (RtI) process. They joined the school from another district with various test results in their files, such as Oral and Written Language Scales (OWLS-II), Woodcock Johnson Tests of Achievement III (Kofi), Learning Accomplishment Profile-Diagnostic Standardized Assessment (LAP-D, and Battelle Developmental Inventory-2 (BDI 2) (Sam). Based on universal screening, RtI (16-week outcomes on CBMs after Tier 2 and Tier 3, and initial instruction in general education setting), they were diagnosed with LD.

Table 2
Students' Demographics, IQ, and STAR[a] Reading Achievement

Student	Age	Gender	Race	Special Education	IQ	STAR Score
John	9.6	M	AA[b]	LD	87	145
Sam[c]	10.7	M	AA	LD	Not tested	197
Kofi	9.9	M	AA	LD	Not tested	233

[a]The STAR reading assessment is used in the school as the universal screening tool three times a year. The expected STAR reading benchmark was 352 at the time of the study; [b]African American; [c]Sam was retained in kindergarten.

The third student (John) was diagnosed with LD in reading based on a discrepancy between his IQ performance and reading achievement on tests conducted by the school district psychometrist, following the RtI process.

A discrepancy between performance on an IQ test and an achievement test in reading, writing, or math is a long-established practice for identifying students with a learning disability in the United States (Zumeta, Zirkel, & Danielson, 2014). Thus, the IQ-achievement discrepancy between an overall IQ score and an achievement score in a particular domain such as reading, writing, or mathematics is assumed to indicate an unexpected underachievement, reflecting a specific learning disability. The specific cut-off scores vary within and across the country (Dean & Burns, 2002).

Debates surrounding the merit of using a discrepancy model led to establishing the RtI as another means of identifying a learning disability (e.g., Kavale, Spaulding, & Beam, 2009; Scanlon, 2013; Tannock, 2013; Zirkel, 2017). When the Individuals With Disabilities Education Act (IDEA) was reauthorized in 2004 (United States Code, 2006), it included statutory language related to identification of a learning disability (Zumeta et al., 2014). States are still required to establish criteria for a learning disability identification, but those criteria neither mandate nor exclude the use of the discrepancy model. Instead, additional criteria permit the use of another process to evaluate a child's learning disability that is based on his or her response to scientific, research-based intervention (Zumeta et al., 2014). Although the IDEA does not specifically use the term *RtI*, it is widely used in the U.S. to denote such a process.

Experimental Design

We employed a single-subject multiple-baseline design across participants (Kazdin, 2010; Kratochwill et al., 2010), with instruction in the written summarization of fictional and nonfictional texts. Assessment included the following dependent variables: total words written (TWW), words spelled correctly (WSC), and correct word sequence (CWS) (Deno, 2003; Deno, Marston, & Mirkin, 1982; Parker, Tindal, & Hasbrouck, 1991). We also included writing in response to a prompt as an additional (albeit non-experimental) indicator of possible improvement in writing (Walker, Shippen, Alberto, Houchins, & Cihak, 2005). In addition, we administered three subtests of the Test of Written Language (TOWL-3; Hammill & Larsen, 1996), Contextual Conventions (CC), Contextual Language (CL), and Story Construction (STC), to assess writing outcomes before and after the intervention.

Dependent variables. To measure the students' potential improvement in writing, we employed curriculum-based measurement (CBM). Developed by Deno and colleagues (Deno et al., 1982; Deno, 2003), CBM provides an alternative to traditional writing assessments. Progress monitoring of writing with CBMs is recommended, especially for students with writing difficulties, as a tool to examine writing growth over time (Espin, Weissenburger, & Benson, 2004).

Dependent variables such as total words written (TWW), words spelled correctly (WSC), and correct word sequence (CWS) have been established as valid and reliable measures of overall writing performance for elementary students with a learning disability as well as students without disabilities (Deno et al., 1982; Parker et al., 1991; Tindal, 2013). TWW indicates the total number of words written regardless of spelling; WSC is used to measure the number of words spelled correctly in the writing sample; and the CWS measure is used to evaluate whether two adjacent words are grammatically, syntactically, and orthographically correct. For example: ^He^is^going^to^the^race^. = 7 correct word sequences, while another example with errors in writing indicates: ^He^is gong to rase. = 2 correct word sequences.

Pretesting and Instructional Materials

Pretest. Students were pretested on the STAR reading assessment used as a universal screening tool in the participating school at the beginning, middle, and end of the school year. Table 1 displays the students' reading scores on the STAR assessment at the time of the study. Pretest results related to writing are presented in baselines in Figures 1, 2, and 3, including data points across participants for Probe 1 and the percentile ranks before the intervention on TOWL-3 (see Table 5).

Instructional materials. Given the increased attention to reading and writing about informational texts in addition to narratives and stories in the Common Core State Standards (Graham & Harris, 2013; National Governors Association & Council of Chief State School Officers, 2010; Shanahan, 2015), we selected a series of short informational and fictional texts that students were to summarize in writing. Informational texts used in the present study addressed the state's science standard related to the personal perspective of science. The texts describe electrical safety, illustrated with various characters. For example, the bride of Frankenstein is depicted putting a fork into a toaster, young Dracula standing in a puddle of water with a hair dryer in his hand, and young Igor flying a kite near a telephone pole. Thus, the characters are depicted in situations to which the students could relate their own life experiences (e.g., do not lift toast out of a toaster using a fork). The students were instructed to summarize for other third graders important points about electrical safety. (For the text samples and characters, see Frankenstein's Lightning Laboratory [2017] at The Atoms Family exhibit at the digital Miami Science Museum: http://oldintranet.puhinui.school.nz/Topics/Science/AtomFamily/frankenstein/index.html.)

The students also summarized fictional texts based on Greek mythology designed as digital books (Source: www.starfall.com, section: *I'm Reading* [Starfall Education Foundation, 2017]). The following stories may be found under "Greek Myths": The Maze, Wings, Pegasus, Midas Touch, The Woman Runner, and The Wooden Horse. The layout of the digital books consists of short texts presented on the right-hand page of the book, with a built-in text-to-speech function providing easy access to information. The characters in the story are presented on the left-hand page. Clicking on the left page sets the characters in motion to illustrate the scenes in the text with added animation.

We also employed a free app, *Dragon Dictation*, as an instructional tool. Nuance Communications, a computer software company, developed the *Dragon Dictation* app based on the computer software *Dragon Naturally Speaking*, which provides speech-to-text features. While the *Dragon Speaking Naturally* software is adaptive (and not free of charge), the *Dragon Dictation* app does not include adaptive features with the student participants, but accurately records dictations provided by the authors of the present study. To illustrate, a student would say: /ma/ /næm/ /z/ … John, while the instructor would pronounce the same sentence as: /mai/ neim/ /iz/… John. We therefore recorded students' thoughts, and their recorded thoughts served as their first drafts.

General Procedures

Upon obtaining approvals from the school principal, the district school board, and the University of Mississippi Institutional Review Board, as well as parental consents and students' assent, we conducted intervention sessions in the special education resource room, three times a week for about 30 minutes each, from the end of January to the beginning of May. While we display graphs with the outcomes measured weekly based on their final drafts (Figures 1 through 3), the total number of intervention sessions ranged from 23 to 24 per student.

Baseline. The baselines were established by engaging each student in reading and summarizing short texts starting with electrical safety and alternated with Greek mythology stories. Baselines were analyzed across TWW, WSC, and CWS to establish a starting point for the intervention for each student. By convention, at least three data points are required to establish dependent measure stability (Kazdin, 2010). We started the instruction after three data points because those did not show variability, and had a downward trend. The first participant achieved an increase of above 20% of his baseline mean across dependent variables prior to implementation of intervention with the second participant; the same criterion was applied to the beginning of the intervention with the third student (Wolery & Dunlap, 2001).

Probes. The first probe was administered before the beginning of the intervention; the second probe was administered at the midpoint of intervention, and the final probe upon the completion of intervention. Because the students read texts in different genres, it was of interest to see whether writing to open-ended prompts would show similar or different trends in comparison to data collected during other phases. The first and the third probes were modeled after Fuchs and Fuchs (2007). We inserted the second probe because it snowed in a state (Mississippi) where children seldom experience snow. It seemed that the probe related to a "snow day" (Probe 2) would constitute a good starting point for narrative writing, implied in the following probes:

(1) One day I entered a dark cave …
(2) When I woke up yesterday morning …
(3) One day I was lost in a forest …

We made sure that we started and ended with similar probes (e.g., "One day …") as opposed to beginning or ending with a probe leading to an event (snow day) that our students actually had experienced (Probe 2: When I woke up yesterday morning …).

Intervention

The intervention was characterized by three main features: (a) explicit instruction in process writing based on the POWER strategy (Englert et al., 1991) through the steps of planning, organizing, writing, editing, and revising; (b) student dictation of the text summarization was recorded on the *Dragon Dictation* app on iPads; and (c) the summarizations recorded on iPads serving as first drafts to be written, edited, and revised, on computers. All students had received instruction in keyboarding and were proficient in writing electronic texts.

The first author/instructor instructed the students in the writing task: The task was for students to learn about electrical safety and to summarize the main points for their third-grade peers. The students also were instructed to read and summarize digital books to be recommended to their teacher. While the first author worked individually with one student at the very beginning, the second author reviewed with the students their previous and current work. The classroom was relatively large, with few students, four computers, and four iPads; consequently, instruction with one student did not interfere with working with other students and vice versa.

After completing the baseline, the students were individually instructed in different texts they were about to read and summarize through the intervention steps as follows.

Modeling.

1. In the first step, the instructor modeled the summarization of the text by first reading the text, followed by think-aloud: "In order to summarize the text, we need to identify the main ideas. The question is: What can we learn about this text on electrical safety? For example, what is the Bride of Frankenstein doing wrong? It looks like she is trying to get her toast out of the toaster with a fork. Is that wrong? I think – yes. Because the fork can take electricity from the toaster and shock her, this is not a good idea. What are the suggestions here on what to do if something is stuck inside the toaster? It says: We need to unplug it and ask parents or adults for help. The main message is here in bold letters: 'Don't stick anything metal inside a toaster or any electrical appliance'."
2. The second step was to review the adapted POWER think sheet consisting of the following steps: plan and record your thoughts on the *Dragon Dictation* app; organize; write; edit, revise. (The instructor reviews the steps.)
3. The third step was to model planning for summarization (planning for a title, introduction, main ideas, and conclusion). The instructor plans a title: "Electrical safety." Introduction: "We need to be careful with electrical appliances." **[**What are some of the main ideas?**]** "For example, if a piece of toast is stuck in a toaster, we should not try to get it out with a fork. If we insert something metal in electrical appliances, we might get an electrical shock. We should first of all unplug the toaster, or try to get help from our parents or other adults." Conclusion: "We should not stick anything metal inside a toaster or any other electrical appliance."
4. The fourth step involved recording the thoughts on the *Dragon Dictation* app on an iPad.
5. The fifth step was to read the recorded thoughts, plan for text organization (modeling through think-aloud), and start writing the second draft on the computer. The instructor changes the third and the fourth sentence to: "It is dangerous to insert a fork into a toaster. It is also dangerous to insert other metal pieces into electrical appliances be-

cause the electricity might shock us. The best thing to do is to unplug the toaster. If we are not sure how to do that, we call our parents or other adults for help."

6. The final step was to edit and revise the text (e.g., the instructor checks the mechanics and usage of language: checks spelling, verb-tense agreement, subject-verb agreement).
7. The instructor examines the overview of the text organization and whether anything is in need of revision.

Instruction. With guided practice to check whether a student understands the steps on the POWER sheet, the procedure involved in recording on the *Dragon Dictation*, and questions to guide summarization, organization, edits, and revisions, instruction consisted of the following steps:

1. The student reads the text.
2. The student reviews the adapted POWER sheet.
3. The student plans the summarization of the text.
4. The summarization is recorded on the *Dragon Dictation* app, making the student's thoughts about the text" under construction" visible to him.
5. The student reads his own thoughts and plans for organization of the text (e.g., plans for introduction, main ideas) before engaging in writing on the computer.
6. The student writes the summary of the text on the computer.
7. The student edits and revises the text and reviews the final draft.

As the *Dragon Dictation* app did not adequately record the students' thoughts, the instructor recorded their thoughts (recorded 99% correctly; a word not recorded adequately was *Icarus*). The students' thoughts were recorded as presented sentence-by-sentence. The students seemed excited and pleased upon seeing their thoughts recorded and spelled correctly. The students never copied the recorded texts verbatim (as indicated by their spelling mistakes in their drafts), but used the recorded text as a basis for planning their summaries. No time limit was imposed on students' writing, but the students never wrote for more than 5 to 8 minutes. By the time reading and recording were completed, there was not enough time to finish all of the steps of the summarization writing task (e.g., editing and revising), the same day, so the activities were resumed in the next sessions.

Post-intervention/maintenance. During the post-intervention and the maintenance session three weeks after the intervention, the students wrote without the help of the *Dragon Dictation* app and the POWER instructional steps, as was the case during the baselines. The first pre-intervention and the last post-intervention summarizations were based on the informational texts about electrical safety. Following the pattern of alternating the texts, as during the baseline, the students summarized a story during the maintenance session. They also wrote in response to a final narrative prompt (Probe 3, very similar to Probe 1). Finally, we evaluated the students' writing based on three subtests of the TOWL-3 standardized test for writing.

Fidelity of Treatment

To make sure that the intervention was delivered as designed, the following procedures were implemented (Horner, Carr, Halle, McGee, Odom, & Wolery, 2005): (a) before the start of the study, the authors reviewed all the instructional steps in detail to establish the same level of understanding of the steps; (b) the authors met before and after each instructional lesson to debrief; and (c) the authors discussed anticipated and unanticipated events (e.g., school events that might interfere with the instruction).

When the special education teacher was present in the classroom, she was not involved in instruction, but served as a fidelity treatment assessor. We instructed her in the design of the study, for example, the necessity of establishing baselines before starting the intervention. Using a fidelity checklist, the teacher and the second author conducted fidelity checks for 20% of the baselines and 25% of the intervention sessions across each participant. The inter-assessor reliability was established at 100% across baselines, and 98% for instructional phases across participants.

Results

In the following, we first provide the visual results of the data. Second, we report the percentage of nonoverlapping data (PND). Third, we report the baseline means, post-intervention, and maintenance data. Fourth, we report the results of writing to a prompt. Fifth, we present the pretest and posttest results on the standardized writing test TOWL-3.

Visual Results

We provide the visual results in Figures 1 through 3. As represented by the post-intervention data and the maintenance data gathered three weeks after the instruction, the writing skills of all three students improved. The probes also indicate an upward trend. Although there are variations within the outcomes, the trend data clearly indicate systematic growth on TWW, WSC, and CWS. (Note one less data point for Kofi, who was absent during the fourth week of instruction.)

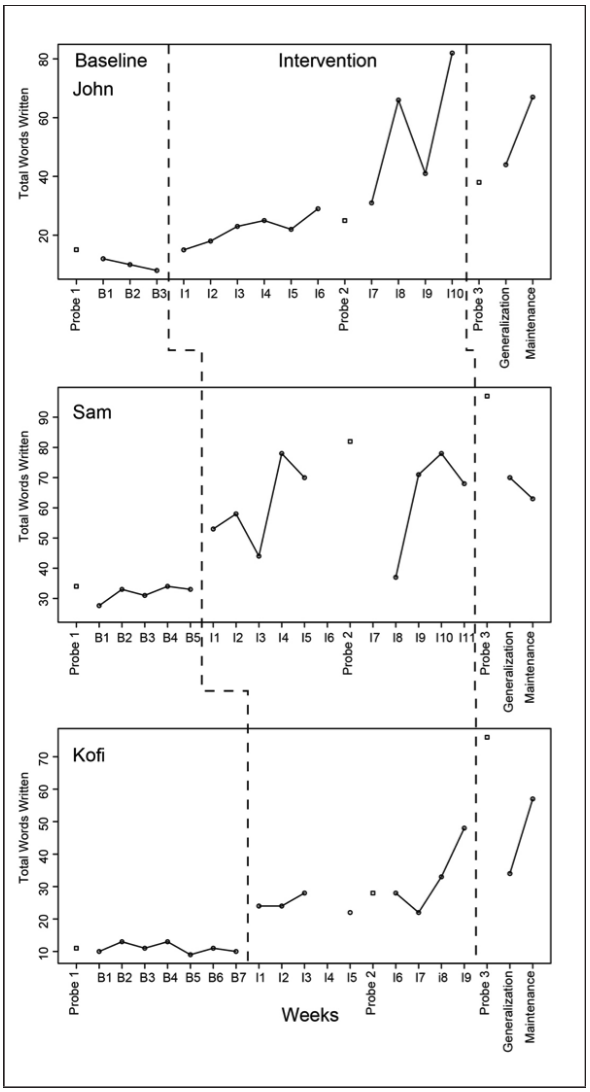


Figure 1. Total words written (TWW).

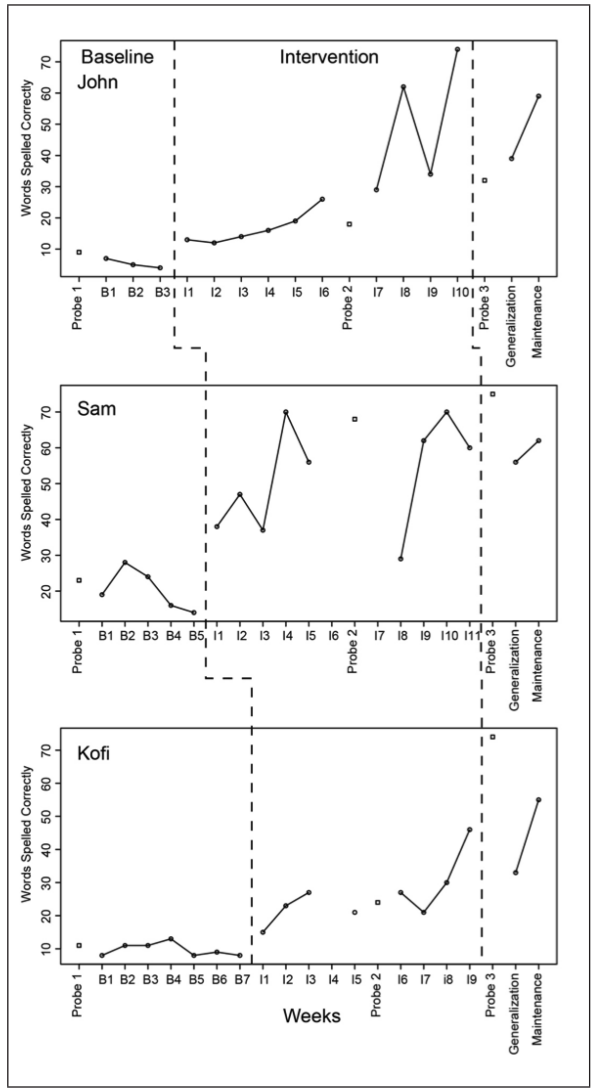


Figure 2. Words spelled correctly (WSC).

Percentage of Nonoverlapping Data (PND)

PND is often used to evaluate the effects of an intervention in single-subject research (Campbell, 2004; Scruggs & Mastropieri, 1998; Scruggs, Mastropieri, & Casto, 1987). Scruggs and Mastropieri (1998) suggested that PND scores of >90% represent highly effective interventions, between 70% and 90%, moderately effective to effective interventions, and between 50% and 70%, questionable interventions. The highest data point in the baseline as well as the percentage of points during the intervention exceeded this level. The analysis of PND on total words written (TWW) indicates that the intervention was highly effective for all three students (100% exceeding baseline); the analysis of words spelled correctly (WSC) indicates that the intervention was also highly effective for all three students (100%);

and, the analysis of correct word sequence (CWS) indicates that the intervention was effective, ranging from 89% for the third student, Kofi, to 100% for the other two students.

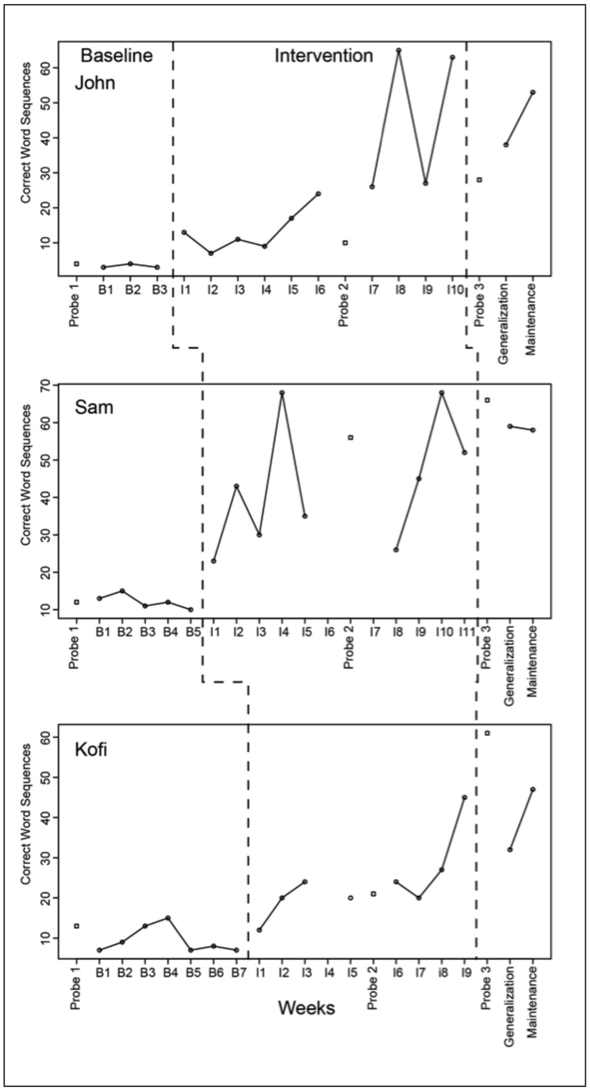


Figure 3. Correct word sequences (CWS).

Baseline Means, Post-Intervention, and Maintenance

In Table 3 we report the means (*M*) of the baseline scores, post-intervention data, and maintenance data three weeks after the intervention across the participants.

Writing to a Prompt

Table 4 presents the writing scores on CBMs across three writing prompts administered at the beginning of the study, at the midpoint of the study, and upon completion of the intervention.

TOWL-3 Results

All three students showed gains on the posttest of Contextual Language (CL); two students showed gains in Story Construction (SC), and two students' scores improved on Contextual Conventions (CC) at posttest. The results expressed in percentiles are summarized in Table 5.

Social Validity

The instructor conducted a brief open-ended interview with each student individually at the end of the intervention. The first question was: "Do you think you have improved as a writer?" The students reviewed some samples of their writing, and it seemed clear to them that their writing improved, as each responded with a resounding: "Yes!" When probed further (e.g., "Why do you think so?"), the students specifically responded as follows:

Kofi: "Now I write ... in more complete sentences." When asked the same question, John seemed impressed by how much he had written: "Look how much I wrote!" During the instruction, John asked for copies of his writing to share those with his father; by comparison, at the beginning of instruction his writing consisted of one sentence.

When asked what they thought helped them to become better writers, the students seemed to appreciate seeing their ideas spelled correctly on the *Dragon Dictation* app. For example, John noted: "You can put your ideas on *Dragon*, and it shows you the words and it shows you how to spell it. I liked *Dragon* the best." Also, Kofi: "*Dragon* with planning [POWER sheet] was helpful;" and "*Dragon* was kind of fun … It teaches you how to write." Similarly, when asked explicitly whether practicing writing with the *Dragon Dictation* helped, Sam responded: "Yes, it shows you the words, and you can put your ideas on *Dragon*."

We also asked the special education teacher to share her impressions of the students' writing skills. After reviewing the students' writing samples, she observed that their writing had improved considerably. She noted that based on her observations, the students seemed to enjoy animated digital books. She also commented that the students seemed to enjoy working with the *Dragon Dictation* app.

Table 3
Baseline Means, Post-Intervention, and Maintenance Across TWW, WSC, and CWS

	Baseline			Post-Intervention			Maintenance		
Student	TWW	WSC	CWS	TWW	WSC	CWS	TWW	WSC	CWS
John	10	5.6	3.3	44	39	38	67	59	53
Sam	30.8	20.8	13.8	70	56	59	63	62	58
Kofi	11.4	10.1	9.7	34	33	32	57	55	47

Note. TWW = total words written; WSC = words spelled correctly; CWS = correct word sequence.

Table 4
Students' Performances Across Three Probes

	Probe One			Probe Two			Probe Three		
Student	TWW	WSC	CWS	TWW	WSC	CWS	TWW	WSC	CWS
John	15	9	4	25	18	10	38	32	28
Sam	34	23	13	82	68	56	97	75	66
Kofi	11	11	13	28	24	21	76	74	61

Note. TWW = total words written; WSC = words spelled correctly; CWS = correct word sequence.

Table 5
TOWL-3 Pretest and Posttest Outcomes Measured in Percentiles

	Pretest %tile			Posttest %ile		
Student	CC	CL	SC	CC	CL	SC
John	16	9	37	63	37	63
Sam	9	1	5	63	25	37
Kofi	50	25	37	50	63	37

Note. CC = Contextual Conventions; CL = Contextual Language; SC = Story Construction.

Discussion

The purpose of the present study was to develop students' writing skills in a reading/writing context. Based on research by Englert and her colleagues (e.g., Englert, 2009; Englert et al., 1991) and Graham and Hebert (2010), we developed a writing intervention focused on explicit instruction, process writing, and dictation. Designing instruction with the *Dragon Dictation* app allowed students' thoughts to become visible to them and made writing less challenging, especially in terms of transcription and text organization. Transcription skills are considered important for young writers (Hayes & Olinghouse, 2015; McMaster, Du, & Pétursdóttir, 2009). Graham and his colleagues (e.g., Gillespie & Graham, 2014; Graham et al., 2015) reported that dictation is an effective intervention in writing for students with LD. The present study confirmed the effectiveness of dictation in writing instruction for the three African American male third graders with LD.

With explicit instruction, modeling, and feedback, the students improved from writing one incomplete sentence at baseline to writing one or two short paragraphs at post-intervention. All three students significantly improved across the CBM measures of total words written (TWW), words spelled correctly (WSC), and correct word sequence (CSW), as indicated by a visual inspection of data. Based on an analysis of the percentage of nonoverlapping data, the effect sizes ranged from a moderate PND of 89% to a large PND of 100%.

The intervention phase shows some variability toward the end of the instruction. Some variability might be related to student absences (e.g., Kofi). Also, the students seemed to perform better on summarizing stories they read than on informational texts. Digital stories had built-in text-to-speech function providing easy access to the texts read. The students seemed to appreciate the added animation in the digital books, which might have increased their motivation to participate in their writing tasks. It also seems, based on their writing to three narrative prompts, that the students were able to transfer their writing summarization skills to writing narratives. Finally, all three students showed gains on the standardized TOWL-3 writing test for the Contextual Language (CL); two students showed gains in Story Construction (SC); and two students' scores improved in Contextual Conventions (CC).

African American students with LD face multiple challenges in acquiring literacy skills (Craig, Zhang, Hensel, & Quinn, 2009). First, as with many other students with LD, they frequently spend little time on planning and revising (Graham & Harris, 2003). Consequently, their writing might lack coherence and clarity (Gillespie & Graham, 2014). The planning sheet with the POWER steps seemed to be helpful. Yet, it was somewhat surprising to see that the students still made spelling mistakes although they were provided with their thoughts spelled correctly on the *Dragon Dictation* app. The recorded texts on the *Dragon Dictation* app nonetheless provided a good base for text organization and coherence. The open-ended interviews with the students following the intervention indicated that they thought they much improved their writing, and reported that the *Dragon Dictation* app was helpful.

Because all three students wrote with spelling mistakes in spite of their thoughts being displayed correctly on *Dragon Dictation*, we analyzed a sample of their writing to examine whether it displayed any features of African American English Vernacular (AAVE) (although such analysis was not specified originally as a research question). The students' writing did display some AAVE features across baselines, intervention, probes, and post-intervention as follows: (a) morphosyntactic features such as the use of preterite "had": *The boy had got on the horse; … they had thought* [*thrown*] *a lot of snowball*[*s*] *at me and then I had got them back;* (b) zero past tense: *"The king wish that everything that he touch was turn to gold." So they all danced and celebrated because they free by* [were freed from] *a wicked king*; (c) reduction of consonants: *"The boy was bored and he went to a wise ole man…"* (AAVE feature also shared with the dialect of the South where the study was conducted); (d) use of auxiliary verbs: "*About mrs. Frakestin* [Mrs. Frankenstein] *was put a fock* [put a fork] *in her torst* [toaster]" (Washington et al., 2014). Toward the end of intervention and during the post-intervention and maintenance sessions, there were fewer inadequate features and, consequently, the outcome measures showed an improvement in writing competence.

As our focus was on writing, we did not measure reading achievement outcomes. We nonetheless believe that our study contributes to intervention studies in reading/writing contexts. Kang et al. (2015) conducted a literature review on integrated reading and writing interventions for students with LD and academic difficulties and identified only 10 studies over a period of 40 years that met their inclusion criteria. In view of potential benefits of connected reading and writing activities (Graham & Hebert, 2010; Shanahan, 2009), it seems important to further explore interventions in reading/writing contexts.

Limitations

Some of the limitations of the current study are inherent in single-subject design studies, in which participants are often not randomly selected, as is the case in the present study. While our study meets multiple criteria for experimental design (Kratochwill et al., 2013) and the visual analysis clearly indicates strong evidence of a causal relation in terms of level, immediacy of the effect, and trend, there is some variability in the data, as discussed earlier (Kratochwill et al., 2013).

Taking into consideration that writing to three probes (at the beginning, middle, and end-point of the study) does not meet multiple-baseline research design standards, we cannot claim that students' improvement in writing of narratives is one of the outcomes of our study beyond the reported improvement on CBM measures. We also had no control over activities taking place in the general education language arts class and the possible impact of those activities on the students' improved writing.

Implications and Future Directions

The challenges that African American students with LD face in developing literacy are not unique to them. The same or similar challenges face students with various disabilities and students whose "culture-specific literacy practices differ from the mainstream" (Washington et al., 2014, p. 217). With a focus on reading and writing activities as connected in the Common Core State Standards (CCSS), we recommend that more educators explore interventions in reading/writing contexts (Shanahan, 2015). Because writing has a positive effect on reading (Graham & Hebert, 2010), interventions in reading/writing contexts might be particularly beneficial for students with LD in reading and writing.

Washington et al. (2014) pointed to converging evidence indicating that speaking an AAVE dialect is negatively related to literacy achievement among children in PreK- 5. However, not all children who speak AAVE are at risk for literacy failure. Children with strong language skills (e.g., strong syntax production in oral language) perform better in reading achievement in first through third grades (Craig, Connor, & Washington, 2003). The findings related to the benefits of strong language oral skills are consistent with studies on second-language acquisition. Thus, the English language learners who exhibit strong language proficiency during their schooling are often those with strong native language skills (Bialystok, 2007).

Although research on writing has expanded in recent decades, research on writing instruction with K-12 students, and especially African American students, has remained limited (Juzwik, Curcic, Wolbers, Moxley, Dimling, & Shankland, 2006). Teachers continue to feel underprepared to teach writing to diverse learners (Graham, Gillespie, & McKeown, 2013; Matsuda, Cox, Jordan, & Ortmeier-Hooper, 2006). The accreditation agencies in the field of teacher preparation recommend that English language arts teachers have knowledge not only about phonology, syntax, and semantics, but also about language use, patterns, and dialects (Council of Chief State School Officers, 2013; Curzan, 2013). At the same time, the linguistic features of dialects are seldom addressed in teacher education programs, and dialects remain a neglected topic in spite of increased attention to culturally responsive pedagogy (Curzan, 2013).

Finally, as an increasing number of elementary schools have LCD projectors in their classrooms along with Internet connections and iPads, there are many possibilities to design supportive and engaging learning environments. A number of studies describe increased student motivation and achievement with various apps (e.g., Hutchison, Beschorner, & Schmidt-Crawford, 2012). While most research on the use of apps in literacy instruction is informative, not many studies meet evidence-based standards (What Works Clearinghouse, 2012). In the present study we used one free app and freely available digital texts. Future studies should further explore the ways in which apps and digital texts may be incorporated in literacy instruction.

References

Bialystock, E. (2007). Acquisition of literacy in bilingual children: A framework for research. *Language Learning, 57*(1), 45-77. doi:10.1111/j.1467-9922.2007.00412.x

Campbell, J. M. (2004). Statistical comparison of four effect-sizes for single-subject design. *Behavior Modification, 28*(2), 234-246. doi:10.1177/0145445503259264

Center for Applied Special Education Technology. (2018). Available at http: www.cast.org.

Council of Chief State School Officers. (2013). *InTASC model core teaching standards and learning progressions for teachers 1:0: A resource for ongoing teacher development.* Washington, DC: Author.

Craig, H. K., Connor, C. M., & Washington, J. A. (2003). Early positive predictors of later reading comprehension for African American students: A preliminary investigation. *Language, Speech & Hearing Services in Schools, 34*(1), 31- 43. doi:10.1044/0161-1461(2003/004)

Craig, H. K., Zhang, L., Hensel, S. L., & Quinn, E. J. (2009). African American English-speaking students: An examination of the relationship between dialect shifting and reading outcomes. *Journal of Speech, Language, and Hearing Research, 52*(4), 839-855. Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/83314

Curzan, A. (2013). Linguistic matters: Resistance and relevance in teacher education. *Language, 89*(1), 1-10. doi:10.1353/lan.2013.0016

Dean, V. J., & Burns, M. K. (2002). Inclusion of intrinsic processing difficulties in LD diagnostic model: A critical review. *Learning Disability Quarterly, 25*(3), 170-176. doi:10.2307/1511300

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184-192. doi:10.1177/00224669030370030801

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48*(4), 368-371.

Englert, C. S. (2009). Connecting the dots in a research program to develop, implement, and evaluate strategic literacy interventions for struggling readers and writers. *Learning Disabilities Research & Practice*, *24*(2), 104-120. doi:10.1111/j.1540-2009.01277.x

Englert, C. S., Raphael, T. E., Anderson, L., Stevens, D. D., & Fear, K. L. (1991). Making writing strategies and self-talk visible: Cognitive strategy instruction in writing in regular and special education classrooms. *American Educational Research Journal, 29,* 337-372. doi:10.2307/1162944

Espin, C. A., Weissenburger, J. W., & Benson, B. J. (2004). Assessing the writing performance of students in special education. *Exceptionality, 12*(1), 55-66.doi:10.1207/s15327035ex1201_5

Fletcher, J. M., Lyon, R. G., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). New York, NY: Guilford.

Fogel, H., & Ehri, L. C. (2000). Teaching elementary students who speak Black English Vernacular to write in Standard English: Effects of dialect transformation practice. *Contemporary Educational Psychology, 25*, 212-235. doi:10.1006/ceps.1999.1002

Frankenstein's Lightning Laboratory. (2017). *Electrical safety*. Retrieved from http://oldintranet.puhinui.school.nz/Topics/Science/AtomFamily/frankenstein/index.html

Fuchs, L. S., & Fuchs, D. (2007). *Using CBM for progress monitoring in written expression and spelling.* Available at http://files.eric.ed.gov/fulltext/ED519251.pdf

Gillespie, A., & Graham, S. (2014). A meta-analysis of writing interventions for students with learning disabilities. *Exceptional Children, 80*(4), 454-473. doi:10.1177/0014402914527238

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing: An Interdisciplinary Journal, 26*(1), 1-15. doi:10.1007/s11145-012-9395-2

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323-344). New York, NY: Guilford.

Graham, S., & Harris, K. R. (2009). Almost 30 years of writing research: Making sense of it all with the *Wrath of Khan*. *Learning Disabilities Research and Practice*, *24*(2), 58-68. doi:10.1111/j.1540-5826.2009.01277.x

Graham, S., & Harris, K. R. (2013). Common Core State Standards, writing, and students with LD: Recommendations. *Learning Disabilities Research and Practice, 28*(1), 28-37). doi:10.1111/ldrp.12004

Graham, S., Harris, K. R., & Beard, K. (in press). Teaching writing to young African American male students using evidence-based practices. *Reading & Writing Quarterly.* doi: 10.1080/10573569.2018.1535775

Graham, S., Harris, K. R., & Santangelo, T. (2015). Research-based writing practices and the Common Core: Meta-analysis and meta-synthesis. *The Elementary School Journal, 115(*4), 498-522. doi:10.1086/681964

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve. Carnegie Corporation Time to Act Report*. Washington, DC: Alliance for Excellent Education.

Graham, S., MacArthur, C. A., & Hebert, M. (Eds.). (2019). *Best practices in writing instruction* (3rd ed.). New York, NY: Guilford.

Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools – A report to Carnegie Corporation of New York.* Washington, DC: Alliance for Excellent Education.

Hammill, D. D., & Larsen, S. C. (1996). *Test of Written Language (TOWL-3).* Austin, TX: Pro-Ed.

Hayes, J. R., & Olinghouse, N. G. (2015). Can cognitive writing models inform the design of the Common Core State Standards? *The Elementary School Journal, 115*(4), 480-497. doi:10.1086/681909

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165-179. doi:10.1177/001440290507100203

Hughes, C. A., Morris, J. R., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities Research & Practice, 32*(3), 140-148. doi:10.1111/ldrp.1242.

Hutchison, A., Beschorner, B., & Schmidt-Crawford, D. (2012). Exploring the use of the iPad for literacy learning. *The Reading Teacher, 66*(1), 15-23. doi:10.1002/TRTR.01090

Juzwik, M., Curcic, S., Wolbers, K., Moxley, K., Dimling, L., & Shankland, R. (2006). Writing into the twenty-first century: A map of contemporary research on writing, 1999-2004. *Written Communication*, *23*(4), 451-476. doi:10.1177/0741088306291619

Kang, E. Y., McKenna, J. W., Arden, S., & Ciullo, S. (2015). Integrated reading and writing interventions for students with learning disabilities: A review of the literature. *Learning Disabilities Research & Practice, 31*(1), 222-33. doi:10.1111/ldrp.12091

Kavale, K. A., Spaulding, L. S., & Beam, A. P. (2009). A time to define: Making the specific learning disability definition prescribe specific learning disability. *Learning Disability Quarterly, 32*(1), 39-48. doi:10.2307/25474661

Kazdin, A. E. (2010). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd.pdf

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38. doi:10.1177/0741932512452794

Mason, L. H., Snyder, K. H., Sukhram, D. P., & Kedem, Y. (2006). TWA+PLANS strategies for expository reading and writing: Effects for nine fourth-gradestudents. *Exceptional Children, 73*(1), 69-89. doi:10.1177/001440290607300104

Matsuda, P. K., Cox, M., Jordan, J., & Ortmeier-Hooper, C. (Eds.). (2006). *Second language writing in the composition classroom: A critical sourcebook*. Boston, MA: Bedford/St. Martin's and National Council of Teachers of English.

McMaster, K. L., Du, X., & Pétursdóttir, A.-L. (2009). Technical features of curriculum-based measures for beginning writers. *Journal of Learning Disabilities*, *42*, 41-60. doi:10.1177/0022219408326212

National Center for Education Statistics. (2012). *The nation's report card: Writing 2011* (NCES 2012–470). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Commission on Writing. (2003). *The neglected "R": The need for a writing revolution.* New York, NY: College Entrance Examination Board.

National Governors Association and Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors. Retrieved from www.corestandards.org

Parker, R., Tindal, G., & Hasbrouck, J. (1991). Progress monitoring with objective measures of writing performance for students with mild disabilities. *Exceptional Children, 58,* 61-73. doi:10.1177/001440299105800106

Rodgers, W., Lewis, T., O'Neill, R., & Vannest, K. (2018). *Policy and position statement on single case research and experimental designs*. Arlington, VA: Council for Exceptional Children, Division for Research.

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology, 100*(4), 879-906. doi:10.1037/0022-0663.100.4.879

Scanlon, D. (2013). Specific learning disability and its newest definition: Which is comprehensive? and which is insufficient? *Journal of Learning Disabilities,46*(1), 26-33. doi:10.1177/0022219412464342

Schumaker, J. B., & Deshler, D. D. (2009). Adolescents with disabilities as writers: Are we selling them short? *Learning Disabilities Research and Practice*, *24*(2), 81-92. doi:10.1111/j.1540-5826.2009.00282.x

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22*(3), 221-242. doi:10.1177/01454455980223001

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. doi:10.1177/074193258700800206

Shanahan, T. (2009). Connecting reading and writing instruction for struggling learners. In G. A. Troia (Ed.), *Instruction and assessment for struggling writers: Evidence-based practices* (pp. 113-131). New York, NY: Guilford.

Shanahan, T. (2015). Common core state standards: A new role for writing. The *Elementary School Journal, 115*(4), 464-479. doi:10.1086/681130

Starfall Education Foundation. (2017). *I'm reading.* Retrieved from www.starfall.com

Tannock, R. (2013). Specific learning disabilities in DSM-5: Are changes for better or for worse? *The International Journal for Research in Learning Disabilities, 1*(2), 2-30.

Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from 1970s to the present. *ISRN Education,* 1-29. doi:10.1155/2013/958530

United States Code. (2006). 20 U.S.C. § 1414(b)(6).

Walker, B., Shippen, M. E., Alberto, P., Houchis, D. E., & Cihak, D. F. (2005). Using *Expressive Writing* program to improve the writing skills of high school students with learning disabilities. *Learning Disabilities Research and Practice, 20*(3), 175-183. doi:10.1111/j.1540-5826.2005.00121.x

Washington, J. A., Patton-Terry, N., & Seidenberg, M. S. (2014). Language variation and literacy learning: The case of African American English. In C. Addison Stone, Silliman, E. R., Ehren, B. J., & Wallach, G. P. (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 204-221). New York, NY: Guilford.

What Works Clearinghouse. (2012). *WWC evidence review protocol for K-12 students with learning disabilities, Version 2.2*. Retrieved from http://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/sld_protocol_v2.2.pdf

Wolery, M., & Dunlap, G. (2001). Reporting on studies using single-subject experimental methods. *Journal of Early Intervention, 24*(2), 85-89. doi:10.1177/105381510102400201

Zirkel, P. A. (2017). RTI and other approaches to SLD identification under the IDEA: A legal update. *Learning Disability Quarterly, 40*(3), 165-173. doi:10.1177/0731948717710778

Zumeta, R. O., Zirkel, P. A., & Danielson, L. (2014). Identifying specific learning disabilities: Legislation, regulation, and court decisions. *Topics in Language Disorders, 34*(1), 8-24.

43rd Annual Conference
Hersonissos, Heraklion, Crete, Greece

The International Academy for Research in Learning Disabilities

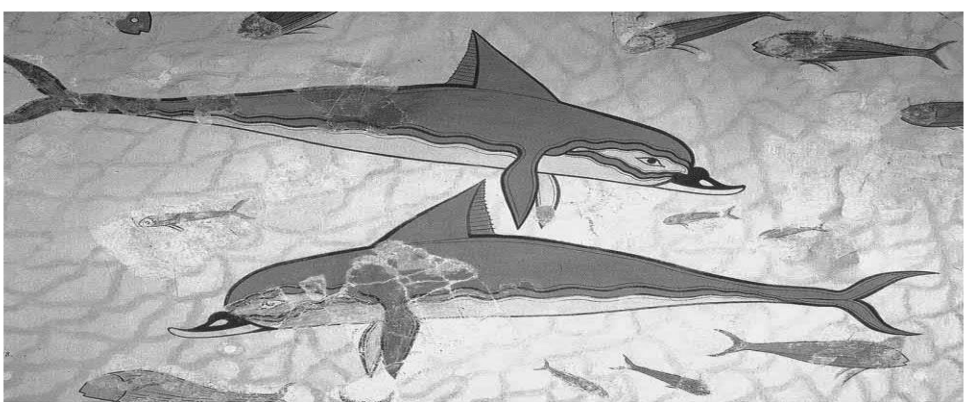
Preconference Activities – June 23, 2019
Conference – June 24 and 25, 2019
Aldemar Knossos Royal, Hersonissos, Crete, Greece

Conference Chairs:
Professor Georgios Manolitsis and Dr. Angeliki Mouzaki
(College of Education, University of Crete)

Cruickshank Memorial Lecture:
by Prof. David L. Share,
Head, Department of Learning Disabilities, Faculty of Education &
Edmond J. Safra Brain Research Center for the Study of Learning Disabilities
University of Haifa

"Climbing out of the Anglocentric, Eurocentric and Alphabetocentric Trenches"

Early Registration Ends April 30, 2019
Register online at https://iarld2019.synedry.com/home.aspx



Guided Tour: June 23, 2019 – City of Heraklion and the Archaeological Museum

For questions, contact:

Conference Chair
Angeliki Mouzaki
amouzaki@edc.uoc.gr

or

Conference Coordinator
Maria Leventi
info@diazoma.net
Tel. +30 2810321494
Mob. +30 6908215112

Manuscript Submission Guidelines

English is used for submissions to the journal, correspondence and publication. All submissions must be formatted consistent with the 6th edition of the *Publication Manual of the American Psychological Association* (APA). Manuscripts must include a 100- to 150-word abstract summarizing the contents.

A critical concern in the learning disabilities field is the definition of the population. Therefore, authors are expected to operationally define the study participants in accordance with professional standards (see CLD Research Committee: Rosenberg et al., 1993. Minimum standards for the description of participants in learning disabilities research. *Learning Disability Quarterly, 26*(4), 210-213). In addition, parameters of the setting in which the research took place are to be clearly delineated. Such descriptions facilitate replication and application of results. Manuscripts that fail to specify participant and setting variables will be rejected or returned to the authors for clarification. Authors of research manuscripts are encouraged to include brief (e.g. one to two sentences) explanations of why specific procedures and analysis methods were employed. Peer reviewers will evaluate the appropriateness of this and all aspects of the reported study.

Manuscripts are not to exceed 35 double-spaced pages, consistently employing a 12-point font (including references, tables, figures and appendices). Please limit tables and figures to those essential to conveyance of your content. Please present figures and tables in portrait format and use grey scale rather than colored images.

The manuscript submission and review process will be conducted electronically. To submit your manuscript digitally, you must use one of the following formats: Microsoft Word (.doc or .docx), RTF or PDF. Manuscripts must be saved as "letter" ("U.S. letter") page length/paper size. Submissions in Word or RTF format will be converted into a PDF before being sent for blind peer review. If you submit a PDF file, it is your responsibility to ensure it can be read and printed by others. To avoid delays, please embed all fonts and use Adobe PDF Distiller instead of PDF Writer to ensure that others can view the article exactly as intended. No email attachment should exceed 15 MB.

Each manuscript must be accompanied by a cover letter that communicates (a) that the manuscript is an original work, (b) that the manuscript is not under consideration by any other journal, and (c) any other disclosures as required by the APA (e.g., ethical treatment of participants, financial relationship disclosures). A cover page that provides the name, affiliation, mailing address, phone number, fax and email address of each author must also accompany each manuscript. All communications will be with the lead author. No author identifying information should be included directly in manuscripts submitted for review (e.g., explicit reference to previous publications, acknowledgments, author biographies). Please send the required cover letter, separate cover page with author identifying information, and manuscript as three separate attachments to a single email to: **ijrld@bc.edu** (please do not send tables, figures, appendices or other supporting materials in separate files). Manuscripts may **not** be submitted by fax or in paper form.

For further information, please refer to:
http://www.iarld.com/wp-content/uploads/2011/08/IJRLD-Call-for-Manuscripts.pdf

International Journal for Research in Learning Disabilities
Lynch School of Education, Boston College
140 Commonwealth Avenue
Chestnut Hill, MA 02467
USA

The IARLD (International Academy for Research in Learning Disabilities) is an international professional organization dedicated to conducting and sharing research about individuals who have learning disabilities.

The IARLD is an elected group of premier scientists, educators and clinicians in the field of learning disabilities throughout the world. The Academy was formed in 1976 by Dr. William Cruickshank (United States of America) and Dr. Jacob Valk (The Netherlands), meeting in Canada with the intention of providing a forum for the exchange of information and the advancement of knowledge regarding learning disabilities.

Since its inception, the Academy has realized its mission of being a professional, international, interdisciplinary consortium of scientists. The Academy currently has a membership of nearly 200 distinguished scholars, representing 26 different countries and thirty disciplines.

IARLD members represent:

- distinguished researchers,
- distinguished practitioner/clinicians,
- young researchers, and
- promising doctoral students.

www.iarld.com

ISSN: 2325-565X (print)
ISSN: 2329-3764 (online)