International Journal for Research in Learning Disabilities
formerly Thalamus
journal of the International Academy for Research in Learning Disabilities

David Scanlon, Editor, Boston College, USA
Kirsten McBride, Copy Editor, Overland Park, Kansas, USA

Editorial Review Board
James Chapman, New Zealand
Cesare Cornoldi, Italy
Merith Cosden, United States
Gad Elbeheri, Egypt
Linda Elksnin, United States
Paul Gerber, United States
Pol Ghesquière, Belgium
Carol Goldfus, Israel
Lorraine Graham, Australia
Steve Graham, United States
Michael Grosche, Germany
Daniel Hallahan, United States
Ian Hay, Australia
Charles Hughes, United States
Olga Jerman, United States
Asha Jitendra, United States
Sunil Karande, India
Lynda Katz, United States
Barbara Keogh, United States
Youn-Ock Kim, South Korea
Che Kan Leong, Canada
Joseph Madaus, United States
Lynn Meltzer, United States
Ana Miranda, Spain
Maria Chiara Passolunghi, Italy
Henry Reiff, United States
Georgios Sideridis, Greece
Linda Siegel, Canada
H. Lee Swanson, United States
Rosemary Tannock, Canada
Masayoshi Tsuge, Japan
Annmarie Urso, United States
Delinda van Garderen, United States
Christa van Kraayenoord, Australia;
Judith Wiener, Canada

2015 IARLD Executive Board
Christa van Kraayenoord, President,
University of Queensland
Linda Siegel, President-Elect,
University of British Columbia
Judith Wiener, Immediate Past President,
University of Toronto
Margaret Pierce, Treasurer,
Stonehill College
Douglas Dexter, Treasurer,
Pennsylvania State University
Evelyn Johnson, Secretary,
Boise State University
Asha Jitendra, Vice President for Fellows,
University of Minnesota
Anya Evmenova, Vice President for Members,
George Mason University
Henry B. Reiff, Vice President for Student
Members, McDaniel College
Malka Margalit, Vice President for International
Development, Tel Aviv University
Lynn Meltzer, Chair of Conference Programs,
Research Institute for Learning and
Development; Harvard University
Brian Bryant, Academy Historian,
University of Texas, Austin
Annmarie Urso, Publications Chair,
State University of New York, Geneseo
Delinda van Garderen, By-Laws and Constitution
Committee Chairperson,
University of Missouri

Members at Large
Rasa Barkauskiene, Vilnius University
James Chapman, Massey University
Marija Kavkler, University of Ljubljana
# Table of Contents

The Effects of Blended Text-Processing and Linguistic Comprehension Interventions Among Struggling Middle-School Readers  
By Amy E. Barth, Sharon Vaughn, and Elisabeth V. McCulley  

Improving Reading Comprehension Using Digital Text: A Meta-Analysis of Interventions  
By Sheri Berkeley, Leigh Ann Kurz, Andrea Boykin, and Anya S. Evmenova  

Use of the Randomization Test in Single-Case Research  
By Matthias Grünke, Richard T. Boon, and Mack D. Burke  

Developing a Comprehensive Mathematical Assessment Tool to Improve Mathematics Intervention for At-Risk Students  
By Jonathan Brendefur, Evelyn S. Johnson, Keith W. Thiede, Everett V. Smith, Sam Strother, Herbert H. Severson, and John Beaulieu  

The Role of Mentoring in Fostering Executive Function, Effort, and Academic Self-Concept  
By Lynn Meltzer, Surina Basho, Ranjini Reddy, and Katelyn Kurkul
Reading comprehension is an essential academic skill (Nash & Snowling, 2006; National Reading Panel, National Institute of Child Health and Human Development, 2000). Yet, among students in the eighth grade, approximately 64% of all students and 91% of students with disabilities do not read at proficient levels (National Center for Education Statistics [NCES], 2013). This suggests that when reading grade-level texts, a large percentage of middle-grade readers are not able to accurately connect important ideas in text, form inferences that integrate information in text with general knowledge of the topic, and synthesize common ideas across various texts (NCES, 2013). These data highlight the need for intensive reading interventions that explicitly teach middle-grade struggling readers how to comprehend grade-level texts and acquire content knowledge from the texts they read.

**Theoretical Explanations for Reading Failure**

In the reading comprehension literature, two classes of models have been proposed to explain how readers comprehend text. One class – component skills models – hypothesizes
that a set of reading component skills underlies reading comprehension. For example, the Simple View of Reading (SVR; Gough & Tunner, 1986; Hoover & Gough, 1990) hypothesizes that reading comprehension is the product of word reading and linguistic comprehension. According to the SVR, word reading retrieves semantic information at the word level, and linguistic comprehension then uses this semantic information to derive sentence- and discourse-level interpretations when listening or reading (Hoover & Gough, 1990). Within this model, both word reading and linguistic comprehension are necessary for comprehension to occur. However, by middle school, linguistic comprehension is the largest determinant of reading comprehension (Adlof, Catts, & Little, 2006; Catts, Hogan, & Adlof, 2005).

A second class of models – process models – suggests that reading comprehension is an iterative and dynamic process whereby the reader integrates information within text and between text and general knowledge to form a coherent mental representation of the situation described (Graesser, Singer, & Trabasso, 1994; McKoon & Ratcliff, 1992; McNamara & Magliano, 2009; van den Broek, Young, Tzeng, & Linderholm, 1999; van Dijk & Kintsch, 1983; Zwan & Radvansky, 1998). As such, process models suggest that a coherent representation of text is established by engaging memory-based and constructionist cognitive processes. Memory-based processes are fast acting and translate print into a literal representation of text that is easily accessible in working memory. Constructionist processes are strategic in nature and are engaged in by the reader to improve comprehension if the literal representation of text is not sufficient or does not meet the reader’s goal for reading (van den Broek, 2005). Process models suggest that linguistic comprehension skills (e.g., inference making, comprehension monitoring, and word and world knowledge) are not only essential for constructing a literal representation of text but are particularly important when the reader’s goal is to learn from text (Cain & Oakhill, 1999; Graesser, Singer, & Trabasso, 1994; Kintsch, 1998; van den Broek, 1990; 1997).

**Linguistic Comprehension as a Mechanism for Improving Reading Comprehension**

Although product and process models of reading comprehension represent different levels of explanation and investigation (Barnes, Ahmed, Barth, & Francis, in press), both suggest that linguistic comprehension (i.e., translation of semantic information to derive sentence and discourse interpretations) is essential for reading comprehension. The importance of linguistic comprehension is also supported by latent modeling of the SVR, indicating that listening comprehension and reading comprehension represent a unitary construct among middle-grade readers (Adlof et al., 2006). Further, more complex models of reading comprehension that have blended component- and process-oriented skills suggest that linguistic comprehension processes such as inference making, strategy use, and word and world
knowledge directly as well as indirectly impact comprehension among middle- and high-school students (Ahmed et al., 2014; Cromley & Azevedo, 2007).

These cross-sectional data suggest that listening to language for the purpose of comprehending oral discourse and reading language for the purpose of understanding text are highly inter-related skills. In both instances, semantic information is used to derive sentence- and discourse-level interpretations (Hoover & Gough, 1990). In the case of listening, acoustic-based, semantic information is received through the ear and is used to understand oral discourse; in the case of reading, graphic-based, semantic information is received through the eye and is used to understand written discourse (Hoover & Gough, 1990). Because listening and reading comprehension both require efficient access, retrieval, and integration of semantic information to derive sentence and discourse interpretations, interventions that affect listening comprehension should result in significant gains in reading comprehension and vice versa (Clarke, Snowling, Truelove, & Hulme, 2010; Gilliam, Gilliam, & Reece, 2012; Hulme & Snowling, 2011; Stuart, Stainthorp, & Snowling, 2008). However, little intervention research has explicitly targeted listening comprehension as a mechanism for improving reading comprehension among older struggling readers.

Recent Syntheses of Adolescent Reading Intervention Research

Although listening comprehension among middle-grade struggling readers has not been specifically targeted, several recent syntheses have reported on reading practices for older struggling readers (grades 4-12) (Edmonds et al., 2009; Elleman, Lindo, Morphy, & Compton, 2009; Hall, 2015; Kamil et al., 2008; Scammacca, Roberts, Vaughn, & Stuebing, 2013; Solis et al., 2012; Wanzek et al., 2013). Collectively, these syntheses note small-to-moderate effects of intervention favoring students in the treatment conditions on proximal measures related to the intervention and researcher-developed and standardized measures of reading comprehension. In addition to strategy-based approaches, older struggling readers consistently benefited from instructional practices that explicitly taught readers how to (a) access or build word and world knowledge; (b) formulate main ideas or summaries of text; and (c) actively engage in text-based discourse. Further, a recent synthesis of the effects of explicitly teaching inference making revealed moderate-to-high effects of measures of inference making and standardized measures of reading comprehension (Hall, 2015).
Listening Comprehension as an Intervention Target

A limited body of literature has directly examined the effect of explicitly teaching listening comprehension and oral language discourse on the language and reading comprehension of struggling readers. Among existing studies, Clarke and colleagues (2010) demonstrated that 20 weeks of oral language training was more effective than text-comprehension training or combined text comprehension and oral language training at improving later reading comprehension performance among children age 8-9 years old with specific reading comprehension difficulties. The oral language intervention targeted expressive language and listening comprehension through conversation between children and a tutor. Students were explicitly taught vocabulary, read and discussed narrative texts, and used figurative language.

Fricke, Bowyer-Crane, Haley, Hulme, and Snowling (2013) demonstrated that a 30 week oral language intervention significantly improved the oral language skills and spoken narrative skills of preschool children and led to significant improvements on a standardized assessment of reading comprehension administered six months post treatment. However, a more recent adaptation of this oral language intervention among 6-year old children with dyslexia failed to demonstrate significant effects on oral language and reading comprehension following nine weeks of intervention (Duff et al., 2014).

In summation, listening comprehension interventions targeting preschool children at risk for reading failure or early-elementary-grade children with specific reading comprehension difficulties have been associated with positive gains on measures of language comprehension and reading comprehension following 20-30 weeks of instruction but not for 9 weeks of instruction. However, there are relatively few of these types of studies, and the effects of these interventions on reading comprehension in later elementary or among middle-grade struggling readers remain uninvestigated.

Summary

In summary, an overview of several recent syntheses of effective practices for older struggling readers and the existing literature on the impact of explicitly targeting listening comprehension revealed that struggling readers in the middle grades can benefit from interventions that include the following instructional practices: (a) accessing or building word and world knowledge; (b) generating inferences within text and between text and general knowledge; (c) formulating main ideas or summaries of text; and (d) engaging in text-based discourse. In addition, systematic and explicit use of listening comprehension and oral language discourse may support and increase the efficacy of these practices. While word and
world knowledge, summarizing text, inference making, and text-based discourse have been examined in previous interventions, the research line summarized in the following differs from other studies in the purposeful use of listening comprehension and oral language discourse to scaffold these activities in support of improved reading comprehension among middle-grade struggling readers.

Listening Comprehension as an Intervention Target
Among Middle-Grade Struggling Readers

Through funding from the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100013 to The University of Texas at Austin as part of the Reading for Understanding Research Initiative, two intervention studies have been conducted with the goal of generating empirical evidence about the use of listening comprehension as a mechanism for impacting reading comprehension among middle-grade struggling readers. For both studies, the following research question was addressed: What are the effects of a text-processing reading comprehension intervention that targets listening comprehension through text-based discussions of grade-level informational texts on the vocabulary, inferencing, listening comprehension, and reading comprehension performance of middle-grade struggling readers? Both studies hypothesized that explicit practice in listening comprehension and oral language discourse around text would (a) build up the language processes that restrict middle-grade struggling readers’ ability to synthesize semantic information to form the central idea of connected text and (b) lead to improved inference making, listening comprehension, and reading comprehension.

Method and Results

Study 1

Participants and screening procedures. Participants were drawn from one middle school located in the southwestern region of the United States. Students were eligible for the study if they performed at or below one half of a standard error of measurement above the passing score on the Reading State of Texas Assessment of Academic Readiness (STAAR; Texas Education Agency, 2012) in the previous year. Students consenting to participate were randomized in a 1:1 ratio. The final sample included 59 middle-grade students (n = 30 treatment; n = 29 control), with an average age of 14.85 years. Participants were 76% Hispanic, 22% White, and 2% other. Approximately 26% of students were identified as English as a second language (ELLs) and 5% as receiving special education.
The Effects of Blended Text-Processing and Linguistic Comprehension Interventions by Amy E. Barth, Sharon Vaughn, and Elisabeth V. McCulley

**Intervention.** Students in the treatment condition received approximately 28 hours of intervention delivered in small groups of 3-5 students. Key components of the treatment included (a) accessing and building background knowledge through explicit instruction and repeated exposure to the target words throughout each unit and across multiple units; (b) explicit practice in the formation of inferences that required integration of information within text as well as between text and general knowledge; (c) summarization of text that required students to identify key words and important details in text and then integrate this information into a concise summary that was shared orally and received targeted feedback from the tutor; and (d) practice answering overarching questions on the unit’s content that required students to integrate information across the unit’s texts.

The intervention explicitly and systematically used oral language discourse and listening comprehension to scaffold reading comprehension. All texts were read orally to students. Tutors then engaged students in discussions about the text and scaffolded the summarization of text. Oral responses made transparent student retrieval and integration of information in text and relevant background knowledge, which provided tutors access to the students’ comprehension process. If a student’s oral summary was incomplete or in some way incorrect, the tutor directed the student back to the text in order to identify correct information, helped the student to retrieve relevant background knowledge, and assisted the student in reprocessing relevant information in order to derive a concise and correct summary of the text.

**Results.** Main-effects analyses using pretest as a covariate (ANCOVA) were conducted on measures of vocabulary, reading comprehension, inference making, and language comprehension. These findings are reported fully in McCulley (2015). Results yielded significant effects on the Curriculum-Based Measure-Vocabulary, $F(1, 52) = 8.21, p < .01, d = .78$.

No statistically significant effects were found on unstandardized or standardized measures of reading comprehension, inference making, or language comprehension, although the adjusted means favored the treatment condition on eight of nine measures: Woodcock-Johnson III-Passage Comprehension subtest (WJ-III, Woodcock, McGrew, & Mather, 2001), $F(1, 52) = .92, p > .05, d = .26$; STAAR-Reading, $F(1, 52) = 2.64, p > .05, d = .44$; Curriculum-Based Measure-Summarization of Text, $F(1, 52) = 1.17, p > .05, d = .29$; Curriculum-Based Measure-Inference, $F(1, 56) = .07, p > .05, d = .07$; Test of Language Competence-Expanded Edition Listening Comprehension, Making Inferences subtest, (TLC; Wiig & Secord, 1998) ($F(1, 52) = 2.48, p > .05, d = .43$; Clinical Evaluations of Language Fundamentals-5 Formulating Sentences subtest (CELF-5; Semel, Wiig, & Secord, 2013), $F(1, 52) = 1.09, p > .05, d = .28$; Clinical Evaluations of Language Fundamentals-5 Recalling Sentences subtest, $F(1, 52) = .94, p > .05, d = .26$; Woodcock-Johnson III-Oral Comprehension subtest, $F(1, 52) < .001, p > .05, d = 0$. 

Table 1

Demographics for Study 1

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hispanic</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>Caucasian</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Special Education</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>English as a Second Language</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Reduced-Price/Free Lunch Status</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Grade</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Male</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 2

Treatment Effects on Outcome Measures for Study 1

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
<th>Cohen's $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted $M$</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLC-Inference</td>
<td>7.71</td>
<td>7.34</td>
<td>1.97</td>
</tr>
<tr>
<td>CELF-5 FS</td>
<td>8.57</td>
<td>8.38</td>
<td>3.35</td>
</tr>
<tr>
<td>CELF-5 RS</td>
<td>7.76</td>
<td>7.59</td>
<td>1.80</td>
</tr>
<tr>
<td>WJ-III OC</td>
<td>95.01</td>
<td>95.48</td>
<td>11.96</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WJ III-PC</td>
<td>85.75</td>
<td>84.38</td>
<td>11.86</td>
</tr>
<tr>
<td>STAAR</td>
<td>1515.20</td>
<td>1511.59</td>
<td>67.96</td>
</tr>
<tr>
<td>Proximal Measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBM-Summary</td>
<td>12.80</td>
<td>12.93</td>
<td>7.11</td>
</tr>
<tr>
<td>CBM-Inference</td>
<td>3.56</td>
<td>3.55</td>
<td>1.96</td>
</tr>
</tbody>
</table>

Notes. TLC-Inference = Test of Language Competence, Listening Comprehension, Making Inferences; CELF-5 FS = Clinical Evaluation of Language Fundamentals-Formulating Sentences; CELF-5 RS = Clinical Evaluation of Language Fundamentals-5 Recalling Sentences; WJ III-OC = Woodcock Johnson-III Oral Comprehension; WJ III-PC = Woodcock Johnson III-Passage Comprehension; STAAR = State of Texas Assessment of Academic Readiness; CBM-Summary = Curriculum-Based Measure of Summary; CBM-Inference = Curriculum-Based Measure of Inference; CBM-Vocabulary = Curriculum-Based Measure of Vocabulary.

* $p < .05$. ** $p < .01$. 
Study 2

Participants and screening procedures. Students were drawn from three middle schools located in the mid-western region of the United States. Students were eligible for the study if they performed at or below Basic on the Missouri Assessment Program – Reading Test (MAP; Missouri Department of Elementary and Secondary Education, 2014). Students consenting to participate were randomized in a 2:1 ratio. The final sample consisted of 134 middle-grade students (n = 83 treatment; n = 51 control), with an average age of 12.90 years. Participants were 84% White, 9% African American, 3% Hispanic, and 5% Other. No students were identified as ELLs; 31% were identified as receiving special education.

Intervention. Students in the treatment condition received approximately 17 hours of intervention delivered in small groups of 4-6 students. Key components of the treatment included (a) identifying key words and main ideas through text-based discourse; (b) synthesizing information within a single text for summarization and making inferences; and (c) integrating information across multiple texts. Texts gradually increased in length (several sentences to several paragraphs). Tutors gradually released targeted corrected feedback to improve the quality of students’ main idea statements.

The intervention explicitly and systematically used listening comprehension and oral language discourse around text to scaffold students’ comprehension of grade-level expository texts. All texts were read aloud either by the tutor or a student. Tutors then engaged students in discussions about the text in order to check for understanding, identify key words in text, and identify relevant information to include in a brief summary of the text. Oral responses made transparent how accurately students retrieved and integrated information in text and integrated this text-based information with relevant background knowledge on the topic. If a student’s oral summary was incomplete or inaccurate, the tutor directed the student back to the text in order to identify relevant information and helped the student to access, retrieve, and integrate relevant background knowledge with this text-based information.

Results. Main-effects analyses using pretest as a covariate (ANCOVA) were conducted on measures of reading comprehension, inference making, language, and recall of vocabulary. Results yielded significant effects in terms of linguistic comprehension. These findings are reported fully in Barth et al. (in press). Specifically, we found significant treatment effects on the Test of Language Competence-Reasoning, $F(1, 119) = 5.34, p = 0.023, \eta^2 = .043, Hedges g = .33$, and Curriculum-Based Measure-Vocabulary, $F(1, 131) = 7.00, p = .009, \eta^2 = .051, g = .39$. We also found significant effects on the Curriculum-Based Measure-Key Word and Main Idea $F(1, 125) = 6.36, p = .013, \eta^2 = .048, g = .45$. No statistically significant differences were found on standardized measures of reading comprehension (i.e., Woodcock Johnson III-Passage Comprehension, $F(1, 125) = .062, p = .804, \eta^2 = .002, g = -.06$, and Gates MacGinitie Reading
Test (Gates, 2000), $F(1, 124) = .329, p = .567, \eta^2_p = .001, g = .00$) or listening comprehension (i.e., Woodcock Johnson III-Listening Comprehension, $F(1, 126) = .084, p = .773, \eta^2_p < .001, g = .03$). After applying the Benjamini-Hochberg correction to control for Type I error, the Test of Language Competence-Reasoning did not remain significant.

**General Discussion**

Overall, these two studies provide preliminary support for integrating listening comprehension and oral language discourse around text into text-based reading comprehension interventions for struggling readers in the middle grades. Across both intervention studies, the content of the intervention and learning goals were chosen to reflect our interest in understanding whether improvements in listening comprehension and oral language discourse around grade-level informational texts would lead to improved listening comprehension and reading comprehension. As illustrated, we found small-to-moderate effects of the intervention on skills explicitly modeled and practiced in the intervention such as understanding target vocabulary and synthesizing concise main ideas or summaries of text.

We hypothesized that improvements in the linguistic skills targeted in the intervention would generalize to students’ general listening and reading comprehension. However, significant effects on proximal measures closely aligned to the intervention did not transfer to more global standardized measures of comprehension. In both studies, the interventions utilized informational science texts to facilitate oral language discourse. Intervention texts and the oral language discourse surrounding these texts did not include the general topics, narrative text structure, or literal question types and formats (i.e., multiple choice or short answer) that are frequently used to probe understanding on standardized assessments of listening or reading comprehension. Although effects were not statistically significant on standardized measures of listening or reading comprehension, the adjusted means generally favored the treatment condition following a small number of intervention hours (i.e., 30 hours in Study 1 and 17 hours in Study 2). Our interpretation of this trend is that there is a practical effect of improved linguistic comprehension on general listening and reading comprehension, but the effect is small. Further, both studies were underpowered to detect small, significant effects.

We consider the findings of Study 1 and 2 to be promising given that the interventions lacked substantial practice (greater than 75 intervention sessions). Substantial practice is likely required to facilitate transfer to global measures of comprehension due to the complex nature of linguistic comprehension. Thus, linguistic comprehension requires the ability to efficiently integrate semantic information into sentence- and discourse-level translations. This semantic information includes knowledge of word forms (i.e., grammatical class, spellings, and
The Effects of Blended Text-Processing and Linguistic Comprehension Interventions by Amy E. Barth, Sharon Vaughn, and Elisabeth V. McCulley

pronunciations) as well as meanings (Perfetti, 2007). Effective practice (i.e., reading) is then required to learn how to accurately and quickly engage this information in order to understand the central message of a text, form inferences, and make generalizations across texts. Because struggling readers read significantly less text than their typically developing peers (Foorman et al., 2006; Kuhn & Schwanenflugel, 2009), their process of accessing, retrieving, and integrating semantic information into larger meaningful units is both less accurate and less efficient, even when struggling readers in the middle grades have the requisite knowledge base to do so (Barnes et al., 2015; Barth, Barnes, Francis, Vaughn, & York, 2015). For this reason, significant effects on proximal measures following only 17-30 hours of intervention are promising.

Limitations and Future Research

**Limitations.** Results of Study 1 and 2 are subject to several limitations. First, the comparisons are underpowered by the small sample size. Second, samples differed across the studies. Selection criteria for participation as well as the sample demographics varied considerably, with Study 1 including a significantly larger number of ELLs than Study 1 (26% vs. 0%). Third, although significant effects were found on proximal measures closely aligned to the interventions, those measures differed across studies both in nature and psychometric properties. Finally, the intervention was of short duration, limiting the amount of practice available to students for using listening comprehension and oral language discourse as a method of building general listening and reading comprehension.

**Future research.** First, although vocabulary was not a major focus of either study, students quickly learned target words through multiple exposures to the words in text and in oral discourse. Future research is needed to understand how implicit instruction may be used to facilitate development of other language targets. Second, Study 1 included a large sample of ELL students. Future research may explore how various subgroups of students such as ELLs or students with learning disabilities respond to interventions that use linguistic comprehension as a scaffold to support comprehension. Third, both studies represent multicomponent interventions; therefore, it is not possible to determine the relative impact of the individual components. Future research should isolate the particular effects of various components in order to gain an understanding of whether components are active or inactive in facilitating comprehension among middle-grade struggling readers. Fourth, the results of the two studies suggest that instruction using oral language discourse and listening comprehension as instructional scaffolds holds promise for improving general listening and reading comprehension. Future research is required to understand whether substantial practice leads to transfer on general comprehension measures. Finally, several recent randomized control trials demonstrate small-to-moderate gains on standardized measures of comprehension (Edmonds et al., 2009; Elleman et al., 2009; Hall, 2015; Kamil et al., 2008; Scammacca et al.,
Future research is needed to understand whether blending linguistic comprehension into these well-conceptualized text-based approaches facilitates greater transfer to global measures of comprehension.

Table 3

Demographics for Study 2

<table>
<thead>
<tr>
<th></th>
<th>Control N = 51</th>
<th>Treatment N = 83</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Caucasian</td>
<td>45</td>
<td>67</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Special Education</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>English as a Second Language</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Reduced-Price/Free Lunch Status</td>
<td>38</td>
<td>63</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Male</td>
<td>22</td>
<td>49</td>
</tr>
</tbody>
</table>

Table 4

Treatment Effects on Outcome Measures for Study 2

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
<th>Hedge’s g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted M</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Listening Comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WJ III-OC</td>
<td>97.66</td>
<td>97.51</td>
<td>9.06</td>
</tr>
<tr>
<td>TLC-Reasoning</td>
<td>100.09</td>
<td>101.46</td>
<td>13.38</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WJ III-PC</td>
<td>90.72</td>
<td>91.19</td>
<td>8.90</td>
</tr>
<tr>
<td>GM</td>
<td>481.35</td>
<td>482.87</td>
<td>23.88</td>
</tr>
<tr>
<td>Proximal Measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>13.29</td>
<td>13.39</td>
<td>5.28</td>
</tr>
<tr>
<td>Key Word and Main Idea</td>
<td>6.74</td>
<td>6.73</td>
<td>2.57</td>
</tr>
</tbody>
</table>


* p < .05. ** p < .01.
The Effects of Blended Text-Processing and Linguistic Comprehension Interventions by Amy E. Barth, Sharon Vaughn, and Elisabeth V. McCulley

References


The Effects of Blended Text-Processing and Linguistic Comprehension Interventions by Amy E. Barth, Sharon Vaughn, and Elisabeth V. McCulley


This research was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant R305F100013 to the University of Texas-Austin as part of the Reading for Understanding Research Initiative as well as K08 HD068545-01A1 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD, the National Institutes of Health, or the Institute of Education Sciences, U.S. Department of Education.
Improving Reading Comprehension Using Digital Text: A Meta-Analysis of Interventions

Sheri Berkeley
Leigh Ann Kurz
Andrea Boykin
Anya S. Evmenova
George Mason University

Abstract

Much is known about how to improve students’ comprehension when reading printed text; less is known about outcomes when reading digital text. The purpose of this meta-analysis was to analyze research on the impact of digital text interventions. A comprehensive literature search resulted in 27 group intervention studies with 16,513 participants. The overall weighted mean effect size for interventions designed to provide basic access to text was small ($ES = -.03$, range $-.49$ to $1.18$), whereas a moderate effect size was obtained for interventions that served as instructional enhancements for digital text ($ES = .51$, range $-.35$ to $1.57$). These findings were consistent across grade level (elementary vs. secondary) and student type (with disabilities vs. without).

Comprehension, or constructing meaning from text, is the ultimate goal of reading and vitally important for student success in school and later life. Providing appropriate interventions to ensure that all students comprehend what they read is more challenging than ever before because general education classrooms consist of students with a wide range of learning needs, including students with learning disabilities (LD; Hock, Schumaker, & Deshler, 1999; Parsons, Dodman, & Burrowbridge, 2013).

An LD is a neurological disorder that affects the brain’s ability to receive, process, store, and respond to information (National Center for Learning Disabilities, 2014). According to the Diagnostic and Statistical Manual-5th Edition (DSM-5; American Psychiatric Association...
Improving Reading Comprehension Using Digital Text by Sheri Berkeley, Leigh Ann Kurz, Andrea Boykin, and Anya S. Evmenova

In order to be diagnosed with a specific learning disorder, a student must meet four criteria: (a) difficulty learning academic skills for at least six months despite academic intervention, (b) academic performance that is below expectation based on the student’s age, (c) difficulties must be apparent in the early years or when higher-level skills are demanded in school, and (d) the learning difficulties are not due to intellectual disabilities, language differences, lack of appropriate instruction, visual or hearing impairments, psycho-social disorders, or mental disorders. In addition, LD occurs in students who have normal intelligence and may also occur in gifted students.

Due to these deficits, students with LD often have problems with several facets of academics, including reading, and may struggle with word decoding, reading fluency, and comprehension. Nevertheless, over half of this population spends 80% or more of the school day in general education settings (U.S. Department of Education [USDOE], 2012) where, especially in the upper grades, students are expected to learn independently from text. However, 33% of fourth graders and 24% of eighth graders in the United States perform below a basic level of reading indicating that they have only partial mastery of the prerequisite knowledge and skills needed to be proficient with grade-level work (National Center for Education Statistics, 2011). Additionally, many students with LD are not performing well at even the most basic reading levels (Denton, Wexler, Vaughn, & Bryan, 2008).

A body of research has shown that reading comprehension strategies can help these students understand more of what they read by helping them prevent or repair problems with comprehension they encounter while reading (see reviews by Gajria, Jitendra, Sood, & Sachs, 2007; Gersten, Fuchs, Williams, & Baker, 2001). Such strategies help students actively engage in behaviors that promote reading comprehension, such as activating background knowledge, identifying text structure, vocabulary word learning, visualizing, self-questioning, and summarizing. Effect sizes from meta-analyses of this research literature have been consistently large (see Berkeley, Scruggs, & Mastropieri, 2010; Edmonds et al., 2009; Swanson, 1999), and best practices in reading comprehension instruction have been identified. These practices include systematic instruction, use of teacher think-aloud, guided and independent practice, explicit corrective feedback, clearly stated objectives, specific teaching sequences, explicit statements about the importance of strategy use, monitoring student performance, promoting student self-questioning, encouraging appropriate attributions, and teaching for generalized use of a given strategy.

The Role of Digital Text

Increasingly, students are exposed to text in digital formats, including e-books and e-textbooks, in addition to traditional print (Kelly, McCain, & Jukes, 2009; Parker, 2010;
Schrum & Levin, 2012). As a result, it is necessary to understand how this type of text affects student comprehension. Evidence to support practices that improve reading comprehension of e-text is growing, but it is not yet as robust as the research base for reading comprehension strategies with traditional print. In general, however, technology is thought to increase possibilities for helping students, students with LD in particular, succeed with reading demands in the general education curriculum (Bryant, Bryant, & Ok, 2014). The unique nature of this format allows students to interact with curriculum content in ways that are not possible with printed text, which can ultimately support students’ reading comprehension (Edyburn, 2010; Hall, Hughes, & Filbert, 2000; Marino, 2009).

Currently, various technology centers (e.g., National Center for Supported E-Text, the Accessible Instructional Materials Centers) have issued recommendations for how students can access the curriculum through the use of technology, as well as how technology enhancements can help students progress towards various standards. For example, supported e-text can be used to transform text in several ways, including embedded supports (e.g., definitions of terms, another language, embedded questions/tutorials), multiple modalities (e.g., text-to-speech, graphics, animation, sound), and links to useful resources (e.g., concept maps, note-taking tools, media to augment the main text) (Anderson-Inman & Horney, 2007; Dunleavy, Dexter, & Heinecke, 2007). Further, digital text features can be embedded that provide both basic access to text and instructional enhancements.

**Digital Text Features**

Some digital features are intended to support student comprehension by providing access to the text in a digital format. The conversion of printed text to a digital format provides students with basic access to the text by helping students to overcome some of the barriers of paper format (Anderson-Inman & Horney, 2007). These basic access features, such as electronic format and text-to-speech, are included within a new provision of the Individuals with Disabilities Education Act (2004) that requires that “accessible instructional materials” be provided in a timely manner to students with print-related disabilities so they may successfully participate in the general education curriculum. In addition to students with visual impairments and blindness and students with physical disabilities, students with LD qualify for and may benefit from these accessible instructional materials (NCLD, 2014).

Paper text converted into a digital format allows for modifications, including enlarged font size, navigation, and text-to-speech, as well as additional supports for the perceptual needs of students who need text presented in a different format (Anderson-Inman & Horney, 2007). These supports are often utilized to provide differentiation and encourage independence for students with varying ability levels (Kennedy & Deshler, 2010). However, specifically for
Improving Reading Comprehension Using Digital Text by Sheri Berkeley, Leigh Ann Kurz, Andrea Boykin, and Anya S. Evmenova

students with LD, text-to-speech, which enables students to hear words or phrases that are difficult to decode (Montali & Lewandowski, 1996), may be especially beneficial because many textbooks are written above students’ reading level (Scruggs, Mastropieri, Berkeley, & Graetz, 2000).

In addition to allowing for basic access of text, digital features are now available that can serve as enhancements to instruction for students. Instructional enhancements are supports embedded in digital text to facilitate understanding of the text (Anderson-Inman & Horney, 2007). Often, such enhancements support student understanding of the text by purposely embedding strategies commonly found to be effective in reading research of print text, such as dictionaries, animations, videos, graphic organizers, specific strategies, and strategy supports. These supports incorporate best practices in reading comprehension research (e.g., Mastropieri, Scruggs, & Graetz, 2003) into technology-based environments.

In 2000, Hall and colleagues conducted a systematic review of computer-assisted instruction (CAI) and the impact of these supports on various reading measures, including word recognition, vocabulary, and comprehension for students with LD. The synthesis found that CAI, which consists of a computerized application that teaches students skills for reading, had positive outcomes for reading measures, including comprehension. The synthesis also found that the most effective computerized programs employed effective teaching practices such as elaborate feedback in a digital format. However, rather than evaluating supports for digital text specifically, the synthesis included digital programs for reading instruction (packaged programs) and external supports (such as Internet applications) to support reading skills.

Purpose of the Current Study

Much is known about how to improve the comprehension of students when reading printed text; however, less is known about outcomes with digital texts. Therefore, the purpose of the current investigation was to conduct a meta-analysis of the existing research on the impact of digital text interventions on students’ reading comprehension. As shown in Table 1, for the purposes of the current study, these typologies have been categorized based on function, either (a) basic access to text (e.g., text-to-speech, simplified text, audio with print text); or (b) instructional enhancement (e.g., animation, dynamic visuals, embedded strategy prompts).

Specifically, the current study sought to answer the following research questions: (a) How effective are interventions designed to provide basic access to digital text for improving student reading comprehension? (b) How effective are interventions that embed instructional enhancements within digital text for improving student reading comprehension? and (c) Are there differences in the quantity and/or effectiveness of interventions for elementary and secondary students?
Table 1

Typology Categorization for the Current Study

<table>
<thead>
<tr>
<th>Typology</th>
<th>Description</th>
<th>Basic Access</th>
<th>Instructional Enhancements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentational</td>
<td>Stagnant digital text, static images, modifiable font, navigation</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Translational</td>
<td>Digital text supported with audio and/or text-to-speech with or without dynamic highlighting; simplified text</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Illustrative</td>
<td>Digital text enhanced with sounds, video, animation, dynamic visuals</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Summarizing</td>
<td>Digital text enhanced with table of contents, concept maps</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Instructional</td>
<td>Digital text enhanced with prompts, questions, strategies, definitions</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

*Note. Created using the typology of digital text features framework (Anderson-Inman & Horney, 2007).*

**Method**

Procedures included a systematic search of numerous online databases (PsycINFO, Education Full Text, Web of Science, ERIC, Education Research Complete) using the following keywords: books (computer-, electronic-, audio-, digital-, talking/interactive-), text (electronic-, digital-), reading (comprehension, fluency), read-aloud (text-to-speech, e-readers, computer-assisted), online, supported e-text, hypertext, accommodation, students, computer-assisted instruction, and special education. Ancestry searches were conducted for relevant articles and existing literature reviews (e.g., Hall et al., 2000; Lai & Berkeley, 2012; Lan, Lo, & Hsu, 2014). Finally, hand searches were conducted for the most recent year of journals where relevant articles had previously been located (e.g., Journal of Special Education Technology).

**Inclusion and Exclusion Criteria**

Initially, our research inclusion criteria were specific to interventions for students with LD; however, because of the limited number of studies focusing on students with LD, the search was expanded to include all students. Thus, participants could include students with and/or without disabilities. In addition, studies were included in the sample if (a) the study was an intervention study that utilized a group design; (b) the sample included participants who were kindergarten through 12th grade (with at least 50% of the sample within this range); (c)
the study primarily investigated outcomes for reading (specifically comprehension), rather than writing or math (e.g., calculators); and (d) the study investigated technology (e.g., computers, audio/visual) for general access or enhancement of text (e.g., read-aloud, hyperlinks, embedded strategy prompts). Any e-text formats could be utilized for instruction and/or testing (i.e., accommodations) in any setting (e.g., school, home, clinic).

Studies were excluded if they (a) were adaptive in nature (e.g., Braille readers, switches, video captioning; Anderson-Inman, Terrazas-Arellanes, & Slabin, 2009); (b) included a specific instructional program targeting basic reading skills (e.g., Chambers, Slavin, & Madden, 2011; Coyne, Pisha, Dalton, Zeph, & Smith, 2012); (c) were online or utilized online tools independent from text (e.g., Clay, Zorfass, Brann, Kotula, & Smolkowski, 2009; Hsin-Yuan, 2010); or did not use connected text (e.g., Jong & Bus, 2004). In addition, studies were excluded if they only contained measures of vocabulary (e.g., Korat, 2010; Verhallen & Bus, 2010). Studies were also excluded if they were not peer reviewed (e.g., reports, dissertations, masters theses), if they were written in a language other than English, and/or if they were published prior to 2000 (because technology prior to this date is now considered outdated). We also excluded studies where the students who are English learners were the primary target (i.e., greater than 50% of the sample). Finally, we excluded studies that did not provide sufficient data to calculate a standardized mean-difference effect size (e.g., Shamir, Korat, & Shlafer, 2011).

**Coding Conventions**

A two-part coding manual was created. Part I captured information about basic article characteristics, characteristics of the sample, and the intervention. Basic article information included authors of the study, journal name, and year of publication. Characteristics of the sample included (a) student age in years; (b) gender (number of males and females); (c) grade level (elementary = K to 6th, secondary = 7th to 12th); (d) disability status (student with disability or general education) and type of disability; (e) ethnicity; (f) socio-economic status; (g) IQ; and (h) student achievement information. Characteristic of the intervention included (a) digital text form (audio book, digital text); (b) digital text features (e.g., text-to-speech, hyperlinks, reading strategies, animation); (c) type of text (narrative, expository, combination); (d) level of text (grade level, instructional level); (e) academic content area (language arts, content area); and (f) number of sessions (1, 2-5, or >5).

Part II of the coding manual captured information related to the rigor of the research methodology, including (a) location of the study; (b) description of interventionist; (c) type of samples (dependent, independent); (d) number of measures; (e) quality of the measures (one measure of a single construct; two or more measures of the same construct; two or more measures of multiple related constructs); (f) inclusion of vocabulary measure; (g) scope of the measures
(immediate criterion referenced, norm-referenced, generalization, maintenance); (h) fidelity of implementation (interventionist and technology use as intended); (i) feedback to students; and (j) statistical significance of findings. Effect sizes were calculated using the following formula:

$$g = d \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right)$$

For each study, interventions were determined to be targeting accessibility of text, support of instruction, or both (due to multiple conditions). For each study, only one effect size was calculated for comparisons that evaluated basic access to text, and one effect size was calculated for comparisons that evaluated instructional supports. These categories were not combined in the final analysis. In instances where multiple measures of comprehension were included, effect sizes were averaged, resulting in one effect size for the target comparison. Each calculation was computed by two trained researchers and reconciled to 100%.

**Procedures.** Coding conventions were established for each variable, and two doctoral students were trained to complete all coding. For Part I coding conventions, a third doctoral student was trained to complete reliability of coding conventions for 30% of research studies. Reliability with Coder 1 = 93.02% (range = 88.37% to 97.67%) and with Coder 2 = 94.32% (range = 90.70% to 100%). For Part II coding conventions, the two primary coders coded all items and reconciled discrepancies to 100% agreement.

**Results**

**Characteristics of the Data Set**

The final sample consisted of 27 group studies published in 25 articles from 2001 to 2013 in 25 peer-reviewed journals. The number of studies increased over time with an average of 1.3 studies per year in 2000 through 2006 (range = 0 to 3), and an average of 2.6 studies per year in 2007 through 2013 (range = 1 to 4). A majority (88%) of studies were conducted in the United States.

The data set included 16,513 participants. Slightly more studies were conducted at the elementary level (56%) than the secondary level (44%). Characteristics of the participants cannot be reliably reported because basic study descriptors were reported in a limited number of studies: age (52%), ethnicity (44%), socio-economic status (30%), and IQ (7%). In addition, some sort of achievement level of students was reported in only 63% of studies. Further, a description of the interventionist was provided in only 37% of studies. Because basic study descriptors were so limited, analyses based on these participant characteristics was not possible.

Ten studies (37%) included students with LD in the sample (4 included students with other disabilities); the remaining studies contained students who were typically developing or at risk.
(including students identified in studies as “struggling readers”). Other disabilities included in study samples were as follows: emotional disorder (ED), speech/language disorder (S/LD), intellectual disability (ID), autism, other health impairment (OHI; only attention deficit hyperactivity disorder was specified), hearing impairment (HI), and orthopedic impairment (OI). Table 2 presents a more detailed description of sample criteria reported in studies with students with LD.

Table 2

**Participant Selection Criteria Reported in Studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant Disability</th>
<th>Classification Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boyle et al. (2003)</td>
<td>LD, ED, OHI, ADHD</td>
<td>Students were identified with a disability and received specialized accommodations for secondary history content on their IEP.</td>
</tr>
<tr>
<td>Crawford et al. (2004)</td>
<td>LD, OHI, ED, TBI, VI, ID, autism, S/LD</td>
<td>Majority of students in special education were identified with a learning disability, and all students with IEPs were partially or fully included in the general education classroom.</td>
</tr>
<tr>
<td>Dolan et al. (2005)</td>
<td>LD</td>
<td>Students had active IEPs, and were partially or fully included in general education classes.</td>
</tr>
<tr>
<td>Flowers et al. (2011)</td>
<td>LD, ID, OHI, other disability (not specified)</td>
<td>Students were identified with a disability and eligible for read-aloud accommodation on state assessments.</td>
</tr>
<tr>
<td>Kim et al. (2006)</td>
<td>LD</td>
<td>Students were legally identified as having a disability.</td>
</tr>
<tr>
<td>Ko et al. (2011)</td>
<td>LD</td>
<td>Participants were identified by the local education agent in southern Taiwan as students with learning disabilities who also had difficulty reading.</td>
</tr>
<tr>
<td>Laitusis (2010)</td>
<td>LD</td>
<td>Classification criteria not specified.</td>
</tr>
<tr>
<td>Srivastava et al. (2012)</td>
<td>LD</td>
<td>Students receiving special education services during the data collection phase.</td>
</tr>
<tr>
<td>Twyman et al. (2006)</td>
<td>LD</td>
<td>Identified with learning disabilities in reading and writing and also had IEPs that included goals for reading and writing.</td>
</tr>
</tbody>
</table>

Note. LD = learning disability; ED = emotional disability; OHI = other health impairment; ADHD = attention deficit hyperactivity disorder; TBI = traumatic brain injury; VI = visual impairment; ID = intellectual disability; S/LD = speech/language disorder.

**Intervention characteristics.** Ten studies included interventions targeting basic access to text – 6 at the elementary level (see Table 3) and 4 at the secondary level (see Table 4). Eighteen studies targeted instructional supports – 10 at the elementary level (see Table 5) and 8 at the secondary level (see Table 6). One of the studies (Ertem, 2010) included both a basic skills intervention and an instructional support intervention.

The forms of digital text investigated included e-text (67%), e-books (22%), or audiobooks used with print text (11%). Most studies reported the type of text that was used in the intervention: expository text (41%), narrative text (22%), or a combination of narrative and
expository text (22%), but four (15%) did not report type of text. Most studies utilized grade-
level text (74%), 22% utilized instructional-level text, and 4% did not report the level of text
used. Two thirds of the studies involved language arts materials; the remaining third utilized
materials for science or social studies. Duration of study length varied: 1 session (41%), 2 to 5
sessions (22%), >5 sessions (30%), and unreported length (7%).

Table 3

Studies Investigating Comprehension Outcomes From Basic Access to Digital Text at the
Elementary Level

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Sample (N)</th>
<th>Grades</th>
<th>Intervention &amp; Comparison</th>
<th>Typology &amp; Digital Text Features</th>
<th>Text Type</th>
<th>Measures</th>
<th>Effect Size [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crawford et al. (2004)</td>
<td>338 (47 LD; 3 ID; 13 S/LD; 1 ED; 8 OHI; 1 HI)a</td>
<td>4-5</td>
<td>I: Audio plus print text (testing) C: Print text</td>
<td>Translational: audio (through video) test administration</td>
<td>NR</td>
<td>• Comprehension Test (MC)</td>
<td>.29* [.08, .51]</td>
</tr>
<tr>
<td>Dundar et al. (2012)</td>
<td>20</td>
<td>5</td>
<td>I: e-text (static) C: Print text</td>
<td>No features (stagnant digital text)</td>
<td>Exp.</td>
<td>• Comprehension Test (open ended)</td>
<td>.29** [-.59, 1.17]</td>
</tr>
<tr>
<td>Jeong (2012)</td>
<td>56</td>
<td>6</td>
<td>I: e-book (static) C: Print book</td>
<td>No features (stagnant digital text)</td>
<td>NR</td>
<td>• Comprehension Test (MC)</td>
<td>-.43** [-.80, -.06]</td>
</tr>
<tr>
<td>Laitusis (2010)</td>
<td>2,028 (903 LD)a</td>
<td>4-8</td>
<td>I: Audio plus print text (testing) C: Print text</td>
<td>Translational: audio (CD-ROM) test administration</td>
<td>NR</td>
<td>• Gates-McGinitie Reading Test (GMRT)</td>
<td>.30** [.18, .43]</td>
</tr>
<tr>
<td>Sorrel et al. (2007)</td>
<td>12 (4 LD)a (8 SR)</td>
<td>2-5</td>
<td>I: e-text C: e-text (static)</td>
<td>Translational: text-to-speech</td>
<td>Both</td>
<td>• Accelerated Reader Quizzes</td>
<td>-.30 [-1.1, .51]</td>
</tr>
</tbody>
</table>

Note. LD = learning disability; ED = emotional disability; OHI = other health impairment; S/LD = speech/language disorder; HI = hearing impairment; ID = intellectual disability; SR = struggling reader; Nar. = narrative; Exp. = expository; NR = not reported.
a special education data not disaggregated in results.
mixed findings. **statistically significant findings.
### Table 4

**Studies Investigating Comprehension Outcomes From Basic Access to Digital Text at the Secondary Level**

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Sample (N)</th>
<th>Grades</th>
<th>Intervention &amp; Comparison</th>
<th>Typology &amp; Digital Text Features</th>
<th>Text Type</th>
<th>Measures</th>
<th>Effect Size [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boyle et al. (2003)</td>
<td>67 (43 LD; 8 OHI; 6 ED; 4 ID; 1 SLD; 1 OI; 1 autism)</td>
<td>9-12</td>
<td>I: Audio plus print textbook C: Print textbook</td>
<td>Translational: audio (CD-ROM)</td>
<td>Exp.</td>
<td>• Comprehension Tests (matching, MC)- Unit &amp; Chapter</td>
<td>.87*** [.22, 1.52]</td>
</tr>
<tr>
<td>Dolan et al. (2005)</td>
<td>9 (9 LD)</td>
<td>11-12</td>
<td>I: e-text (testing) C: Print text</td>
<td>Translational: text-to-speech test administration</td>
<td>Exp.</td>
<td>• Released NAEP Test Items</td>
<td>1.18 [.18, 2.18]</td>
</tr>
<tr>
<td>Flowers et al. (2011)</td>
<td>12,699 (all: LD, ID, OHI, or other)</td>
<td>7-8</td>
<td>I: e-text (testing) C: Print text (with adult read-aloud)</td>
<td>Translational: text-to-speech test administration</td>
<td>Both</td>
<td>• State Test for Reading</td>
<td>-.49** [.61, -.36]</td>
</tr>
<tr>
<td>Schmitt et al. (2011)</td>
<td>25 (25 SR)</td>
<td>6-8</td>
<td>I: e-text C: e-text (static)</td>
<td>Translational: text-to-speech</td>
<td>NR</td>
<td>• Comprehension Test (MC)</td>
<td>.23 [-.32, 0.79]</td>
</tr>
</tbody>
</table>

*Note. LD = learning disability; ED = emotional disability; OHI = other health impairment; SLD = speech/language disorder; OI = orthopedic impairment; ID = intellectual disability; SR = struggling reader; Nar. = narrative; Exp. = expository; NR = not reported.

*Comparable ES resulted when using subset of scores used in propensity selection for secondary analysis (ES = -.495).
*Mixed findings. **Statistically significant findings.

### Table 5

**Studies Investigating Instructional Enhancements for Comprehension of Digital Text at the Elementary Level**

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Sample (N)</th>
<th>Grade</th>
<th>Intervention &amp; Comparison</th>
<th>Typology &amp; Digital Text Features</th>
<th>Text Type</th>
<th>Measures</th>
<th>Effect Size [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dalton et al. (2011)</td>
<td>106</td>
<td>5</td>
<td>I: UDL multimedia e-text (vocab + comprehension) C: UDL multimedia e-text (vocab only)</td>
<td>Translational: text-to-speech Instructional: vocabulary &amp; comprehension strategy training</td>
<td>Nar.</td>
<td>• Gates-McGinitie Reading Test (GMRT) • ICON</td>
<td>.51* [.03, 1.00]</td>
</tr>
<tr>
<td>Study</td>
<td>Observer</td>
<td>Grade</td>
<td>Treatment</td>
<td>Dominant Medium</td>
<td>Illustrative</td>
<td>Narrative</td>
<td>Comprehension</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td>Ko et al. (2011)</td>
<td>30</td>
<td>5-6</td>
<td>I: multimedia e-text (with person or robot avatar)</td>
<td>C: Paper text</td>
<td>Instructional: comp. aids</td>
<td>Exp.</td>
<td>• Comprehension Test (MC)</td>
</tr>
</tbody>
</table>

Note. LD = learning disability; SR = struggling reader; Nar. = narrative text; Exp. = expository text; CI = confidence interval; NR = not reported.; UDL = universal design for learning.
*Estimated grade based on reported student age; **this study contained an e-text and a narration (no animation) condition; comparison to print book was ES = .27; *effect size differed greatly by measure.
*Mixed findings. **Statistically significant findings.
### Table 6

**Studies Investigating Instructional Enhancements for Comprehension of Digital Text at the Secondary Level**

<table>
<thead>
<tr>
<th>Study (Year)</th>
<th>Sample (N)</th>
<th>Grade</th>
<th>Intervention &amp; Comparison</th>
<th>Typology &amp; Digital Text Features</th>
<th>Text Type</th>
<th>Measures</th>
<th>Effect Size [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuevas et al. (2012)</td>
<td>145</td>
<td>10</td>
<td>I: e-text C: Print text</td>
<td>Instructional: comprehension aids, definitions</td>
<td>Both</td>
<td>• Gates-McGinitie Reading Test (GMRT)</td>
<td>.04 [-.47, .55]</td>
</tr>
<tr>
<td>Gegner et al. (2009) Study 1</td>
<td>122</td>
<td>11</td>
<td>I: Multimedia e-text C: e-text (static)</td>
<td>Instructional: comprehension aids, reference materials Illustrative: narrated descriptions, animation</td>
<td>Exp.</td>
<td>• Comprehension Test (MC)</td>
<td>.77** [.39, 1.15]</td>
</tr>
<tr>
<td>Gegner et al. (2009) Study 2</td>
<td>97</td>
<td>10</td>
<td>I: Multimedia e-text C: e-text (static)</td>
<td>Instructional: comprehension aids, reference materials Illustrative: narrated descriptions, animation</td>
<td>Exp.</td>
<td>• Comprehension Test (MC)</td>
<td>.81** [.39, 1.22]</td>
</tr>
<tr>
<td>Johnson-Glenberg (2005)</td>
<td>20 (20 SR)</td>
<td>6-7</td>
<td>I: Multimedia e-text C: e-text (static)</td>
<td>Instructional: comprehension strategy training Illustrative: 3-D interactive reading</td>
<td>Both</td>
<td>• Comprehension Test (open-ended)</td>
<td>.46 [-.16, 1.09]</td>
</tr>
<tr>
<td>Kim et al. (2006)</td>
<td>34 (28 LD; 6 ED, OHI, &amp; S/LD)</td>
<td>6-8</td>
<td>I: Multimedia e-text C: Business as usual (print text)</td>
<td>Instructional: comprehension strategy training</td>
<td>Exp.</td>
<td>• Woodcock Reading Mastery Test- Revised (WRMT-R) • CSR Test (main idea, question)</td>
<td>.75** [.05, 1.45]</td>
</tr>
<tr>
<td>McNamara et al. (2006)</td>
<td>39</td>
<td>8-9</td>
<td>I: Multimedia e-text with avatar C: e-text (video training w/ no multimedia)</td>
<td>Instructional: comprehension strategy training, practice &amp; feedback (via virtual agents) Illustrative: animated demonstrations</td>
<td>Exp.</td>
<td>• Comprehension Test (open-ended)</td>
<td>.40 [-.23, 1.04]</td>
</tr>
<tr>
<td>Srivastava et al. (2012)</td>
<td>39 (14 LD)*</td>
<td>8</td>
<td>I: e-text C: Print text</td>
<td>Instructional: hyperlinked vocabulary</td>
<td>Both</td>
<td>• Comprehension Test (MC, open-ended)</td>
<td>-.12* [-.56, .33]</td>
</tr>
<tr>
<td>Twyman et al. (2006)</td>
<td>24 (24 LD)</td>
<td>11-12</td>
<td>I: e-text C: Print textbook</td>
<td>Translational: simplified text, hyperlinked glossary, optional text-to-speech Instructional: comprehension aids Summarizing: hyperlinked graphic organizers</td>
<td>Exp.</td>
<td>• Concept MAZE Task</td>
<td>.16 [-.37, .70]</td>
</tr>
</tbody>
</table>

**Note.** LD = learning disability; SR = struggling reader; ED = emotional disability; OHI = other health impairments; S/LD = speech/language disorder; Nar. = narrative text; Exp. = expository text.

*Special education data not disaggregated in results. *Mixed findings. **Statistically significant findings.
**Study design characteristics.** More studies consisted of independent samples (59%) than dependent samples (41%). Studies included as many as seven measures of a range of constructs (e.g., reading comprehension, vocabulary, strategy use, word recognition, fluency, eye fatigue, reading motivation, time reading, reading enjoyment); however, most studies (90%) contained one to three measures. Two studies (7%) measured comprehension with multidimensional and robust measures; seven (26%) were of moderate rigor; however, most (67%) were singular and one-dimensional. Three (11%) of these studies included a vocabulary element within the assessments of comprehension. All studies included immediate criterion- or norm-referenced measures of student performance; only one had an additional maintenance measure. Fourteen studies (50%) found statistically significant findings on all measures, seven (25%) had mixed findings, and seven (25%) did not demonstrate any statistically significant outcomes.

Of studies that included an interventionist role in facilitation of the intervention (rather than using technology alone), only a small number described fidelity procedures with (19%) or without (22%) supporting outcome data; the remaining studies did not mention fidelity at all (56%). Similar findings occurred with reporting of fidelity of student use of the technology. Only a small number (11%) both described fidelity procedures and reported supporting outcome data or reported a description of fidelity procedures without supporting outcome data (15%); the remaining studies did not mention fidelity at all (74%). One third of studies (33%) provided a description of the type of feedback or instruction provided to students while using the technology.

**Basic Access of Text vs. Instructional Supports**

The overall weighted mean effect size for interventions that investigated the effects of basic access technologies on comprehension was small ($ES = -.03$, range -.49 to 1.18) even after studies that compared digital text with no features to print text were removed from the analysis ($ES = -.02$). The weighted mean effect size for instructional enhancement interventions was moderate ($ES = .51$, range -.35 to 1.57). As illustrated by the confidence intervals reported in Tables 2 through 5, variability was pronounced in the majority of studies in the analysis.

**Basic access studies.** Further investigation revealed that the weighted mean effect size for intervention studies that investigated the impact of basic access technologies on comprehension was small at both the elementary ($n = 6$) and the secondary level ($n = 4$), $ES = .24$ and $ES = -.39$, respectively. Studies with both typically developing students ($n = 4$) and students with disabilities ($n = 6$) yielded small effect sizes ($M = -.07$, and $M = -.02$, respectively). Although caution needs to be used in interpreting these findings due to the small sample sizes (six or fewer studies), small effects seem to be robust both by grade level and type
of student. While there are too few studies to aggregate findings by smaller ranges of grades, visual inspection shows that the two studies conducted at the high school level had large overall effect sizes (Boyle et al., 2003; Dolan, Hall, Banerjee, Chun, & Stangman, 2005) whereas the two studies at the middle school level had overall small effect sizes (Flowers, Do-Hoing, Lewis, & Davis, 2011; Schmitt, McCallum, & Mauck, 2011).

Extreme caution should be used when drawing conclusions from this observation both because of the very small number of studies and the large variability within study findings. To illustrate, in the study of 11th- and 12th-grade students with LD conducted by Dolan et al. (2005), the confidence interval for the large overall effect size of 1.18 ranged from .18 to 2.18. In the study of sixth, seventh, and eighth graders conducted by Schmitt et al. (2011), the confidence interval for the small overall effect size of .23 ranged from -.32 to .79. Similar variability was present in studies with students at the upper-elementary level. For example, in a study of fifth graders conducted by Dundar and Akcayir (2012), the confidence interval for the small to medium overall effect size of .29 ranged from -.59 to 1.17.

**Instructional enhancement studies.** Table 7 lists the reading strategies featured in reading comprehension strategy training interventions within digital text. As illustrated, the weighted mean effect sizes of interventions with instructional enhancements at both the elementary (n = 10) and the secondary (n = 8) level suggest moderate effect (respectively, $M = .58$, and $M = .43$). The mean weighted effect size was also moderate for studies with typically developing students (n = 14, $M = .52$) as well as for studies that included students with disabilities (n = 4, $M = .45$). Moderate effects seem to be robust both by grade level and type of student. The small number of studies did not permit further analysis by upper (grades 4-6) and lower (grades K-3) elementary levels.

Finding for studies with students with LD should be interpreted with caution. Only four studies included students with LD, and findings were mixed – with two studies reporting moderate-to-large overall effect sizes: $ES = .75$ (Kim et al., 2006) and $ES = 1.57$ (Ko, Chiang, Lin, & Chen, 2011), and two reporting small overall effect sizes: $ES = -.12$ (Srivastava & Gray, 2012) and $ES = .16$ (Twyman & Tindal, 2006). As for the basic access studies, there was large variability within studies as well. For example, in the 2006 study by Kim and colleagues with 34 students with LD and other mild disabilities, the confidence interval ranged from .05 to 1.45. By comparison, in the 2006 study by Twyman and Tindal with 24 students with LD, the confidence interval ranged from -.37 to .70.
Table 7
Reading Strategies Featured in Reading Comprehension Strategy Training Interventions Within Digital Text

<table>
<thead>
<tr>
<th></th>
<th>Activating Prior Knowledge</th>
<th>Prediction</th>
<th>Text Structure</th>
<th>Visualization</th>
<th>Questioning</th>
<th>Comprehension Monitoring</th>
<th>Summarizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dalton et al. (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Gegner et al. (2009)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Johnson-Glenberg (2005)</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et al. (2006)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McNamara et al. (2006)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sung et al. (2008)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The value of technology for delivering instruction has been questioned for years. As noted by Clark (1983), “media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition” (p. 457). Research is needed to support or refute this kind of claim. However, our review indicates that the research base on the use of connected digital text (whether for basic access or with instructional enhancements) is very limited, particularly for research including students with LD. Across studies reviewed in this analysis, effect sizes were small for interventions designed to help students gain basic access to digital text, and effect sizes were moderate for interventions that also included instructional enhancements. However, variability within both categories of interventions was pronounced.

Digital Features for Basic Access to Text

According to Anderson-Inman and Horney (2007), while electronic text may help students with reading difficulties to overcome substantial barriers imposed by printed materials, the usefulness of electronic text by itself is rather limited. Although only two studies (Dundar &
Akcayir, 2012; Jeong, 2012) investigated the effects of static digital text vs. print text, findings support the claim that simply putting text in a digital format for display on a computer screen does not result in improved performance (ES = .29 and ES = -.43, respectively). For example, in the study by Dundar and Akcayir (2012), the paper-based classroom textbook was converted into a digital format, which was read by the students on a tablet without any additional features or supports; the control group read the same text in a paper format. Although the text was converted into a digital format, it did not support student comprehension of the material.

Studies that included other features to support basic access, such as audio and text-to-speech, were also consistently small. In a study included within the meta-analysis (Sorrell, Bell, & McCallum, 2007), struggling readers (students reading below grade level) in the experimental group read digital text with support of text-to-speech software. When their performance was compared to that of students who read the same digital text without text-to-speech, no differences in comprehension were found between the two groups.

The minimal impact of text-to-speech on student comprehension in studies included within this meta-analysis is unexpected because it is well established that decoding ability is related to reading comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Goff, Pratt, & Ong, 2005). The ability to decode with automaticity frees up cognitive resources that can be allocated to text comprehension (Shinn, Good, Knuston, Tilly, & Collins, 1992), which becomes increasingly important as text increases in complexity. As such, digital text features are increasingly being used as a test accommodation, including high-stakes state assessments.

Testing accommodations are generally considered effective when a “differential boost” occurs; that is, when accommodations lead to greater score improvements for students with disabilities than for students without disabilities (Thompson, Blount, & Thurlow, 2002). Only four studies investigated the effects of digital text features for students with disabilities within testing contexts, but results were inconsistent. Specifically, two studies of audio accommodation with elementary-level students with LD (Crawford & Tindal, 2004; Laitusis, 2010) both had small to moderate effects. At the middle and high school levels, two studies investigated the effects of text-to-speech as a testing accommodation and found very small (Flowers et al., 2011) and very large effects (Dolan et al., 2005).

Although based on a very limited number of studies, these findings are consistent with mixed results reported in previous reviews of the accommodation literature for read-aloud accommodations for students with LD (e.g., Lai & Berkeley, 2012). Variability in the data indicated that the digital text features were helpful for some students, but not all. As suggested by Lai and Berkeley (2012), “until this body of research develops, IEP teams should reference evidence on the effectiveness of this accommodation for an individual student in classroom-testing situations before making a determination whether to provide it to the student in a high-
stakes testing situation” (p. 167). Further, regardless of the inconclusive evidence, it is clear that more attention needs to be devoted to the quality of technology training and opportunities for sufficient student practice with technology-based tools to ensure students are ready to successfully use digital text prior to test administration.

Digital Features to Enhance Instruction With Digital Text

Based on the effectiveness of reading strategies on student reading comprehension of print text (e.g., Berkeley et al., 2010; Edmonds et al., 2009; Swanson, 1999), researchers (e.g., Okolo, 2005) have logically conjectured that reading comprehension can be further supported by adding instructional enhancements to digital text such as concept maps, text-to-speech, embedded dictionaries, and alternative representations of information. Further, findings from previous research syntheses have shown positive outcomes for digital supports on the reading performance of both general education students (Lan et al., 2014) and students with LD (Hall et al., 2000). Therefore, the mixed findings in this analysis of digital text studies were surprising – especially for the studies with students with LD who presumably would be more likely to benefit from supports.

Differences in findings across syntheses may be due to differences in the target populations and the nature of the intervention reviewed. For example, Lan and colleagues (2014) included college-age students in their selection criteria. Specifically, half of the studies included in the synthesis were conducted with undergraduate students, and the researchers noted generally positive reading outcomes for this age group. However, struggling readers did not benefit from the digital instructional enhancements. In addition, this meta-analysis only included studies with digital metacognitive strategies and did not determine the impact of other digital texts features, such as text-to-speech, stagnant digital text, or animations, on comprehension. Based on the nature of the interventions selected and the inclusion of college-aged students, it is possible that digital supports are more effective for certain age groups of readers and as a teaching medium rather than a built-in digital support to facilitate comprehension.

Differences may also be related to the type of technology targeted for synthesis. A narrative review of the research found that CAI interventions with instructional enhancements resulted in positive outcomes in various reading skills for students with LD (Hall et al., 2000). However, Hall et al.’s synthesis included computer programs for reading instruction and external supports, such as Internet applications, to support reading skills, whereas the present study focused on built-in supports that impact the comprehension of digital text specifically. Interventions with instructional enhancements in this meta-analysis all incorporated digital features to aid in increasing the understanding of the text, such as digital text with built-in comprehension aids and reference materials (Gegner, Mackay, & Mayer, 2009), text-to-speech combined with comprehension strategy training (Dalton, Proctor, Uccelli, Mo, & Snow, 2011),
e-books with animation and music (Ertem, 2010), virtual agents that provide comprehension support (Kim et al., 2013), and digital text with comprehension aids (Kim et al., 2006).

Given the limited number of studies and the degree of similarity among the intervention components, it is difficult to arrive at a definite explanation for the inconsistent findings; however, we speculate that they may be due, at least in part, to the demands of the text itself. For example, each of the LD studies with limited effectiveness (i.e., Srivastava & Gray, 2012; Twyman & Tindal, 2006) contained hyperlinked text. Hypertext requires making choices among multiple potential paths (Lee & Tedder, 2003), which can lead to cognitive overload, particularly for students with LD who are known to have challenges with reading comprehension caused by deficits in working memory (Backenson et al., 2015; Swanson, 1994; Swanson & Alexander, 1997). In addition, students with LD are less likely to self-regulate their learning and persist with tasks (Gersten et al., 2001), so non-linear reading tasks are likely to be more challenging for them and explicit instruction in how to approach these types of text is necessary for them students to be successful.

Although results were mixed for instructional enhancements on comprehension in the present synthesis, variability in the findings might also be explained – at least in part – by the variability of rigor in the research methodologies of the studies. The majority of the reviewed studies did not meet the quality indicators for special education research (Gersten et al., 2005; Odom et al., 2005). Quality indicators for special education technology research require assessment of both surface and quality; in other words, both the technology-based intervention itself and how well it is implemented (Gersten & Edyburn, 2007).

Studies in the current analysis were particularly lacking in description of the sample, components of the intervention (both the implementation and the characteristics of the technology itself), and fidelity of treatment (including feedback provided to students). These areas are especially important in technology research when digital features are optional to students (e.g., text-to-speech, digitized realistic narrations, dynamic highlighting, hyperlinks) because the outcome is likely directly related to frequency of use.

Implications for Research

According to Gersten and Edyburn (2007) “the use of technology in special education has been advanced on the basis of marketplace innovations and federal policy initiatives rather than on a compelling research base” (p. 3). As such, statements about the effectiveness of digital text must be made with great caution. Studies investigating the effectiveness of digital text for improving reading comprehension are of mixed quality and outcome. Further, the research base on the effectiveness of digital text (whether for basic access or with instructional enhancements) for students with LD is very limited. It is imperative that more empirical evidence be obtained
that meets the quality indicators for special education technology research to support the use of digital text. However, this is a daunting task considering that technology advances generally outpace research. Compounding the issue is the fact that unlike other disciplines where knowledge accumulates over time, much of the research on technology becomes irrelevant as older technologies become obsolete. For this reason, it is particularly important for researchers to meticulously document the characteristics of the participants as well as the salient features of the intervention – to include type of text, instructional procedures, and fidelity of implementation, in addition to thorough descriptions of the digital text features utilized.

References

*References marked with an asterisk indicate studies included in the meta-analysis.


Improving Reading Comprehension Using Digital Text by Sheri Berkeley, Leigh Ann Kurz, Andrea Boykin, and Anya S. Evmenova


Improving Reading Comprehension Using Digital Text by Sheri Berkeley, Leigh Ann Kurz, Andrea Boykin, and Anya S. Evmenova


Use of the Randomization Test in Single-Case Research

Matthias Grünke
University of Cologne, Germany

Richard T. Boon
The University of Texas at San Antonio

Mack D. Burke
Texas A&M University

Abstract

The purpose of this study was to illustrate the use of the randomization test for single-case research designs (SCR; Kratochwill & Levin, 2010). To demonstrate the application of this approach, a systematic replication of Grünke, Wilbert, and Calder Stegemann (2013) was conducted to evaluate the effects of a story map to improve the reading comprehension skills of five elementary students with learning disabilities in Germany. A multiple-baseline design (Baer, Wolf, & Risley, 1968) was used to evaluate the effectiveness of the story mapping instruction in teaching students the key story grammar elements in a reading passage. However, unlike in traditional multiple-baseline designs, intervention and withdrawal phases were applied at randomly determined points. Results indicated that the use of story maps increased the students’ recall and comprehension of the stories from baseline to intervention, and continued during maintenance. A randomization test confirmed that the differences between baseline and intervention were statistically significant. Findings, limitations, and implications of the use of randomization tests in SCR are discussed.
needs of students with disabilities are increasingly being supported by research from the special education literature (Cook & Cook, 2013; Cook, Tankersley, & Landrum, 2013). One particular area of research is the importance of establishing quality standards that determine what constitutes an EBP (Cook et al., 2014; Cook, Tankersley, & Landrum, 2009). In order to be counted as an EBP, an approach must have documented effectiveness. Specifically, it has been suggested that to qualify as an EBP, an intervention must use high-quality research designs consisting of at least two group design experiments (Gersten et al., 2005) or a series of at least five single-case studies (Horner et al., 2005).

With regard to students with learning disabilities (LD), single-case research (SCR) is playing an increased role in establishing the evidence base for practices. SCR uses the data from one participant or from a very small number of subjects to establish the existence of cause-and-effect relationships (Riley-Tillman & Burns, 2009). In other words, SCR is used to determine whether there is a functional relationship between manipulation of the independent variable(s) and corresponding changes in the dependent measure(s). Further, in SCR designs, individuals serve as their own controls by providing a baseline prior to implementation of an approach. Thus, an intervention can be progress monitored and responsiveness to intervention can be judged against prior progress. As a result, researchers are able to make inferences regarding whether a given intervention is instrumental in fostering different skills for a specific student (Riley-Tillman & Burns, 2009).

Several meta-analyses in recent years have focused on interventions for students with LD that are mainly based on SCR (e.g., Codding, Burns, & Lukito, 2011; Lee & Kim, 2013; Zheng, Flynn, & Swanson, 2013). One reason why this research methodology seems to be especially applicable for evaluating interventions for students with LD is that repeated measurement of respective success criteria (e.g., reading fluency, automatic recall of math facts, spelling skills) is often relatively easy. To frequently record school performance-related variables like reading fluency, automatic recall of math facts, or spelling skills in an objective, reliable, and valid way is generally a lot less complicated than to register emotional or social parameters (self-esteem, interpersonal skills, or psychological resilience). Through systematic measurement of the dependent measure, intervention responsiveness can be evaluated.

However, despite the prominent role that SCR occupies in the research-based literature for students with LD, many scholars still have reservations about this approach (Matson, Turygina, Beighleya, & Matson, 2012). A major reason for such reservation involves the common method of analyzing data from SCR, which relies on visual analysis. Visual analysis consists of graphing a given data set and then appraising the differences between phases for changes in level, trend, and variability. While widespread in the SCR literature, visual analysis of SCR data is viewed by some researchers as being biased (Dugard, File, & Todman, 2012).
The inter-rater reliability for this approach is alarmingly low, rarely exceeding .50 (Brossart, Parker, Olson, & Mahadevan, 2006). Further, the effect of training and experience on visual inspection seems to be negligible. For example, Harbst, Ottenbacher, and Harris (1991) demonstrated that long-time journal reviewers performed little better than completely untrained raters at graph judgment tasks.

For these reasons, several attempts have been made to apply statistical inferential procedures to analyzing data from SCR designs (Callahan & Barisa, 2005; Campbell & Herzinger, 2010; Ferron, 2002; Garthwaite & Crawford, 2004; Janosky, Al-Shboul, & Pellitieri, 1995; Manolov, Arna, Solanas, & Bono, 2010; Parker, Vannest, Davis, & Sauber, 2011; Scruggs, Mastropieri, & Regan, 2006). Unfortunately, applying common parametric tests for group comparisons is generally unsuitable in this context (Campbell & Herzinger, 2010), primarily due to the small sample size and the repeated measures that are used in SCR. Thus, we need to look for alternatives for common parametric methods to provide researchers with tools to analyze their data from SCR designs in a way that is equally objective.

Randomization Tests for SCR

One remedy for making visual analysis more acceptable is to apply randomization tests for phase designs (Kratochwill & Levin, 2010). Traditionally, in SCR, it is recommended that baseline observations continue until a stable pattern of measurements has been established (Gast & Ledford, 2010). However, when researchers wish statistically to analyze their data with a randomization test for phase designs, a different approach is needed – one that first introduces a random assignment scheme into the experiment (Ferron, Foster-Johnson, & Kromrey, 2003).

One way to introduce randomization in phase designs consists of defining a total number of measurement points, a reasonable minimum and maximum number of measurement points for each phase a priori, and subsequently selecting the beginning of the intervention phase by chance within the preset range of options (Edgington, 1992). Randomization tests work by taking into account all possible permutations of the data. For example, if a simple AB design with 30 observations and at least 5 measurement points in each phase was applied, the intervention could commence after the 5th, 6th, 7th, 8th, … or the 25th baseline observation. This would add up to 21 possible intervention starting points. One of these options is drawn by chance, and the study is executed. Following the collection of all data, a test statistic is computed, and the results are compared to the test statistic of the 20 theoretically possible permutations that could have occurred. The test p-value is the proportion of test statistic values greater than or equal to the observed test statistic. In the current example, if the observed test statistic exceeded the other 20 options, the probability of such an outcome would be 1/21 = 0.048 (Todman, 2002).
Other designs and computational procedures are often much more complicated than this simple example. However, in every instance, researchers do not rely on a standard test statistic distribution (e.g., \(t\)- or \(F\)-distribution), but act on the premise that all the information needed to perform inferential statistics is included the data set itself, and derive a \(p\)-value from permutations of the data. In the aforementioned example, this can easily be done with a pocket calculator. However, such a procedure can otherwise become very complex. For instance, for an AB multiple-baseline design across five participants again with 21 possible intervention starting points per subject, there would be \(21^5 = 4,084,101\) permutations. Enumerating such a large number of possibilities requires fast and intensive computing.

Until recently computational power to perform such randomization tests has not been readily available, which explains why these methods have not yet played a major role in analyzing data from SCR designs in practical research (Dugard et al., 2012). Nowadays, Monte Carlo randomized tests, which compute an approximate \(p\)-value of the observed test statistic for a random sample of all possible data arrangements, are commonly used. The test \(p\)-value is the proportion of test statistic values in the random sample as large as the observed test statistic. Monte Carlo randomized tests are available in the familiar environments of IBM® SPSS and Microsoft® Excel (Dugard, 2013), or SCDA, a software application for analyzing single-case designs implemented in R, which includes a randomization test module, SCRT (Bulté & Onghena, 2013).

**Story Mapping**

A story map is a visual strategy designed to promote comprehension of the main parts of a story. It has been shown to be an effective intervention for elementary students with LD (e.g., Boulineau, Fore, Hagan-Burke, & Burke, 2004; Gardill & Jitendra, 1999; Idol, 1987; Idol & Croll, 1987; Johnson, Graham, & Harris, 1997; Stagliano & Boon, 2009; Taylor, Alber, & Walker, 2002; Wade, Boon, & Spencer, 2010). According to Davis and McPherson (1989), a story map is “… a graphic representation of all or part of the elements of a tale and the relationships between them” (p. 232). It is a form of a graphic organizer that makes the structure of concepts and relationships between them apparent by creating a systematic schema to connect prior knowledge with the content of a text that a learner is reading (Anderson & Pearson, 1984; Ausubel, 1960, 1968).

Story maps reduce the amount of semantic information a student has to process in order to extract meaning (Jitendra & Gajria, 2011), thus decreasing the potential for cognitive overload (O’Donnell, Dansereau, & Hall, 2002). Figure 1 shows a sample story map template from Idol (1987, p. 199) completed with all the main components of the fairy tale *The Frog Prince* (Grimm & Grimm, 2013).

Purpose

The primary purpose of the present study was to illustrate the use of the randomization test for SCR designs (Dugard, 2013; Kratochwill & Levin, 2010). Since this approach has only been applied four times (Grünke & Calder Stegemann, 2014; Grünke et al., 2013; Mastropieri, Scruggs, Mills, et al., 2009; Regan, Mastropieri, & Scruggs, 2005), our study extends the current special education research base on its application to LD populations. A secondary purpose of the study was to conduct a systematic replication of a previous study by Grünke et
al. (2013) examining the effects of a story map to improve the reading comprehension skills of elementary students with LD, which also extends current research on the use of a randomization test with a story mapping intervention.

**Method**

**Participants**

**Students.** Five elementary students were recruited to participate in the study. All had been identified with an LD by a multidisciplinary team. In Germany, where the study took place, the main criterion for diagnosing a student with LD is generalized school failure. Such a label indicates that a student shows deficits in one or more psychological processes that manifest themselves in an insufficient ability to perform at grade level in most or all academic core courses compared to typically achieving peers. However, these students do not meet the criteria for a mild intellectual disability, even though often they exhibit low average intelligence (Al-Yagon, Cavendish, Cornoldi, et al., 2013).

Participants included two female (Asena and Julia) and three male students (Eman, Leon, and Marvin). Two of the participants, Asena and Eman, had an immigrant background: Asena’s parents were from Turkey, while Eman’s parents were from Bosnia-Herzegovina. Both of them were bilingual and spoke German and either Turkish or Bosnian, respectively. All students were fluent in academic German. Participants’ ages ranged from 8 years 1 month to 10 years 1 month ($M = 9$ years 2 months). Their intelligence quotient (IQ) scores, as measured by the German Number Combination Test (ZVT; Oswald & Roth, 1987), ranged from 87 to 99 ($M = 93.6$). The students’ fluency level according to the German Salzburg Reading and Orthography Test II (SLRT II; Moll & Landerl, 2010) was above average, with all students’ scores exceeding the 75th percentile (range = 76-96%), indicating that the students were proficient decoders and fluent readers. However, the students’ scores on the German Reading Comprehension Test for First to Six Graders (ELFE 1-6; Lenhard & Schneider, 2006), which measures students’ ability to understand what they read, were remarkably low. Thus, all participants scored in the lowest third of their population with a percentage between 6 and 33%. According to the manual, the reliability of the ZVT varies between .84 and .97 (test-retest correlation). Comparisons between results from the ZVT and the Culture Fair Intelligence Test (CFT 3; Cattell, 1966) show $r = .83$. The retest-reliability of the SLRT II ranges between .80 and .97. Comparisons between results from the SLRT II and the Salzburg Reading Screening Instrument (SLS; Mayring & Wimmer, 2003) amount to .75. For the ELFE 1-6, the test-retest correlation averages .91. Comparisons between results from the ELFE 1-6 and teacher appraisals amount to $r = .70$. Table 1 presents a summary of students’ demographic information.
Table 1

Student Demographic Information (With Percentiles for Fluency and Comprehension Scores)

<table>
<thead>
<tr>
<th>Student</th>
<th>Gender</th>
<th>Age (year-month)</th>
<th>IQ</th>
<th>Fluency Score&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Comprehension Score&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asena</td>
<td>F</td>
<td>9-8</td>
<td>90</td>
<td>96%</td>
<td>16%</td>
</tr>
<tr>
<td>Eman</td>
<td>M</td>
<td>8-6</td>
<td>96</td>
<td>83%</td>
<td>33%</td>
</tr>
<tr>
<td>Julia</td>
<td>F</td>
<td>8-1</td>
<td>87</td>
<td>76%</td>
<td>6%</td>
</tr>
<tr>
<td>Leon</td>
<td>M</td>
<td>10-1</td>
<td>99</td>
<td>88%</td>
<td>18%</td>
</tr>
<tr>
<td>Marvin</td>
<td>M</td>
<td>9-10</td>
<td>96</td>
<td>76%</td>
<td>11%</td>
</tr>
</tbody>
</table>

<sup>a</sup>ZVT: Number Combination Test (Oswald & Roth, 1987).
<sup>b</sup>SLRT II: Salzburg Reading and Orthography Test (Moll & Landerl, 2010).
<sup>c</sup>ELFE 1-6: Reading Comprehension Test for First to Sixth Graders (Lenhard & Schneider, 2006).

**Interventionists.** Two graduate students in special education from a university in the western part of Germany served as the interventionists and administered all instructional sessions across conditions. Prior to the start of the study, the first author trained the interventionists on the instructional procedures for teaching the use of the story mapping procedure. To ensure treatment fidelity, the interventionists were provided with a detailed script to follow and to assess their implementation of the strategy. In addition, the first author was in regular contact with them via e-mail and phone. Finally, during this intervention period, he held four formal research meetings to discuss issues related to implementation of the intervention in the schools.

**Setting**

The students were enrolled in two schools in North Rhine-Westphalia, Germany. Three of the students, Asena, Eman, and Julia, attended an inclusive elementary school in an outlying district of a major city, while Leon and Marvin were enrolled in a rural special school for slow learners. In both schools, the study took place in a room outside the students’ classrooms during a daily period of independent class work.

**Materials**

Eighteen stories from three German storybooks (Wölfel, 1974, 2010a, 2010b) were used. The stories were short and modified to consist of exactly 150 words and contain all of the key story grammar elements. Ten story grammar comprehension questions were generated for each story, expressed in such a way that only one answer was possible. The level of reading difficulty of the pool of questions was assessed with 10 low-achieving students between 9 and 10 years of age. Based on this assessment, questions that were not answered correctly by at least five students were replaced or rephrased.
General Procedures

The study was conducted for 15 consecutive school days and 3 additional days that were evenly dispersed over 3 weeks following the intervention. Procedures identical to those used in the Grünke et al. (2013) study were implemented. All instructional sessions across conditions were carried out in a 1:1 format. During each lesson, the students read a story and provided written answers to 10 comprehension questions. Stories were presented in random order. The students did not receive any assistance or performance feedback on their answers to the comprehension questions.

Experimental Procedures

Baseline. During the baseline phase, the students silently read a story, rehearsed its content, and answered 10 comprehension questions within a 15-minute session. They were provided with a pencil, scratch paper, and a copy of the story. A timer was set to monitor the duration of the session. During silent reading of the story, the participants were allowed to consult any aid of their choice to memorize and make themselves familiar with the content (e.g., take notes, rehearse the information verbally, draw pictures). After the students finished reading and stated they were familiar with the story, the interventionists collected the copy of the story and any student-generated aids (e.g., notes, pictures). Next, the participants completed 10 comprehension questions related to the story grammar elements within the story. At the end of the 15-minute period, the students were asked to submit their responses to these questions to the interventionists.

Intervention. During the intervention phase, the participants were taught to use a story map using a procedure similar to Idol (1987), which consisted of three phases: a Model phase, a Lead phase, and a Test phase. Students read a story, completed a story map, and answered 10 comprehension questions in a 30-minute session. At the beginning of the intervention sessions, the students were provided with a pencil and a copy of a story. During the Model and Lead phase instructional sessions, the students also received a blank copy of a German version of a story map (see Figure 1 for an example).

During the Model phase, students were shown how to use a story map while reading a text. First, the interventionists sat next to the students, displayed a German version of the story map on a sheet of paper, and provided the students with a copy of a story and a blank story map. Next, the interventionists read the story aloud as the students followed along. While reading the story, the interventionists paused when a relevant story grammar element was identified in the text, filled out the appropriate parts of the story map, and asked the students to do the same on their own copy of the story map. Upon completion of the story, the participants turned in the reading passage and the completed story map to the interventionists.
In the Lead phase, the students read a story independently and completed a story map with minimal support from the interventionists. However, assistance was provided to scaffold the process, provide feedback on how the students were to complete the story map, answer story-related questions, remind students to be mindful of the key story grammar elements within the story, and to review their completed story map. If needed, the interventionists also assisted the students in finishing the story map. After completing the reading, the students turned in their copy of the story, filled out a story map, and answered 10 comprehension questions.

Finally, during the Test phase, the students read a story independently and completed their own story map on a piece of scratch paper, while the interventionists loosely monitored their work, answered questions pertaining to the story grammar elements, and provided support when students explicitly asked for help or if it was evident that they needed assistance to identify story grammar elements in the story.

All students completed two Model phase sessions. After the second lesson, they had reached a basic level of proficiency in the story mapping strategy and were able to move on to the next step, the Lead phase. For the Lead phase, Eman and Marvin received two instructional sessions, Asena received four, and Julia and Leon each received five. The criterion to advance from the Lead phase to the Test phase required the students to correctly complete the story maps with no assistance for two consecutive Lead phase sessions with 90% accuracy. All participants continued in the Test phase until they had completed the predetermined number of intervention sessions.

Maintenance. During the maintenance phase, procedures identical to those described in the baseline phase were implemented.

Experimental Design

A multiple-baseline design across participants (Baer, Wolf, & Risley, 1968) was used to determine the effects of the story mapping strategy. This approach demonstrates experimental control by systematically introducing the intervention in a time-lagged manner (Gast & Ledford, 2010). As mentioned above, introducing a random procedure into the design is an essential prerequisite for applying a randomization test, but it also strengthens the explanatory power of the whole study. Carrying on with baseline measurements until the variability in the data stabilizes may bias the design toward a particular intervention effect, thus compromising internal validity. The baseline data path may have shown more variability if baseline observations were allowed to continue. Moreover, several high and low random data points in the baseline might be mistaken for baseline stabilization (Todman, 2002). Introducing the intervention at random in a multiple-baseline design with a predetermined number of probes and a minimum number of baseline and intervention sessions controls for potentially systematic error and strengthens the internal validity of the findings (Dugard, 2013; Marascuilo & Busk, 1988).
The following specifications were established for the study:

1. A total number of 15 daily sessions were chosen for the baseline and the intervention sessions.
2. The baseline phase had to consist of at least three probes and the intervention phase had to consist of at least five probes, yielding eight possible intervention starting points, from the 4th to the 11th session (after the 3rd, 4th, 5th, 6th, 7th, 8th, 9th, or 10th baseline probe).
3. The starting point for each participant was randomly selected from the eight possible intervention starting points. Asena’s intervention started after the fourth baseline probe, for Eman after the eighth, for Julia and Leon after the fourth, and after the fifth for Marvin.
4. To assess the continuation of the intervention effects, all the students received three maintenance probes after completion of the intervention phase.

Inter-Rater Reliability

Student responses to the comprehension questions were independently scored by the two interventionists for all probe sessions. A point was awarded for each question that was answered correctly. The maximum number of possible points earned in each probe was 10. Initial mean inter-rater agreement was 97%. Disagreements in scoring were extremely rare; however, when they happened, they were discussed and resolved by consensus to reach a 100% inter-rater reliability agreement.

Data Analysis

The data were analyzed through visual inspection examining the level, trend, and variability within and between phases (Gast & Spriggs, 2010). In addition, a one-tailed Monte Carlo randomization test for multiple-baseline designs across participants (AB) at a 0.05 significance level was applied to participants’ baseline and intervention score data using a Microsoft® Excel macro downloaded from http://www.routledge.com/books/details/9780415886932/ (Dugard et al., 2012). As mentioned, our design involved an A, a B, and another A phase. However, there is no way to statistically analyze data from such a procedure using a randomization test (P. Dugard, personal communication, February 12, 2012). As a result, we had to limit ourselves to just considering the first A and the B phase for this part of the analysis.

The macro was set to generate 2,000 arrangements of the data at random. The sum of differences between intervention and baseline means was selected as the test statistic because the comprehension scores were expected to increase with the introduction of the intervention. With eight possible intervention starting points and five participants, there were $8^5 = 32,768$ arrangements of the data. Thus, the lowest possible $p$-value would be $1/32,768 = 0.00003$ if an exact
A randomization test was conducted; with a Monte Carlo randomization test with 2,000 random samples, the lowest estimate p-value would be 0.0005. This created very favorable conditions for detecting an intervention effect in case it actually existed.

Finally, the improvement rate difference (IRD) for single-case research designs was calculated for all students (Parker, Vannest, & Brown, 2009). This effect size measures “… the difference in successful performance between baseline and intervention phases” (Alresheed, Hott, & Bano, 2013, p. 10). The IRDs were computed using the IRD calculator available at http://www.singlecaseresearch.org/calculators/ird.

Results

Table 2 and Figure 2 present the number of correctly answered comprehension questions during the baseline, intervention, and maintenance phases.

Table 2
Overview of Study Results per Phase

<table>
<thead>
<tr>
<th>Student</th>
<th>Baseline N (Probes)</th>
<th>Model Phase</th>
<th>Lead Phase</th>
<th>Test Phase</th>
<th>Maintenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asena</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Scores 3; 0; 1; 1</td>
<td>4; 3</td>
<td>10; 8; 6</td>
<td>9; 8; 6; 10</td>
<td>10; 9; 6</td>
</tr>
<tr>
<td></td>
<td>M 1.25</td>
<td>3.50</td>
<td>8.00</td>
<td>8.40</td>
<td>8.33</td>
</tr>
<tr>
<td></td>
<td>IRD -/-</td>
<td>0.50</td>
<td>0.83</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>Eman</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Scores 2; 3; 0; 1; 3; 1; 2</td>
<td>10; 10</td>
<td>10; 10</td>
<td>10; 9; 10</td>
<td>10; 8; 10</td>
</tr>
<tr>
<td></td>
<td>M 1.88</td>
<td>10.00</td>
<td>10.00</td>
<td>9.67</td>
<td>9.33</td>
</tr>
<tr>
<td></td>
<td>IRD -/-</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Julia</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Scores 2; 0; 5; 1</td>
<td>3; 6</td>
<td>8; 9; 7; 6; 10</td>
<td>8; 8; 5; 10</td>
<td>10; 10; 8</td>
</tr>
<tr>
<td></td>
<td>M 2.00</td>
<td>4.50</td>
<td>8.00</td>
<td>7.75</td>
<td>9.33</td>
</tr>
<tr>
<td></td>
<td>IRD -/-</td>
<td>0.50</td>
<td>0.86</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td>Leon</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Scores 4; 0; 5; 2</td>
<td>5; 7</td>
<td>9; 8; 10; 8; 8</td>
<td>10; 9; 8; 10</td>
<td>9; 9; 8</td>
</tr>
<tr>
<td></td>
<td>M 2.75</td>
<td>6.00</td>
<td>8.60</td>
<td>9.25</td>
<td>8.67</td>
</tr>
<tr>
<td></td>
<td>IRD -/-</td>
<td>0.50</td>
<td>0.86</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>Marvin</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Scores 2; 1; 3; 0; 2</td>
<td>10; 8</td>
<td>10; 10</td>
<td>10; 10; 10; 10</td>
<td>9; 10; 10</td>
</tr>
<tr>
<td></td>
<td>M 1.60</td>
<td>9.00</td>
<td>10.00</td>
<td>10.00</td>
<td>9.67</td>
</tr>
<tr>
<td></td>
<td>IRD -/-</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*aBaseline vs. Model phase. bBaseline vs. Model + Lead Phase. cBaseline vs. Model + Lead + Test Phase. dBaseline vs. Maintenance Phase.
Figure 2. Number of correctly answered comprehension questions across baseline, intervention, and maintenance phases for Asena, Julia, Leon, Marvin, and Eman.
Results of the randomization test applied to the baseline and intervention phases showed that the differences between these two phases in the comprehension skills of the five students were statistically significant \( (p < .001; \text{one-tailed}) \). Furthermore, visual inspection of the data support the findings of the randomization test as all five students showed an increase in the number of story grammar elements identified in the story after the interventionists introduced the story mapping strategy. On average, during the baseline phase, participants were only able to answer fewer than three of the comprehension questions correctly, with mean scores ranging from 1.25 to 2.75. By contrast, in the intervention phase, students’ overall mean comprehension scores ranged from 7.27 to 9.86. The average comprehension scores of Asena, Julia, Leon, and Marvin increased from the Model to the Lead phase as follows: from 3.50 to 8.00 for Asena, from 4.50 to 8.00 for Julia, from 6.00 to 8.60 for Leon, and from 9.00 to 10.00 for Marvin. The remaining student, Eman, scored 10 out of 10 during both the Model and Lead phases. From the Lead to the Test phase, the mean comprehension scores of Asena and Leon continued to improve, from 8.00 to 8.40 for Asena, and from 8.60 to 9.25 for Leon. However, a slight decrease in average performance was noted for Eman (from 10.00 to 9.67) and Julia (from 8.00 to 7.75). Alternatively, Marvin continued to answer all of the comprehension questions correctly in the Test phase. In the course of the maintenance phase, all students had average scores ranging from 8.33 to 9.67, with Eman, Julia, Leon, and Marvin scoring n 8-10 comprehension questions answered correctly.

Moreover, across all participants, IRD scores ranged from 0.50 to 1.00 between baseline and Model phase, and from 0.83 to 1.00 between baseline and Model phase + Lead phase. The overall IRD between baseline and intervention (Model phase + Lead phase + Test phase) ranged from 0.75 to 1.00. Between baseline and maintenance, IRD scores were 1.00 for all students (see Table 2). According to Alresheen et al. (2013), IRD scores between 0.70 and 0.75 are considered large or very large.

Overall, the results of the randomization test, the visual data inspection, and the IRD scores indicate that the use of a story map was an effective strategy for learning and acquiring the key story grammar elements in a story passage for five elementary students with LD. Finally, the IRD scores and visual analysis between baseline and maintenance phases for all students suggest that the effect of the story mapping strategy on the students’ reading comprehension skills continued after the completion of the intervention.

**Discussion**

Over the last few years, the importance of SCR in identifying EBPs has increased, and researchers and practitioners alike now often resort to this kind of quality control.
when evaluating the effects of a given intervention. However, in order to make sound instructional decisions, it is crucial to analyze and interpret a data set as objectively as possible. Conventional visual inspection often leaves too much to the discretion of the researcher in terms of drawing conclusions about the efficacy of an intervention. Thus, recent research has attempted to supplement traditional visual inspection of data from single-case research designs with statistical analysis (see Gage & Lewis, 2013, for a review).

The purpose of this study was to illustrate the use of a randomization test using a story map to promote the reading comprehension skills of five elementary students with LD. Results indicated that the participants increased the number of story grammar elements answered correctly from the baseline (overall $M = 1.90$) to the intervention (overall $M = 8.53$) and maintenance phases (overall $M = 9.07$). Moreover, the newly obtained skills were sustained after the instruction was withdrawn at levels near to those obtained in the Test phase during intervention. Based on the students’ data, the results of the randomization test showed a significant difference between the baseline and the intervention mean scores at a 1% significance level, which rejects the null hypothesis of no intervention effect. Moreover, the two other procedures applied to measure the effectiveness of the intervention (i.e., visual inspection and effect size calculation) also suggested that story mapping is an effective strategy to recall and comprehend the key story grammar elements within a story passage. Together, the evidence suggests a large effect that positively impacted student comprehension outcomes related to story grammar.

**Limitations**

The study’s findings show promising results; however, several limitations must be considered in this replication study. First, the sample size was small, which is a general limitation of SCR. The present study consisted of five elementary students with LD, which limits the generalizability of the findings. Second, differences between the duration of the baseline and maintenance sessions (15 minutes) and intervention sessions (30 minutes) may have positively affected students’ reading comprehension scores during the intervention phase. Nevertheless, participants’ comprehension levels in the maintenance phase were similar to those achieved during the Test phase. Third, inter-rater reliability measures were conducted by the two interventionists, which may have introduced a bias in scoring. It would have been more appropriate to have external observers independently score the students’ answers to the comprehension questions. Fourth, no formal procedural reliability measures were conducted. However, the first author met on a regular basis with the interventionists to ensure that the procedures were consistently delivered as planned. Finally, although the randomization test was able to substantiate a strong intervention effect triggered by the treatment, an important limitation of this method of analyzing data from randomized SCR designs must be considered.
in relation to slope effects. Wilbert (2014) was able to demonstrate that the randomization test is not sufficiently sensitive to slope effects. This limitation did not apply to our study, because the participants responded quickly to the intervention. However, in cases where the indicators of a treatment effect change more slowly, randomization tests might not be adequate for analyzing data from SCR designs. Thus, even though a significant effect of the story mapping strategy was observed in our study through the use of a randomization test, such an approach may not always be feasible for evaluating methods that elicit rather gradual responses.

Implications

In summation, the present study confirms insights from previous research on the positive effects of using a story map to improve the reading comprehension skills of students with LD. In addition, our findings also corroborate the results of Grünke et al. (2013) indicating that elementary students with moderate information-processing and poor comprehension skills, but with proficient decoding and fluency abilities can benefit from a rather short story mapping intervention of less than 12 sessions.

This entails some important implications for working with students comparable to the ones in our study. Specifically, it is possible to significantly assist students with average or above-average reading fluency but poor comprehension skills if key skills are targeted related to comprehension outcomes. Only about 10 lessons were needed to master the phase using the story mapping strategy. In most cases, such a short intervention can be embedded in typical reading instruction, thus preventing students from falling behind their classmates.

With regard to the randomization test, our study mainly focused on illustrating the technique. Thus, we demonstrated that this procedure can easily be implemented when conducting single-case analyses. In our case, the number of probes for the baseline and the intervention phase was only 15. Thus, the expenditure of time needed to conduct the experiment was minimal while still yielding a statistically significant result. As Dugard (2013) pointed out, the benefits of this method derive not only from the chance to verify statistically significant treatment effects, but the explanatory power of visual inspection also improves because the procedure requires the researcher to choose the intervention point at random.

Todman and Dugard (1999) noted that interpreting SCR data is problematic, leading to false conclusions when determining the differences between phases if no random assignment of treatments or conditions is used. The randomization test is one of the few methods that have the potency to validly detect statistically significant effects in data from SCR designs. Since this procedure can now easily be performed in familiar computational environments, it is hoped the approach will become more widely used among researchers who are trying to identify effective interventions.
References


Wilbert, J. (2014, August). *Which technique is appropriate for analyzing single-case AB designs?* Paper presented at the biannual Special Educational Needs Conference of the European Association for Research on Learning and Instruction (EARLI), Zürich, Switzerland.


Developing a Comprehensive Mathematical Assessment Tool to Improve Mathematics Intervention for At-Risk Students

Jonathan Brendefur, Evelyn S. Johnson, and Keith W. Thiede
Boise State University

Everett V. Smith
University of Illinois, Chicago

Sam Strother
Boise State University

Herbert H. Severson
Oregon Research Institute

John Beaulieu
Visual Health Information, Tacoma, WA

Abstract

Students who complete kindergarten with an inadequate knowledge of basic mathematics concepts and skills will continue to experience difficulties with mathematics throughout their elementary and secondary years and may be at increased risk for math disabilities. There is a critical need to identify students experiencing difficulties in mathematics in the early elementary grades and to provide immediate and targeted instruction to remediate these deficits. Most early math screening tools focus on only a single skill, resulting in an incomplete picture of student performance and limited predictive validity. To address this need, we are developing a multiple-gating system of math assessment, the Primary Math Assessment (PMA), that both screens and provides diagnostic information in six domains. In this study, we present the results of the development and validation of items across the domains that will comprise
the PMA. Multidimensional Rasch models were used to estimate theoretically plausible
dimensionality structures. Parsimony fit indices supported the six-dimensional model as the
most generalizability model for the PMA data and supports reporting of six separate scores.

A recent review of early math screeners reported that virtually all screeners for the
primary grades rely on assessing aspects of number sense (Gersten et al., 2012). These
screening tools are used to identify students in need of more intensive math intervention as
well as to identify students at risk for math disability. Two important issues present potential
constraints on the efficacy of number sense screeners to adequately serve these purposes.

First, adoption of the Common Core State Standards (CCSS) in Mathematics has led to an
expanded math curriculum that includes a much broader set of math content and process standards,
especially in the primary grades (National Governors Association Center for Best Practices &
Council of Chief State School Officers, 2010). The Common Core is a set of high-quality academic
standards outlining what a student should know and be able to do at the end of each grade. In math,
the standards draw on the most effective international models for mathematical practice and include
not only number and operations, but also algebraic thinking, measurement and data analysis, and
geometry standards. Many primary-grade teachers do not have the mathematics knowledge to
accurately identify students’ needs across these more complex math skills (Hill, Rowan, & Ball,
2005) and, therefore, rely on established screening instruments to inform their decisions. Because
the validity of instructional decisions depends to an extent on the alignment between the screening
measures and the content standards on which classroom instruction is based (Irvin, Park, Alonzo,
& Tindal, 2012), screening instruments that only assess number sense will not inform teachers of a
student’s skills across the other dimensions.

Second, math disability (MD) has typically been described as a core deficit in processing
numerical quantity (Butterworth, Varma, & Laurillard, 2011) and number sense (Piazza et al.,
2010; Wilson & Dehaene, 2007). Consistent with this description, the development of tools to
identify children at risk for math disabilities has concentrated primarily on aspects of number
sense (Gersten et al., 2012). However, a recent review of cognitive theories and functional imaging
studies suggest that a model of number sense deficits as the unitary source of math disability is an
oversimplification (Ashkenazi, Black, Abrams, Hoeft, & Menon, 2013). Specifically, in addition
to a deficit in number sense, MD has also been described as (a) a specific impairment in symbolic
processing and visual-spatial reasoning (Rousselle & Noel, 2007); (b) a domain-general deficit in
working memory (Geary, 2004; Swanson, Howard, & Saez, 2006); and (c) a hybrid of impairments
representing and manipulating numerical magnitude on an internal number line and in working
memory and attention (Ashkenazi et al., 2013). Additionally, a majority of children with MD have
comorbid math and reading disabilities that result in significant difficulties with word problems (von Aster & Shalev, 2007). Each of these subtypes of math disability may result in impaired performance across a variety of math constructs and applications that are critical to identify in order to develop and implement interventions that meet an individual student’s area of need.

Taken together, an increased understanding of various subtypes of math disability as well as significant changes to the math curriculum in the early grades call for screening and diagnostic tools that are more comprehensive than those currently available.

**Improving Early Math Screening: The Primary Mathematics Assessment**

The Primary Mathematics Assessment (PMA) (Brendefur & Strother, 2010) is being developed to address the major limitations of current early math screeners. The PMA is an assessment system designed for use in grades K-2 to identify students at risk for poor math outcomes across six dimensions and to provide further diagnostic information to guide intervention decisions. The PMA is designed as a multiple-gating system, in which students are first screened using the Primary Mathematics Assessment-Screener (PMA-S). Students who are identified as at risk on the screener are further assessed using the PMA-Diagnostic (PMA-D), a diagnostic assessment that provides a more complete evaluation of student performance to support intervention planning.

Multiple-gating approaches for identifying deficits in mathematics offer a promising solution to the problems with the “direct route” model of screening, in which intervention decisions are made based on the results of a screening test (Johnson, Jenkins, Petscher, & Catts, 2009). The goal of multiple gating is to administer a series of sequential assessments, in order to quickly assess a large population and identify students who have a high probability of being at risk for poor math performance. More in-depth evaluations are then used to confirm initial screening results and to provide a comprehensive analysis of a student’s needs, which can then be used to inform intervention efforts. Multiple-gating approaches have been successfully applied to behavioral screening (Walker, Small, Severson, Seeley, & Feil, 2014) but are less commonly applied to academics. This is unfortunate because use of multiple-gating systems has demonstrated a reduction in intervention resource consumption by as much as 58% compared to single-stage screening procedures (Loeber, Dishion, & Patterson, 1984).

The PMA is hypothesized to measure six dimensions of math – number sequencing, operations (number facts), contextual problems, relational thinking, measurement, and spatial reasoning – that align closely with the Common Core State Standards in math and have been found to be highly predictive of later math achievement. A more complete review of the research supporting their importance for successful math achievement is presented elsewhere (Brendefur, Thiede, & Strother, 2012). Below, we provide a definition and brief synopsis of the research supporting each of the dimensions as critical components of early mathematics achievement.
**Number sense/sequencing.** Number sense has been suggested as the most important area of mathematical learning in early childhood (Clements & Sarama, 2007). This domain includes subitizing small quantities without counting, counting items in a set and knowing the final count word tells how many, discriminating between small quantities, comparing numerical magnitudes, and transforming sets of five or less by adding or taking away items (Jordan, Glutting, & Ramineni, 2008). A key component of number sense is counting or sequencing (Baroody, 1987). Counting has been described as the bridge between innate number sense and more advanced arithmetic abilities (Butterworth, 2004; Desoete, Ceulemans, Roeyers, & Huylebroeck, 2009). Given its role as a bridge to more advanced mathematics, sequencing is considered a prerequisite for future mathematical strategies such as basic operations (Blöte Lieffering, & Ouwehand, 2006; LeFevre et al., 2006). Several researchers (Geary, 2010; Geary, Hoard, & Hamson, 1999; Jordan, Glutting, & Ramineni, 2010) have found that difficulty in counting and other number sense deficits in early childhood is strongly predictive of later math achievement and should, therefore, be included as a key dimension on early math screeners.

**Number facts.** Basic math operations include the ability to add, subtract, multiply, and divide single digit numbers to 10. Fluency with math operations is critical for math achievement throughout students’ school careers. Students with or at risk for math disabilities often have difficulty with fact retrieval, accurate computations (Geary, 2004) and flexibility, or the ability to solve problems in a variety of ways (Beishuizen & Anghileri, 1998). Immature calculating strategies, problems retrieving facts (Geary, 2004), and executive deficits (Passolunghi & Siegel, 2004) can prevent students from developing fluency with number facts (Geary, 2004), and result in more severe math learning challenges throughout school.

**Contextual problems.** Accurately solving contextualized problems is a key factor in early mathematics achievement, and word problems are a significant part of elementary math curricula. Contextualized problems serve as a means of developing students’ general problem-solving skills and can promote proficiency with whole-number arithmetic (Verschaffel, Greer, & DeCorte, 2007). As described by Jitendra et al. (2013), contextualized problem solving requires the ability to understand the underlying problem type and related problem-solving procedures for that class of problems (Hatano, 2003), strong metacognitive skills (Montague, 2007), and the ability to distinguish relevant information (related to mathematical structure) from irrelevant details (Van Dooren, de Bock, Vleugels, & Verschaffel, 2010). Students at risk for learning disabilities that impact both math and reading (e.g., comorbid MD + RD) typically have difficulty solving word problems (Ashkenazi et al., 2013).

**Relational thinking.** Relational thinking describes the thinking of students who use number and operation sense to reflect on mathematical expressions as objects rather than as arithmetic procedures to be carried out (Carpenter, Franke, & Levi, 2003). As such, relational thinking is a
precursor to the development of algebraic thinking. Sarama and Clements (2008) stressed the importance
of recognition and analysis of patterns in the early years to bring order and facilitate generalizations
in math. An example of relational thinking provided by Stephens (2006) suggests that a student who
is thinking relationally is able to recognize the equivalence of \(3(x+4)\) and \(3x+12\) by attending to their
structures without the need to solve the problem. The ability to find and extend numerical patterns to
develop relational thinking is heavily dependent on how students are taught and should be a significant
component of children’s learning of mathematics (Sarama & Clements, 2009). Briefly, relational
thinking is developed through helping students understand the equal sign (Driscoll, 1999), recognizing
that equality is preserved if equivalent transformations are made on both “sides” of an equation, and can
be fostered by posing true/false and open number sentences (Carpenter et al., 2003).

**Measurement.** Measurement of length has a direct link to understanding fractions and
decimals because measurements often do not use complete units (Cramer, Post, & del Mas, 2002;
Lehrer, 2003; Watanabe, 2002). For example, a table can be 3-½ feet wide. Students must make
sense of the part of the unit left over after the three complete units are counted. Through this process,
students develop a model for the continuous nature of rational numbers, which supports learning about
fractions and ratios in later grades (Lehrer, Jaslow, & Curtis, 2003; McClain, Cobb, Gravemeijer, &
Estes, 1999). Measurement tasks also support stronger proportional reasoning, which in turn supports
understanding of geometry, numeracy, and data analysis (National Research Council, 2001). The
underlying principles of measurement are unit iteration, partitioning, comparative measurement,
and the meaning of measurement. Unit iteration is the act of repeating a unit to measure an object’s
attributes. Partitioning is the act of breaking an object into equal-sized measuring units (Lehrer, 2003).
Finally, comparative measurement is the process of using a known measurement from one part of an
object to find an unknown measurement (Kamii & Clark, 1997).

**Spatial reasoning.** Spatial reasoning is strongly correlated with achievement in math
(Battista, 1981; Clements & Sarama, 2007; Gustafsson & Undheim, 1996). Students who perform
well on spatial tasks also perform well on tests of mathematical ability (Geary, Hoard, Bryd-
Spatial reasoning involves (a) spatial visualization, or the ability to mentally manipulate, rotate,
twist, or invert pictures or objects; (b) spatial orientation, or the ability to recognize an object
even when its orientation changes; and (c) spatial relations, or the ability to recognize spatial
patterns, understand spatial hierarchies, and imagine maps from verbal descriptions (Lee, 2005).
Recent evidence indicates that spatial reasoning training can have transfer effects on mathematics
achievement, particularly on missing term problems (e.g. \(7 + _\) = 15), which are important for
developing algebraic understanding (Cheng & Mix, 2014).

As demonstrated through this review of math constructs, a multidimensional measure
of early math ability that can be efficiently administered and interpreted by elementary teachers
would allow more children with math deficits across a variety of important areas to be identified for intervention before these deficits begin to negatively affect math achievement. Additionally, a comprehensive measure such as the PMA would assist with intervention planning for children with specific deficits in critical math dimensions by providing a better match of intervention strategy to the demonstrated need.

Purpose of the Study

To develop a screening and diagnostic tool that adequately addresses the issues with existing math screeners, several phases of research have to be conducted. First, the psychometric qualities of the data, including estimates of reliability and investigations of dimensionality, need to be established. Next, the predictive validity of the screening tool needs to be established so that decision rules about performance can be tied to meaningful outcomes. Finally, the treatment validity of the assessment needs to be determined; that is, the extent to which the assessment contributes to positive outcomes (Gersten, Keating, & Irvin, 1995). Thus, there must be a clear and unambiguous relationship between the assessment data collected and the intervention that is recommended.

The current study reports on the development and validation of items for the PMA-S and PMA-D. The specific aims of the research were to:

1. Develop and determine the best set of items for assessing student ability within each of the six dimensions.
2. Assess the dimensionality of the PMA.
3. Determine the reliability of each of the six subscales.

Method

Participants

Students in kindergarten through second grade from seven schools within three districts in the Mountain Northwest participated in this project. All schools qualify for schoolwide Title 1 programs, with 40% or more of the student population eligible for free or reduced-price lunch; the number of students in kindergarten through sixth grade ranged from 350 to 450. All students were invited to participate, and those who returned parent consent forms participated. Across schools, between 70-74% of eligible students participated.

To prevent over-testing, not all participants responded to all of the questions. Classrooms were randomly assigned to complete two of the six dimensions. Three of the schools completed the number facts dimension in addition to their assigned dimensions. This
resulted in an uneven number of students completing the items to various dimensions. The demographics for the sample of students who participated in the data collection are presented in Table 1 by dimension. Students receiving special education services (i.e., students with individualized education program plans [IEPs]) were served under the category of speech or language impaired.

Table 1
Demographics by Subscale

<table>
<thead>
<tr>
<th></th>
<th>Number Sequencing</th>
<th>Number Facts</th>
<th>Contextual Problems</th>
<th>Relational Thinking</th>
<th>Measurement</th>
<th>Spatial Reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students (total)</td>
<td>97</td>
<td>232</td>
<td>124</td>
<td>112</td>
<td>119</td>
<td>131</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade Level</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td>28</td>
<td>61</td>
<td>37</td>
<td>31</td>
<td>36</td>
<td>39</td>
</tr>
<tr>
<td>First Grade</td>
<td>35</td>
<td>106</td>
<td>37</td>
<td>45</td>
<td>41</td>
<td>53</td>
</tr>
<tr>
<td>Second Grade</td>
<td>34</td>
<td>65</td>
<td>37</td>
<td>36</td>
<td>42</td>
<td>39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>42%</td>
<td>39%</td>
<td>47%</td>
<td>46%</td>
<td>47%</td>
<td>55%</td>
</tr>
<tr>
<td>Female</td>
<td>58%</td>
<td>61%</td>
<td>53%</td>
<td>54%</td>
<td>53%</td>
<td>45%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unspecified</td>
<td>14%</td>
<td>34%</td>
<td>27%</td>
<td>8%</td>
<td>29%</td>
<td>20%</td>
</tr>
<tr>
<td>American Indian</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Asian</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>Black</td>
<td>2%</td>
<td>0%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Native</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>White</td>
<td>66%</td>
<td>43%</td>
<td>56%</td>
<td>70%</td>
<td>49%</td>
<td>50%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>16%</td>
<td>19%</td>
<td>14%</td>
<td>17%</td>
<td>16%</td>
<td>24%</td>
</tr>
<tr>
<td>Multiracial</td>
<td>0%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
</tr>
</tbody>
</table>

| ELL                  | 0%                | 0%           | 0%                  | 1%                 | 0%          | 1%                |
| IEP                  | 0%                | 0%           | 1%                  | 4%                 | 0%          | 3%                |

*Note.* ELL = English language learners. IEP = students with individualized education program plans (i.e., special education).
**Procedures**

The development and testing of the PMA took place over a nine-month period. A total of 148 items were created and dispersed across the six dimensions as follows: (a) 23 number sequence items, (b) 34 number fact items, (c) 10 context items, (d) 25 relational thinking items, (e) 25 measurement items, and (f) 31 spatial reasoning items. Items were administered to between 97 to 232 students in grades K-2, depending on the dimension. Rasch analysis (described in more detail in the data analysis section) allowed us to determine whether items fit the model requirements. Items were distributed across the six dimensions, with some items linked across grade levels as outlined in Table 2. Linked items are common items administered to more than one grade level so the calibration process would be able to place all items and persons across grades on a common metric.

### Table 2

**PMA Items Linked by Grade Level**

<table>
<thead>
<tr>
<th>Items per test/grade level</th>
<th>Sequencing</th>
<th>Facts</th>
<th>Relational Thinking</th>
<th>Context</th>
<th>Measurement</th>
<th>Spatial Reasoning</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items linked K – 1st grade</td>
<td>9</td>
<td>6</td>
<td>22</td>
<td>10</td>
<td>19</td>
<td>29</td>
<td>148</td>
</tr>
<tr>
<td>Items linked 1st – 2nd grade</td>
<td>1</td>
<td>4</td>
<td>21</td>
<td>10</td>
<td>24</td>
<td>29</td>
<td>27</td>
</tr>
<tr>
<td>Items linked K – 1st-2nd grade</td>
<td>0</td>
<td>1</td>
<td>17</td>
<td>10</td>
<td>18</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

**Measures**

The PMA is designed as a multiple-gating system for students in kindergarten through second grade. The item bank will be used to develop a PMA-Screener and a PMA-Diagnostic. Students whose performance on the PMA-Screener indicates they may be at risk for poor performance in one or more dimensions will be evaluated using the PMA-Diagnostic in order to get a fuller evaluation of their abilities. Currently, we have developed and validated a total of 148 items across each of the six dimensions.

**Data Analysis**

Rasch models were used to analyze and evaluate the data. The use of Rasch models allows items and persons to be arranged in order of difficulty and ability, respectively, along a common metric, which in turn enables direct comparisons both between and across individuals and items. The metric can also be maintained across time points, which is necessary for understanding
which students require intervention and for calculating meaningful change scores over time once intervention has been provided. In addition, the Rasch model provides fit indices that aid in identifying items that may not contribute to measurement of the underlying dimension or latent trait measured. Finally, Rasch analysis provides person reliability indices that are analogous to internal consistency coefficients (KR20 and alpha, see Smith [2001] for why person reliability is a more accurate estimate of internal consistency than traditional estimates).

To determine the best fitting items to include on the PMA, the Rasch model were used to identify items that did not contribute to a unidimensional construct. For each subscale, we used WINSTEPS v3.8 (Linacre, 2014) to fit the data to a dichotomous Rasch (1-Parameter Logistic) model. This allowed us to (a) evaluate whether the items measured the desired constructs using item fit statistics (misfitting items were then revised or eliminated), (b) establish the internal consistency reliability of the subscales, and (c) use item Wright maps, provided by WINSTEPS to analyze the distribution of item difficulty with respect to the distribution of children’s ability and remove items that were too difficult or too easy (i.e., poor targeting between item difficulty and children’s ability).

To investigate the overall dimensionality of the data, multidimensional Rasch models for dichotomous data (Smith & Smith, 2004) were used for all Rasch modeling. The program used to estimate parameters for the multidimensional models was Conquest (Wu, Adams, & Wilson, 1997). We hypothesized the six dimensions of the PMA were statistically distinguishable. However, we also tested whether the dimensions could be combined to form a theoretically supported two-dimensional model consisting of Measurement and Spatial items as Dimension 1 and Sequencing, Facts, Contextualized Problems, and Relational Thinking as Dimension 2. Both hypothesized structures as well as a unidimensional model were evaluated using Conquest, which is capable of fitting multidimensional extensions of most basic unidimensional Rasch models using the Multidimensional Random Coefficient Multinomial Logit (MRCML) Model (Wu et al., 1997). Specifically, all three models were estimated and compared for relative model fit. However, the statistically best fitting model (i.e., minimizing the -2 log likelihood) does not mean that the identified model will be the model that generalizes the best as the model could overfit the data. As such, fit indices that take into account model complexity (e.g., number of parameters and/or number of observations) were implemented. These indices have collectively been labeled parsimony fit indices. The four parsimony fit indices employed for the dimensionality assessment were Akaike’s Information Criterion (AIC; Akaike, 1974), the sample corrected Akaike Information Criteria (AIC-C; Burnham & Anderson, 2002), the Bayesian Information Criterion (BIC; Kass & Wasserman, 1995), and the Consistent Akaike Information Criterion (CAIC; Bozdogan, 1987). Lower values for these fit indices indicate the best tradeoff between model fit and generalizability.
**Results**

For each subscale, we first identified and removed misfitting items and investigated targeting issues. Item fit indices are expressed as mean squares, which represent the average value of squared residuals for each item, calculated from the difference between Rasch-predicted item performance and actual item performance in the observed data (Bond & Fox, 2013). Thus, larger mean square values represent poorer item fit with the Rasch model. The unstandardized unweighted mean square fit (MNSQ outfit) values have an expected value of 1. Values less than 1 indicate possible item redundancy or model overfit, whereas values greater than 1 indicate unpredictability or model underfit. In standardized form (ZSTD outfit), the expected value is 0 and approximates a unit normal distribution. Items for which the MNSQ outfit statistic was <1.3 or >0.7 and for which the ZSTD outfit was <-2.00 or >-2.00 were considered to be fitting satisfactory (Bond & Fox, 2013). We had an uneven number of participants across the various dimensions; however, mean square statistics have been found to be relatively independent of sample size when using polytomous data (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2008).

Once the final set of items was developed, Wright maps displaying the item difficulty and student ability for each dimension were created (see Figures 1-6). As a means of explanation, we interpret the Wright map for measurement (see Figure 5). It displays the student ability measures expressed in logits (short for “log odd units,” which result from applications of Rasch models) in a histogram on the left and the item difficult parameters on the right of the scale. The M, S, and T on either side of the vertical axis represent the mean, one standard deviation, and two standard deviations, respectively. The higher the student ability measure, the more able the student; the higher the item measure, the more difficult it is to get the item correct. Figure 5 indicates the items provide good coverage (e.g., targeting) for this sample of students, which helps contribute to a relatively small standard error of measurement for the student ability parameters. When item difficulties did not cover the range of student abilities, items were reviewed. For example, if there were too many students at the top of the map (ceiling effects) and too few items targeted toward these higher ability students, more difficult items were constructed.
Figure 1. Wright map for number sequencing.
**Figure 2.** Wright map for number facts.
Figure 3. Wright map for contextualized problems.
Figure 4. Wright map for relational thinking.
### Figure 5. Wright map for measurement.

```
<table>
<thead>
<tr>
<th>MEASURE</th>
<th>PERSON - MAP - ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>M27 F26</td>
</tr>
<tr>
<td></td>
<td>T</td>
</tr>
<tr>
<td></td>
<td>?20 M25</td>
</tr>
<tr>
<td>2</td>
<td>M28 M28</td>
</tr>
<tr>
<td></td>
<td>F15 M16 F25 M25 F25 F25</td>
</tr>
<tr>
<td></td>
<td>M15 M15 ?20 M25 M25</td>
</tr>
<tr>
<td>3</td>
<td>Meas25</td>
</tr>
<tr>
<td></td>
<td>M05 F05 ?10 M15 M15 M15 F25</td>
</tr>
<tr>
<td>0</td>
<td>M06 F06 F06 F15 F05 F05 M05 M15 M15 F15</td>
</tr>
<tr>
<td></td>
<td>M06 F05 F05 M05 M15 ?20 M25 F25</td>
</tr>
<tr>
<td></td>
<td>F06 F05 F05 F15 F05 F05 F15 F05 M05 M15 M05 F15</td>
</tr>
<tr>
<td>1</td>
<td>Meas30 Meas38</td>
</tr>
<tr>
<td></td>
<td>M06 F06 F06 ?10 F00 M06 ?10 F16</td>
</tr>
<tr>
<td></td>
<td>?00 F05 ?10 ?20</td>
</tr>
<tr>
<td>-2</td>
<td>M05 F05 M05</td>
</tr>
<tr>
<td></td>
<td>?00 ?00 T</td>
</tr>
<tr>
<td>-1</td>
<td>M06 M06</td>
</tr>
<tr>
<td>-2</td>
<td>M06 M06</td>
</tr>
<tr>
<td>-3</td>
<td>T00</td>
</tr>
<tr>
<td></td>
<td>&lt;less&gt;..&lt;frequent&gt;</td>
</tr>
</tbody>
</table>
```
Figure 6. Wright map for spatial reasoning.

Rasch reliability. Rasch estimates of internal consistency reliability for items (see Table 3) and students (see Table 4) were also used to determine the quality of the data. For both persons and items, reliability of .70 to .79 is considered acceptable, .80 to .89 is good, and .90 or greater is excellent (Duncan, Bode, Lai, & Perera, 2003). As illustrated, in the current study, reliability for the various dimensions ranged from .74 to .87 for persons, and from .69 to .93 for items.
Table 3

Summary Statistics for Person Measures by Subscale on the Primary Math Assessment

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Infit</th>
<th>Outfit</th>
<th>Separation</th>
<th>Person Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MNSQ</td>
<td>ZSTD</td>
<td>MNSQ</td>
<td>ZSTD</td>
</tr>
<tr>
<td>Number Sequencing</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 97)</td>
<td>1.00</td>
<td>.0</td>
<td>1.06</td>
<td>.1</td>
</tr>
<tr>
<td></td>
<td>.17</td>
<td>.7</td>
<td>.60</td>
<td>.9</td>
</tr>
<tr>
<td>Number Facts</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 232)</td>
<td>1.00</td>
<td>.0</td>
<td>1.02</td>
<td>.1</td>
</tr>
<tr>
<td></td>
<td>.30</td>
<td>1.0</td>
<td>.69</td>
<td>1.0</td>
</tr>
<tr>
<td>Context (n = 124)</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.96</td>
<td>-.2</td>
<td>1.11</td>
<td>.5</td>
</tr>
<tr>
<td></td>
<td>.13</td>
<td>1.0</td>
<td>.37</td>
<td>1.0</td>
</tr>
<tr>
<td>Relational Thinking</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 112)</td>
<td>1.00</td>
<td>.0</td>
<td>1.06</td>
<td>.0</td>
</tr>
<tr>
<td></td>
<td>.20</td>
<td>1.0</td>
<td>.56</td>
<td>1.0</td>
</tr>
<tr>
<td>Measurement</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 119)</td>
<td>1.00</td>
<td>.0</td>
<td>1.00</td>
<td>.0</td>
</tr>
<tr>
<td></td>
<td>.15</td>
<td>.8</td>
<td>.26</td>
<td>.9</td>
</tr>
<tr>
<td>Spatial Reasoning</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 131)</td>
<td>.99</td>
<td>.1</td>
<td>.99</td>
<td>.1</td>
</tr>
<tr>
<td></td>
<td>.23</td>
<td>.7</td>
<td>.47</td>
<td>.8</td>
</tr>
</tbody>
</table>

Note. MNSQ = unstandardized unweighted mean square fit values. ZSTD = standardized unweighted mean square fit values.

Table 4

Summary Statistics for Item Measures by Subscale on the Primary Math Assessment

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Infit</th>
<th>Outfit</th>
<th>Separation</th>
<th>Person Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MNSQ</td>
<td>ZSTD</td>
<td>MNSQ</td>
<td>ZSTD</td>
</tr>
<tr>
<td>Number Sequencing</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 23)</td>
<td>.97</td>
<td>-.1</td>
<td>1.02</td>
<td>.1</td>
</tr>
<tr>
<td></td>
<td>.25</td>
<td>1.2</td>
<td>.58</td>
<td>1.1</td>
</tr>
<tr>
<td>Number Facts</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 34)</td>
<td>.00</td>
<td>.33</td>
<td>1.01</td>
<td>.0</td>
</tr>
<tr>
<td></td>
<td>1.15</td>
<td>.14</td>
<td>.19</td>
<td>1.2</td>
</tr>
<tr>
<td>Relational Thinking</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 25)</td>
<td>.00</td>
<td>.25</td>
<td>-.2</td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>.98</td>
<td>.02</td>
<td>1.7</td>
<td>.63</td>
</tr>
<tr>
<td>Measurement</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 25)</td>
<td>.00</td>
<td>.21</td>
<td>.99</td>
<td>-.1</td>
</tr>
<tr>
<td></td>
<td>.75</td>
<td>.01</td>
<td>.19</td>
<td>2.1</td>
</tr>
<tr>
<td>Spatial Reasoning</td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 31)</td>
<td>.00</td>
<td>.23</td>
<td>1.00</td>
<td>.0</td>
</tr>
<tr>
<td></td>
<td>.86</td>
<td>.02</td>
<td>.21</td>
<td>1.6</td>
</tr>
</tbody>
</table>

MNSQ = unstandardized unweighted mean square fit values. ZSTD = standardized unweighted mean square fit values.
**PMA construct structure evaluation.** We hypothesized that a six-dimensional model would best fit the data. However, based on current conceptualizations of math disability as primarily related to number sense, we also hypothesized a two-dimensional model, in which sequencing, facts, context, and relational thinking would reflect one dimension around the construct of number sense and measurement and spatial reasoning would reflect a second dimension related to visual-spatial processing (R. Gersten, personal communication, June 2, 2014).

Table 5 lists the results of the multidimensional model comparisons. As illustrated, all four parsimony fit indices (i.e., AIC, AIC-C, BIC, and CAIC) favored the six-dimensional model, indicating that, among the models evaluated, the six-dimensional model is the most generalizable model for the PMA data. This finding was critical to supporting the conceptual framework of the PMA – the existence of six statistically distinguishable dimensions that can be used to inform a student’s math ability and ultimately inform instruction and intervention needs across these distinct dimensions.

Table 5

<table>
<thead>
<tr>
<th>Dimensionality of the PMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>6 dimensions</td>
</tr>
<tr>
<td>2 dimensions</td>
</tr>
<tr>
<td>Context</td>
</tr>
</tbody>
</table>

Note. AIC = Akaike’s Information Criterion; AIC-C = corrected Akaike Information Criteria; BIC = Bayesian Information Criterion; and CAIC = Consistent Akaike Information Criterion.

**Discussion**

As the evidence underscoring the importance of strong mathematics achievement continues to grow, more schools are realizing the need to begin instruction and intervention programs in the early grades to support students’ growth and performance in math. Many early elementary teachers do not have strong foundations in teaching math, and lack the ability to accurately assess their students’ instructional needs across a range of dimensions. This increases the likelihood that teachers use results from screening tools to inform their instruction. If the tools they use are one-dimensional (e.g., focus on number sense only), this may have the unintended consequence of restricting early mathematics instruction and intervention, leaving students unprepared for the demands of math instruction in later grades and possibly at increased risk of developing significant math learning challenges. Given the comprehensive nature of the
Common Core State Standards in Mathematics and our increased understanding of multiple subtypes of math disability, assessment tools that align with a broader set of mathematical constructs may provide an important way to help teachers ensure that their instruction is meeting the needs of their young students.

To address these concerns, we developed the Primary Math Assessment (PMA). The goal of the PMA is to create a multiple-gating assessment system to support the need for an efficient, accurate but comprehensive evaluation of K-2 students’ math ability so appropriate instructional and intervention decisions can be made. The development of a multiple-gating assessment system requires several stages, including developing items and confirming that the dimensional model is consistent with the theoretical framework of the test, establishing the reliability and the validity of the screening and diagnostic results, and determining whether the use of the system has adequate treatment validity. The specific goals of this phase of PMA development were to (a) develop and determine the best set of items for assessing student ability within each of the six dimensions, (b) assess the dimensionality of the PMA, and (c) determine the reliability of each of the six subscales.

With regard to these goals, our analyses showed that the person reliabilities for relational thinking and measurement were greater than .80, and .81 for number facts, thus indicating that these dimensions can be reliably measured with the current items on the PMA. Other dimensions had person reliabilities ranging from .74 (spatial reasoning) to .78 (sequencing), suggesting that more items may be needed to reliably measure these dimensions. Our analysis of the dimensionality of the PMA indicated that the best fitting model for the PMA is a six-dimensional model, supporting the PMA’s theoretical framework that a comprehensive assessment of multiple dimensions is important for informing student ability and subsequent intervention.

These findings are important in two ways. First, they provide evidence that there are ways to assess a broad set of skills that underlie critical math dimensions predictive of later math success. Second, as a first step in creating a system that will provide a quick, yet comprehensive assessment of student performance across a broad range of skills to inform instruction and intervention, the results reported here represent an encouraging improvement to currently available tools to address the needs of students at-risk for math disability. Being able to assess math performance in the very early elementary grades means that efforts to intervene will likely be more successful.

Limitations

Although the results reported in this manuscript are very encouraging, it is important to note that the data collection was conducted within one state and included seven schools from three districts. Although the demographics of the participating students includes high
percentages of students eligible for free or reduced-price lunch, English language learners (ELL) and a high percentage of Latino students, the demographics of the sample do not reflect those in other areas in the nation. As research and development of the PMA continues, a broader participant pool will be recruited.

Conclusion

It is evident that students in the early grades are not adequately prepared in mathematics (NCES, 2013). Using large data sets and nationally representative samples, several researchers have demonstrated that students who complete kindergarten with an inadequate knowledge of basic math concepts and skills will continue to experience difficulties with math throughout their elementary and secondary years (Duncan et al., 2007). This points to a critical need for early identification of students who are experiencing difficulties in math and subsequently to provide immediate and targeted intervention in order to build foundational skills and knowledge (Chernoff, Flanagan, McPhee, & Park, 2007). However, current screening tools tend to focus on few or single dimensions and are thus inadequate to fully inform early elementary teachers’ instructional planning.

There is a great need and demand for reliable, efficient, and valid primary level math screening and diagnostic tools to identify students with math deficiencies so teachers can intervene with differentiated lessons in order to remediate student deficiencies. The results of this initial development indicate that assessment of the six dimensions included in the conceptual framework of the PMA may be reliably measured in the early grades to promote strong assessment and instructional planning to improve students’ math proficiency. Next steps include developing and assessing new items and creating cut scores for the PMA-S and PMA-D that reliably identify students at risk for poor math achievement.

References


Primary Math Assessment by Jonathan Brendefur, Evelyn S. Johnson, Keith W. Thiede, Everett V. Smith, Sam Strother, Herbert H. Severson, and John Beaulieu


The Role of Mentoring in Fostering Executive Function, Effort, and Academic Self-Concept

Lynn Meltzer  
Research Institute for Learning and Development (ResearchILD), Lexington, MA

Surina Basho  
Morrissey-Compton Educational Center, Redwood City, CA

Ranjini Reddy  
Research Institute for Learning & Development (ResearchILD), Lexington, MA

Katelyn Kurkul  
Merrimack College, North Andover, MA, and ResearchILD

Abstract

This exploratory study examined the impact of an in-school intervention program that blends peer mentoring with executive function strategy instruction for at-risk learners. More specifically, the study focused on evaluating the effects of the SMARTS Executive Function and Mentoring intervention on students’ strategy use, effort, academic self-concept, and resilience. The final sample consisted of 34 at-risk students in grades 9 and 10 from an urban high school. Findings showed that students in engaged peer mentoring relationships demonstrated significantly higher levels of effort and used more executive function strategies in their classwork, homework, and studying. Findings provide preliminary support for blending peer mentoring with executive function strategy instruction for at-risk learners in schools.
Students who are at risk for academic failure, including those diagnosed with learning disabilities (LD), often struggle to complete their work and to perform at a level that reflects their ability and their range of knowledge in the classroom. These weaknesses become increasingly evident in the higher grades when academic performance depends more heavily on students’ use of executive function strategies as well as their ability to maintain the motivation, effort, and persistence needed to master and implement executive function strategies (Meltzer, 2010, 2014; Meltzer & Krishnan, 2007). In addition to their academic struggles, many of these students have limited self-understanding and low self-concept so that they feel isolated from their peers as well as their school communities. For these struggling students, peer support and peer-assisted programs often improve their self-concepts, motivation, attitudes, and academic performance (Fuchs, Fuchs, & Burish, 2000; Mastropieri et al., 2001; Rhodes, 2008; Rhodes, Reddy, Roffman, & Grossman, 2005). School-based peer mentoring programs that are structured and blended with executive function strategy instruction, therefore, provide a potentially powerful approach to improving students’ metacognitive awareness and teaching them executive function strategies that are critically important for the demands of 21st-century classrooms.

This exploratory study was designed to investigate the effects of an in-school intervention program that blends peer mentoring and executive function strategy instruction and its impact on students’ strategy use, effort, academic self-concept, and resilience. This intervention, known as the SMARTS Executive Function and Mentoring program, promotes five core outcomes for students: Strategies, Motivation, Awareness, Resilience, Talents, and Success (Meltzer, Reddy, Kurkul & Greschler, 2013). The goal of SMARTS is to initiate a cycle of success through peer mentoring so that students are motivated to make the effort to use executive function strategies in their schoolwork and to persist in order to make academic gains.

Theoretical Framework

Executive function. Over the years, a broad range of definitions and models have been proposed to explain executive function (Denckla, 2007; Gioia, Isquith, Kenworthy, & Barton, 2002; Goldstein & Naglieri, 2014; Meltzer, 2010). In general, theorists and researchers agree that executive function is an all-encompassing construct or “umbrella term” for the complex cognitive processes that underlie flexible, goal-directed responses in novel or difficult situations (Anderson, 2002).

In the present study, we defined executive function in terms of the cognitive processes underlying goal-directed behavior, namely, goal-setting, organizing, prioritizing, shifting approaches flexibly, accessing information in working memory, and self-monitoring (Meltzer, 2007, 2010, 2014). These executive function processes become increasingly important in middle and high school, when the curriculum content constrains students’ interests and
challenges them at a higher level (Hidi, Renninger, & Krapp, 2004). Many bright and talented students, especially those with learning and attention difficulties, no longer make the effort needed to master the academic load and consequently become less productive or drop out of school (Denckla, 2007).

However, success can be attainable when students use executive function strategies to set realistic goals, focus their effort on reaching those goals, and self-regulate their cognitive, attentional, and emotional processes appropriately (Dawson & Guare, 2010; Denckla, 2007). All students, particularly students with learning difficulties, therefore, need to become strategic learners with strong metacognitive awareness who understand how they think and how they learn (Scruggs, Mastropieri, Berkeley, & Graetz, 2010; Swanson, 2001). They need to be taught executive function and learning strategies explicitly and systematically in the context of reading, writing, math, and the content areas (Deshler, Ellis, & Lenz, 1996; Mastropieri, Scruggs & Marshak, 2008; Meltzer & Basho, 2010; Swanson, 2001).

These strategies are beneficial for all students, but they are essential for students with learning difficulties (Scruggs et al., 2010; Swanson, 2001). In fact, in their meta-analysis of studies focused on research-based strategies to support students in content-area classes, Scruggs et al. (2010) find high effect sizes for systematic teaching of learning and memory strategies, concluding that strategies help students to “think more systematically about the content to be learned.”

**Peer mentoring and peer-assisted learning.** Students who struggle academically often have difficulty making meaningful connections within their school environments, putting them at risk for social rejection, school failure, school dropout, and delinquency (Achilles, McLaughlin, & Croninger, 2007). Many of these students also lose confidence and feel isolated, which often leads to low motivation, inconsistent effort, and poor academic performance.

Peer mentoring and peer tutoring provide a powerful forum for intervening and helping these students (Karcher, 2005; Rhodes & Spencer, 2010). Thus, numerous studies over the past decade have demonstrated the strong influence of peer support and social relationships on students’ motivation, effort, and achievement (Fuchs et al., 2000; Fuchs, Fuchs, Mathes, & Martinez, 2002; Harris & Meltzer, 2015; Karcher, 2005; Regan, Evmenova, Mastropieri, & Scruggs, 2015; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003). Specifically, findings have shown that peer mentoring influences students’ self-esteem, social skills, and how connected they feel to school (Karcher, 2005, 2008, 2009). Peer mentors also help students to understand their struggles in school so that they no longer need to prove that they are “smart” (Karcher, 2005, 2008, 2009; Plata, Trusty, & Glasgow, 2005).

To date, there has been a paucity of formal mentoring programs that address the specific needs of at-risk learners or students with diagnosed learning disabilities (Meltzer, Greschler, Kurkul, & Stacey, 2015; Regan et al., 2015; Scruggs, Mastropieri, & Marshak, 2012). Therefore,
there is a need for more school-based programs that include peer mentoring to help at-risk students to become part of an accepting social community.

The beneficial effects of peer mentoring are often dependent on the quality of the mentor-mentee relationship and there has been an emphasis on the need for more studies that examine these connections (Grossman & Rhodes, 2002; MENTOR/National Mentoring Program, 2005). The limited number of studies to date have shown that youth in high-quality mentoring relationships reported significantly fewer symptoms of depression, higher self-esteem, and less substance abuse problems than youth in low-quality mentoring relationships (Whitney, Hendricker, & Offutt, 2011). Furthermore, in high-risk college students, studies have shown that mentees in positive mentoring relationships (i.e., friendly, respectful bonds and mutual agreement on goals) were more likely to participate in class, seek help from teachers, and persist in school than students with less positive mentoring relationships (Larose, Chaloux, Monaghan, & Tarabulsy, 2010).

Together, these studies suggest that the quality of mentoring relationships is critically important for intervention efficacy and that more research is needed to address this issue. The current study was designed to address this gap by measuring the quality of mentoring relationships in an in-school program and assessing the interactions among students’ use of executive function strategies, level of effort, academic self-concept, and resilience.

The relationships among executive function strategies, effort, academic self-concept, academic performance, and resilience. Over the past 10 years, our understanding of students with learning difficulties has moved beyond a deficit model to one that emphasizes the importance of fostering academic success and resilience. This, in turn, has led to research focused on the intrinsic and extrinsic processes that buffer the potential impact of risk factors in students (Margalit, 2003, 2004). As a result, different research strands have identified specific internal processes (motivation, self-concept, effort) and external factors (social support, connections with peers and family, teachers’ judgments, school placement) that impact academic performance and resilience in at-risk learners (Dweck & Master 2008; Dweck & Molden, 2005; Margalit, 2003, 2004; Raskind, Goldberg, Higgins, & Herman, 1999).

In our previous studies, we have investigated the interactions among self-understanding, self-concept, executive function strategies, effort (internal processes), teachers’ perceptions (an external process), and academic performance (Meltzer, Katzir, Miller, Reddy, & Roditi, 2004a; Meltzer et al., 2004b; Meltzer, Reddy, Pollica, & Roditi, 2004c). Overall, our results showed that a cyclical relationship exists among academic self-concept (students’ perceptions of themselves as competent learners), effort (students’ willingness to work hard), executive function strategy use (students’ use of strategies in their schoolwork), and academic performance (see Figure 1).
Struggling learners with positive academic self-concepts were more likely to work hard and to use learning strategies in their schoolwork than students who showed negative academic self-concepts (Meltzer et al., 2004a, 2004b, 2004c). These findings suggest that when students succeed academically as a result of their effort and strategy use, they begin to value these learning strategies and feel empowered to work harder, which, in turn, leads to positive academic self-concept, a willingness to persist, stronger academic performance, and resilient mindsets (see Figures 1 and 2).
**The SMARTS Executive Function and Mentoring Intervention**

The SMARTS intervention is anchored in a theoretical paradigm (see Figure 2) that blends the teaching of selected executive function strategies (i.e., organizing, prioritizing, shifting flexibly, memorizing, and self-monitoring) with peer mentoring. Such a blended executive function and mentoring program initiates a positive cycle in which students are more motivated to make the effort to use executive function strategies in their schoolwork. This, in turn, builds positive academic self-concept in students and a strategic approach to learning which results in more efficient performance and academic gains.

The current SMARTS intervention study consisted of two components. The first focused on the explicit teaching of executive function strategies that were linked with the curriculum content and that students could apply to their classwork, homework, and test-taking (Meltzer, 2007, 2010; Meltzer & Basho, 2010). The second component focused on the creation of a peer mentoring system to enhance students’ learning and application of these executive function strategies. The major goals of the intervention were to improve students’ academic self-concept, effort, strategy use, and resilience by building a supportive peer mentoring community in schools.

This exploratory study evaluated these interactions in highly engaged vs. less engaged mentor-mentee pairs. The following questions were investigated:

1. Do students in strongly engaged peer mentoring relationships use executive function strategies more frequently in their academic work than students in weak peer mentoring relationships?
2. Do students in strongly engaged peer mentoring relationships show higher levels of effort and willingness to work hard in school compared to students in weak peer mentoring relationships?
3. Do students in strongly engaged peer mentoring relationships show higher academic self-concept and resilience than students in weak peer mentoring relationships?

**Method**

**Participants**

The students in this study attended a small urban public high school in a Northeastern city in the United States. The school had a majority of African American (50.5%) and Hispanic (38.5%) students. Annual student dropout rates were 14%, and student mobility was 31.3%. According to the school’s demographic data, 69% of the students were eligible for free or reduced-price lunch and 29% needed some form of special education. This school was selected for the current study because 30% of the school population comprised students with special
Mentoring, Executive Function, Effort, Self-Concept by Lynn Meltzer, Surina Basho, Ranjini Reddy, and Katelyn Kurkul

needs who were referred by the school district. All the procedures for subject selection and other aspects of the study were consistent with ResearchILD’s IRB policies and were approved by an external IRB committee.

All 9th- and 10th-grade students who were available during the weekly 80-minute advisory/home room period in school took part in the study. The sample included all students with high-incidence disabilities (LD and attention deficit/hyperactivity disorder) as well as students who were struggling academically but had not been formally diagnosed as having learning disabilities. Students with more severe disabilities (autism spectrum disorders, developmental delays) were not included.

At the start of the study, 87 students were randomly assigned to either the SMARTS intervention or the control group. However, a number of challenging situations resulted in an unequal assignment of students to these groups by the end of the intervention. First, student absenteeism increased due to gang violence. In addition, budget cuts by the superintendent at the end of the first semester led to the decision that the school would be closed at the end of the year. As a result, a number of families decided to transfer their children to other schools in the spring. All these factors resulted in a high attrition rate and an uneven number of students in the intervention and control groups. Further, the resulting morale problems for teachers and students had a negative effect on compliance, particularly for the control group teachers whose students were not benefiting from the intervention. Specifically, these control group teachers were not invested in completing their own surveys or distributing surveys to their students. As a result, the control group had an attrition rate of 41% and 13% of the data were incomplete due to incomplete surveys. For the intervention group, there was a 19% attrition rate.

Because of the high rate of student attrition and teacher noncompliance, particularly in the control group, we decided to focus our analyses on pre-/post comparisons for the students who were initially assigned to the SMARTS intervention group and to omit the control group students. This resulted in a final analysis sample of 34 students (see Table 1).

1 The decision was made to remove the students with missing data rather than performing multiple imputations.
Table 1

Descriptive Statistics for the Sample

<table>
<thead>
<tr>
<th>Demographics</th>
<th>SMARTS Program (N = 34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
</tr>
<tr>
<td>Male</td>
<td>27</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
</tr>
<tr>
<td>9th</td>
<td>19</td>
</tr>
<tr>
<td>10th</td>
<td>15</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>14.85 (.82)</td>
</tr>
<tr>
<td>Unweighted GPA (10th Graders)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>2.39 (.95)</td>
</tr>
<tr>
<td>First Marking Period (9th Graders)</td>
<td></td>
</tr>
<tr>
<td>Humanities</td>
<td>Mean D grade</td>
</tr>
<tr>
<td>Math</td>
<td>Mean C- grade</td>
</tr>
<tr>
<td>Special Education Status</td>
<td>14.7%</td>
</tr>
</tbody>
</table>

Note. Unweighted GPA scores were available for 14 tenth graders. First marking period grades were available for 18 of the ninth graders.

Because of the challenging situations that were outside our control, the final analysis sample included an uneven distribution of males and females and an uneven mentor-mentee ratio: 19 mentees (5 females and 14 males) and 15 mentors (2 females and 13 males). Grades from report cards during the first marking period as well as teacher reports indicated that a majority of the students in this study were “at-risk learners.” In fact, 50% of the 9th graders and 31% of the 10th graders had failing grades in the sciences. Similarly, 31% of the 9th graders and 25% of the 10th graders were failing in their humanities courses. Nevertheless, most of these students did not have individualized education program plans (IEPs) (i.e., were not enrolled in special education) due to the school policy whereby students with mild to moderate learning disabilities were taken off their pre-existing IEPs when they entered this particular high school. Instead, the school continued to provide these students with accommodations in their general education classes as part of the regular school day. For all these reasons, we refer to the students in our sample as “at-risk learners” throughout this article.

2 The attrition in the sample also resulted in the formation of two mentor-mentee triads. As the analysis was performed at the level of the student (and not mentor-mentee pair), these triads were included in the sample.

3 Four mentoring pairs were female and 13 were male.
Procedure

The SMARTS intervention. The SMARTS intervention was designed to promote students’ metacognitive awareness and use of executive function strategies. Instructional methodology and content were developed and incorporated into a six-month curriculum. The paradigm that guided the intervention emphasized the importance of metacognitive awareness as the foundation for teaching executive function strategies, which can be strengthened with peer mentoring (see Figure 3).

The SMARTS intervention comprised four components: (a) matching of mentor-mentee pairs, (b) mentor training, (c) executive function strategy instruction, and (d) application of executive function strategies to a school-based group project.

Matching of mentor-mentee pairs: Mentor-mentee matching was based on the criteria developed by Rhodes and Kupersmidt (2009) for establishing strong mentor-mentee relationships, with an emphasis on students’ gender, interests, as well as their self-reported strengths and weaknesses. Before the intervention, 10th-grade students were required to complete a brief survey to assess their interest in becoming mentors. They were also familiarized with the expectations for their roles as mentors, including the importance of a one-year commitment to their mentoring relationships. A number of additional surveys assessed their skills, interests, and self-understanding of their academic strengths and weaknesses. Using all this information, peer mentoring dyads were assigned in ways that maximized students’
abilities to learn from each other. Students with similar profiles of strengths and weaknesses in strategy use and educational performance were matched so that mentors could provide support and modeling to help their mentees apply the executive function strategies they were learning to their weekly homework. To maximize the possibility that peer dyads would develop positive bonds, students were also matched on the basis of their shared interests (e.g., sports, music, video games), an element that has been shown to be critically important for matching mentors with mentees (Rhodes & Kupersmidt, 2009).

Two 10th-grade mentors were transferred to another school after the school closing was announced, with the result that we had to pair two 9th-grade mentees with a 10th-grade mentor who already had an assigned mentee. Even though this decision affected the design of the study, we needed to address the teacher requests that these mentees not be excluded from the intervention program.

**Mentor training:** The first four weeks of the intervention focused on providing systematic mentor training to the 10th-grade mentors while simultaneously building a sense of community for the 9th-grade mentees. Mentor training procedures were developed using the evidence-based guidelines developed by the National Mentoring Partnership (2005). Mentors were initially introduced to the elements of successful mentoring through role-playing; for example, commitment, supporting mentees, and mentor responsibilities. They were also involved in activities that helped promote students’ understanding of their individual learning profiles and feeling of belonging in a community of learners. As a starting point, one of the lessons focused on the Know Yourself strategy sheets, which provided a structured system for encouraging students to think about their strengths and weaknesses and to create visual displays of their learning profiles (see Figure 4).
Mentoring, Executive Function, Effort, Self-Concept by Lynn Meltzer, Surina Basho, Ranjini Reddy, and Katelyn Kurkul

Figure 4. Know Yourself strategy sheet for promoting students’ metacognitive awareness and understanding of their strengths and weaknesses.

Mentors were paired with one another during role-playing sessions so that they could help each other strengthen their self-understanding and metacognitive awareness. They discussed their learning profiles with one another and coached one another as they reflected on the specific strategies that had been helpful to them in different areas. Mentors then discussed ways in which they could use these same approaches and strategies when they were paired with their mentees.

Executive function strategy instruction: This component of the SMARTS intervention provided direct and explicit strategy instruction in key executive function processes: goal-setting, organizing, prioritizing, cognitive flexibility/shifting flexibly, accessing information in working memory, and self-monitoring/self-checking. Each strategy was taught directly and explicitly, after which student mentors and mentees worked together to apply the strategies to school tasks during interactive group activities. Mentors also coached their mentees during activities that were
designed to teach students to apply the specific executive function strategies to their classwork and homework. For example, during the lessons on cognitive flexibility, students were taught strategies for shifting flexibly in the context of reading, writing, and math tasks.

Students were taught the Triple Note Tote strategy (see Figure 5) to help them to organize, prioritize, and shift flexibly from the main ideas to the supporting details and back. They then applied the strategy to school tasks where they needed to extract information from textbooks, homework assignments, and class notes (Meltzer, Pollica, & Barzillai, 2007). After the mentors and mentees were taught the Triple Note Tote strategy in their separate groups, they worked in their peer mentoring pairs to apply the strategy to writing assignments that required summarizing, planning, and note-taking. When studying for history or science tests, mentors were also encouraged to work with their mentees to create study guides using the Triple Note Tote strategy to help them shift between the main ideas and details. Similarly, when completing math homework or studying for math tests, mentors coached their mentees to shift between the different steps for solving word problems.

Figure 5. The Triple Note Tote Strategy: A strategy for shifting flexibly between major concepts or main ideas and relevant details (Meltzer, Greschler, Kurkul, & Stacey, 2015).
Application of executive function strategies to a group project: In the last month of the SMARTS intervention, mentors worked with their mentees to apply these executive function strategies to a school-based group project. Throughout the project, which required students to take notes and to synthesize themes and details, there was an emphasis on collaborative learning and inquiry while applying executive function strategies (goal-setting, organizing and prioritizing, shifting flexibly, memorizing, checking) to the components of the project. Because of the morale problem created by the imminent school closing, the students and SMARTS staff decided that the final academic project would comprise multimedia presentations focused on the theme of giving students a voice in their school.

The SMARTS intervention was implemented once a week during the 80-minute home room period in school, beginning in the fall (October). Students met in two designated classrooms, each staffed by two SMARTS teachers and two assistant teachers. For the first month of the program, 9th and 10th graders were separated and remained in their grade-level groups. The first four sessions focused on mentor training for the mentors and community building for the mentees. For the remaining SMARTS sessions, 10th-grade mentors were paired with 9th-grade mentees for every session.

For the next seven months (23 sessions), weekly SMARTS sessions focused on teaching students selected executive function strategies. Thirteen strategies were taught in the five core executive function areas: goal-setting, organizing, prioritizing, accessing working memory, shifting flexibly, and self-monitoring. Students learned strategies that addressed the key processes relevant for reading comprehension, writing, math problem-solving, completing homework, studying, and taking tests.

For the first half of each SMARTS session (approximately 30 minutes), the SMARTS teacher taught the weekly executive function strategy in the context of academic tasks such as reading comprehension and math. The mentor-mentee pairs then worked together to apply these strategies to structured activities related to their academic work. For example, goal-setting strategies were modeled by mentors, who helped their mentees to identify the steps needed to achieve their goals. Mentoring pairs collaboratively estimated the amount of work involved in major school projects and open-ended homework assignments and selected specific strategies for breaking down tasks into manageable parts, especially when there were multiple deadlines for different assignments. The supportive nature of peer mentoring helped to make goal-setting strategies more meaningful for students. Mentors evaluated their mentees’ goals in an accepting and nonjudgmental way, and offered advice about ways of coping with obstacles. For example, when using the CANDO goal-setting strategy (see Figure 6), students shared ideas about clear, manageable goals as they worked together to set their goals.
Figure 6. CANDO goal-setting worksheet for mentors and mentees to use together.

Measures

Student and teacher surveys were administered for pretesting before the intervention and for posttesting at the end of the school year. Semi-structured interviews were also conducted by SMARTS staff after the final surveys had been completed by students and their teachers.

Measures for Students and Teachers

Metacognitive Awareness System (MetaCOG). The Metacognitive Awareness System (MetaCOG), for use with 9- to 18-year-olds, is a criterion-referenced assessment system that compares students’ and teachers’ perceptions of students’ metacognitive awareness, strategy use, and academic self-concept. Several studies have consistently shown high levels of reliability for the MetaCOG (Meltzer et al., 2004a, 2004b, 2004c; Meltzer & Krishnan, 2007; Miller, Meltzer, Katzir-Cohen, & Houser, 2001).

The MetaCOG comprises five rating scales that allow teachers to compare their own ratings and judgments of students’ effort, strategy use, and academic performance with their students’ self-ratings of the same processes. These strategy ratings focus on academic areas that depend on executive function processes and include written language, homework, studying, and taking tests (Meltzer, Katzir-Cohen, Miller, & Roditi, 2001; Miller et al., 2001). In this study, three of the five MetaCOG surveys were used (ME survey and PR survey for students, TPSE survey for teachers).
The MetaCOG surveys help teachers to understand their students’ learning profiles and help students to develop an understanding of their strengths and weaknesses. Such self-awareness is the foundation for metacognitive awareness and use of executive function strategies.

**Student Measures**

**Motivation and Effort Survey (ME).** The Motivation and Effort Survey (ME), part of the MetaCOG, comprises three surveys that assess students’ self-ratings of their motivation, effort, and strategy use in school. Part 1 consists of 19 items that assess selected components of students’ academic goals, motivation, and effort. Students rate themselves on a 5-point scale (1 = never to 5 = always), with higher scores indicating higher levels of motivation and effort. Part 2 of the ME focuses on students’ self-ratings of their executive function strategy use. Here students rate themselves along a 5-point scale (1 = poor to 5 = strong) on 14 items that sample key executive function strategies in relation to academic tasks (e.g., homework, tests, long-term projects, organization, checking, and making a plan before starting schoolwork). The final part of the ME assesses students’ self-ratings of their academic competence with a single item, “I would rate myself as a … student” (ratings ranging from poor to strong). Higher scores on the ME indicate stronger motivation, effort, use of executive function strategies, and academic self-concept.

An overall level of effort was derived from 14 items on Part 1 of the ME (see also Meltzer et al., 2004a, 2004b, 2004c), which included items such as, “In general, I am a hard worker;” “I spend as much time as I need to get my work done,” and “I don’t give up even when the work is difficult.” This variable was used to represent students’ effort and persistence in their schoolwork. Reliabilities for this subscale were high at pre- and posttest (α = .89 at T1, α = .94 at T2). Single items from Part 2 of the ME were used to index how well the students thought they were using strategies in their schoolwork (i.e., overall executive function strategy use, organization, and checking). A final variable from Part 3 of the ME was used to assess students’ academic self-concept (i.e., whether or not they thought they were good students). Higher values on these variables indicated more positive outcomes. Findings in this study again indicated high overall reliabilities for the ME (α = .94 at T1, α = .94 at T2), which were consistent with previous findings (Meltzer et al., 2001, 2004a, 2004b, 2004c).

**Persistence and Resilience Survey (PR).** For use in this exploratory school-based study, the Persistence and Resilience Survey (PR) was adapted from a longer survey. The PR survey included 11 items that assessed students’ ratings of their persistence in relation to daily challenges in school. On a 5-point rating scale (1 = never and 5 = always), higher scores indicated higher levels of persistence and resilience. The last item was open-ended and required students to indicate what grade they hoped to earn at the end of the school year on selected subjects. Using factor analysis, a resilience index (labeled “resilience”) was created based
on the following items, “I do not let problems stop me from reaching my goals,” “I’m good at bouncing back from a bad grade,” and “When I have a setback, I am optimistic that I can figure it out” ($\alpha = .79$ at T1, $\alpha = .83$ at T2).

In addition to these two MetaCOG surveys, students completed strategy reflection sheets and participated in structured interviews.

**Strategy reflection sheets.** Students completed strategy reflection sheets on four occasions over the course of the SMARTS program. Strategy reflection sheets required students to reflect and describe the strategies they had used the previous week for their classwork, homework, or test preparation. These strategy reflection sheets incorporated a multiple-choice format, structured questions, and open-ended questions that required students to explain their strategy use (see Figure 7). By completing and sharing strategy reflection sheets, students began to understand which strategies worked well for them as well as why, where, when, and how to apply specific strategies.

**Figure 7.** Strategy reflection sheets for writing and test preparation: Multiple-choice and open-ended question formats.
Qualitative responses on the strategy reflection sheets were coded by two raters using a 0-1 rating scale (1 if identified or applied strategy correctly; 0 if not). Inter-rater reliability was high (K = 85% for application and K = 85% for identification). Scores for strategy identification, application, and overall strategy use were each summed across the four different time points, resulting in three outcome variables.

Student Interviews
Students’ perceptions of their mentoring relationships were examined using one-to-one semi-structured interviews with 20 students who were randomly selected at the end of the intervention. Of these students, three were not included in the analysis as they did not complete surveys at follow-up, leaving a final sample of 17 students. The semi-structured interviews were coded by two members of the SMARTS research team for indices of students’ reports about the quality of their mentor-mentee relationships. These interviews were also coded for indices of students’ effort, persistence, executive function strategy use, and metacognitive awareness. Inter-rater reliability was established at 88%.

Teacher Measures
Teacher Perceptions of Students’ Effort Survey (TPSE). Teachers completed the MetaCOG Teacher Perceptions of Students’ Effort Survey (TPSE) for students in the intervention study. The TPSE items were identical to the ME items for students and followed a similar format. Part 1 required teachers to rate students’ effort in the various academic domains on a 5-point Likert scale (1 = never to 5 = always). In Part 2, teachers rated how well the students used executive function strategies and how they performed in reading, writing, math, homework, tests, and long-term projects, all academic tasks that rely on executive function processes (e.g., “He spends as much time as needed to get his work done;” “She does not give up even when the work is difficult”). Teachers also rated students’ overall academic performance in response to the question: “If you had to assign a grade for this student’s overall academic performance, what would this be?” (1 = poor; 5 = strong). Finally, teachers responded to open-ended questions regarding each student’s motivation, effort and academic performance.

An overall index of students’ effort was derived from 14 TPSE items using the same approach as that used for the ME (α = .99 at T1, α = .99 at T2). Thus, teacher-rated outcomes included an overall index of students’ effort, academic self-concept, and strategy use in their schoolwork. Overall reliabilities for the TPSE for the current study were high (α = .98 at T1, α = .99 at T2).

Teacher ratings of quality of mentor-mentee relationships and level of engagement. SMARTS teachers completed a weekly staff session reflection survey in which they rated the engagement of all the mentor-mentee pairs. Each mentor-mentee pair was rated by multiple
raters based on their interactions and their engagement with each other. Interrater reliability was 89%. A score of 1 was given to pairs who worked well together. These student pairs shared more interests, had more frequent positive verbal exchanges, and stayed on task for over 90% of the time. A score of 0 was given to pairs whose interactions were not positive and showed markedly less engagement. These students sat together but did not interact with one another and needed frequent teacher prompting to complete the tasks. Three ratings were obtained for each mentor-mentee pair, and a mean score was obtained for each rating.\(^4\) Ratings were used to categorize mentor-mentee pairs into two groups: those who showed positive relationships with strong engagement and those who did not.

Data Analysis

To study the effects of the quality of mentor-mentee relationships on strategy use, effort, academic self-concept, and resilience, students were classified as being in strongly engaged vs. weakly engaged mentor-mentee relationships. Teacher ratings of mentor-mentee engagement were analyzed separately from students’ self-ratings of their engagement in their mentor-mentee pairs. When teacher ratings were used, the sample comprised all 34 students. When student ratings were used, the sample comprised 17 students who were randomly selected for the semi-structured interviews. More specifically, teachers’ vs. students’ ratings of mentor-mentee engagement were calibrated as follows.

Teacher ratings of mentor-mentee engagement. SMARTS teachers’ observations and ratings of mentor-mentee interactions and engagement were used to categorize students into strongly engaged vs. poorly engaged mentoring relationships. Strong levels of mentor-mentee engagement were defined in terms of the extent to which mentors and mentees bonded, showed interest in each other, and displayed enthusiasm when they worked together to apply executive function strategies to academic tasks. Nineteen students were classified as belonging to “strong” and “engaged” peer mentoring relationships and 15 students were in “weak” or “poorly engaged” relationships, where mentors and mentees did not engage readily with each other. Initial analyses of these subgroups showed no gender or grade-level differences between the groups.

Students’ ratings of their engagement in their mentor-mentee pairs. To assess students’ perspectives about their engagement in their mentor-mentee pairs, analyses focused on the one-on-one semi-structured interviews with the mentors and mentees. The moderating effects of the quality of mentor-mentee relationships on students’ use of executive function strategies in their schoolwork were also analyzed. As mentioned, a subset of 17 students from the 34 students in the SMARTS intervention group were randomly selected to participate in

\(^4\) Initial exploration of the data showed that staff ratings changed relatively little over time. The decision was, therefore, made to average the data over time. Observations of mentor-mentee pairs started once the dyads were formed and started working on curriculum-based projects.
these student interviews. Twelve of these students were from dyads that showed high levels of engagement and worked well together (five mentees and seven mentors). Five of the students were from mentor-mentee pairs that showed limited levels of engagement (four mentees and one mentor).5

Analyses used either parametric or nonparametric methods, depending on the sample sizes. As a first step, we analyzed pre- vs. post-intervention differences across the key variables for the overall intervention sample, using repeated-measures analysis of variance. Analysis of covariance (ANCOVA) was used to compare students in strongly engaged vs. weak peer mentoring relationships, given the pre- vs. post-intervention nature of the outcome variables. The pre-intervention score was entered as a covariate to increase precision of estimates and to control for possible differences at baseline. To examine differences between these student groups for dependent variables derived from the strategy reflection sheets, analysis of variance (ANOVA) was used. Nonparametric analysis, specifically the Mann-Whitney U (MWU) test, was used to test all group differences for teacher-reported and student interview data given the small sample size \((n = 17)\). Outcome variables were converted into gain scores (posttest-pretest) and used as dependent variables in these analyses. In addition, effect sizes (Hedge’s \( g \)) were computed using parameters from the raw data (i.e., means, standard deviations) (DeFife, 2009). In accordance with the approach recommended by the What Works Clearinghouse (2008), effect sizes greater than \(.25\) were interpreted as “substantively important.”

Results

**Question 1: Do students in strongly engaged peer mentoring relationships use executive function (EF) strategies more frequently in their academic work than students in weak peer mentoring relationships?**

Prior to examining differences between students in strong vs. weak peer mentoring relationships, pre-intervention versus post-intervention changes were analyzed for all students in the intervention sample \((N = 34)\). For the combined groups, findings showed a decrease in self-reported EF strategy use at the end of the program year, \(F(1, 33) = 12.58, p<.01\). There were no significant changes in students’ use of organizing strategies, \(F(1, 33) = .03, p>.05\), and checking strategies, \(F(1, 33) = 2.16, p>.05\). Students in strongly engaged peer mentoring relationships were then compared with students in weak peer mentoring relationships.

**Teacher ratings.** When teacher ratings of student engagement were used, findings showed no statistically significant differences in students’ overall use of EF strategies, \(F(1, 31) = .98, p>.05\); Hedge’s \( g = .27, \) checking strategies, \(F(1, 31) = .69, p>.05\); Hedge’s \( g = .25, \)

---

5 Thus, students who were interviewed came from strongly engaged or poorly engaged mentor-mentee pairs.
or organizing strategies, $F(1, 31) = .31, p>.05$; Hedge’s $g = .15$, between students in strongly engaged vs. weakly engaged peer mentoring relationships. However, the effect sizes were moderate for overall use of EF strategies (Hedge’s $g = .27$) and for checking strategies (Hedge’s $g = .25$). These effect sizes were large enough to be considered substantively important even though they were not significant (What Works Clearinghouse, 2008), indicating that students in the strongly engaged peer mentoring relationships used EF strategies more frequently than students who were not engaged.

Analyses also focused on students’ use of strategies in their strategy reflection sheets based on three parameters: (a) the extent to which students could identify and name the strategy they were using, (b) how and when they could use the strategy, and (c) how successfully they had applied the strategy to their classwork or homework. Findings indicated that students in the engaged peer mentoring group were significantly more likely to name the EF strategies correctly, $F(1, 32) = 6.93, p = .01$; Hedge’s $g = .89$, to apply these strategies correctly, $F(1, 32) = 8.34, p <.01$; Hedge’s $g = .97$, and to use general EF strategies more often, $F(1, 32) = 4.57, p <.05$; Hedge’s $g = .72$, in their classwork, homework, projects, studying, and tests.6

**Student ratings.** When students’ self-ratings of their levels of engagement in their mentor-mentee relationships were used, findings indicated that students in strong mentor-mentee relationships used overall EF strategies significantly more frequently (MWU = 15, $z = -2.02, p<.05$). These students also used checking strategies more often; this difference was not statistically significant (MWU = 18, $z = -1.77, p>.05$) but was large enough to be considered substantively important (Hedge’s $g = 1.31$). Their use of organizing strategies did not change significantly during the intervention (MWU = 31.50, $z = -.37, p>.05$). However, the change was large enough to be considered substantively important (Hedge’s $g = .08$). Analyses of these students’ strategy reflection sheets were consistent with the ME findings. Students in strongly engaged mentor-mentee pairs correctly applied strategies significantly more often (MWU = 15.5, $z = -2.03, p<.05$) and reported using these strategies significantly more frequently in their classwork, homework, projects, studying, and test-taking (MWU = 15, $z = -1.98, p<.05$) (see Table 2). Similarly, for the strategy use variable derived from the interview data, students who reportedly had more positive mentor-mentee relationships also used EF strategies significantly more often (MWU = 14.5, $z = -2.05, p<.05$).

---

6 The homogeneity of variance assumption did not hold for the overall EF strategy use variable. The analysis was re-run using a corrected model. The results remained the same.
Table 2

*Students’ Use of EF strategies: Comparison of Students in Strong vs. Weak Peer Mentoring Relationships*²

<table>
<thead>
<tr>
<th></th>
<th>Strong Peer Mentoring Relationship</th>
<th>Weak Peer Mentoring Relationship</th>
<th>Mann-Whitney U</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy Reflection Sheets²</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified EF strategy correctly</td>
<td>10.70</td>
<td>6.57</td>
<td>18.00</td>
<td>-1.71</td>
<td>0.09</td>
</tr>
<tr>
<td>Applied EF strategy correctly</td>
<td>10.95</td>
<td>6.21</td>
<td>15.50</td>
<td>-2.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Total reported EF strategy use</td>
<td>11.00</td>
<td>6.14</td>
<td>15.00</td>
<td>-1.98</td>
<td>0.05</td>
</tr>
<tr>
<td>Interview Composites³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of EF strategies</td>
<td>11.05</td>
<td>6.07</td>
<td>14.50</td>
<td>-2.05</td>
<td>0.04</td>
</tr>
<tr>
<td>Hard work/effort</td>
<td>10.25</td>
<td>7.21</td>
<td>22.50</td>
<td>-1.25</td>
<td>0.23</td>
</tr>
<tr>
<td>Persistence</td>
<td>10.20</td>
<td>7.29</td>
<td>23.00</td>
<td>-1.19</td>
<td>0.23</td>
</tr>
<tr>
<td>Evidence of metacognitive awareness</td>
<td>9.85</td>
<td>7.79</td>
<td>26.50</td>
<td>-.90</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Note.

¹Results are shown only for the students who were interviewed (n = 17). All results are based on analyses using the Mann-Whitney U-test statistic. Mean ranks are displayed.

²For the strategy reflection sheets, results are based on students’ use of EF strategies summed over four time points during the school year.

³For the interview composites, results are based on students’ reports of their EF strategy use, metacognitive awareness, and effort at the end of the school year.

![Figure 8](image-url)  

*Figure 8.* Executive function strategy use, effort, and resilience for students in strong vs. weak peer mentoring relationships.

Note. Results are based on the interview sample (n = 17).

*<.05. **<.01.
**Question 2.** Do students in strongly engaged peer mentoring relationships show higher levels of effort and willingness to work hard in school than students in weak peer mentoring relationships?

Prior to examining differences for students in strong vs. weak peer mentoring relationships, pre-intervention vs. post-intervention changes were analyzed for all students in the intervention sample. Students in the combined group showed less effort at the end of the year, $F(1, 33) = 7.67, p < .05$; Hedge’s $g = .94$. Students in strongly engaged peer mentoring relationships were then compared with students in weak peer mentoring relationships.

**Teacher ratings.** Overall findings indicated a significant between-group difference in students’ effort and willingness to work hard in school. More specifically, students in engaged mentor-mentee relationships reported significantly higher effort, $F(1, 31) = 6.70, p < .05$; Hedge’s $g = .58$, at the end of the intervention program (Mean = 3.61; SD = .67) than those in less engaged peer mentoring relationships (Mean = 3.28; SD = .86).

**Student ratings.** When students’ ratings of engagement were considered, the same results were evident. Students who reported having more positive experiences with their mentors also reported significantly higher levels of effort (MWU = 14, $z = -2.05$, $p < .05$) (see Figure 8).

**Question 3.** Do students in strongly engaged peer mentoring relationships show higher academic self-concept and resilience than students in weak peer mentoring relationships?

When the combined intervention sample was considered, students showed similar levels of academic self-concept over the year, $F(1, 33) = 3.17, p > .05$; Hedge’s $g = .60$. However, they showed lower levels of resilience at the end of the year, $F(1, 33) = 8.61, p < .05$; Hedge’s $g = .99$. Findings were then analyzed separately for students in strongly engaged peer mentoring relationships versus students in weak peer mentoring relationships.

**Teacher ratings.** Findings showed that students in more engaged mentor-mentee dyads demonstrated higher academic self-concept than students in weak mentor-mentee relationships (Hedge’s $g = .27$). This effect size exceeded the .25 effect size benchmark (What Works Clearinghouse, 2008) even though it was not statistically significant, $F(1, 31) = 0.94, p > .05$. When teacher ratings of students’ resilience were used, the group differences were not significant, $F(1, 31) = 0.03, p > .05$; Hedge’s $g = .05$.

**Student ratings.** Students who reported strong engagement in their mentor-mentee relationships did not report significantly higher academic self-concept (MWU = 26, $z = -0.94, p > .01$). However, the Hedge’s $g$ value, calculated using the raw data, was larger than the .25 benchmark (Hedge’s $g = .41$). With respect to resilience, students who reported strong engagement in their mentor-mentee relationships during the interviews also reported
significantly higher levels of resilience (MWU = 7, z = -2.76, \( p < .01 \)) than those who did not (see Figure 8).

Post-Hoc Analyses

Correlational analysis was performed for all students in the program (strong peer mentor-mentee engagement group, \( n = 19 \); weak mentor-mentee engagement group, \( n = 15 \)). We examined the pattern of associations among the variables from the student surveys that directly mapped onto their use of executive function strategies as well as their effort, resilience, and academic self-concept (see Figure 1). All correlations were calculated using spearman rho (see Table 3).

Table 3

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Use of EF Strategies</td>
<td></td>
<td>0.30</td>
<td>0.54*</td>
<td>0.16</td>
</tr>
<tr>
<td>2. Effort</td>
<td>0.25</td>
<td></td>
<td>0.54*</td>
<td>0.47*</td>
</tr>
<tr>
<td>3. Resilience</td>
<td>0.08</td>
<td>0.16</td>
<td></td>
<td>0.60**</td>
</tr>
<tr>
<td>4. Academic Self-Concept</td>
<td>-0.36</td>
<td>0.68**</td>
<td>-0.07</td>
<td></td>
</tr>
</tbody>
</table>

Note. Spearman-Rho correlations above the diagonal represent students in strong peer mentoring relationships (\( n = 19 \); groups defined by SMARTS staff). Correlations below the diagonal represent students in weak peer mentoring relationships (\( n = 15 \)). *\( p < .05 \). **\( p = .01 \).

As is evident from Table 3, there were moderately strong correlations among the four outcome variables in the success paradigm for students in engaged mentor-mentee pairs. On the other hand, for students who were not engaged in their mentor-mentee pairs, only students’ general effort and academic self-concept were significantly correlated (\( r = .68, \ p = .01 \)).

Teacher ratings of students’ strategy use were available for a subset of the students (\( n = 17 \) in the strongly engaged peer mentoring group; \( n = 11 \) in the weak mentor-mentee engagement group). Teachers’ ratings of students’ strategy use on the TPSE were also analyzed using gain scores (T2-T1) and the Mann-Whitney U test. Findings indicated a “substantively important” effect size when the raw data were analyzed (What Works Clearinghouse, 2008) for students’ use of organizational strategies (Hedge’s \( g = .46 \)) and overall use of EF strategies (Hedge’s \( g = .36 \)). However, these differences did not meet the criteria for significance (MWU = 71.5, \( z = -1.10, \ p > .05 \) for organizational strategies and MWU = 77.50, \( z = -.85, \ p > .05 \) for overall use of EF strategies).
Discussion

Findings from this exploratory study provide preliminary support for an intervention approach that blends peer mentoring in schools with explicit instruction in executive function strategies. When subgroups of students in engaged mentor-mentee pairs were compared with mentor-mentee pairs who were not engaged with one another, significant group differences emerged. More specifically, with the exception of teacher-rated use of EF strategies, students in strong peer mentoring relationships used executive function strategies more often and displayed significantly higher levels of effort than students in weak peer mentoring pairs. On a number of additional qualitative measures (e.g., their completion of strategy reflection sheets), students in stronger peer mentoring relationships were also significantly more likely to identify and correctly use executive function strategies in their classwork, homework, projects, studying, and tests.

These findings suggest that students who connect well with their peer mentors feel more confident academically, are more willing to work hard in school, and are more open to learning executive function strategies as well as using these strategies for homework and tests. This generalizability of strategy use is extremely important given that many students use strategies in the classroom when their teachers explicitly direct them to do so but often do not generalize these strategies to their homework and studying. Furthermore, these results begin to address Karcher’s (2005) recommendation in his review of cross-age peer mentoring research that it is imperative to evaluate the effects of mentoring programs when mentors use structured activities with their mentees. In the current study, the combination of a structured executive function curriculum with a strong mentoring program met Karcher’s guidelines and had a positive effect on academic outcomes.

Another important finding was that, based on student ratings, students in strong peer mentoring relationships showed significantly higher levels of resilience in comparison with students in weak peer mentoring relationships (e.g., “I do not let problems stop me from reaching my goals”). This suggests that the social support offered by mentors who could connect with their mentees’ social and emotional needs helped mentees to feel more confident and better equipped to deal with the many academic and other challenges in school. These results are consistent with the findings of Karcher (2005) and Parra, DuBois, Neville, and Pugh-Lilly (2002) that mentors’ and mentees’ self-esteem and perceived self-efficacy are central mediators of the impact of mentoring.

The findings of the current study provide additional support for our theoretical paradigm and for the strong interactions among academic self-concept, effort, executive function strategies, resilience, and academic success (review Figures 1 and 2). Students in more engaged mentor-mentee pairs (compared to the weak peer mentoring pairs) reported
significantly higher levels of effort and executive function strategy use in their schoolwork as well as higher levels of resilience, a finding that is consistent with our Academic Success Cycle (review Figure 1). These results suggest that strong peer mentoring relationships may be an “extrinsic factor” (within the resiliency framework) that can help motivate students to work hard and to commit to using executive function strategies, which initiates a positive learning cycle. This support for our theoretical model aligns with the recommendations of DuBois, Portillo, Rhodes, Silverthorn, and Valentine (2011), who emphasized the need for mentoring studies that are grounded in a theoretical framework that explicitly links the processes occurring at the level of the mentoring program with the processes occurring at the level of the mentor-mentee relationships.

Overall, this exploratory study represents an important first step towards establishing and evaluating the efficacy of school-based programs that address the cognitive (i.e., executive function strategies) and socio-emotional (i.e., peer mentoring) needs of at-risk learners. More specifically, our findings suggest that students’ performance can be strengthened when they are explicitly taught executive function strategies that are linked with the academic curriculum so that they can apply these strategies to their classwork, homework, and tests. Further, when a supportive peer mentoring community is created in schools, engaged mentoring relationships can strengthen students’ motivation to work hard and to use these executive function strategies in their classwork and homework. Lastly, a blended executive function and mentoring program can build students’ academic self-concept and foster resilience, processes that are critically important for academic and life success (Margalit, 2003, 2004; Raskind et al., 1999).

Limitations

A number of the factors that affect our findings have been frequently documented in a wide range of youth mentoring programs (DuBois et al., 2011). First, school-based programs are extremely difficult to implement reliably because the practical realities of school schedules and curricula often interfere with tightly designed intervention programs. As a result, year-long intervention programs such as SMARTS are often reduced to only 3-6 months due to last-minute cancellations of sessions by school staff for field trips, special events, state standardized tests, snow storms, holidays and school vacations. These challenges often affect the quality of mentor-mentee relationships, which need to extend beyond a one-year relationship in order to ensure a strong impact (Rhodes, 2008). In this study, the fact that significant, albeit small, differences were identified when the scheduled mentoring events were reduced to 6 months’ duration, suggests strongly that the SMARTS intervention should be extended and its impact evaluated in greater detail.
Other factors outside the control of SMARTS staff often affected program implementation as well as student engagement. First, student absenteeism was increased by gang membership. Second, a high attrition rate and an uneven number of students in the intervention vs. control groups resulted from the announcement that the school would be closed at the end of the year. Third, our data collection efforts at the beginning and end of the SMARTS intervention program were challenging because students often did not complete all sections of the surveys, resulting in large amounts of missing data. Further, many of the mentees in the program seemed to have such severe writing deficits that they were unable to complete some of the measures. As a result, we refined our data collection efforts to gather more qualitative data (e.g., interviews, qualitative ratings of students’ performance). Therefore, our statistical analyses were restricted by the inconsistent sample size and missing data. The small subgroup sample sizes of the current study also raise the possibility of both Type I and Type II errors. The limited statistical power may have affected the findings so that existing group differences may not have been detected. Furthermore, school administrators had promised us access to students’ standardized test scores and final grades, but they did not comply despite our many attempts to obtain these data. As a result, we had to refine our data collection efforts on an ongoing basis so that we could use a multi-pronged approach to collecting qualitative and quantitative data.

**Recommendations for Future Research**

There is a major need for more research on the impact of peer mentoring in schools, as emphasized in previous reviews of cross-age peer mentoring studies (Chan et al., 2012; Karcher, 2005). Future studies should ideally involve school administrators and teachers who are trained to implement a blended executive function and mentoring program systematically. This would ensure that executive function strategies are taught systematically and embedded into the day-to-day curriculum as well as homework and tests. In addition, outcome measures in school-based peer mentoring programs should include semi-structured interviews, portfolio analyses, video ratings, and qualitative data analyses so that the impact of school-based peer mentoring can be effectively assessed. Therefore, school-based peer mentoring programs need to incorporate practical, easy-to-use pre- and post-measures, which are less time consuming or detailed than the measures we used in the current study. Finally, an interesting trend in our study was that the benefits of mentoring relationships appeared to be as strong or stronger for mentors than mentees. Future research is needed to evaluate the transformative effects of mentoring relationships on older students who begin to take on more responsibility.
Conclusions

The fast pace in our 21st-century classrooms and the expanding influence of technology on classroom instruction have placed increasing pressure on students to use executive function strategies in order to set goals, organize, prioritize, think flexibly, and self-monitor. As a result, teachers are beginning to recognize the benefits of promoting metacognitive awareness and teaching executive function strategies to all students, regardless of whether or not they exhibit any learning challenges. Peer mentoring is a powerful technique that teachers can use to extend and deepen the effects of teaching these executive function strategies. When executive function strategy instruction is combined with peer mentoring, teachers provide students with a strong foundation for building self-concept, persistence, and resilience, the gateways to academic and life success.

References


Mentoring, Executive Function, Effort, Self-Concept by Lynn Meltzer, Surina Basho, Ranjini Reddy, and Katelyn Kurkul


Acknowledgments
Thanks to our ResearchILD colleagues for their help and support: Bethany Roditi, Mimi Ballard, Jamie Cutler, Wendy Stacey, Michael Greschler, Elizabeth Ross, Elizabeth Brach, Aaron Gunning, Abigail DeMille, and Nicole Ashby.

Author Note
This study was made possible through support, in part, from the Oak Foundation, Cisco Foundation, Verizon, Shire, and Bain Children’s Charities.
CALL FOR PROPOSALS

The International Academy for Research in Learning Disabilities invites you to submit proposals for the

40th Annual IARLD Conference

Austin, Texas

June 28-30, 2016

Thompson Conference Center on the Campus of
The University of Texas at Austin

PROPOSAL DEADLINE: January 30, 2016

INTRODUCTION
This year’s conference will include three presentation formats:

- Interactive posters
- Symposia
- Roundtable discussions

Each of these formats allows time for in-depth discussion among participants. We will also have the William Cruickshank Memorial Lecture, an invited address in honor of our founder.

All proposals should be submitted electronically. Proposals should indicate the preferred format; however, the program committee reserves the right to change the requested session format in order to create a balanced and integrated program. All proposals will be reviewed for the 2016 by conference committee members and/or by Fellows in the Academy with expertise in the specific content area.

For more details, go to IARLD.com
Manuscript Submission Guidelines

English is used for submissions to the journal, correspondence and publication. All submission 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:
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)