School Psychology Review,
2017, Volume 46, No. 1, pp. 122–140

Research Into Practice

Efficacy of Peer-Mediated Incremental Rehearsal for English Language Learners

David A. Klingbeil
University of Wisconsin–Milwaukee

Mariola Moeyaert
State University of New York–Albany

Christopher T. Archer
University of Wisconsin–Milwaukee

Tatenda M. Chimboza
University of Wisconsin–Milwaukee

Scott A. Zwolski Jr.
Loyola University Chicago

Abstract. School psychologists will likely become more involved in supporting the reading achievement of English language learners (ELLs). This requires evidence-based interventions that are validated for ELL students. Incremental rehearsal (IR) is an evidence-based intervention for teaching words, but the resource intensity often precludes its use. Using peers as interventionists may increase the contextual validity of IR while maintaining the benefits when compared with other drill techniques. This efficacy study examined if (a) peer-mediated IR (PMIR) was effective for teaching ELL students high-frequency words and (b) improvements in word reading generalized to changes in students' oral reading fluency. Five ELL students participated in a randomized multiple-

This research was partially funded by a Research and Creative Activities Support Award from the University of Wisconsin–Milwaukee Office of Research. This research was also partially funded by the Institute of Education Sciences (R305D150007). The opinions expressed are those of the authors and do not represent the views of the Institute of Education Sciences.

Correspondence concerning this article should be addressed to David A. Klingbeil, Department of Educational Psychology, University of Wisconsin–Milwaukee, PO Box 413, Milwaukee, WI 53201; e-mail: davidak5@uwm.edu

baseline design across participants. Results indicated that PMIR was functionally related to an increase in word reading for all 5 participants. Effect sizes estimated using Tau*U* and multilevel modeling indicated that PMIR had a large effect on sight-word reading. No functional relationship between PMIR and oral reading fluency was observed. PMIR was generally acceptable to target students and peer tutors. Limitations and potential implications of the results are discussed.

English language learners (ELLs), or youth who do not speak English at home and whose English proficiency limits their ability to access grade-level material, represent one of the fastest growing populations in the United States (Kena et al., 2014). Poor reading achievement has been a long-standing concern for ELL students (e.g., U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2015), suggesting that school psychologists will be increasingly involved with supporting ELL students' reading achievement (August, McCardle, & Shanahan, 2014). This underscores the need for evidence-based interventions (EBIs) validated for ELL youth (Moore & Klingner, 2014).

Despite the progress in the identification of EBIs over the past decade, data suggest the use of EBIs remains limited in schools (e.g., Crosse et al., 2011; Kretlow & Helf, 2013). Bridging the gap between research and practice requires efficacious interventions that are contextually valid. The amount of teacher time and number of resources needed to deliver EBIs are critical threats to contextual validity (Skinner, McCleary, Skolits, Poncy, & Cates, 2013). Modifying existing EBIs is a promising approach to increase the contextual validity of these practices. The purpose of this study was to examine the efficacy of one such modification, using peer interventionists to implement incremental rehearsal (IR; Tucker, 1989).

READING DEVELOPMENT AND INTERVENTION FOR ELL STUDENTS

Reading is a complex process that incorporates multiple core skills—phonemic awareness, phonics, fluency, comprehension, and vocabulary (National Institute of Child Health and Human Development, 2000). ELL students will generally benefit from systematic, intensive instruction in these core areas (August et al., 2014; Goldenberg, 2010). Word reading, or the ability to recognize single words by sight, is a foundational literacy skill for native English speakers (Hudson, Torgesen, Lane, & Turner, 2012; Samuels & Flor, 1997) and ELL students (August & Shanahan, 2006). Word reading is a predictor of later reading fluency (Hudson et al., 2012; Speece & Ritchey, 2005), and word reading skills may mediate the relationship between phonological awareness and reading fluency (Yaghoub Zadeh, Farnia, & Geva, 2012). Early deficits in word reading skills predict later delays for ELL students, suggesting that word reading is an important intervention target (Lesaux, Rupp, & Siegel, 2007; Vaughn, Mathes, Linan-Thompson, & Francis, 2005).

Intensive instruction for ELL students should be explicit and direct, maximize opportunities to respond, provide increased repetition of foundational skills, and provide immediate corrective feedback (Albers & Martinez, 2015; August & Shanahan, 2006). School psychologists should search for interventions that incorporate these aspects when working with ELL students. One such intervention is IR (Tucker, 1989).

IR involves the teaching of unknown items interspersed with a high percentage of known items. Interventionists provide direct and explicit teaching of unknown material when introducing unknown items. IR allows for substantial repetition and reinforcement of unknown and known vocabulary words. IR includes a large amount of scaffolding through the modeling of unknown items and immediate corrective feedback during practice. Moreover, interventionists can closely control the instructional level by adjusting the ratio of

unknown to known words used per session and can adjust the amount of new information presented based on a student's acquisition rate. Providing interventions at the appropriate instructional level and avoiding exceeding a student's acquisition rate can lead to increased learning and retention of unknown material (Burns, 2007; Helman & Burns, 2008).

EVIDENCE SUPPORTING THE USE OF IR FOR WORD READING

There is relatively robust evidence supporting the use of IR for teaching unknown words. In a recent meta-analytic review, Burns, Zaslofsky, Kanive, and Parker (2012) found that IR was associated with a moderate effect on word reading across 12 studies using single-case or group designs, $\phi = .62$, 95% confidence interval (CI) [.54, .70]. In addition, using IR to teach unknown words had a moderate effect on students' oral reading fluency (ORF) in three studies, $\phi = .61$, 95% CI [.36, .86]. Only one of the studies included in the review of Burns et al. (2012) targeted ELL students. Matchett and Burns (2009) found that using IR to teach words led to increases in word reading fluency and ORF for an ELL student.

Further support for the use of IR with ELL students comes from three recent studies targeting letter–sound expression. Two studies found that traditional IR (i.e., using flashcards) was effective for teaching ELL students letter sounds (Peterson et al., 2014; Rahn et al., 2015). In addition, a computerized version of IR was effective for teaching letter–sound expression to ELL students in kindergarten and first grade (DuBois, Volpe, & Hemphill, 2014).

Intervention Efficiency

Previous research has suggested that IR, and flashcard interventions more generally (Tan & Nicholson, 1997), is an effective practice for teaching words to native English speakers. Research demonstrating that IR was effective for improving letter–sound expression (DuBois et al., 2014; Peterson et al., 2014; Rahn et al., 2015) and word reading (Matchett & Burns, 2009) lends support for using IR with ELL students. However, IR is relatively inefficient compared to other flashcard interventions, which is a critical consideration for practitioners.

Procedures among flashcard interventions are varied. Traditional drill (TD) includes the rehearsal of a set of unknown items, whereas IR incorporates the practice of known items. Interspersing known items may double the time necessary for the intervention (Burns & Sterling-Turner, 2010) or reduce the number of items taught within a controlled period (Joseph, Eveleigh, Konrad, Neef, & Volpe, 2012).

When intervention duration is fixed, TD will result in more words being taught and consequently more words being initially retained. However, IR is generally associated with enhanced long-term maintenance and generalization of material (Burns & Sterling-Turner, 2010; Joseph et al., 2012; Nist & Joseph, 2008). This pattern is consistent with research on memory and learning that underscores the importance of spacing within the practice of unknown material (Varma & Schleisman, 2014). Practitioners should carefully balance findings related to initial retention, maintenance, and generalization with the time and resources necessary to provide either TD or IR.

MODIFYING IR TO PROMOTE EFFICIENCY

Given the need for efficient interventions that increase maintenance and generalization, researchers have recently examined the use of computerized IR to improve ELL students' letter–sound expression and fluency. IR delivered via computer retained the interspersal of known items and high number of opportunities to respond but enhanced the efficiency of IR by (a) increasing the number of students who could receive the intervention in a given time frame, (b) decreasing the number of materials needed, and (c) automating data collection (DuBois et al., 2014; Volpe, Burns, DuBois, & Zaslofsky, 2011). Adult interventionists were still needed to provide immediate corrective feedback during intervention sessions.

Another potential method for reducing the intervention intensity of IR is using peer-mediated instruction. Peer-mediated instruction includes strategies that use peers as teachers to provide individualized instruction, practice, and repetition (Utley & Mortweet, 1997). For example, the well-researched Peer-Assisted Learning Strategies program trains peer tutors to model oral reading, provide support using comprehension strategies, and provide corrective feedback (Fuchs & Fuchs, 2005). Peer-mediated instruction can be used to facilitate best practices in teaching ELL students reading, including providing additional opportunities to respond, increased practice opportunities, and immediate corrective feedback (Greenwood, Arreaga-Mayer, Utley, Gavin, & Terry, 2001).

Several meta-analytic reviews have underscored the empirical support for using peers as teachers. For example, Hattie (2009) found that the average overall peer tutoring effect was moderate ($d = 0.55$) across 14 meta-analyses that included 767 original studies. Similarly, Bowman-Perrott et al. (2013) synthesized single-case research on peer tutoring programs and found that peer tutoring had moderate to large effects on reading outcomes (e.g., sight-word acquisition, reading) across 10 studies. Specific to ELL students, a review by the Institute of Education Sciences found strong evidence for providing structured activities that pair students with different English proficiencies (Gersten et al., 2007). Peer-mediated IR (PMIR) could provide a structured format for pairs of ELL students to practice unknown material. Training peers to deliver IR would maintain key causal mechanism components of IR such as frequent opportunities to respond, immediate corrective feedback, and expanded practice of known material (Varma & Schleisman, 2014). PMIR does not reduce the materials necessary, but it could reduce the adult time necessary to implement the intervention.

PURPOSE

Although IR may enhance maintenance and generalization of unknown material, the resource intensity of the intervention greatly limits its contextual validity. Modifications to reduce the resource intensity of IR, such as using peer interventionists, may improve the efficiency of IR and its usefulness for applied settings. The purpose of this study was to investigate the efficacy of PMIR for teaching high-frequency words to ELL students. We had three major research questions:

1. Is there a functional relationship between PMIR and an increase in the level of students' word reading on previously unknown words?
2. Is there a functional relationship between PMIR and an increase in the level of students' ORF (i.e., would the effects of PMIR on word reading generalize to ORF)?
3. To what extent are the words taught maintained after discontinuing PMIR?

We hypothesized that PMIR would have a large, immediate effect on students' word reading. On the basis of the link between word reading and ORF (Hudson et al., 2012; Speece & Ritchey, 2005) and a meta-analytic review of IR (Burns et al., 2012), we hypothesized that PMIR would have a small, delayed effect on students' ORF level. Finally, we hypothesized that the effects of PMIR on word reading would be maintained.

METHOD

This study was conducted at a public charter school in Milwaukee, Wisconsin. The school served 432 students during the 2013–2014 school year. Approximately 91% of these students were identified as Hispanic, 4% as Black, 4% as White, and 1% as other. Nearly all students (98%) were eligible for free or reduced-price lunch. Five percent of students received limited English proficiency services. Pseudonyms for participants are used throughout the article. The Institutional Review Board at the University of Wisconsin–Milwaukee approved the study prior to initiation of all research activities.

Target Students

Classroom teachers nominated second-grade (Iker, Sofia, and Tomas) and third-grade (Martin and Lucia) students who could benefit from additional reading support based on their reading performance in the classroom. All five nominees participated in the study, after we obtained parent consent and student assent. None of the participants received limited English proficiency services. Data regarding language proficiency were not available for these students. All students completed the winter administration of the Measures of Academic Progress (Northwest Evaluation Association, 2009) 2 weeks prior to the study. Scores from the Measures of Academic Progress are presented in Rasch unit (RIT) values that are comparable across grade levels.

Martin

Martin was an 8-year-old, Hispanic boy in third grade. Martin had attended the school for the past 1.5 years and had moved to Wisconsin from Mexico the previous spring. His reading performance on the winter administration of the Measures of Academic Progress was at the first percentile nationally (RIT = 135).

Lucia

Lucia was an 8-year-old, Hispanic girl in third grade. Lucia was in her first year of attendance at the school. She was born in the United States and previously attended a public school that had a bilingual program. Lucia's reading performance on the winter administration of the Measures of Academic Progress was at the 10th percentile nationally (RIT = 168).

Iker

Iker was a 7-year-old, Hispanic boy in second grade. He had attended the school for 1.5 years and moved to Wisconsin from Mexico in the spring of 2014. Iker's reading performance on the winter administration of the Measures of Academic Progress was at the first percentile nationally (RIT = 138). Iker's class received informal bilingual reading instruction from the school's music teacher approximately once per week.

Sofia

Sofia was a 7-year-old, Hispanic girl in second grade. Sofia emigrated from Mexico, and this was her first year attending school in the United States. Her reading performance on the winter administration of the Measures of Academic Progress was at the first percentile nationally (RIT = 126). Sofia's class received informal bilingual reading instruction from the school's music teacher approximately once per week.

Tomas

Tomas was a 7-year-old, Hispanic boy in second grade. Tomas emigrated from Mexico, and this was his first year attending school in the United States. Tomas' reading performance on the winter administration of the Measures of Academic Progress was at the first percentile nationally (RIT = 148). Tomas' class received informal bilingual reading instruction from the school's music teacher approximately once per week.

Peer Tutors

Teachers were also asked to nominate high-achieving students with well-developed English proficiency who they believed could effectively deliver PMIR under the supervision of an adult. Teachers identified four students in third grade who were asked to participate after receiving parent consent. Target students and peer tutors were not matched based on grade or gender. Target students worked with at least two tutors; however, we could not randomly assign target tutors to students because of classroom schedules.

Rosa

Rosa was an 8-year-old, Hispanic girl who had attended the school for 2 years. She reported that she primarily spoke Spanish at home. Rosa's reading performance on the winter administration of the Measures of Academic Progress was at the 17th percentile nationally (RIT = 181).

Maria

Maria was an 8-year-old, Hispanic girl who had attended the school for 2 years. Maria

reported that she primarily spoke Spanish at home. Her reading performance on the winter administration of the Measures of Academic Progress was at the 36th percentile nationally (RIT = 190).

Carlos

Carlos was an 8-year-old, Hispanic boy who had attended the school for 2 years; however, he moved back to Mexico for part of his first-grade year. Carlos reported speaking Spanish and some English at home. His reading performance on the winter administration of the Measures of Academic Progress was at the 79th percentile nationally (RIT = 208).

Michael

Michael was an 8-year-old, Hispanic boy who had attended the school for 2 years. Michael reported primarily speaking Spanish at home. Michael's reading performance on the winter administration of the Measures of Academic Progress was at the 66th percentile nationally (RIT = 202).

Measures

Prior to the baseline phase, we assessed target students' ability to read words from the Fry list of high-frequency words. A word was considered known if the participant correctly pronounced the word within 3 s. A word was considered unknown if the participant took longer than 3 s to read it correctly or read the word incorrectly. We created a set of 30 known words and a set of 80 unknown words for each target student. We wrote each word on the front of 3 × 5-in. index cards using a black marker. The word was also written on the back using a red (unknown words) or green (known words) colored pencil to help tutors differentiate between known and unknown words during intervention sessions.

Word Reading

The primary dependent variable was word reading accuracy. We assessed students' ability to read words individually by shuffling the selected index cards and presenting each word to the target student. The student was then asked to read each word. A word was considered correct if the participant read the word correctly within 3 s. A word was not considered correct if the participant took longer than 3 s to read the word correctly, read the word incorrectly, or did not provide a response. No corrective feedback was given during word reading assessments. Sessions were scheduled so that the number of days between each session ranged from 0 (i.e., conducted the next day) to 1.

We assessed students' ability to read three words that were randomly selected from their unknown set during each baseline and intervention session. Previous research found that the average acquisition rate of ELL students with very limited English proficiency (M = 3.2) was lower than the acquisition rate of students with moderate (M = 5.5) or well-developed English proficiency (M = 7; Burns & Helman, 2009). Without access to data regarding students' English proficiency, and considering the length of time each participant had received English instruction, we conservatively chose to teach three words per session to avoid exceeding students' acquisition rates. We randomly selected three words from the student's set of unknown words prior to each intervention session. We assessed students' ability to read the words that were taught the previous session, prior to teaching three new words. Words were only taught once during the intervention phase.

In order to have comparable values across baseline and intervention phases, we assessed students' ability to read three words during each baseline session. Prior to each baseline session, we randomly selected three words from the student's set of unknown words. Words could only be selected once during the baseline phase (i.e., sampling without replacement). These baseline data allowed us to predict the extent to which students' ability to read previously unknown words would improve prior to introduction of PMIR.

Oral Reading Fluency

In order to determine if improvements in word reading generalized to connected text reading, we used AIMSweb reading curriculum-based measures (CBMs) to assess stu-

dents' ORF. According to the technical manual (Pearson, 2012), resulting ORF data have sufficient alternate-form and test–retest reliability (r = .94). Evidence of validity was provided by moderate correlations between AIMSweb scores and state-test performance (r = .67-.70), as well as norm-referenced reading assessments (r = .67-.70). The variable of interest was the number of words read correctly per minute (WRCM).

Graduate assistants administered one CBM probe at the beginning of each session, resulting in three ORF data points per week. Each target student was assessed using the progress monitoring probes for the grade in which he or she was enrolled. The 28 second-grade passages and 26 third-grade passages were administered in numerical order. Once participants had been assessed with the final grade-level passage, we randomly selected passages for the remaining sessions.

Maintenance

We assessed students' ability to read all of the words taught during PMIR sessions approximately 10 days after the conclusion of the intervention phase. The number of words taught ranged from 54 to 75. We created a list of the words taught to each participant in Microsoft Word (black ink, 14-point font). Each participant was asked to read each word aloud. A word was considered learned if the participant read it aloud accurately within 3 s. A word was not considered learned if the participant took longer than 3 s to read the word correctly, read the word incorrectly, or did not provide a response. No corrective feedback was given during maintenance assessments. Students received a small tangible item (i.e., pencil or fruit snacks) and verbal praise after completing the assessment.

Treatment Acceptability

In order to examine treatment acceptability, target students were asked to rate if the intervention helped them learn more words on a scale of 1 to 10. To provide a visual representation of this scale, a horizontal line with numbers 1 through 10 was drawn on a whiteboard with three cartoon faces above it. The face above numbers 1 through 3 had a frown, the face above the numbers 4 through 7 had neutral affect, and the face above numbers 8 through 10 had a smile. Target students were also told that 1 meant *not at all*, 5 meant *some*, and 10 meant *yes definitely*. Finally, target students were asked open-ended questions regarding what they liked and disliked about PMIR and if they would want to use PMIR again to learn new words.

Peer tutors were asked if PMIR was easy to administer (yes or no) and if they thought PMIR helped the target students learn new words, using the same 1 to 10 visual scale described above. Peer tutors were also asked open-ended questions about what they liked and disliked about PMIR and if they would want to use PMIR again to teach other classmates new words.

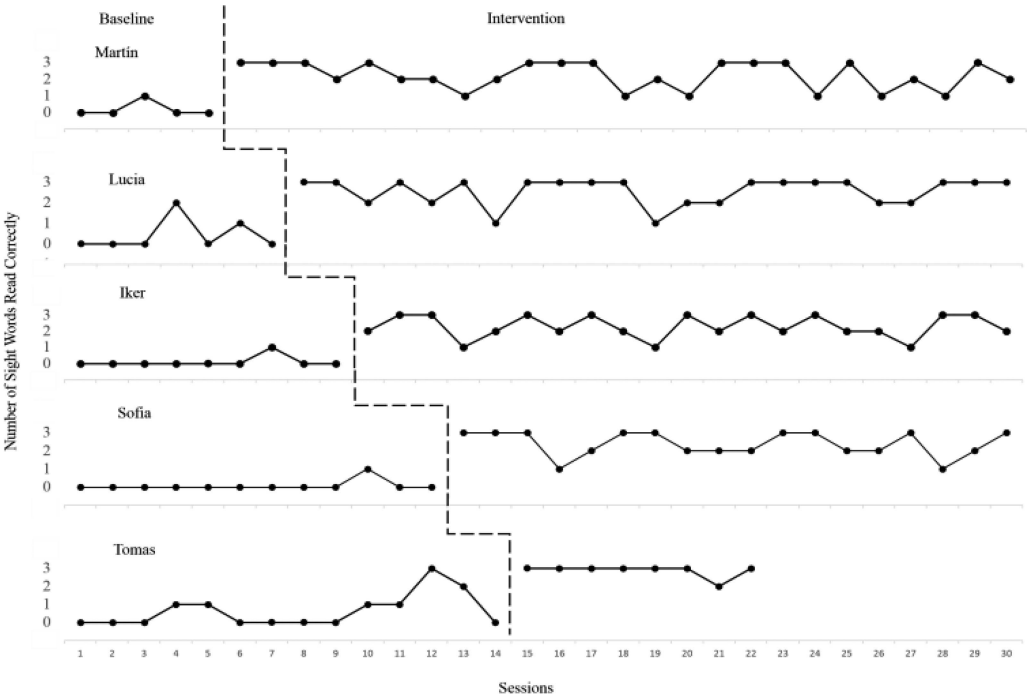
Experimental Design

We examined the efficacy of PMIR using a randomized, concurrent multiple-baseline design (MBD) across participants. We randomly assigned participants to the order in which the intervention was introduced (i.e., panels in Figures 1 and 2). We also randomly assigned the intervention start points (with a minimum two-session lag) using the random number function in Microsoft Excel. Using randomization to determine intervention start points, instead of the traditional response-guided approach, can reduce threats to internal validity such as experimenter bias (Kratochwill & Levin, 2010). The use of randomization limits the flexibility found in response-guided approaches (e.g., determining when to implement the intervention), but because this was an efficacy trial of PMIR, we sought to maximize the internal validity of the findings.

Procedures

The Institutional Review Board at the University of Wisconsin–Milwaukee approved this study prior to initiation. Two research assistants administered all assessments and conducted interventionist training after receiving training from the first author. Both students were completing their second semester in a school psychology program. The first graduate

Figure 1. Number of Words Read for Target Students



Note. During both phases, words were randomly selected from an unknown set of words identified during the prebaseline assessment. During baseline, students were assessed using three randomly selected words. During intervention sessions, students were assessed on words taught during the previous session.

student was a White man and a native English speaker. The second graduate student was a Black woman who spoke Shona and English fluently. Assistants were required to demonstrate 100% fidelity and 90% agreement on the ORF measures prior to study initiation. All sessions took place in an unused classroom.

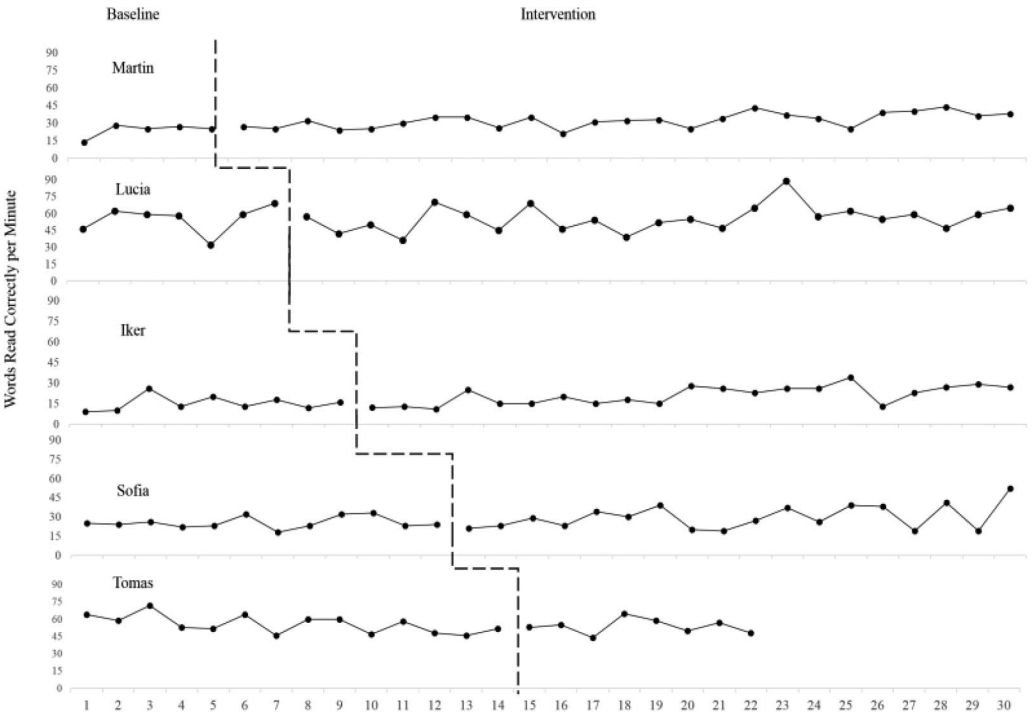
Tutor Training

Graduate assistants trained each peer tutor during three to four 20-minute sessions. Training included six major components. First, graduate assistants modeled the intervention for the peer interventionist by delivering the intervention to the student. Second, interventionists were trained to introduce unknown words (i.e., "This word is cat. What word is this? Good, this word is cat.") and provide standard corrective feedback (i.e., "This word is dog. What word? Good, this word is dog."). Peer tutors had an opportunity to practice both of these components separately. Third, tutors were trained to present unknown words interspersed with known words. Fourth, tutors were trained to remove the last known word from the deck after they presented all seven known words and then to introduce the second unknown word. Fifth, tutors were given opportunities to practice administering the procedure to one of the graduate assistants while the other provided immediate corrective feedback. Sixth, tutors were given a script with blanks for each of the unknown words that were to be taught. They were also given a 3 × 5-in. index card with the corrective feedback procedure stated clearly. Tutors were able to use these materials during all intervention sessions. Tutors were considered sufficiently trained when they administered IR to a graduate assistant with 100% integrity.

Baseline

Graduate assistants collected ORF data using one probe at the start of each session,

Figure 2. ORF During Baseline and Intervention Phases



Note. Data were collected three times per week. During each assessment, students read one AIMSweb ORF progress monitoring probe under standardized conditions. ORF probes were at the student's grade level. ORF = oral reading fluency.

resulting in a maximum of three data points per week. Next, word reading was assessed using three randomly selected unknown words. Baseline sessions took approximately three minutes per student. The number of baseline sessions ranged from 5 to 14 ($M = 9.4$, $Mdn = 9$).

Intervention

Prior to each intervention session, graduate assistants randomly selected three unknown and seven known words. Each unknown word was rehearsed with seven known words to maintain an appropriate ratio for drill tasks. Prior to the intervention session, graduate assistants asked peer tutors to read each word and prompted them to write the unknown words on a script. At the start of each session, graduate students collected ORF and word reading data while the tutors organized the known and unknown items in the correct order.

Peer tutors administered IR following standardized procedures. In brief, tutors began each session by introducing IR and presenting the first unknown word. The target student was then asked to read the word aloud with the tutor providing immediate error correction if the target student responded incorrectly. The tutor presented the unknown (U) word and a sequence of known (K) words (i.e., U1, K1; U1, K1, K2; U1, K1, K2, K3; U1, K1, K2, K3, K4; U1, K1, K2, K3, K4, K5; U1, K1, K2, K3, K4, K5, K6; U1, K1, K2, K3, K4, K5, K6, K7). The tutor provided standard error correction any time an error was made. If an error was made on a known word, the tutor started again with U1. For example, if the target student made a mistake on K5, the tutor provided standard error correction and then started again with the sequence U1, K1, K2, K3, K4, K5 before continuing on. Once the unknown word was presented with all known words, the tutor removed K7 from the deck and introduced a second unknown item. This process

was completed until three unknown words were taught. Intervention sessions took approximately ten minutes per student. The number of intervention sessions ranged from 8 to 25 ($M = 19$, $Mdn = 21$). The fifth target student to start the intervention moved to another state after 9 weeks.

Treatment Integrity and Interrater Agreement

We created a fidelity checklist with eight discrete tasks necessary for the administration of IR. The steps included briefly introducing the task, correctly introducing the unknown words, correctly completing the interspersal task, correctly adding another unknown word and removing a known word, and correctly using the standard error correction procedure. Graduate assistants observed 38% of the intervention sessions and recorded if each intervention step was completed correctly. Treatment integrity was calculated by dividing the number of correctly completed steps by the total number of steps and multiplying the result by 100. Average integrity was 97.4% (range = 75% to 100%). Graduate assistants provided corrective feedback after any session in which a step was conducted incorrectly.

We collected assessment fidelity and interrater agreement data during 32% of baseline sessions ($n = 15$) and 34% of intervention sessions ($n = 32$). We created an assessment fidelity checklist for assessing target students' word reading. The checklist included three steps: correctly presenting and prompting students to read the unknown word, correctly scoring words as known or unknown, and accurate recording of student performance. Word reading assessment fidelity was 100%. We used the AIMSweb Accuracy of Implementation rating scale to measure ORF fidelity. The measure includes 14 steps. Assessment fidelity for ORF averaged 99.5% (range = 92% to 100%).

We calculated interrater agreement by dividing the number of agreements by the total agreements and disagreements. For word reading, interrater agreement was 100% during baseline and 97% during intervention. Average interrater agreement for ORF data was 97.2% ($SD = 2.2\%$) during baseline and 95.7% ($SD = 3.4\%$) during intervention.

Data Analyses

Data analyses followed What Works Clearinghouse guidelines for evaluating intervention effects in single-case designs (Kratochwill et al., 2013). Visual analyses were used to determine if there was a functional relationship between PMIR and each dependent variable (i.e., word reading and ORF). If a functional effect was present, we proceeded to estimate the effect sizes using quantitative methods.

Visual Analyses

Visual analyses were conducted for each participant and outcome variable. First, we examined the level, trend, and stability of data within each phase. We estimated within-phase stability by calculating the percentage of data points within 20% of the phase median. Second, we examined the immediacy of effect, consistency of data patterns, and overlap of data between baseline and PMIR phases. We examined between-phase changes in level by evaluating the relative, absolute, median, and mean level change (Lane & Gast, 2014). Three demonstrations of an intervention effect are necessary for establishing a functional relationship (Kratochwill et al., 2013).

Across-Case Effect Sizes

If visual analysis indicated a functional relationship, we proceeded to estimate across-case effect sizes using Tau*U* and multilevel modeling. Tau*U* provides a nonparametric index of the percentage of data that do not overlap minus the percentage of data that overlap between baseline and control phases (Parker, Vannest, & Davis, 2014). Tau*U* is based on Kendall's rank correlation and the Mann–Whitney test for two groups, both of which follow the S distribution. Tau*U* improved upon previous indexes of nonoverlap by following a known distribution and allowing for the control of trends in baseline data

(Parker et al., 2014). Parker and Vannest (2009) provided tentative guidelines for interpreting the nonoverlap of all pairs statistic, which can be transformed into Tau*U*. Effect sizes from 0 to 0.31 are small; from 0.32 to 0.84, medium; and from 0.85 to 1.0, large (Parker & Vannest, 2009).

We calculated Tau*U* for each target student and a weighted, across-case Tau*U* using an online calculator (Vannest, Parker, & Gonen, 2011). In brief, we assessed the trends in the baseline data for each student. Significant trends in the baseline data ($p < .15$) were controlled for prior to calculating the phase contrast between baseline and PMIR phases for each participant. Finally, we used the calculator to estimate the weighted average Tau*U* across all participants.

We also used multilevel modeling to estimate the effects of PMIR across cases. In single-case designs, data are hierarchically structured: Measurements are nested within participants, and participants in turn are nested within studies. Ignoring the multilayered nature of MBD data can adversely impact the conclusions of a single-case analysis (Van den Noortgate, Opdenakker, & Onghena, 2005) as standard error estimates will be too small, resulting in an inflated number of Type I errors when used in statistical tests. Multilevel modeling is a recommended approach to analyze data characterized by a two-level data structure (Ferron, Bell, Hess, Rendina-Gobioff, & Hibbard, 2009). After checking the tenability of regression assumptions, we used the multilevel model suggested by Van den Noortgate and Onghena (2003, 2008) to estimate (a) the overall average effect of PMIR across participants and (b) the within- and between-participant variability. Using an MBD across 5 participants, with 30 measurement occasions, provided sufficient power ($> .80$) to detect large treatment effects (Ferron, Moeyaert, Van Den Noortgate, & Beretvas, 2014). We conducted all statistical analyses using SAS Version 9.4 (SAS Institute, 2011–2014) and used a significance level of .05 for all statistical tests.

RESULTS

In the following sections, we describe the visual analysis results for each student. Then, when visual analysis indicated that a functional relationship exists, we discuss the across case effect sizes.

Question 1: Effect of PMIR on Word Reading

We hypothesized that the implementation of PMIR would be functionally related to an increase in the level of students' word reading. Results are shown in Figure 1.

Martin

As shown in the first panel of Figure 1, Martin's performance during baseline was low ($M = 0.2$, $Mdn = 0$) and stable (i.e., 80% of data within ± 20% of phase median). There was no apparent trend in baseline performance. Martin's performance during PMIR increased in level ($M = 2.24$, $Mdn = 2$). His performance was highly variable (i.e., 28% of data points within ± 20% of phase median), and the relative level of Martin's performance decreased slightly across time. The observed between-phase change in performance was immediate and consistent (e.g., median level change = 2). There was very little overlap of the data between the two phases. Despite the increased variability of performance during PMIR, results indicated that PMIR had an effect on word reading for Martin.

Lucia

Lucia's performance during baseline was low ($M = 0.43$, $Mdn = 0$) and relatively stable (i.e., 71% of data within ± 20% of phase median). There was an abrupt increase in her performance during the middle of the baseline phase followed by a negative trend. During PMIR, Lucia's performance increased in level ($M = 2.56$, $Mdn = 3$) and variability (i.e., 65% of data within ± 20% of phase median). Her performance during PMIR decreased slightly between Sessions 9 and 13 but increased between Sessions 19 and 23. There was an immediate and consistent change in Lucia's performance between phases. There was one overlapping data point between the two phases. Results

indicated that PMIR had an effect on Lucia's word reading despite the change in trend during the PMIR phase.

Iker

Iker's performance during baseline was low ($M = 0.11$, $Mdn = 0$) and stable (i.e., 89% of data within ± 20% of phase median). No trends in his performance were evident during baseline. During PMIR, Iker's performance increased ($M = 2.29$, $Mdn = 2$) and became more variable (i.e., 43% of data within ± 20% of phase median). No trends were apparent during PMIR. Iker's performance between phases increased immediately and was consistent (e.g., median level change = 2). There was no overlap between the two phases. PMIR had an effect on Iker's word reading.

Sofia

Sofia's performance during baseline was low ($M = 0.08$, $Mdn = 0$) and stable (i.e., 92% of data within ± 20% of phase median). Her performance during PMIR increased ($M = 2.39$, $Mdn = 2.5$) and was generally consistent (i.e., 89% of data within ± 20% of phase median). There was no trend apparent in her performance during either the baseline phase or the PMIR phase. Between-phase change in Sofia's performance was immediate and consistent (e.g., median level change = 2.5). There was no overlap in the data between phases. PMIR had an effect on Sofia's word reading.

Tomas

Tomas demonstrated low performance during baseline ($M = 0.64$, $Mdn = 0$); however, his performance was relatively variable (i.e., 57% of data within ± 20% of phase median). Tomas' performance improved during the baseline phase. During PMIR, Tomas' performance increased in level ($M = 2.28$, $Mdn = 3$) and consistency (i.e., 87.5% of data within ± 20% of phase median). No trends emerged during the PMIR phase. Although the observed change in performance was consistent (e.g., median level change = 3), the immediacy of the effect was unclear because of the improvement during Baseline Sessions 9 and 13. Two of the data points overlapped between phases. Despite the change in performance during baseline, the relative consistency in performance during intervention indicated that PMIR had an effect on Tomas' word reading.

Similar patterns occurred for all 5 participants with one caveat. Tomas' performance increased when PMIR was introduced for Iker; however, the increase was brief and his performance decelerated before PMIR was initiated. Moreover, Tomas' performance immediately improved during PMIR, and this improvement was consistent. Visual analyses indicated that there was a functional relationship between PMIR and sight-word reading.

Across-Case Effect Size

In order to estimate an omnibus effect size across all 5 participants, we calculated individual Tau*U* values for each participant and calculated a weighted average. The significant baseline trend ($p < .15$) was controlled for Tomas. Across all 5 participants, PMIR had a large effect on sight-word reading, Tau$U = 0.91$, 95% CI [0.69, 1.13]. This effect was statistically significant ($p < .001$).

We also estimated the effect of PMIR on word reading using multilevel modeling. For Question 1, the dependent variable (i.e., number of words read correctly out of 3) represents a proportion for which a two-level logistic model is recommended over a Poisson model (Moeyaert, Ferron, Beretvas, & Van den Noortgate, 2014). Consistent with our hypothesis, visual analysis indicated that PMIR resulted in a change in level and few trends emerged during baseline or intervention phases. Thus, we decided to model changes in level but not slope. Moreover, baseline observations were stable and we did not correct for time trends. The resulting Level 1 model was as follows:

$$\log\left(\frac{\hat{\varphi}}{1-\hat{\varphi}}\right) = \beta_{0j} + \beta_{1j}Phase_{ij} \quad (1)$$

in which $\hat{\varphi}$ indicates the expected proportion of retained words within Session i for

Subject j: $\hat{\varphi} - \frac{y_{ij}}{3}$. $Phase_{ij}$ is a dummy-coded variable equaling zero if Measurement Occasion i nested within Participant j belongs to the baseline, one otherwise. As a consequence, β_{0j} and β_{1j} indicate the predicted number of retained words for Case j during the baseline phase and the treatment phase, respectively. In order to estimate the effects across cases, a second level was added by allowing the Level 1 coefficients from Equations 1 and 2 (i.e., β_{0j} and β_{1j}) to vary:

$$\beta_{0j} = \theta_{00} + u_{0j}$$

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

$$\text{with } \begin{bmatrix} u_{0jk} \\ u_{1jk} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u_0} & \sigma^2_{u_0u_1} \\ \sigma_{u_1u_0} & \sigma^2_{u_1} \end{bmatrix} \right) \qquad (2)$$

The residuals at the second level are assumed to be multivariate normally distributed.

For Question 1, we were especially interested in the estimate of σ_{00} and σ_{10}, indicating the overall average baseline level and treatment level across participants. In addition, the multilevel analysis allows estimation of the variability of these estimates across cases within the same study. These coefficients are obtained by estimating the diagonal elements of the covariance matrix in Equation 2. The $\sigma^2_{u_0}$ and $\sigma^2_{u_1}$ reflect the between-case variance in baseline level and treatment level, respectively.

Participants recalled a low number of words during the baseline phase. After back-transforming from the log odds (−2.333), the average number of retained words during baseline was 0.265, $t(4) = 2.17$, $p = .095$. The average number of retained words during the intervention phase was significantly higher. With application of the same back-transformation to the resulting log odds (3.901), the average number of retained words during the intervention phase was 2.94, $t(4) = 10.88$, $p < .001$. Thus, the across-case effect of PMIR on the reading of words equals 2.675, $t(4) = 2.15$, $p < .001$. Observed between-participant variability in the effect of PMIR was low, 0.075 ($SE = 0.166$), and statistically significant.

Question 2: Effect of PMIR on ORF

The second question examined whether the observed effect of PMIR would generalize to improvements in ORF (see Figure 2). We hypothesized that PMIR would have a delayed, small effect on ORF. As discussed below, results of the visual analysis revealed that there was no functional relationship between PMIR and ORF. Therefore, we did not estimate across-case effect sizes. The percentiles presented for each student were calculated using the AIMSweb national norms for the winter administration based on the student's current grade level.

Martin

During baseline, Martin's ORF ($Mdn = 25$) was between the first and second percentiles for third-grade students. Martin's ORF was relatively stable (80% of data within ± 20% of phase median). There was a positive trend in his ORF during baseline. During PMIR, Martin's ORF ($Mdn = 33$) was between the second and third percentiles. Although his performance varied (i.e., 72% of data points within ± 20% of phase median), there was a small positive trend in his ORF performance. When Martin's performance between phases was examined, there was no immediate change in performance and there was substantial overlap in the data. There was a small positive change in level (2 WRCM). Taken together, visual analysis did not indicate that PMIR had an effect on Martin's ORF.

Lucia

Lucia's ORF during baseline ($Mdn = 59$) was at the 11th percentile for third-grade students. There was a large absolute level change (23 WRCM) and a positive trend in her performance during baseline. Her performance during baseline was variable (i.e., 71% of data within ± 20% of phase median). During PMIR, Lucia's ORF ($Mdn = 55$) was between the 10th and 11th percentiles for third-grade students. Her performance was variable (i.e., 74% of data within ± 20% of phase median). The relative level change and absolute level change were positive and sug-

gestive of a small positive trend in her ORF during PMIR.

There was an immediate decrease in Lucia's ORF after the implementation of PMIR. Relative, absolute, and median level changes reflected a negative change in ORF between phases. The positive trend in ORF during baseline decelerated after implementation of PMIR, and there was substantial overlap between phases. Visual analysis did not indicate that PMIR had an effect on Lucia's ORF.

Iker

Iker's ORF during baseline (*Mdn* = 13) was at the second percentile for second-grade students. His performance during baseline was varied (i.e., 56% of data within ± 20% of phase median). The relative and absolute level changes during baseline were positive, and there was a small, positive trend in performance. During PMIR, Iker's ORF (*Mdn* = 23) was between the fourth and fifth percentiles for second-grade students. Iker's performance was varied (i.e., 52% of data within ± 20% of phase median). The relative level change and absolute level change were positive, and there was a small positive trend in his ORF performance during PMIR.

PMIR was associated with an immediate decrease in Iker's ORF. However, this change was not consistent with positive relative, median, and mean level changes suggesting improvement from baseline to PMIR. The positive trend in ORF during baseline accelerated after implementation of PMIR, but there was substantial overlap between phases. Visual analysis did not indicate that PMIR had an effect on Iker's ORF.

Sofia

Sofia's baseline ORF (*Mdn* = 24) was at the fifth percentile for second-grade students. Sofia's performance during baseline was somewhat varied (i.e., 75% of data within ± 20% of phase median). Estimates of trend were inconsistent with negative relative and absolute level changes observed but a small positive trend in her ORF slope. During PMIR, Sofia's median ORF (*Mdn* = 28) was at the sixth percentile for second-grade students. Sofia's performance was varied (i.e., 52% of data within ± 20% of phase median), and this variability appeared to increase throughout the PMIR phase. The relative level change (14 WRCM) and absolute level change (31 WRCM) were large, and there was a positive trend in her ORF performance during PMIR.

PMIR was associated with an immediate decrease in Sofia's ORF (3 WRCM). Her relative level change was negative, while the median and mean level changes were positive. The positive trend in ORF during baseline accelerated after implementation of PMIR. There was substantial overlap in the data between phases. Visual analysis did not indicate that PMIR had an effect on Sofia's ORF.

Tomas

During baseline, Tomas' median ORF (*Mdn* = 56) was at the 18th percentile for second-grade students. Tomas' performance during baseline was stable (i.e., 93% of data within ± 20% of phase median). Within-phase relative and absolute level changes were negative, and his ORF performance decreased over time. During PMIR, Tomas' ORF (*Mdn* = 54) was at the 16th percentile for second-grade students. His ORF during PMIR was stable (100% of data within ± 20% of phase median). There was a small, negative relative level change (−0.5 WRCM) and large absolute level change (5 WRCM). His ORF slightly decreased during PMIR.

There was a small, positive improvement in ORF between phases. This positive effect was not consistent, with small, positive changes in the relative and absolute level changes but negative changes in the median and mean levels. The decreasing trend in ORF during baseline decelerated after implementation of PMIR but remained negative. There was substantial overlap between phases. Visual analysis did not indicate that PMIR had an effect on ORF for Tomas.

Overlap of Words Taught

Because of the lack of effect on ORF, we examined the overlap between the unknown words taught and the words contained

in the AIMSweb passages. While some unknown words were common across students, each word was only included once for these analyses. Iker, Sofia, and Tomas were assessed using second-grade passages. A total of 132 of the 155 unknown words taught (81%) appeared at least once in an AIMSweb passage. The median number of passages that included an unknown word was 5 (range = 1 to 26), and the average number of appearances per page was 1.67 (range = 1 to 6.11). The unknown words appeared a total of 1,467 times, which represents approximately 22.3% of the total words in the second-grade passages.

Lucia and Martin were assessed using third-grade passages. A total of 95 of the 142 unknown words taught (67%) appeared at least once in an AIMSweb passage. The median number of passages that included a given unknown word was 4 (range = 1 to 27). The average number of appearances per page was 1.50 (range = 1 to 7.33). The unknown words appeared a total of 1,016 times, which represents approximately 12.6% of the total words in the third-grade passages. Although we did not analyze the placement of the words in the passage to determine if the student reached the words, these data indicate that there was little overlap between the words taught and the words contained in the AIMSweb passages.

Question 3: Maintenance of PMIR Effects

The third research question pertained to the durability of any observed PMIR effects. Given the lack of a functional relationship between PMIR and ORF, we did not discuss maintenance data for ORF. We assessed maintenance by examining the number of words the student read correctly from all of the words taught during PMIR. This assessment occurred 10 days after the final intervention session for all students, except Martin, who completed the assessment 8 days after the final intervention session because of absences.

Sofia (47 of 54; 87%) and Lucia (53 of 69; 77%) were able to read more than 75% of the words taught during PMIR sessions after intervention completion. Martin (51 of 75; 68%) and Iker (42 of 63; 67%) correctly read a smaller percentage of words taught during PMIR but still more than 50%. Overall, participants correctly read an average of 74.6% of the words taught (SD = 9.4%) during PMIR.

Treatment Acceptability

Treatment acceptability data were collected from all target students except for Tomas. The four target students reported that PMIR helped them learn new words (M = 9). The target students reported generally liking PMIR, although one participant reported not liking reading the ORF passages prior to each session. All target students reported they would participate in PMIR again to learn new words. Tutors reported that PMIR helped other students learn other words (M = 8.8). All tutors reported liking teaching the target students and reported that they would want to use PMIR to teach other students again. One tutor did comment on the perceived difficulty of the interspersal procedure, reporting that she felt like she made too many mistakes during PMIR.

DISCUSSION

Previous research has supported the use of IR for teaching words; however, the intensity of the intervention likely precludes its use in most situations. Methods to decrease the resource intensity of IR while maintaining the causal mechanisms may increase the relevance of IR for schools that serve large populations of ELL students. Using peer interventionists appears to be a promising modification of IR for ELL students, but this modification of IR has yet to be empirically investigated. To fill this gap in the literature, this efficacy study investigated the effects of PMIR for ELL students.

Results from this study indicated that there was a functional relationship between PMIR and word reading. Visual analyses suggested that PMIR had an effect on word reading for all 5 participants. Large across-case effect sizes were found using Tau$U$ and multilevel modeling of raw data. The estimated effect sizes were consistent across all 5 par-

ticipants. Results also suggested that the improvements in word reading were maintained over time. Approximately 10 days after the intervention, participants retained a high percentage of taught words.

The large effect of PMIR on word reading was generally consistent with the average treatment effect of IR implemented by adults found in 12 studies (Burns et al., 2012). In addition, the peer interventionists indicated that PMIR was easy to implement after some practice. These findings are promising because they suggest peers can deliver IR with a similar degree of effectiveness to adults, although we did not test that hypothesis here. Reducing the adult time necessary to deliver IR while maintaining the causal mechanisms of the intervention could increase the contextual validity of IR. Moreover, the use of PMIR is consistent with the Institute of Education Sciences recommendations for providing structured instructional activities for ELL students with peers of differing English proficiencies (Gersten et al., 2007).

Contrary to our hypothesis, IR did not have an effect on participants' ORF. Recent meta-analytic research indicated that IR was associated with a moderate to large effect on students' ORF in three studies (Burns et al., 2012). Multiple hypotheses may explain the lack of effect. First, we used grade-level passages to assess ORF. Given students' baseline performance ($<$ 11th percentile), these passages were likely at the frustration level. Using instructional level passages may have been more sensitive to small changes in ORF performance (Shapiro, 2011). Second, we did not use IR to preteach words contained in the passages. In previous studies of IR, researchers taught unknown words that were contained in the passage and then examined differences in ORF (e.g., Burns, 2007; Burns, Dean, & Foley, 2004). There was little overlap in the unknown words and the AIMSweb passages in this study. Third, nine words being taught per week may have been too few to affect students' ORF. We taught three words per session to avoid exceeding students' acquisition rates. Exceeding a student's acquisition rate can reduce learning and reading (Helman & Burns, 2008), a result that may have led us to conclude that peers could not effectively deliver IR. Because of the small number of words taught per week, word reading fluency may have provided a better measure of generalization.

Sight-word reading is an important precursor to upper-level reading skills such as fluency and comprehension (National Institute of Child Health and Human Development, 2000). ELL students with early deficits in word reading skills are likely to have later delays in word reading skills (Lesaux et al., 2007). Improving word reading is an important intervention target for students with a limited sight-word vocabulary (Hudson et al., 2012; Vaughn et al., 2005). However, these improvements are unlikely to generalize to improvements in ORF without incorporating practice of connected text reading (Daly, Neugebauer, Chafouleas, & Skinner, 2015). Future research could examine the efficacy and efficiency of using peers to deliver IR within a complex intervention package that also targets fluency.

Limitations and Future Directions for Research

Results from this study must be interpreted in the context of the limitations. First, the use of peer tutors is a promising approach for increasing the contextual validity of IR. Our findings suggest that peers can deliver IR with similar effectiveness to adults. Documenting the efficacy of PMIR was a necessary precursor to comparing PMIR with other interventions, but it is important to note that these data do not answer questions regarding the comparative effectiveness or the comparative efficiency of PMIR. Future research is necessary to examine the efficiency of PMIR compared with other interventions delivered by adults or other peers.

Second, single-case designs have strong internal validity but replication is needed to establish an intervention as evidence based (Kratochwill et al., 2013). In this initial efficacy study, the effects of PMIR were replicated across 5 participants with small variability in the effects between participants. Additional research across participants and outcomes is needed before PMIR can be considered evidence based.

For example, future research could examine if PMIR is effective for teaching words to ELL students who speak a primary language other than Spanish, if PMIR is effective for teaching skills other than word reading (e.g., math facts), or if PMIR can be used as a reciprocal peer tutoring intervention.

Third, we used randomization to select intervention start points prior to the study. Incorporating randomization into single-case research may improve the credibility of the results (Kratochwill & Levin, 2010) but decreases the flexibility found in the traditional response-guided approach. In this study, Tomas began PMIR while his baseline performance was variable. We started PMIR during the session that was randomly assigned, but a response-guided approach would have allowed us to extend the baseline until Tomas' performance was stable. We prioritized the benefits of randomization because this was an efficacy trial of PMIR, but this approach did introduce ambiguity into the visual analysis.

Fourth, we combined data from 5 participants in order to estimate the overall average effect of PMIR. Previous research on the multilevel modeling of single-case data (Moeyaert et al., 2014) indicated that our sample size was sufficient for obtaining an unbiased and precise estimate of the fixed (i.e., treatment) effects, but caution is warranted when interpreting between-case variance estimates as they may be biased. More research on PMIR using MBDs would allow for more accurate estimates of variance components.

CONCLUSION

School psychologists are likely to be involved in supporting ELL students' reading achievement. This requires evidence-based interventions, studied with ELL students directly, that are contextually valid. PMIR had similar effects to previous studies of IR delivered by adults. These results also provide additional support for using IR procedures with ELL students. Using peer interventionists has promise for maintaining the benefits of IR while increasing the contextual validity of the intervention. PMIR could also be used as a structured learning activity for peers of different English proficiency levels. More research is needed before PMIR can be considered an evidence-based practice, but practitioners could consider its use for teaching high-frequency words to ELL students. As with any intervention, practitioners must monitor short-term and long-term outcomes carefully to ensure PMIR is having the desired effects.

REFERENCES

Albers, C. A., & Martinez, R. S. (2015). *Promoting academic success with English language learners.* New York, NY: Guilford Press.

August, D., McCardle, P., & Shanahan, T. (2014). Developing literacy in English language learners: Findings from a review of the experimental research. *School Psychology Review, 43*, 490–498. doi:10.17105/SPR-14-0088.1

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth.* Mahwah, NJ: Erlbaum.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*, 39–56.

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. *School Psychology Quarterly, 22*, 297–313. doi:10.1037/1045-3830.22.3.297

Burns, M. K., Dean, V. J., & Foley, S. (2004). Preteaching unknown words with incremental rehearsal to improve reading fluency and comprehension with children identified as reading disabled. *Journal of School Psychology, 42*, 303–314. doi:10.1016/j.jsp.2004.04.003

Burns, M. K., & Helman, L. A. (2009). Relationship between language skills and acquisition rate of sight words among English language learners. *Literacy Research and Instruction, 48*, 221–232. doi:10.1080/19388070802291547

Burns, M. K., & Sterling-Turner, H. E. (2010). Comparison of efficiency measures for academic interventions based on acquisition and maintenance. *Psychology in the Schools, 47*, 126–134. doi:10.1002/pits.20458

Burns, M. K., Zaslofsky, A. F., Kanive, R., & Parker, D. C. (2012). Meta-analysis of incremental rehearsal using phi coefficients to compare single-case and group designs. *Journal of Behavioral Education, 21*, 185–202. doi:10.1007/s10864-012-9160-2

Crosse, S., Williams, B., Hagen, C. A., Harmon, M., Ristow, L., DiGaetano, R., . . . Derzon, J. H. (2011). *Prevalence and implementation fidelity of research-based prevention programs in public schools: Final report.* Washington, DC: U.S. Department of Education.

Daly, E. J., Neugebauer, S., Chafouleas, S., & Skinner, C. H. (2015). *Interventions for reading problems* (2nd ed.). New York, NY: Guilford.

DuBois, M. R., Volpe, R. J., & Hemphill, E. M. (2014). A randomized trial of a computer-assisted program tar-

geted letter-sound expression. *School Psychology Review, 43*, 210–221.

Ferron, J., Moeyaert, M., Van Den Noortgate, W., & Beretvas, S. (2014). Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods, 19*, 493–510. doi: 10.1037/a0037038

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

Fuchs, D., & Fuchs, L. S. (2005). Peer-assisted learning strategies: Promoting word recognition, fluency, and reading comprehension in young children. *Journal of Special Education, 39*, 34–44. doi:10.1177/00224669050390010401

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades: A practice guide* (NCEE 2007-4011). Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides

Goldenberg, C. (2010). Reading instruction for English language learners. In M. Kamil, P. D. Pearson, E. Moje, & P. Afflerbach (Eds.), *Handbook of reading research* (Vol. 4, pp. 648–710). doi:10.4324/9780203840412.ch28

Greenwood, C. R., Arreaga-Mayer, C., Utley, C. A., Gavin, K. M., & Terry, J. (2001). Classwide peer tutoring learning management system. Applications with elementary-level English language learners. *Remedial and Special Education, 22*, 34–47. doi: 10.1177/074193250102200105

Hattie, J. A. (2009). *Visual analysis. A synthesis of over 800 meta-analyses relating to achievement.* London, England: Routledge.

Helman, L. A., & Burns, M. K. (2008). What does oral language have to do with it? Helping young English-language learners acquire a sight word vocabulary. *The Reading Teacher, 62*, 14–19. doi:10.1598/RT.62.1.2

Hudson, R. F., Torgesen, J. K., Lane, H. B., & Turner, S. J. (2012). Relations among reading skills and subskills and text-level reading proficiency in developing readers. *Reading and Writing, 25*, 483–507. doi: 10.1007/s11145-010-9283-6

Joseph, L., Eveleigh, E., Konrad, M., Neef, N., & Volpe, R. (2012). Comparison of the efficiency of two flashcard drill methods on children's reading performance. *Journal of Applied School Psychology, 28*, 317–337. doi:10.1080/15377903.2012.669742

Kena, G., Aud, S., Johnson, F., Wang, X. Zhang, J., Rathburn, A., . . . Kristapovich, P. (2014). *The condition of education 2014* (NCES 2014-083). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2014/2014083.pdf

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

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*, 122–144. doi:10.1037/a0017736

Kretlow, A. G., & Helf, S. S. (2013). Teacher implementation of evidence-based practices in Tier 1. A national survey. *Teacher Education and Special Education, 36*, 167–185. doi:10.1177/0888406413489838

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

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: Findings from a 5-year longitudinal study. *Journal of Educational Psychology, 99*, 821–834. doi:10.1037/0022-0663.99.4.821

Matchett, D. L., & Burns, M. K. (2009). Increasing word recognition fluency with an English language learner. *Journal of Evidence Based Practices in Schools, 10*, 194–209.

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

Moore, B. A., & Klingner, J. K. (2014). Considering the needs of English language learner populations: An examination of the population validity of reading intervention research. *Journal of Learning Disabilities, 47*, 391–408. doi:10.1177/0022219412466702

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

Northwest Evaluation Association. (2009). Technical manual for Measures of Academic Progress and Measures of Academic Progress for primary grades. Lake Oswego, OR: Author.

Nist, L., & Joseph, L. M. (2008). Effectiveness and efficiency of flashcard drill instructional methods on urban first-graders' word recognition, acquisition, maintenance, and generalization. *School Psychology Review, 37*, 294–308.

Parker, R. I., & Vannest, K. L. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*, 357–367. doi:10.1016/j.beth.2008.10.006

Parker, R. I., Vannest, K. L., & Davis, J. L. (2014). Nonoverlap analysis for single-case research. In T. R. Kratochwill & J. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 127–151). doi:10.1037/14376-005

Pearson, Inc. (2012). Aimsweb technical manual. Bloomington, MN: Author.

Peterson, M., Brandes, D., Kunkel, A., Wilson, J., Rahn, N. L., Egan, A., & McComas, J. (2014). Teaching letter sounds to kindergarten English language learners using incremental rehearsal. *Journal of School Psychology, 52*, 97–107. doi:10.1016/j.jsp.2013.11.001

Rahn, N. L., Wilson, J., Egan, A., Brandes, D., Kunkel, A., Peterson, M., & McComas, J. (2015). Using incremental rehearsal to teach letter sounds to English language learners. *Education and Treatment of Children, 38*, 71–92. doi:10.1353/etc.2015.0000

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Read-*

ing and Writing, 13, 107–121. doi:10.1080/1057356970130202

Shapiro, E. S. (2011). *Academic skills problems* (4th ed.). New York, NY: Guilford Press.

Skinner, C. H., McCleary, D. F., Skolits, G. L., Poncy, B. C., & Cates, G. L. (2013). Emerging opportunities for school psychologists to enhance our remediation procedure evidence base as we apply response to intervention. *Psychology in the Schools, 50*, 272–289. doi:10.1002/pits.21676

Speece, D. L., & Ritchey, K. D. (2005). A longitudinal study of the development of oral reading fluency in young children at risk for reading failure. *Journal of Learning Disabilities, 38*, 387–399. doi:10.1177/00222194050380050201

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology, 89*, 276–288. doi:10.1037/0022-0663.89.2.276

Tucker, J. A. (1989). *Basic flashcard technique when vocabulary is the goal.* Unpublished teaching materials.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2015). *National Assessment of Educational Progress (NAEP) 1990–2013 mathematics and reading assessments.* Retrieved from http://www.nationsreportcard.gov/reading math 2015/#/student-groups

Utley, C. A., & Mortweet, S. L. (1997). Peer-mediated instruction and interventions. *Focus on Exceptional Children, 29*, 1–23.

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

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence Based Communication Assessment and Intervention, 2*, 142–151. doi:10.1080/17489530802505362

Van den Noortgate, W., Opdenakker, M., & Onghena, P. (2005). The effects of ignoring a level in multilevel analysis. *School Effectiveness and School Improvement, 16*, 281–303. doi:10.1080/09243450500114850

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

Varma, S., & Schleisman, K. B. (2014). The cognitive underpinnings of incremental rehearsal. *School Psychology Review, 43*, 222–228.

Vaughn, S., Mathes, P. G., Linan-Thompson, S., & Francis, D. J. (2005). Teaching English language learners at risk for reading disabilities to read: Putting research into practice. *Learning Disabilities Research & Practice, 20*, 58–67. doi:10.1111/j.1540-5826.2005.00121.x

Volpe, R. J., Burns, M. K., DuBois, M., & Zaslofsky, A. F. (2011). Computer-assisted tutoring: Teaching letter sounds to kindergarten students using incremental rehearsal. *Psychology in the Schools, 48*, 332–343. doi:10.1002/pits.20557

Yaghoub Zadeh, Z., Farnia, F., & Geva, E. (2012). Toward modeling reading comprehension and reading fluency in English language learners. *Reading and Writing, 25*, 163–187. doi:10.1007/s1145-010-9252-0

Date Received: September 8, 2015
Date Accepted: April 5, 2016
Associate Editor: Elizabeth Meisinger ■

David A. Klingbeil, PhD, is an assistant professor in the School Psychology Program at the University of Wisconsin–Milwaukee. His research interests include identifying and modifying evidence-based practices to improve their usefulness for urban schools.

Mariola Moeyaert, PhD, is an assistant professor of statistics and research design in the Department of Educational Psychology and Methodology at the State University of New York. Her major research interests and publications are in the field of multilevel analysis, meta-analysis, and single-case experimental data.

Christopher T. Archer, MS, is a graduate student in the School Psychology Program at the University of Wisconsin–Milwaukee. His research interests include mood disorders and social and emotional learning.

Tatenda M. Chimboza, MS, is a graduate student in the School Psychology Program at the University of Wisconsin–Milwaukee. Her research interests include issues related to improving school climate through positive behavioral interventions and promoting home–school collaborations with urban students.

Scott A. Zwolski Jr., MEd, is a doctoral student in the School Psychology Program at Loyola University Chicago. His research interests include examining the impact of poverty-related stress and identifying efficient ways of promoting resilience, grit, and a growth mindset within the context of schools.