Running head: WORD OVERLAP

Effects of Word Overlap on Generalized Gains from a Repeated Readings Intervention

Brian K. Martens

Syracuse University

Nicholas D. Young

Medical College of Wisconsin and Children's Hospital of Wisconsin

Michael P. Mullane, Emily L. Baxter, Samantha J. Sallade, David Kellen, Stephanie J. Long,

William E. Sullivan, Allison J. Womack, and Joseph Underberg

Syracuse University

(in press, *Journal of School Psychology*)

Please address correspondence to Brian K. Martens, Department of Psychology, Syracuse University, 430 Huntington Hall, Syracuse, NY 13244-2340. Email: bkmarten@syr.edu.

Abstract

We compared children's gains in oral reading fluency after applying a standard fluency-building intervention to 3 training passages that differed in word overlap (high, low, and multiple exemplar) with an untrained generalization passage. Participants were 132 White and Hispanic third-grade children from 2 schools in the northeast and mountain west. Children were randomly assigned within classrooms to the 3 word overlap conditions, pre-tested on their assigned training and a common generalization passage, received a fluency-building intervention on their assigned training passage, and then post-tested on the same 2 passages. Regression analyses were conducted to examine the effects of word overlap condition on the children's fluency gains after controlling for pre-test fluency and classroom. Results revealed significantly larger priming and generalization effects for the multiple exemplar versus both the low- and high-word overlap conditions. Survival curves showed that a significantly larger proportion of children in the multiple exemplar condition survived as generalized responders at all generalization levels relative to the other two conditions. Implications for assessing and promoting generalized oral reading fluency in response-to-intervention models and directions for future research are discussed.

Key words: oral reading fluency, repeated readings, word overlap, generalization, survival analysis

Effects of Word Overlap on Generalized Gains from a Repeated Readings Intervention

Curriculum-based measurement of oral reading fluency (CBM-R) is a key component of response-to-intervention (RtI) models (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013a; Christ & Silberglitt, 2007). CBM-R involves the administration of equivalent grade-level passages to identify children at risk for reading failure (screening), to establish oral reading fluency standards predictive of performance on high-stakes tests (benchmarking), and to measure children's growth over time in response to increasingly more intensive instruction (progress monitoring). CBM-R probes are uniquely suited for these purposes as they are inexpensive, quick and easy to administer, involve sensitive rate measures of child responding, assess an important, foundational dimension of reading competence (oral reading fluency), and can be administered repeatedly (Ardoin et al., 2013; Good, Simmons, & Kame'enui, 2001).

When monitoring children's progress in an RtI model, it is common for intervention to be applied to one set of passages (e.g., a basal curriculum series or a "practice" set) while gains in oral reading fluency are measured on a different, standardized set of global outcome measures (e.g., AIMSweb or DIBELS passages; Ardoin & Christ, 2009). This is done for several reasons. First, the goal of intervention is not to increase fluency on a particular trained passage, but to promote fluent passage reading in general. As such, assessing children's oral reading fluency on untrained passages is an indicator of more robust intervention effects (e.g., Ardoin, McCall, & Klubnik, 2007; Daly, Martens, Hamler, Dool, & Eckert, 1999). Second, standardized CBM-R probes are believed to represent most basal curricula and therefore serve as global indicators of reading competence (Good et al., 2001). Third, CBM-R global outcome measures allow intervention teams to aggregate children's progress data at different levels (e.g., class, school, district), because the same pool of equivalent, grade-level passages are used at different sites throughout the school year.

Testing children on passages that differ from those on which they are trained has potential to create both an instructional design problem and an assessment problem in RtI models (Martens, Daly, & Ardoin, 2015). The instructional design problem occurs when choosing passages for children to practice during intervention sessions. For example, one empirically-supported intervention for increasing children's oral reading fluency is repeated readings (RR) combined with rate-contingent reinforcement (CR), modeling via listening passage preview (LPP), and phrase drill (PD) error correction (e.g., Begeny, Yeager, & Martinez, 2012; Chard, Vaughn, & Tyler, 2002; Jones et al., 2009; Martens et al., 2007). To date, research comparing generalized gains in oral reading fluency from reading low-word overlap (LWO), high-word overlap (HWO), and multiple-exemplar (ME) passages has produced mixed results. For example, Faulkner and Levy (1994) had 46 third-grade children read the same passage (RR), a HWO passage, a paraphrased LWO passage that shared the same story line, plot, and characters, or an unrelated LWO passage before reading a second passage. For good readers, reading times on the second passage were fastest after reading the same (RR) or a paraphrased LWO passage, indicating a positive priming effect from reading these types of passages first. Prior reading of the HWO and unrelated LWO passage produced the slowest reading times on the second passage (i.e., a negative priming effect). For poor readers, reading times on the second passage were fastest in the RR and HWO conditions.

Daly, Martens, Dool, and Hintze (1998) conducted a brief experimental analysis comparing LPP, RR, CR, and PD error correction components. If generalized increases were not observed on HWO passages, they were explicitly programmed by applying the most effective combination of intervention components to both the training and HWO passages (i.e., a sequential modification strategy; Stokes & Baer, 1977). One child showed generalized increases in WRCM when intervention was applied to only the training passage, but the other two children required sequential modification on the HWO passage.

In Ardoin et al. (2007), only half of the participants showed larger generalized gains after LPP, RR, CR, and PD error correction were applied to one LWO passage (RR) versus two LWO passages which the authors characterized as multiple exemplar (ME) training. Ardoin, Binder, Zawoyski, Foster, and Blevins (2013b) found that children's mean reading times for a generalization passage decreased whether they read that same passage four times (RR) or four different passages each one time (ME). However, times on the fourth reading were significantly shorter for children assigned to the RR condition. Finally, Silber and Martens (2010) found that children showed similar gains on a 16-sentence untrained generalization passage whether they practiced a 4-sentence passage containing representative key words and sentence structures (their version of ME training) or a 16-sentence HWO passage.

The assessment problem in RtI models is created when intervention teams attempt to establish criteria for generalized gains in oral reading fluency to identify treatment responders. On the one hand, if fluency gains on untrained global outcome measures depend on their similarity to trained passages, then different generalization criteria will be needed for different word overlap conditions. To our knowledge, this issue has not been examined in previous research. On the other hand, if expected gains on global outcome measures are not observed, it is unclear whether the intervention failed to promote generalization or if the CBM-R passage used was insensitive to the child's generalized gains. If the latter, effective treatments might be terminated prematurely rather than assessing for generalization more systematically, potentially leading to wasted time and resources (Ardoin et al., 2013a).

To date, the relative effects of different word overlap conditions on children's generalized oral reading fluency have not been compared in the same study. In addition, prior studies that employed ME passages to promote generalization had children read multiple LWO passages once or twice (e.g., Ardoin et al., 2013b; Ardoin et al. 2007) or read representative key words and sentence structures of a generalization passage repeatedly (Silber & Martens, 2010). An alternative approach to designing ME passages might involve changing words in a training passage to be exemplars of words in the generalization passage, and then have children repeatedly read the ME training passage. This would allow children to attempt to read unfamiliar words in the generalization passage by blending known onsets and rimes practiced in the ME training passage. Such a reading-by-analogy strategy would be expected to better promote generalization than practicing different words in a LWO passage or the same words in different order in a HWO passage. To examine these effects, we conducted an analogue study in which one dose of an evidence-based fluency-building intervention (LPP, RR, CR, and PD error correction) was applied to three passages that differed systematically in their amount and type of word overlap with an untrained generalization passage. The goals of the study were twofold; (a) to examine the effect of word overlap condition on the direct and generalized gains of a fluency-building intervention, and (b) to identify the proportion of children in each word overlap condition who would be classified as generalized responders (i.e., who "survived") at increasingly higher levels of generalization.

Method

Participants and Setting

Participants were 132 third-grade regular education students in nine classrooms from two rural public schools, one in the northeast and one in the mountain west. Demographic information for each experimental group (LWO, HWO, ME) and the total sample is presented in Table 1. About two-thirds of the children were White and the rest Hispanic with the exception of one child who was Multiracial. A slight majority of the children were male. About equal percentages of children read at mastery (100> WRCM) and instructional (70-99 WRCM) levels at pre-test, with a smaller percentage reading at a frustrational (<70 WRCM) level (Shapiro, 2004). The mean age of the northeast sample (n=74) was 9.1 years (age data were unavailable for the mountain west sample). The three experimental groups did not differ significantly in terms of gender ($X^2[2]=1.7$, $p=.42$), ethnicity ($X^2[4]=3.5$, $p=.48$), WRCM on the assigned training passage at pre-test ($F[2,129]=2.26$, $p=.11$), or instructional level composition ($X^2[4]=4.7$, $p=.32$).

Sessions took place in the children's classrooms during regularly scheduled reading instruction, with all pre-testing, fluency training, and post-testing occurring in the same session. Sessions were conducted by a team of 16 experimenters including one faculty member, 13 graduate students in two School Psychology training programs, and two undergraduate research assistants. Pre- and post-test assessments were conducted with each child individually, whereas fluency training was conducted in small groups of two to four children.

Materials

Training and generalization passages. Four third-grade passages, one generalization passage (*Mama Frog*) and three training passages (LWO-*Billy Waiting*, HWO-*Baby Frogs*, ME-*Papa Dog*), were developed for use in the study. Both the generalization and LWO passages were existing passages taken directly from the AIMSweb database (http://www.aimsweb.com). The HWO and ME passages were created by modifying the generalization passage. To construct the HWO passage, we went through the generalization passage sentence by sentence, retaining some words as overlapping words and replacing some nouns and verbs with new words that did not appear anywhere in the original passage. For example, the sentence in the generalization passage, "Those baby frogs just loved to eat black bugs." was changed to, "Her baby frogs loved to munch on earthworms." in the HWO passage (2 new words). Word replacements and rearrangements were restricted to only those that maintained a coherent story line. Because the HWO passage was based largely on the generalization passage, the story had a similar storyline and contained the same characters. To construct the ME passage, we first identified several exemplar words for each of the words in the generalization passage. An exemplar word was defined as any word that shared the same number of syllables, same ending sound, same vowel and vowel sound, and same diphthong and diphthong sound (e.g., “should” and “would”), and in some cases one or more of the same consonants as a word in the generalization passage (e.g., “papa” for “mama”, “pack” for “back”). Because the ME passage contained more exemplar words, the storyline for the ME passage was developed around these exemplar words. The ME storyline was based on providing exemplar word exposure and not on providing exemplar sentence structure exposure.

Both the ME and HWO passages were revised several times in order to meet desired percentages of new, exemplar, and overlapping words based on our definitions regardless of location in the passage. The final passages were all between 152 and 155 words in length with Spache readability indices of 2.9, 2.7, 2.8, and 2.9, for the generalization, LWO, HWO, and ME passage, respectively. Characteristics of the four passages are summarized in Table 2 including total number of words, number and percentage of content and function words, and number and percentage of low- and high-frequency words. High frequency words were based on the first 100 words from the Fry list. Also shown in Table 2 for the LWO, HWO, and ME passages are the number and percentage of new, exemplar, and overlapping words compared to the generalization passage. (Passages are available from the first author upon request.)

Experimenter equipment and protocols. Experimenters were equipped with a clipboard, pencil, and stopwatch as well as a packet of materials including examiner and student copies of each passage, pre- and post-test assessment protocols, and a small-group training protocol. Each experimenter also had a clear plastic bag that contained a variety of school supplies (e.g., pencils, pens, erasers, rulers) for students to choose from as prizes for beating their group's score during fluency training.

Procedures

The study's procedures are summarized in Figure 1. As shown in the figure, children were randomly assigned within each classroom to the three word overlap conditions (LWO, HWO, ME). In one 30-min session for each classroom, a team of eight researchers individually pre-tested all children on the training passage to which they were assigned followed immediately by the common generalization passage (PreT and PreG). Pre-testing was completed in two or three rounds depending on class enrollment. All children were then given a fluency-building intervention on their assigned training passage in small groups, and individually post-tested on the same training and generalization passages (PostT and PostG). Although children in each word overlap condition practiced a different training passage, all children read the same generalization passage.

Pre-testing. Using standard CBM-R procedures (Shapiro, 2004), each child was given a copy of the training passage and told, "When I say begin, start reading aloud at the top of the page and read the whole story. If you come to a word you don't know, I'll tell it to you. Be sure to do your best reading." The examiner said begin, started the stopwatch, and followed along on the examiner copy marking error words. Errors included omissions, mispronunciations, and hesitations for more than 3-s. When the student finished reading the entire passage, the examiner stopped the stopwatch and recorded the elapsed time. The same procedure was then repeated for the generalization passage.

Fluency training. After pre-testing on both passages, all students were trained in groups of two to four on the training passage. Training consisted of choral PD error correction on all words missed at pre-test by any student in the group, LPP by an experimenter, three timed, choral RR's, and CR for increasing WRCM from the first to third RR. PD error correction consisted of the experimenter pointing to each error word and saying, "<error word>. Repeat after me <the phrase containing the error word>." Students in the group were prompted as needed to say each 2- to 4-word phrase a total of three times for each error word. Any errors were corrected and students were praised after every two or three sets of drills. Immediately following PD error correction, the experimenter read the passage aloud to the group (LPP) at a comfortable reading rate (approximately 130 WRCM) while the students followed along reading the words to themselves. The students then practiced reading the passage chorally three times (RR) while the experimenter timed the first and third readings. If an error occurred during choral reading, the experimenter stopped the group, modeled the correct reading of the word, and had the students repeat the word three times. Error corrections were included in the timing of the choral readings. If the group read the passage faster and with fewer errors on the third versus first reading, each child was able to pick a prize from the prize bag (CR).

Post-testing. All children in a classroom were individually post-tested on both the training and generalization passages in that order. Post-test procedures and instructions were identical to those used at pre-test.

Procedural Integrity and Interscorer Agreement

Procedural integrity for a sample of the small-group fluency trainings and pre- and post-test assessment protocols as well as interscorer agreement for WRCM on a sample of passages were assessed across all experimenters and conditions at both sites. Integrity of the small-group fluency trainings (36% of children) was assessed by two independent observers equipped with the same training protocol as the experimenters. Integrity of the small-group fluency trainings was 100%. Integrity of the pre- and post-test assessment protocols (28% of children) was assessed by the same independent observers equipped with assessment protocols. Integrity of the assessment protocols was 99.4% (range 75-100%). Interscorer agreement for WRCM was calculated on a word-by-word basis as the number of agreements divided by the number of agreements plus disagreements multiplied by 100%. Interscorer agreement for WRCM (27% of children) was 99.2% (range 95-100%).

Experimental Design and Dependent Variables

A 3 (word overlap condition) X 4 (testing occasion) split-plot factorial design was used to examine children's WRCM at each of the four testing occasions (PreT, PreG, PostT, PostG). Closer inspection of these data however revealed differences in children's PreG scores depending on which training passage they read first. This suggested that PreG scores were not a true pre-test, but reflected a differential priming or "warm-up" effect from having just read an assigned training passage (e.g., Faulkner & Levy, 1994). To quantify this priming effect on the same scale as the direct and generalized effects of fluency building, we calculated three gain scores for each child based on their PreT score as dependent variables for the regression analyses; (a) gain from PreT to PreG (priming effect from having just read an assigned training passage), (b) gain from PreT to PostT (training effect from fluency building), and (c) gain from PreT to PostG (generalization effect from fluency building).

A secondary goal of the study was to compare the proportion of generalized responders in each word overlap condition. The dependent variable for this analysis was calculated in two steps. First, we identified those children in each condition who showed positive gains of any magnitude on both the training and generalization passages from pre- to post-test. The vast majority of children (113 of 132 or 85.6%) met this criterion, with only 7 removed from the LWO group, 12 from the HWO group, and none from the ME group. Second, we scaled the magnitude of each child's generalization effect as a percentage of their training effect, and removed as outliers those children whose generalization effects were more than three times (i.e., 300% of) their training effects. Only four children were removed as outliers (2 from the ME group and 2 from the LWO group), leaving 109 children for analysis.

Data Analysis

Regression analyses. The regression model for each dependent variable (prime, train, and generalize) included parameters capturing the main effects of word overlap condition (Condition) while controlling for Classroom and WRCM on the assigned training passage at pretest (Pretest WRCM). Condition was dummy coded using the ME condition as the reference group as we hypothesized that children in this condition would show larger gains than children in either the HWO or LWO conditions. We controlled for Pretest WRCM and Classroom as (a) previous research suggested that effects of oral reading fluency interventions are likely to differ depending on children's instructional level (e.g., Bonfiglio, Daly, Martens, Lan-Hsiang, &

Corsaut, 2004; Daly, Martens, Kilmer, & Massie, 1996), and (b) even though we randomly assigned children to Condition *within* classrooms, there was no reason to expect equivalent reading performance *across* classrooms. Regression analyses were conducted using the statistical language R (R Core Team, 2018), and testing was conducted using Type III sums of squares. The residuals of each model were found to be unrelated with the model's predictions and generally in line with the assumption that they were normally distributed.

Survival analysis. Rather than adopt a single, arbitrary cut score to define generalized responders (e.g., 50% of their training effect), we compared groups throughout the entire range of generalization effects (0% to 300% of the training effect) by conducting a Kaplan-Meier survival analysis of the children in each condition. Survival analyses were conducted using SPSS version 24.0.0.0. Differences between experimental groups were evaluated quantitatively using log rank follow-up pairwise comparisons, as well as visual inspection of graphed survival functions (i.e., curves). While the ME condition (n=42) contained slightly more participants than the LWO (n=34) and HWO (n=33) conditions, this difference was not considered large enough to skew survival curve displays or influence sensitivity for detecting differences between groups.

Survival analyses are used to compare the efficacy of alternative medical treatments by identifying the percentage of individuals who have not yet reached a specified end point (i.e., have survived) at increasingly longer durations from the start of treatment. Survival curves are typically graphed with percentage of the sample surviving on the y-axis and time from the start of treatment on the x-axis. As the name suggests, death is a common end point for potentially terminal conditions, but other significant end points may include relapse or tumor progression (e.g., Morita et al., 2009). Treatments associated with higher survival curves are more effective as more patients survived at all points in time from when treatment started.

For the present analysis, we graphed the percentage of children who survived as generalized responders (y-axis) in each condition at increasingly higher levels of generalization effects (x-axis). The condition associated with the highest survival curve would be the most effective at promoting generalization as more children survived as generalized responders at all generalization levels (0% to 300% of their training effects).

Results

Regression Analyses

Children's mean WRCM at each testing occasion by condition are presented in Figure 2. Table 3 summarizes the regression results for Condition as a predictor of the prime, train, and generalize effects including unstandardized beta coefficients and their standard errors after controlling for Classroom and Pretest WRCM as covariates. The full model R^2 (i.e., with all predictors included) is reported at the bottom of each table. Pretest WRCM was not a significant predictor of the prime ($F[1,119]=.07$, $p=.79$), train ($F[1,120]=2.97$, $p=.09$), or generalize ($F[1,120]=1.20$, $p=.27$) effects. Classroom was a significant predictor of the train ($F[8,120]=2.15$, $p=.04$) and generalize ($F[8,120]=2.11$, $p=.04$) effects, but not the prime ($F[8,119]=1.63$, $p=.12$) effect. This merely indicated that the magnitude of the train and generalize effects differed across classrooms.

The regression results reported in Table 3 show very large effects of Condition on the magnitude of the prime and generalize effects, but not on the train effect after controlling for Classroom and Pretest WRCM. The differences in r^2 between a model with all predictors included and one without Condition included were .39 (prime), .30 (generalize), and .02 (train). After controlling for Classroom and Pretest WRCM, for the dependent variable prime students in the ME condition gained 23.90 WRCM more on average than students in the HWO condition, and students in the ME condition gained 14.29 WRCM more on average than those in the LWO condition. After controlling for Classroom and Pretest WRCM, for the dependent variable generalize students in the ME condition gained 25.21 WRCM more on average than students in the HWO condition, and students in the ME condition gained 13.71 WRCM more on average than those in the LWO condition.

Given an experimental design with 132 participants across three Condition levels, our estimated power for small-medium effect sizes (Cohen's $f^2 = 0.1$) was .88.[1] Our estimated effect sizes $f^2_{Prime} = 0.74$ and $f^2_{Generalize} = 0.50$ were very large, which when plugged into a post-hoc power analysis yielded power estimates of over .99.

Survival Analysis

Figure 3 shows Kaplan-Meier survival curves for participants achieving 0% to 300% generalized gains as a percentage of their training gains by experimental condition. The survival analysis showed that a substantially larger percentage of children in the ME condition remained as generalized responders at increasingly higher generalization levels relative to the LWO and HWO conditions. Specifically, ME training resulted in higher percentages of generalized responders at generalization levels up to about double that of observed training gains (i.e., 200%). Beyond this level (200% to 300% of training gains), all conditions produced low and equivalent percentages of generalized responders. Results of the log rank follow-up comparisons confirmed significant differences between the ME group and both the LWO (X^2=7.54, p=.006)

---

[1] For two nested linear models, the effect-size for a set of predictors *B* measure Cohen's $f^2 = \frac{R^2_{AB} - R^2_A}{1 - R^2_{AB}}$, where R^2_A is the explained variance of the nested model with predictor set A, and R^2_A is the explained variance of the encompassing model comprised of predictor sets *A* and *B*. In the present analyses, predictor set *A* corresponded to variables Classroom and PreT WRCM, whereas set *B* corresponded to Condition. Power analyses were conducted using function `pwr.f2.test` from the R package `pwr` (Champely, 2018).

and HWO group (X^2=16.17, p=.001). The difference between the LWO and HWO groups was not significant (X^2=1.90, p=.168).

Discussion

The primary goal of this study was to compare children's gains in oral reading fluency after applying one dose of a standard fluency-building intervention to each of three training passages that differed in word overlap with an untrained generalization passage. A secondary goal of the study was to examine the proportions of children in each word overlap condition who survived as generalized responders throughout the range of generalization effects (0% to 300% of training effects).

Summary of Findings and Possible Explanations

With respect to the first goal, results indicated that the fluency-building package produced similar gains in WRCM on trained passages regardless of word overlap condition. This finding was not surprising given that we equated the training passages for difficulty and applied the same intervention package to each. This finding is also consistent with previous research showing that LPP, RR, CR, and PD error correction is an effective package for increasing children's oral reading fluency (e.g., Begeny et al., 2012; Martens et al., 2007).

Results also showed that fluency training on the ME passage produced significantly larger generalized gains than practicing either the LWO or HWO passages. Practicing the LWO passage also produced significantly larger generalized gains than practicing the HWO passage. A similar pattern of results favoring the ME passage followed by the LWO and then HWO passage was also obtained for the priming effect on the generalization passage at pre-test.

There are several possible explanations for why reading and practicing the ME passage produced a positive priming effect and was better at promoting generalized oral reading fluency than the other two word overlap conditions. Children may attempt to read unfamiliar words using several strategies including; (a) guessing the word from context clues, (b) phonologically recoding (i.e., decoding) the word by blending known letter sounds (e.g., d/o/g), or (c) reading the word by analogy by blending known onsets and rimes (e.g., d/og and fr/og) (Ehri & Robbins, 1992). Because onsets and rimes are larger subunits of words than phonemes, reading unfamiliar words by analogy is believed to be easier for beginning readers. Experienced readers with advanced decoding skills are likely to read unfamiliar words by analogy or by decoding. A larger percentage of words in the ME passage (36%) shared onsets and rimes with words in the generalization passage. In addition, a larger percentage of words in the ME passage were low frequency or content words (66%). This may have allowed children in the ME condition to apply a reading-by-analogy strategy to unfamiliar words, which in turn promoted generalization. Conversely, the HWO and generalization passages had a large percentage of words in common (75.7%) but few exemplar words (3.3%). Even the LWO passage had a higher percentage of exemplar words (16%) than the HWO passage. Moreover, common words between the HWO and generalization passages appeared in different order. Children in this condition would not have benefitted from reading by analogy to the words they just practiced, and seeing the same words in a different order may have hindered generalization, similar to the scrambled words condition in Levy and Burns (1990). The negative priming effect for the HWO condition is also consistent with Faulkner and Levy (1994) who found longer reading times on a second passage after prior reading of a HWO passage versus the same passage by good readers.

With respect to the second goal, rather than adopting a single cut score for defining generalized responders, the survival analysis allowed us to identify proportions of generalized responders throughout the range of generalization effects. Results showed that a larger proportion of children in the ME group survived as generalized responders throughout most of the generalization range (i.e., 0% to 200% of training gains). This finding suggests that training with ME versus HWO or LWO passages will produce more generalized responders regardless of how stringently generalization is defined. The survival curves also allowed us to identify what proportion of children showed generalized gains in oral reading fluency and to what extent under each practice condition. For example, if a generalized responder is defined as someone whose generalized gains are at least half that of their training gains, approximately 30%, 55%, and 90% of children in the HWO, LWO, and ME conditions respectively would be classified as generalized responders.

Limitations of the Study

The present study is limited in several respects and confidence in the findings must be tempered accordingly. First, although we calculated Spache readability indices and detailed the word breakdown for each passage used in the study, we did not pilot test the passages for difficulty or have them rated for readability by reading content experts. Second, children's gains in WRCM were examined on both the training and generalization passages in response to one dose of LPP, RR, CR, and PD error correction immediately after training. As such, ours was an analogue study of the direct and generalized effects of this particular evidence-based intervention under controlled conditions. A different intervention or more extended exposure to the current intervention may produce different results, but this a direction for future research. Although exposing students to one intervention trial increased the internal validity of our comparisons, interventions are typically administered repeatedly and on different passages in response-to-intervention models. Despite receiving only one dose of intervention, children actually read the passage assigned to their word overlap condition a total of 6 times (pre-test, LPP, 3 choral RRs, post-test). The extra practice from pre- and post-testing differed from typical implementation of this intervention, but was required for the study. In addition, generalized effects of training are typically assessed periodically. It would be interesting in future research to probe for generalized gains at multiple points in time, after practicing multiple passages of each type, and/or during an extended intervention trial.

Third, most children in the sample were reading at instructional (42%) or mastery (42%) levels with only 16% reading at a frustrational level. This likely influenced the outcomes of our tests for simple main effects of condition on frustrational level readers. On the one hand, previous research has shown that the fluency-building package used in the present study is most effective when applied to instructionally matched passages (e.g., Bonfiglio et al., 2004; Daly et al., 1996). This suggests that a sample comprised of more instructional than frustrational level readers was appropriate for this type of intervention. On the other hand, children reading at a mastery level (42% of our sample) are not likely to receive a fluency-building intervention.

Fourth, children’s scores on the generalization passage were confounded by prior reading of the assigned training passages at pre-test and possibly also at post-test (i.e., a testing confound). Although this allowed us to analyze priming effects, it would be beneficial in future research to parse out this testing confound by adding a no PreT group.

Implications for Research and Practice

Despite these limitations, the present results may have implications for assessing and promoting generalized oral reading fluency in RtI models. Intervention researchers typically construct HWO passages to assess for generalization. The present findings argue against this practice by suggesting that HWO passages are less sensitive then ME or even LWO passages for assessing generalization. As such, fluency gains on untrained HWO passages may underestimate generalization effects. Relatedly, when monitoring student progress, it is common for intervention to be applied to one set of passages while gains in oral reading fluency are measured on untrained CBM-R probes. When generalized gains in oral reading fluency are not observed on these LWO global outcome measures, it may be informative to assess for generalization using a passage containing words with the same onsets or rimes (i.e., a ME passage) to separate generalized responders from non-responders. When RR, LPP, CR, and PD error correction procedures fail to produce generalized gains in oral reading fluency when applied to LWO passages, practicing phonics-controlled passages or passages containing exemplar words may better promote generalization for instructional level readers by allowing them to read unfamiliar words by analogy. As such, multiple exemplar training appears to hold promise as a practical and effective means of promoting generalized oral reading fluency in children.

References

Ardoin, S.P., Binder, K.S., Zawoyski, A.M., Foster, T.E., & Blevins, L.A. (2013b). Using eye-tracking procedures to evaluate generalization effects: Practicing target words during repeated readings within versus across texts. *School Psychology Review, 42*, 477-495.

Ardoin, S.P., & Christ, T.J. (2009). Curriculum-based measurement of oral reading: Standard errors associate with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review, 38*, 266-283.

Ardoin, S.P., Christ, T.J., Morena, L.S., Cormier, D.C., & Klingbeil, D.A. (2013a). A systematic review and summarization of the recommendations and research surrounding curriculum-based measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51,* 1-18.

Ardoin, S. P., McCall, M., & Klubnik, C. (2007). Promoting generalization of oral reading fluency: Providing drill versus practice opportunities. *Journal of Behavioral Education, 16*, 54–69.

Begeny, J.K., Yeager, A., & Martinez, R.S. (2012). Effects of small-group and one-on-one reading fluency interventions with second grade, low-performing Spanish readers. *Journal of Behavioral Education, 21*, 58-79.

Bonfiglio, C.M., Daly, E.J., Martens, B.K., Lan-Hsiang, R.L., & Corsaut, S. (2004). An experimental analysis of reading interventions: Generalization across instructional strategies, time, and passages. *Journal of Applied Behavior Analysis, 37*, 111-114.

Champely, S. (2018). pwr: Basic functions for power analysis. R package version 1.2-2. URL: https://CRAN.R-project.org/package=pwr.

Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, *35*, 386–406.

Christ, T.J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, *36*, 130-146.

Daly, E.J., Martens, B.K., Dool, E.J., & Hintze, J.M. (1998). Using brief functional analysis to select interventions for oral reading. *Journal of Behavioral Education, 8,* 203-218.

Daly, E.J., Martens, B.K., Hamler, K., Dool, E.J., & Eckert, T.L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32,* 83-94.

Daly, E.J., Martens, B.K., Kilmer, A., & Massie, D.R. (1996). The effects of instructional match and content overlap on generalized reading performance. *Journal of Applied Behavior Analysis, 29,* 507-518.

Ehri, L. C., & Robbins, C. (1992). Beginners need some decoding skill to read words by analogy. *Reading Research Quarterly, 27*, 12–26.

Faulkner, H.J, & Levy, B.A. (1994). How text difficulty and reader skill interact to produce differential reliance on word and content overlap in reading transfer. *Journal of Experimental Child Psychology, 58*, 1-24.

Good, R.H., Simmons, D.C., & Kame'enui, E.J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5,* 257-288.

Jones, K.M., Wickstrom, K.F., Noltemeyer, A.L., Brown, S.M., Schuka, J.R., & Therrien, W.J. (2009). An experimental analysis of reading fluency. *Journal of Behavioral Education, 18*, 35-55.

Levy, B.A., & Burns, K.I. (1990). Reprocessing text: Contributions from conceptually driven processes. *Canadian Journal of Psychology, 6*, 293-323.

Martens, B.K., Daly, E.J., & Ardoin, S.P. (2015). Applications of applied behavior analysis to school-based instructional intervention. In H.S. Roane, J.L. Ringdahl, & T.S. Falcomata (Eds.), *Clinical and organizational applications of applied behavior analysis* (pp. 125-150) .New York: Elsevier.

Martens, B.K., Eckert, T.L., Begeny, J.C., Lewandowski, L.J., DiGennaro, F., Montarello, S., Arbolino, L., Reed, D.D., & Fiese, B.H. (2007). Effects of a fluency-building program on the reading performance of low-achieving second and third grade students. *Journal of Behavioral Education, 16*, 39-54.

Morita, S., Okamoto, I., Kobayashi, K., Yamazaki, K., Asahina, H., Inoue, A.,…Nukiwa, T. (2009). Combined survival analysis of prospective clinical trials of Gefitinib for non-small cell lung cancer with EGFR mutations. *Clinical Cancer Research, 15*, 4493-4498.

R Core Team (2018). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Vienna, Austria. URL http://www.R-project.org/.

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention.* New York: Guilford Press.

Silber, J. M., & Martens, B. K. (2010). Programming for the generalization of oral reading fluency: Repeated readings of entire text versus multiple exemplars. *Journal of Behavior Education*, 19, 30–46.

Stokes, T.F., & Baer, D.M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349-367.

Table 1

Child Demographic Information

Measure	LWO (n = 43)	HWO (n = 45)	ME (n = 44)	Total (N = 132)
Number (%) Female	16 (37%)	23 (51%)	20 (45%)	59 (45%)
Number (%) by Ethnicity				
White	25 (58%)	32 (71%)	30 (68%)	87 (66%)
Hispanic	17 (40%)	13 (29%)	14 (32%)	44 (33%)
Multiracial	1 (2%)			1 (1%)
Mean (SD) Age*	9.1 (.35)	9.1 (.36)	9.1 (.25)	9.1 (.33)
Mean (SD) WRCM at Pretest	98.4 (35.6)	110.2 (38.0)	95.7 (26.7)	101.5 (34.4)
Number (%) by Instructional Level at Pretest				
Frustrational (<70 WRCM)	10 (23%)	4 (9%)	6 (14%)	20 (15%)
Instructional (70-99 WRCM)	15 (35%)	19 (42%)	23 (52%)	57 (43%)
Mastery (100> WRCM)	18 (42%)	22 (49%)	15 (34%)	55 (42%)

*Age data were available for the northeast sample only.

Note. LWO = low word overlap condition; HWO = high word overlap condition; ME = multiple exemplar condition; WRCM = words read correctly per minute.

Table 2

Number (Percentage) of Different Word Types in Each Passage

Word Type	LWO	HWO	ME	Generalization
Total Words	155	152	155	153
Content Words	78 (50.3)	76 (50.0)	102 (65.8)	83 (54.2)
Function Words	77 (49.7)	76 (50.0)	53 (34.2)	70 (45.8)
High Frequency Words	60 (38.7)	78 (51.3)	53 (34.2)	64 (41.8)
Low Frequency Words	95 (61.3)	74 (48.7)	102 (65.8)	89 (58.2)
Overlapping Words	39 (25.2)	115 (75.7)	38 (24.5)	n/a
New Words	91 (58.7)	32 (21.0)	61 (39.4)	n/a
Exemplar Words	25 (16.1)	5 (3.3)	56 (36.1)	n/a

Note. High frequency words were based on the first 100 words from the Fry list (LWO = low word overlap passage; HWO = high word overlap passage; ME = multiple exemplar passage; Generalization = generalization passage).

Table 3

Regression Analysis Results for Condition as a Predictor of the Priming (N = 131), Training (N = 132), and Generalization (N = 132) Effects

Variable	*B*	*SE B*	t-value	p-value
		Priming Effect		
Intercept (ME)	20.41	4.90	4.17	<.001
Condition (HWO)	-23.90	2.56	-9.33	<.001
Condition (LWO)	-14.29	2.53	-5.65	<.001
		Training Effect		
Intercept (ME)	49.76	6.84	7.27	<.001
Condition (HWO)	-4.66	3.64	-1.28	0.20
Condition (LWO)	-4.79	3.59	-1.34	0.18
		Generalization Effect		
Intercept (ME)	35.16	6.11	5.76	<.001
Condition (HWO)	-25.21	3.25	-7.76	<.001
Condition (LWO)	-13.71	3.20	-4.28	<.001

Note. Full model R^2 = .48 for the priming effect, .14 for the training effect, and .39 for the generalization effect.

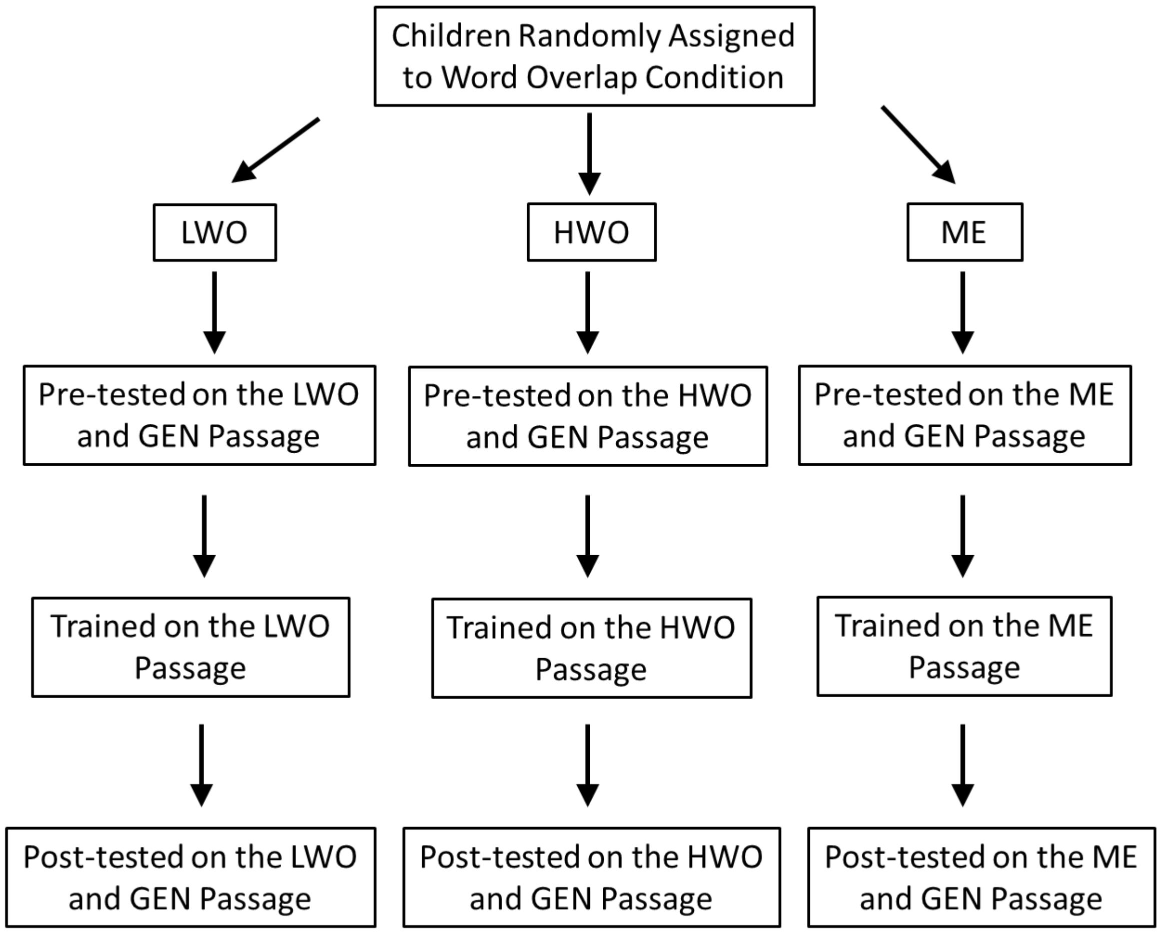

Figure 1. Sequence of testing and training procedures for children assigned to each word overlap condition (LWO = low word overlap, HWO = high word overlap, ME = multiple exemplar, GEN = generalization).

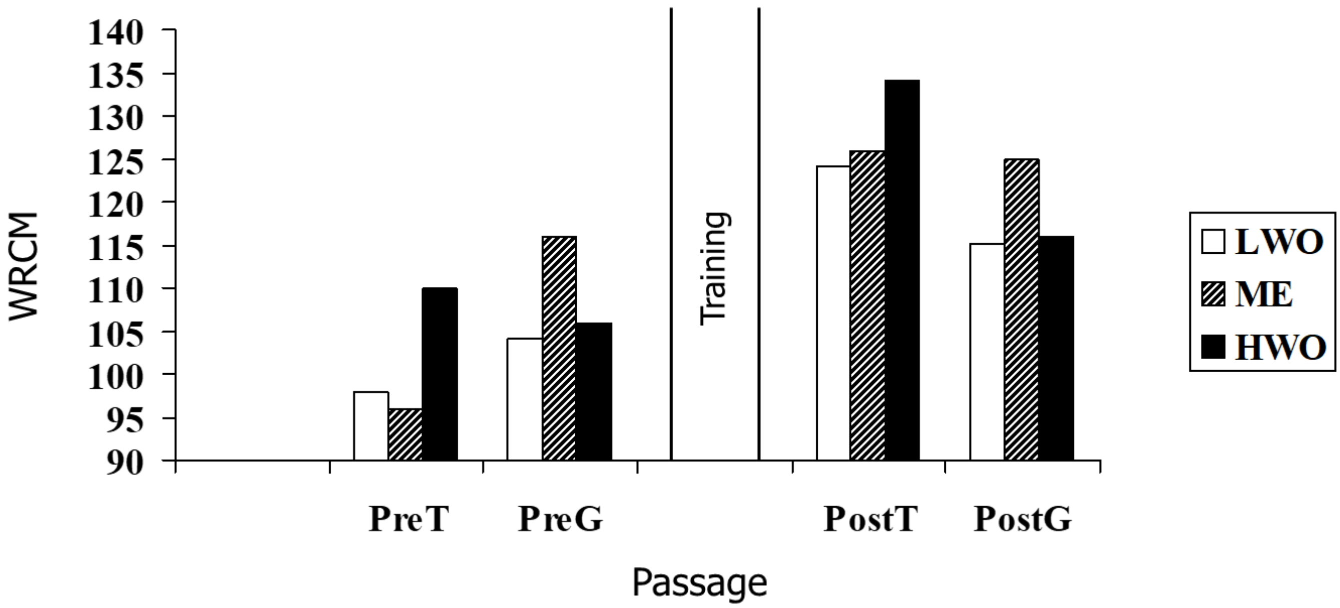

Figure 2. Average words read correctly per minute (WRCM) for participants at each testing occasion by condition (PreT = pre-test on the training passage, PreG = pre-test on the generalization passage, PostT = post-test on the training passage, PostG = post-test on the generalization passage; LWO = low word overlap, HWO = high word overlap, ME = multiple exemplar).

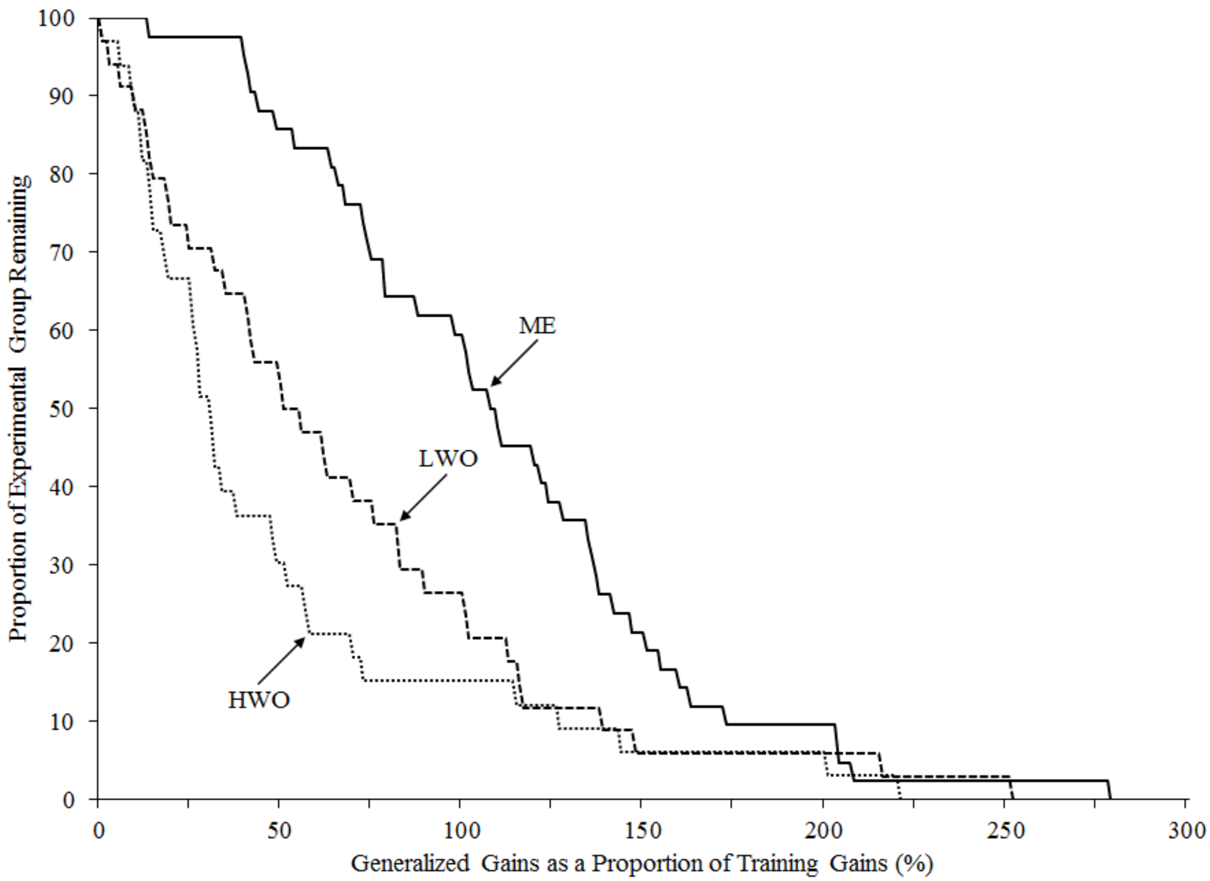

Figure 3. Survival analysis for participants achieving 0% to 300% generalized gains as a proportion of training gains by experimental condition (LWO = low word overlap, HWO = high word overlap, ME = multiple exemplar).