Title:
I'm Not Throwing Away My Shot:
What Alexander Hamilton Can Tell Us about Standard Reading Interventions

Al Otaiba, S., Petscher, Y., Wanzek, J., Lan, P., & Rivas, B. (2018). I'm not throwing away my shot: What Alexander Hamilton can tell us about standard reading interventions. *Learning Disabilities Research & Practice*, 33(3), 156–167. https://doi.org/10.1111/ldrp.12179

Stephanie Al Otaiba, Ph.D.
Southern Methodist University

Yaacov Petscher, Ph.D.,
Florida State University

Jeanne Wanzek, Ph.D.
Vanderbilt University

Patrick Lan, M.Ed.
Southern Methodist University

Brenna Rivas, Ph.D.
Southern Methodist University

Acknowledgements:
The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A150269 to Vanderbilt University and through Grant R324A160132 to Southern Methodist. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

This article summarizes findings from a two-year randomized control trial, focusing on a subset of 194 fourth graders with reading comprehension scores at or below the 15th percentile. Students in the treatment condition received an average of 94 daily 30-min sessions of small group intervention implemented with fidelity by well-trained research staff. Standardized measures of word identification, vocabulary, and comprehension, and an oral reading fluency measure were administered pre- and post-testing. Results indicated no statistically significant differences between students in the treatment or business-as-usual conditions; effect sizes for comprehension were small (0.14 and 0.19); a quantile regression, however, revealed slightly larger effect sizes for students at the 0.25 to 0.50 quantiles. The effect sizes for word identification, fluency, and vocabulary were less than 0.05. We discuss implications of the study, as well as limitations and directions for future research. We conclude with recommendations for intensifying interventions.

I'm Not Throwing Away My Shot: What Alexander Hamilton Can Tell Us about Standard Reading Interventions

In the popular musical about Alexander Hamilton, the main character sings his mantra about persistence – not throwing away his shot at success. For many fourth graders, reading intervention can be their shot at personal and academic success. Expectations for performing complex comprehension tasks have increased as a result of changing standards that aim to build college and career readiness, such as the Com- mon Core Standards for Literacy (CCSS: National Governor's Association Center for Best Practices, 2010; Haager & Vaughn, 2013). Consequently, students in the upper elementary grades are increasingly expected to independently read complex texts across an array of genres, an expectation that is challenging for students whose poor reading performance is a concern (e.g., NCES, 2016). This gap between performance and expectations is of particular concern for the roughly 20 percent of students who are either identified with or at risk for a specific learning disability (SLD; Horowitz, Rawe, & Whitaker, 2017), because of the challenge to identify and remediate the reading skills of older children (Speece et al., 2010). Schools that deliver interventions that are not sufficiently intense for students with persistent reading problems may be throwing away students' opportunity to improve their reading performance.

Research supports the efficacy of reading interventions to prevent and remediate reading problems for *most* children in the early elementary grades (Al Otaiba et al., 2014; Blachman et al., 2004; Gersten et al., 2008; Torgesen et al., 1999; Vellutino et al., 1996). A growing body of research has begun to confirm the efficacy of interventions for struggling readers in the upper elementary grades. To date, the strongest effects for reading comprehension (ranging from moderate to large) have been reported in studies using multi-component comprehensive interventions that included word reading and comprehension (e.g., O'Connor et al., 2002; Ritchey, Silver- man, Montanaro, Speece, & Schatschneider, 2012; Therrien, Wickstrom, & Jones, 2006; Vadasy & Sanders, 2008; Wanzek & Roberts, 2012; Wanzek, Wexler, Vaughn, & Ciullo, 2010). Yet there is a need for additional research to evaluate whether interventions at the upper elementary level lead to meaningful growth - in other words, whether they represent a good shot at reading success. This need is especially important for students with the poorest reading skills, regardless of whether they are receiving special education. Within the somewhat limited research for upper elementary students, reviews of the literature (e.g., Wanzek et al., 2010) indicate that most studies have examined only brief interventions targeting a single skill, though the few studies examining multi-component interventions have the highest effects. The current research also demonstrates that only a few studies have evaluated growth using standardized measures of reading rather than more proximal researcher-made measures, and there is little information about the efficacy of interventions specifically for students whose comprehension skills are extremely poor (i.e., at or below the 15th percentile). As most states are encouraging schools to implement Response to Intervention (RTI) or Multi-Tiered Systems of Support (MTSS) for prevention and remediation of reading difficulties (Zirkel & Thomas, 2010), administrators and teachers need guidance on the appropriateness of standard protocol interventions for students with the most intensive needs.

Empirical Context for the Study

The empirical context for our study was a large project involving two cohorts of students in a randomized control trial (RCT) of Passport to Literacy (Wanzek et al., 2016; Wanzek et al., 2017). Passport to Literacy applies principles of behavioral learning theory and cognitive psychology (Flavell, 1992; Palincsar & Brown, 1984). The program provides sequential, hierarchical instructional progressions in multiple reading components: foundational skills including phonics and word reading strategies, and higher-level processes for gaining understanding from text, including explicit vocabulary and comprehension instruction. Foundational skills and their application to text are emphasized in the first six weeks of lessons, and then the main emphasis shifts to strategies for gaining understanding from text with continual review and practice.

The first empirical study (Wanzek et al., 2016) provided a preliminary examination of the effects of the intervention for the first cohort of 221 fourth-grade students scoring below the 30th percentile in reading comprehension. Students were randomly assigned to receive the standard implementation of the intervention or a business-as-usual (BAU) approach, or typical school services. Project staff provided the intervention to small groups of four to seven students for 30 min, 4 days a week throughout the school year (*M* 90.45 lessons). We found small, significant effects on standardized measures of reading comprehension (ES 0.14 to 0.28). There were no differences between the treatment and comparison conditions on word reading or fluency outcomes. We cautioned, however, that intervention effects differed by students' comprehension abilities; in particular, students with low levels of comprehension demonstrated no increased benefit of the standard implementation relative to the BAU condition. In other words, the multi-component intervention demonstrated average increased outcomes on reading comprehension, but was least effective for students with the lowest comprehension levels, with these students performing similarly whether they were in the treatment or the BAU condition.

In a subsequent study (Wanzek, 2017), we combined data from two years of implementation, involving two cohorts totaling 451 students. The larger sample allowed us to use a partially nested analysis with latent variables to examine the efficacy of intervention on improving word reading, vocabulary, and reading comprehension. Findings were consistent with our initial study; students in the treatment condition significantly outperformed the BAU on reading comprehension, and, with the larger sample size, the magnitude of the effect was moderate (ES 0.38). We again found no differences between conditions on word reading or vocabulary. With this larger sample, we determined that students' initial word reading scores moderated these effects, with students at lower word reading levels benefitting less from the intervention than students beginning with higher word reading scores. Thus, from the previous studies, we had concerns for the students with the most intensive needs, something that motivated the present study and guided our working hypothesis that there would be differences between the conditions favoring the treatment condition, but only on comprehension.

Study Purpose and Research Questions

The primary purpose of the present study was to describe student reading progress overall, and then to evaluate the impact of a standard Tier 2 multi-component intervention relative to a business-as-usual (BAU) comparison condition for a subset of students with very poor comprehension skills. Two research questions guided the study: (1) What progress did students make regardless of treatment? (2) What was the impact of a standard Tier 2 multi-component intervention on the reading growth (comprehension, fluency, and word reading) of students whose pretreatment comprehension suggests that they have the most severe challenges (defined as performance at or below the 15th percentile for comprehension)?

Method

The Larger Study: Participants, Assignment to Condition, Reading Instruction and Intervention

The present study involves a reanalysis of data collected within the prior studies, involving two cohorts of fourth graders in 16 schools (Wanzek et al., 2017). One school district was located in a large, urban metropolitan area; one district was located in a mid-size city; and four districts were located in rural areas. At the end of the first month of school, our staff screened all students whose parents consented for them to participate. Students who scored at or below the 30th percentile on a standardized measure of reading comprehension were selected to participate. They were randomly assigned to treatment, or to a BAU comparison group, using a stratified procedure. Trained project staff provided students in the treatment condition with a comprehensive standard reading intervention, Passport to Literacy, in daily 30-min sessions to groups of 4–7, totaling roughly 100 sessions across the school year. Students assigned to the comparison group received the typical services provided by the school; some received intervention, but the majority did not.

We monitored the fidelity of intervention implementation, and also carefully documented both the type and amount of core (i.e., Tier 1) and supplemental reading intervention students received. To determine adherence to the intervention, we observed our interventionists' fidelity of implementation of program requirements. Our fidelity ratings ranged from a possible score of 0, indicating that the interventionist did not complete elements of a component, to 3, indicating she administered all or nearly all of the required elements. Instructional quality indicators included ongoing monitoring, redirection of off-task behavior, positive and corrective feedback, organization of materials, and appropriate selection of additional items for practice when needed. A total of three observers were required to code to a gold standard with at least 90% reliability, which was scored at 95.3% (Wanzek et al., 2017).

To understand the reading instruction our students received outside of our intervention, our research staff were trained to use a relatively low-inference reading observation tool, Instructional Content Emphasis Instrument-Revised (ICE-R; Edmonds & Briggs, 2003). Data on the amount of time students received instruction in each of the reading components (e.g., phonics/word recognition, vocabulary, and so on) were collected. In addition, observers recorded the level of student engagement (ranging from 3 = high engagement to 1 = low engagement), the type of instructional grouping (i.e., whole group, small group, individual), and the overall quality of instruction (e.g., direct and explicit instruction, providing modeling, feedback and encouragement) (ranging from 1 = weak to 4 = excellent). For the study, interrater reliability across coders and time periods was strong (95.10 percent).

On average, Tier 1 instruction lasted for 75.40 min (*SD* 26.34). The majority of instructional time focused on reading comprehension and vocabulary (35 min, or 46 percent of the time). On average, students participated in fluency- and reading-connected text activities for about 9 min, and participated in varied additional reading activities for about 15 min. On average, however, students received only 1 min of word study instruction and no phonemic awareness instruction. Moreover, the remaining instructional time, an average of 14 min, did not pertain to reading instruction (e.g., transition). In terms of grouping for instruction, observers noted that, on average, only about 8 min of instruction was conducted in small groups or peer pairs; students worked independently for about 10 min, and 45 min of instruction was whole-class.

We asked our classroom teachers to identify any students in the treatment or BAU comparison group who received additional supplemental reading intervention, which we audio-recorded three times per year (fall, winter, and spring). Project staff used the ICE-R measure to describe these interventions. Supplemental intervention did occur, typically for less than 30 min (*M* 28.34, *SD* 13.78), and, as in Tier 1, on average, the majority of time was spent in reading comprehension (12.72 min), followed by reading connected text (6 min), with the least amount of time dedicated to word study and decoding (about 4 min) or fluency (about half a minute). Small-group format was more frequent in intervention than in Tier 1, and it was most often delivered by the classroom teacher, or by another certified teacher. Nearly 20 percent of the reading interventions were provided by a paraprofessional or volunteer, and about 14 percent were provided by speech language pathologists.

Participants in the Current Study

For the current study examining the effectiveness of this intervention with our most impaired readers, we selected a subset from the larger RCT (Wanzek et al., 2017) study. Therefore, our sample included 194 fourth-grade students who scored below the 15th percentile on the reading comprehension subtest of the Gates-MacGinitie Reading Tests (GMRT; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2006). Consistent with the larger study, equal numbers of students had participated in the treatment and comparison conditions (n = 97). Of those 194 students in the sample, 45.36 percent (n = 105) were female. Only 2.6 percent (n = 5) of the total sample was flagged as currently receiving English Language services. All schools provided instruction only in English. With regards to ethnicity, 45.90 percent (n = 89) of the students were identified as Hispanic. The racial composition of the sample was 39.20 percent (n = 76) Black, 39.70 percent (n = 77) White, 20.6 percent n = (40) American Indian, 1.5 percent (n = 3) Asian or Pacific Islander, and .9 percent (n = 2) identified as Hispanic only. The majority of students received free or reduced-price lunch. Ten percent (n = 19) were identified as having a disability. The majority of students with a disability were identified with a specific learning disability, or with a speech language impairment. There were no differences in demographics between the treatment and BAU comparison conditions.

Measures and Data Collection Procedures

At pre- and post-treatment (i.e., fall and spring), our research team administered group tests of vocabulary and comprehension, and individual tests of word reading, fluency, and comprehension. All testers were blind to students' treatment condition assignment. The group-administered vocabulary and reading comprehension tests were part of the GRMT (MacGinitie et al., 2006). As previously noted, in fall, students' reading comprehension scores at or below the 30th percentile determined their eligibility for the larger study, but for the present study, scores at or below the 15th per- centile were used to select the sample. On the vocabulary subtest, students selected the correct meaning of the target word presented in context. On the comprehension sub- test, students read passages and answered multiple-choice questions that included identifying facts, inferencing, and drawing conclusions. Test-retest reliabilities are reported as above .85, and criterion validity estimates range from .79-.81 (MacGinitie et al., 2006).

Students were also assessed individually on four subtests of the Woodcock-Johnson III Tests of Achievement (WJ- III; Woodcock, McGrew, & Mather, 2001). Word level reading was assessed using the letter-word identification subtest, which measures the identification of letter names and reading of real words, and the word attack subtest, which measures pseudo-word decoding. We also assessed their expressive vocabulary using the picture vocabulary subtest, which requires students to name pictured objects. Finally, we assessed reading comprehension using the passage comprehension subtest, a cloze procedure that asks students to supply the missing word in a sentence or passage. Test authors report test-retest reliability for these four subtests at fourth grade as .81, 85, .77, and .86 respectively.

The last assessment, also administered individually, was the oral reading fluency (ORF) subtest of the Dynamic Indicators of Basic Early Literacy Skills -6th Edition (DIBELS; Good & Kaminski, 2002). Specifically, we asked students to read three end-of-grade-level passages for one min each to determine the number of words they could read correctly. Test-retest reliabilities for ORF with elementary age students range from .92 to .97; alternate-form reliability across passages from the same level is reported as .89 to .94 (Good et al., 2004).

Data Analytic Procedures

For research question 1, we estimated means, standard deviations, and correlations among measures. For research question 2, we used a conditional multilevel approach to estimate the impact of a standard Tier 2 multi-component intervention on the reading outcomes, controlling for initial skills tested in fall (comprehension, vocabulary, fluency, and word reading) Seven separate multilevel models were used for the individual measures (GMRT Reading Comprehension, WJ-III Passage Comprehension, WJ-III Letter Word Identification, WJ-III Word Attack, GMRT Vocabulary, WJ-III Picture Vocabulary, and DIBELS ORF). To correct for Type 1 error, we applied a Benjamini-Hochberg linear step-up procedure (Benjamini & Hochberg, 1995) to determine significance.

Following methods suggested by the What Works Clearinghouse (2014) to examine the magnitude of group differences in addition to statistically significant differences, we computed effect sizes for each measure. This procedure involved calculating Hedge's *g* using the formula

$$g = \sqrt{\frac{F(n_1 + n_2)(1 - r^2)}{n_1 n_2}}$$

where F is computed from the covariate-adjusted within-group variance from the multilevel conditional model, n_1 and n_2 are the sample sizes for the given intervention group and the comparison group, and r is the pretest/post-test correlation for the measure.

Results

Descriptive Statistics and Correlations among Measures: Student Progress and Stability of Rank Order of Students across Fourth Grade

Our first research question addressed students' overall reading progress, regardless of treatment. Table 1 provides descriptive statistics for the students in our sample on all study measures at pre- (fall) and post-test (spring). The WJ-III measures are all reported as W scores; ORF is reported as the raw score of words correct per min (wcpm). A preliminary review of the data for missingness (Table 1) showed that complete data were available for the fall GMRT-RC measure (n 194), but missing data rates varied from 1.5 percent to 22.4 percent for other measures. The reason for the high level of missing data on the Fall GMRT Vocabulary measure was that it was not administered in one site in Year 1. Consequently, the design conformed to a planned missing data design, a type of missing completely at random (MCAR) structure. Little's MCAR test suggested that all missing data met reasonable assumptions for MCAR [$\chi^2(35) = 39.11$, $p > .290$]; thus, using expectation-maximization for model estimation was appropriate and would not negatively bias results.

When examining the entire sample of intervention and comparison students, their comprehension scores were relatively poorer than their vocabulary or word reading scores on both fall and spring test scores. Their standard scores and w W scores on all measures increased slightly from fall to spring. Students ended the study with average reading performance, with a mean WJ-III reading comprehension standard score in spring of 85.94 ($SD = 8.27$), roughly one

SD below norms. Their mean spring vocabulary scores were only slightly higher (89.37; *SD* = 11.84). Their mean word reading and word attack scores, however, were higher (M = 92.49; SD = 9.76) and (M = 93.91; SD = 10.03) respectively. The DIBELS ORF measure uses a benchmark score for fourth grade of 91 wcpm in fall and 124 wcpm in spring for a student to be considered not at-risk for reading difficulties; on average, respectively, our students read only 73.54 wcpm in fall (SD = 27.43) and 89.39 wcpm in the spring (SD = 29.80).

Correlations among the fall measures ranged from a low of r .15 between WJ-III picture vocabulary and the GMRT reading comprehension to a high of r .76 between the WJ- III word attack and word identification. In spring, the lowest correlation again was between WJ-III picture vocabulary and the GMRT reading comprehension (r .25), and the highest was again between WJ-III word attack and word identification (r = .82). Notably, the stability coefficients from fall to spring were low for GMRT reading comprehension (r =.16), but were strong for the remaining measures, ranging from r = .63 for GMRT vocabulary to r =.84 for WJ-III word identification. These correlations suggest that the relative rank order of students was relatively stable across the year for vocabulary and word identification.

Impact of Intervention: Strength of Evidence for this Population

The second research question examined the impact of a standard Tier 2 multi-component intervention on the reading growth (comprehension, fluency, and word reading) for students in the treatment and BAU conditions. Table 2 shows the descriptive statistics of all measures by condition, split by W and standard scores. Baseline tests of equivalence showed no significant differences between the treatment and comparison groups (Table 3). Multilevel impact model results (Table 4) showed no statistically significant differences between the two groups on GMRT Reading Comprehension (p = .180), WJ-III Passage Comprehension (p = .153), WJ-III

Letter Word (p = .893), WJ-III Word Attack (p = .737), GMRT Vocabulary (p = .840), WJ-III Picture Vocabulary (p = .772), or DIBELS ORF (p = .570). Despite our hypothesis that we would find significant differences favoring the treatment group, statistically significant results were not observed. This finding may not be surprising, as the analytic sample was a small subset drawn from the original study. Effect size estimates for the outcomes were less than g = .03 for all outcomes, with the exception of WJ-III Passage Comprehension (g = 0.15) and GMRT Reading Comprehension (g = 0.19).

In a manner similar to the methodological approach used by Wanzek et al. (2016), we followed the primary impact models with an exploratory post-hoc analysis on the reading comprehension outcomes, using quantile regression. In contrast with conventional multilevel applications, which are conditional means models, a multilevel quantile regression allows for relations between independent and dependent variables to be uniquely estimated on multiple points of the dependent variable's conditional distribution (see Petscher, 2016; Solari, Denton, Petscher, & Haring, 2017, for examples). The multilevel quantile regression was estimated using the lqmm package in R software and was specified similarly to the multilevel conditional models, such that the predictors of post-test performance were the baseline variable and the treatment indicator. Because the lqmm package only allows for a two-level model to be specified, a student-within-school nesting model was used, given that the multilevel conditional model yielded relatively minor ICCs for the clustering at the classroom level.

Results from the quantile regression model are re- ported in Figure 1 for the GMRT Reading Comprehension (Figure 1, top) and WJ-III Passage Comprehension (Figure 1, bottom) assessments. The x-axis of the figures represents the conditional quantile of the respective outcomes (i.e., .10 to .90), and the y-axis is the estimate of the Passport co- efficient. Recall that in multilevel conditional models from Table 4, the Passport coefficient reflects the fitted deflection of the intervention from the BAU group. For example, the Passport coefficient for GMRT Reading in Table 4 was 4.57. In the same way, Figure 1 shows the Passport coefficients along a range of conditional quantiles rather than only at the conditional mean, as in Table 4. At the .50 quantile, which is close to the conditional mean, the estimated Passport coefficient was 5.12, a value that closely approximates the 4.57 in Table 4. The findings from the multilevel quantile model differ from the multilevel means model of Table 4, such that the former demonstrates that statistically significant effects for the treatment in this sample were observed when measuring comprehension by the GMRT.

Specifically, at the .50 quantile, there was an effect size of $g = 0.22$, and the treatment group outperformed the comparison group by approximately 5 points, controlling for baseline. At lower levels of the GMRT Reading post-test (e.g., .25 to .30 quantiles), there was a larger effect of Passport, where the treatment group outperformed the comparison group by ~9 points ($g = .35$). No significant differences were observed at other estimated quantiles of the GMRT, nor on the WJ-III PC outcomes at any quantile.

Discussion

Summary of Findings and Major Implications

The present study adds uniquely to the research base by examining effects specifically for a subset of students with very poor comprehension skills (who began the Wanzek et al., 2017, larger study at or below the 15^{th} percentile in comprehension). The purpose of this present study was to extend our prior work examining the effects of a widely used multi-component reading intervention (Wanzek et al., 2016; Wanzek et al., 2017), and more generally to add to the relatively limited research base on intensive interventions for students in upper elementary grades. In other words, we examined whether a standard intervention represents the best shot to improve future academic success by improving reading comprehension.

Our first research question addressed overall reading growth of the students in the study, and the second question focused on the main effects of this intervention in comparison to the BAU on measures of reading comprehension, vocabulary, and word reading. The relatively strong stability coefficients from fall to spring on all measures indicate students generally did not change their rank order over time, regard- less of condition, on most measures. The one exception was the GMRT comprehension measure, which had lower coefficients. On average, all students did grow from fall to spring, but as shown in Table 1, students did not close the reading gap, as determined by a comparison of their standard scores (which ranged on average from roughly 86 to 93) relative to norms for their grade level.

In terms of our second research question, as we hypothesized based on our prior studies, group differences and related effect sizes were negligible for vocabulary, fluency, and word reading (less than 0.05). This finding is likely related to the relatively small amount of word work, fluency instruction, or explicit vocabulary instruction, in comparison to the larger emphasis on comprehension instruction that occurred during intervention, and during Tier 1 or any supplemental school- provided interventions. Whereas effect sizes were small for reading comprehension (0.14 on the WJ-III and 0.19 on the GMRT), the quantile regression revealed stronger, albeit small to moderate, effects at the .25-.30 quantile (0.35), and at the .50 quantile (0.22), for reading comprehension. The magnitude of the effects on comprehension in the present study is smaller than in our previous study, with a larger sample of students below the 30th percentile (Wanzek et al., 2017; 0.38 on a latent factor of comprehension), suggesting that students with the most severe reading comprehension difficulties may have received less benefit from the intervention than the broader range of students with reading difficulties.

The small effects in the present study are also smaller than effects reported in Wanzek & Roberts' (2012) synthesis, which described moderate to strong effects of multi- component interventions for comprehension for students in upper elementary, noting O'Connor et al.'s (2002) effect sizes ranging from 1.39 to 1.46, Ritchey et al.'s (2012) effect size of 0.56, and Vadasy and Sanders' (2008) effect size of 0.50. These three sets of effect sizes, however, were only on researcher-made measures rather than standardized measures, which precluded evaluating whether growth indicated a closing of the reading gap with peers.

Limitations and Directions for Future Research

It is always a challenge to conduct rigorous randomized control trials in schools, and there are always limitations that are important to consider when interpreting findings. One limitation was that in the present study, our sample size was much smaller, because of our decision to focus on the subset of students with very poor initial reading comprehension. The small sample size likely reduced our power to detect significant differences between conditions. Results of our quantile regression indicate that the intervention was more robust for students from the .25-.50 quantiles. The small sample size precluded moderation analyses, something that warrants further research, because many of these students also were very dysfluent readers who had relatively low word reading scores. Relatedly, due to the small sample size, we were not able to consider clustering - in other words, treatment students' nesting within intervention groups. Our research team is currently conducting research with a larger sample of readers with comprehension skills that are similar to those of students in the present study; in that new study, we ex- amine the effects of a more intensive version of the Passport intervention.

Another possible limitation relates to our research design, which used research staff rather than teachers to implement the intervention. One advantage of doing so is that it is easier to achieve experimental control over the dosage, and to ensure a high degree of adherence to a standard protocol. Our staff were highly trained and had high fidelity. Yet there are some potential advantages of using teachers, who might have demonstrated some positive infidelity; for example, they may have been more able than staff to individualize intervention, and to add more time related to word recognition and flu- ency. We did not, however, observe much individualization either in Tier 1 or in school-provided interventions. Thus, a future research project could replicate the study with teacher implementers.

A final limitation about interpreting our study findings is related to our design. It was our intention to recruit schools serving students with low socioeconomic backgrounds, and the demographics of our participants reflect the diversity in the schools. Thus, our findings could differ in schools with more resources or different reading curricula, something that warrants further research in other settings.

Implications for Intensive Intervention: Suggestions for Improving Students' Shot

The intervention may not have been sufficient for the most struggling readers. If we consider Hamilton's advice not to throw away one's shot, then Fuchs, Fuchs, & Malone's (2017) Taxonomy provides a framework and process for teachers and schools to intensify intervention. The first dimension, *strength*, relates to the strength of evidence for a validated program. In general, such evidence could derive from findings from randomized control or experimental studies, or from single case design studies, and researchers should con- sider the effect size. Fuchs et al. suggested educators would ask themselves whether there is evidence that a generally effective program has been validated for students with particular needs. Thus, our prior studies with larger samples that included readers at or below the 30th percentile in comprehension indicated the Passport intervention was more effective than a BAU in improving reading comprehension. In the present study, when we focused on a subset of students with comprehension scores at or below the 15th percentile, there were no significant differences, and the effects were generally small for reading comprehension. As in our prior studies (Wanzek et al., 2016; Wanzek et al., 2017), the Passport intervention was not more effective than the BAU in increasing vocabulary, fluency, or word reading.

Furthermore, by the end of the present study, despite having participated in an average of 94 well-implemented small- group intervention sessions, students remained a standard deviation behind national norms in reading comprehension, with commensurately low vocabulary. They showed negligible change in their low average performance on word reading. This result is problematic in light of the *Endrew F. v. Douglas School District* (2017) decision that students must receive interventions that support them in achieving more than minimal growth, and also in light of the high proportion of students who struggle to read on grade level after third grade (e.g., National Center for Educational Statistics, 2016). We are conducting ongoing follow-up with these students to learn more about potential longer-term effects of the intervention. A meta-analysis by Suggate (2014) that examined whether long-term effects of interventions declined or increased across time suggested that from post-test to follow- up, effect sizes grew for comprehension interventions and for students in the upper grades.

Given the challenge of selecting effective reading interventions to help students with an array of intensive reading needs, we believe that Fuchs et al.'s (2017) Taxonomy of Intervention Intensity may provide some promising directions to schools to further intensify the strength of this intervention, or of similar interventions, for students like our participants. One dimension the Taxonomy recommends to increase intensity would be to increase *dosage*. Students may need more intervention in terms of duration or frequency, or in terms of a smaller group size, in order to maximize their opportunities to respond and learn. Another dimension, *alignment*, suggests that an intervention should target the appropriate skills and current levels of performance. Standard interventions may not cover all the skills - for example, all phonetic pat- terns, or adequate fluency practice at an independent reading level - that students need to master, even within a comprehensive program. It might be easier for teachers to fine-tune the *alignment* of the intervention to the needs of the students if groups were smaller and more homogeneous in needs. This adaptation could make it easier to increase students' opportunities to respond and receive more immediate feedback. It would also be easier to adapt the intervention to ensure extra time and practice to master any basic skills (e.g., phonetic patterns or multi-syllabic word or morphological patterns) that students lacked.

Another recommended dimension is to increase intensity through explicitly scaffolding *attention to transfer.* For example, students who mastered a comprehension strategy during one of the lessons or units might need help from their teachers to use that same strategy on a text of a different genre, or they might require additional fluency practice to read well enough to apply that strategy to a more complex text level. For example, according to CCSS, students are expected to comprehend informational text (e.g., history, science) within their grade level. For students in the upper elementary grades, this independent reading level would involve the ability to read texts with Lexile levels within the 770-980 range (Meta-metrics, 2017). Students in our study, however, had average students' Lexile levels in fall of only 366.37 (SD 200.20), and in spring, their average Lexile level had increased to 474.81 (SD 217.45). The large individual differences reflected in the standard deviations are indicative of a challenge teachers face - that of ensuring students can read and access grade level content area, when a fourth-grade classroom might include students reading as low as a first-grade level. Note that our students' Lexile levels were more similar to the recommended range for first graders (i.e., 220–450), or for students in second through third grade (i.e., 450–790). As the reading requirements increase in fifth grade and beyond for academic content, our participating students remain one to two grade levels behind, a gap that will limit their ability to master content-area knowledge and may further impede the ability to accelerate their learning.

Another recommended dimension is to increase intensity through *comprehensiveness*, a term that refers to the importance of interventions following principles of explicit and systematic instruction, (i.e., modeling, immediate corrective feedback, and cumulative review and practice). Our intervention was *comprehensive*, in that it was explicit, was systematic, and incorporated modeling and cumulative practice. The next recommended dimension for intensification is to provide *behavior supports* that train and encourage self-regulation, engagement, and motivation for task completion. For example, additional strategies could have involved self-monitoring or graphing of oral reading fluency progress toward goals, or some form of motivation for successful mastery at the end of adventures. Another direction worth exploring stems from social emotional learning theory regarding persistence (Duckworth, 2016) or building a growth mindset (e.g., Dweck, 2006). In other words, once students have reached the upper elementary grades and they still struggle to learn to read, they may benefit from under- standing that hard work (and what specific strategies work) will gradually help them improve so they do not feel as if they are "not readers."

The final dimension within the Taxonomy, *individualization*, highlights the importance of using data to monitor student performance, and to make ongoing adjustments to interventions as the students' skill levels and academic demands change (data-based individualization; DBI; Fuchs, Fuchs, & Vaughn, 2014; Lemons, Kearns, & Davidson, 2014; Stecker, Fuchs, & Fuchs, 2005). In other words, as teachers begin a systemic manipulation of these dimensions in a standard intervention such as Passport, student response should be closely evaluated through frequent progress monitoring. Then, as needed, teachers should continue adjusting or adapting interventions, as well as using other diagnostic information such as mastery within a curriculum, observations of oral prosody, phonics inventories, and interest inventories and observations of engagement of texts. To ensure that students continue to learn grade level material, modifications and adaptations that might include technology such as e-books, apps, and videos for content information may also be considered.

Conclusion

In the present study, we examined the efficacy of a widely used comprehensive reading intervention as a Tier 2 standard protocol. We found that when adhering to a standard protocol, the intervention was not a strong enough shot to meaningfully accelerate progress for all students with very poor comprehension skills. Yet the intervention was not throwing away a shot for those students scoring at the .25-.50 quantiles; the treatment group did achieve between a 5- and 9-point advantage over the comparison group. We then described several ways in which schools and teachers could further intensify this intervention, or other similar interventions, as they balance the path between implementing with fidelity and moving toward DBI.

References

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., Wagner, R. K. (2014). To wait in tier 1 or intervene immediately: A randomized experiment examining first grade response to intervention (RTI) in reading. *Exceptional Children, 81* (1) 11-27. DOI: 10.1177/0014402914532234.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1). 289-300.

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology, 96*(3), 444-461.

Duckworth, A. (2016). *Grit: The power of passion and perseverance*. New York, NY: Simon and Schuster.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House Incorporated.

Edmonds, M. S., & Briggs, K. L. (2003). Instructional Content Emphasis instrument. In S. R. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for observing teaching and learning* (pp. 31-52). Baltimore, MD: Brookes.

Endrew F. v. Douglas County School District RE-1, 580 U.S.(2017).

Flavell, J. H. (1992). Cognitive development: Past, present, and future. *Developmental Psychology, 28*, 998-1005. doi:10.1037/0012-1649.28.6.998

Fuchs, D., Fuchs, L. S., Malone, A. S. (2017). The taxonomy of intervention intensity. *TEACHING Exceptional Children, 50*(1), 35-43.

doi:10.1177/0040059917703962

Fuchs, D., Fuchs, L. S., Vaughn, S. R. (2014). What is intensive instruction and why is it important? *TEACHING Exceptional Children, 46*(4), 13–18. doi:10.1177/0040059914522966

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., and Tilly, W. D. (2008). Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide. (NCEE 2009-4045). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/ publications/practiceguides/.

Good, R. H., & Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills, 6th edition (DIBELS)*. Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu/

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., & Flindt, N. (2004). *Technical adequacy of DIBELS: Results of the Early Childhood Research Institute on Measuring Growth and Development* (Technical Report, No. 7). Eugene, OR: University of Oregon.

Haager, D. & Vaughn, S. (2013). The Common Core State Standards and reading: interpretations and implications for elementary students with learning disabilities. *Learning Disabilities Research & Practice, 28*(1), 5-16.

Horowitz, S. H., Rawe, J., & Whittaker, M. C. (2017). *The state of learning disabilities: Understanding the 1 in 5*. New York: National Center for Learning Disabilities.

Lemons, C. J., Kearns, D. M., & Davidson, K. A. (2014). Data-based individualization in reading: Intensifying interventions for students with significant reading disabilities. *TEACHING Exceptional Children, 46*(4), 20–29.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2006). *Gates-MacGinitie Reading Tests* (4th ed.). Rolling Meadows, IL: Riverside Publishing.

Meta-metrics (2017). Downloaded from https://lexile.com/educators/measuring-growth-with-lexile/lexile-measures-grade-equivalents/.

National Center for Educational Statistics (2016). *National assessment of educational progress: The nation's report card.* Washington, D.C.: U.S. Department of Education.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards.* Washington, D.C.

O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology, 94*(3), 474-485. doi:10.1037/0022-0663.94.3.474

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117-175. doi: 10.1207/s1532690xci0102_1

Petscher, Y. (2016). Do our means of inquiry match our intentions?. *Frontiers in Psychology*, *7*, 1048.

Ritchey, K. D., Silverman, R. D., Montanaro, E. A., Speece, D. L., & Schatschneider, C. (2012). Effects of a Tier 2 supplemental reading intervention for at-risk fourth-grade students. *Exceptional Children, 78*(3), 318-334. doi: 10.1177/001440291207800304

Solari, E. J., Denton, C. A., Petscher, Y., & Haring, C. (2017). Examining the effects and feasibility of a teacher-implemented Tier 1 and Tier 2 intervention in word reading, fluency, and comprehension. *Journal of Research on Educational Effectiveness*, *11*(2), 1-29.

Speece, D. L., Ritchey, K. D., Silverman, R. D., Schatschneider, C., Walker, C., & Andrusik, K. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review*, *39*(2), 258–276.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795–820. doi:10.1002/pits.20113

Suggate, S. P. (2014) A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, *49*(1), 77-96. doi: 10.1177/0022219414528540.

Therrien, W. J., Wickstrom, K., & Jones, K. (2006). Effect of a combined repeated reading and question generation intervention on reading achievement. *Learning Disabilities Research & Practice, 21*(2), 89-97. doi:10.1111/j.1540-5826.2006.00209.x

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. S., & Conway, T. (1999). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*(1), 33-58.

Vadasy, P. F., & Sanders, E. A. (2008). Repeated reading intervention: Outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology, 100*(2), 272-290. doi: 10.1037/0022-0663.100.2.272

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601-638.

Wanzek, J., & Roberts, G. (2012). Reading interventions with varying instructional emphases for fourth graders with reading difficulties. *Learning Disability Quarterly, 35*(2), 90-101.

Wanzek, J., Petscher, Y., Al Otaiba, S., Kent, S. C., Schatschneider, C., Haynes, M, Rivas, B.K., & Jones, F. G. (2016). Examining the effects of a standardized treatment for fourth graders with reading difficulties. *Journal of Research on Educational Effectiveness 9*(sup1), 45-66.

Wanzek, J., Petscher, Y., Al Otaiba.S., Rivas, B., Jones, F. J., Kent, S., Schatchneider, C., Mehta, P. (2017). Effects of a year-long supplemental reading intervention for students with reading difficulties in fourth grade. *Journal of Educational Psychology, 109*(8), 1108-1119.

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing, 23*(8), 889-912. doi: 10.1007/s11145-009-9179-5

What Works Clearinghouse. (2014). *Procedures and standards handbook* (version 3.0). Washington, D.C.: U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside.

Zirkel, P. A. & Thomas, L. B. (2010). State laws and guidelines for implementing RTI. *Teaching Exceptional Children, 43*(1), 60-73.

Table 1: Descriptive statistics and correlations for study measures (standard scores)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Fall GMRT RC	1.00													
2. Fall WJ PC	0.23	1.00												
3. Fall WJ LWID	0.23	0.65	1.00											
4. Fall WJ WA	0.20	0.54	0.76	1.00										
5. Fall GMRT Voc	0.37	0.50	0.53	0.46	1.00									
6. Fall WJ PV	0.14	0.49	0.28	0.12	0.32	1.00								
7. Fall DIBELS	0.20	0.53	0.72	0.66	0.49	0.14	1.00							
8. Spring GMRT RC	0.16	0.44	0.35	0.30	0.40	0.18	0.34	1.00						
9. Spring WJ PC	0.24	0.69	0.64	0.46	0.57	0.50	0.50	0.38	1.00					
10. Spring WJ LWID	0.22	0.64	0.84	0.71	0.53	0.25	0.70	0.38	0.67	1.00				
11. Spring WJ WA	0.17	0.55	0.77	0.76	0.48	0.21	0.63	0.33	0.55	0.81	1.00			
12. Spring GMRT Voc	0.24	0.54	0.52	0.42	0.63	0.30	0.47	0.66	0.57	0.55	0.51	1.00		
13. Spring WJ PV	0.13	0.52	0.35	0.14	0.40	0.79	0.24	0.17	0.57	0.35	0.26	0.38	1.00	
14. Spring DIBELS	0.21	0.56	0.73	0.63	0.56	0.16	0.90	0.37	0.53	0.71	0.65	0.50	0.26	1.00
Mean	425.17	84.79	93.23	93.41	438.07	88.84	73.54	448.93	85.94	92.49	93.91	454.40	89.37	89.39
Range	363-443	47-111	55-120	56-125	331-506	1-118	7-144	334-517	47-106	61-119	55-127	312-535	26-113	15-183
S.D.	17.50	9.81	10.51	11.09	29.91	12.63	27.43	24.11	8.27	9.76	10.03	29.70	11.84	29.80
N	194	194	194	194	151	194	194	191	190	190	190	191	190	190
% Missing Data	0%	0%	0%	0%	22.4%	0%	0%	2.1%	2.1%	2.1%	2.1%	1.5%	2.1%	2.1%

Note. GMRT RC = Gates-McGinitie Reading Comprehension, WJ PC = WJ-III Passage Comprehension, WJ LWID = WJ-III Letter Word Identification, WJ WA = WJ-III Word Attack, GMRT Voc = Gates-McGinitie Vocabulary, WJ PV = WJ-III Picture Vocabulary. All correlations statistically significant at least $p < .05$.

Table 2
Descriptive statistics of measures by condition

		Passport			Comparison		
Measure	Score Type	N	M	SD	N	M	SD
Fall GMRT RC		97	424.73	18.051	97	425.61	17.019
Fall WJ PC	W	97	479.01	12.434	95	478.37	13.267
	SS	97	85.05	9.519	97	84.53	10.140
Fall WJ LWID	W	97	480.27	21.156	95	480.01	21.000
	SS	97	93.32	10.623	97	93.14	10.453
Fall WJ WA	W	97	486.03	19.530	95	487.35	16.722
	SS	97	93.12	12.021	97	93.70	10.119
Fall GMRT Voc		76	438.25	28.306	75	437.88	31.633
Fall WJ PV	W	97	485.09	15.120	95	485.21	13.365
	SS	97	88.89	13.346	97	88.78	11.931
Fall DIBELS ORF		97	72.26	28.421	97	74.81	26.496
Spring GMRT RC		96	451.09	24.961	95	446.75	23.134
Spring WJ PC	W	96	485.46	10.374	94	483.86	10.413
	SS	96	86.64	8.133	94	85.22	8.385
Spring WJ LWID	W	96	488.77	19.815	94	489.13	19.020
	SS	96	92.47	9.959	94	92.51	9.606
Spring WJ WA	W	96	492.83	16.249	94	493.52	15.787
	SS	96	93.69	10.474	94	94.14	9.612
Spring GMRT Voc		96	453.86	32.742	95	454.95	26.443
Spring WJ PV	W	96	489.95	14.457	94	489.14	12.558
	SS	96	89.78	12.608	94	88.96	11.049
Spring DIBELS ORF		96	86.53	29.899	94	90.30	29.734

Note. GMRT RC = Gates-McGinitie Reading Comprehension, WJ PC = WJ-III Passage Comprehension, WJ LWID = WJ-III Letter Word Identification, WJ WA = WJ-III Word Attack, GMRT Voc = Gates-McGinitie Vocabulary, WJ PV = WJ-III Picture Vocabulary, ORF = Oral Reading Fluency.

Table 3
Baseline equivalence test

	GMRT Reading			WJ-PC			WJ-LWID			WJ-WA			GMRT Voc			WJ-PV			DORF		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Parts																					
(Intercept)	425.61	421.99 – 429.22	<.001	478.23	475.14 – 481.33	<.001	478.83	473.69 – 483.97	<.001	485.80	480.71 – 490.90	<.001	438.50	430.66 – 446.35	<.001	485.64	481.68 – 489.59	<.001	75.07	68.31 – 81.84	<.001
PASSPORT	-0.83	-5.76 – 4.09	.742	0.53	-3.03 – 4.09	.773	0.36	-5.45 – 6.17	.904	-1.39	-6.26 – 3.49	.583	-0.71	-8.09 – 6.66	.852	-0.07	-3.84 – 3.70	.972	-2.85	-10.38 – 4.68	.464
Random Parts																					
N_{TEACH_ID}		43			43			43			43			43			43			43	
N_{SCHOOL}		17			17			17			17			17			17			17	
ICC_{TEACH_ID}		0.000			0.031			0.000			0.000			0.021			0.109			0.003	
ICC_{SCHOOL}		0.008			0.045			0.063			0.137			0.135			0.092			0.068	
Observations		194			192			192			192			194			192			194	

Note. GMRT Reading = Gates-McGinitie Reading Comprehension, WJ-PC = WJ-III Passage Comprehension, WJ-LWID = WJ-III Letter Word Identification, WJ-WA= WJ-III Word Attack, GMRT Voc = Gates-McGinitie Vocabulary, WJ-PV = WJ-III Picture Vocabulary, DORF = Oral Reading Fluency, ICC = intraclass correlation.

Table 4

Multilevel impact model results

	GMRT Reading			WJ-PC			WJ-LWID			WJ-WA			GMRT Voc			WJ-PV			DORF		
	B	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*	*B*	*CI*	*p*
Fixed Parts																					
(Intercept)	446.87	441.01 – 452.73	<.001	483.67	482.07 – 485.26	<.001	488.84	486.62 – 491.05	<.001	492.66	490.43 – 494.90	<.001	454.21	449.20 – 459.22	<.001	489.83	487.69 – 491.97	<.001	88.62	86.00 – 91.25	<.001
Passport	4.57	-1.90 – 11.04	.180	1.55	-0.52 – 3.62	.153	-0.21	-3.23 – 2.82	.893	0.52	-2.47 – 3.51	.737	-0.63	-6.71 – 5.45	.840	0.34	-1.96 – 2.65	.772	-1.09	-4.80 – 2.63	.570
Fall GMRT Reading	0.21	0.03 – 0.40	.035																		
Fall WJ-PC				0.56	0.48 – 0.64	<.001															
Fall WJ-LWID							0.77	0.70 – 0.85	<.001												
Fall WJ-WA										0.67	0.59 – 0.76	<.001									
Fall GMRT Voc													0.71	0.60 – 0.83	<.001						
Fall WJ-PV																0.73	0.65 – 0.82	<.001			
Fall DORF																			0.97	0.90 – 1.04	<.001
Random Parts																					
N_{TEACH_ID}		43			43			43			43			43			43			43	
N_{SCHOOL}		17			17			17			17			17			17			17	
ICC_{TEACH_ID}		0.003			0.000			0.013			0.003			0.000			0.018			0.000	
ICC_{SCHOOL}		0.070			0.017			0.000			0.009			0.038			0.075			0.000	
Observations		194			188			188			188			194			188			194	

Note. GMRT Reading = Gates-McGinitie Reading Comprehension, WJ-PC = WJ-III Passage Comprehension, WJ-LWID = WJ-III Letter Word Identification, WJ-WA= WJ-III Word Attack, GMRT Voc = Gates-McGinitie Vocabulary, WJ-PV = WJ-III Picture Vocabulary, DORF = Oral Reading Fluency, ICC = intraclass correlation.