

# Appendix

## Appendix A1 Study characteristics: Denton, Anthony, Parker, & Hasbrouck, 2004 (randomized controlled trial with differential attrition)

Characteristic	Description
<b>Study citation</b>	Denton, C. A., Anthony, J. A., Parker, R., and Hasbrouck, J. E. (2004). Effects of two tutoring programs on the English reading development of Spanish-English bilingual students. <i>The Elementary School Journal</i> , 104(4), 289–305.
<b>Participants</b>	The report by Denton and colleagues covers two studies: one investigates the impacts of <i>Read Naturally</i> and the other investigates the impacts of <i>Read Well</i> . Ninety-three Hispanic English language learning (ELL) students (45 females) of the 99 students selected to participate completed the study. The participants were striving readers in second through fifth grade (2nd = 22; 3rd = 37; 4th = 28; 5th = 6), and their ages ranged from 7 years to 12 years (with an average age of 9 years). The 36 students in the <i>Read Well</i> <sup>1</sup> study were matched based on their pretest Word Attack WRMT-R scores and were randomly assigned to the <i>Read Well</i> intervention group or control group. Three students in the control group who did not complete the pre- and post-assessments were not included in the analysis. The final sample consisted of 33 students (19 in the intervention group, 14 in the control group). Overall, 82% of the control group completed the study, while 100% of the intervention group did. <sup>2</sup>
<b>Setting</b>	Students attended one of five schools in a Central Texas school district. The district served 13,664 total students, 32% of whom were Hispanic and 56% of whom were identified as economically disadvantaged.
<b>Intervention</b>	The intervention involved pull-out tutoring sessions during school day that did not interfere with normal English instruction. Students in both the <i>Read Well</i> and <i>Read Naturally</i> programs attended an average of 22 tutoring sessions that lasted 40 minutes each. Tutors were randomly assigned to treatments after receiving training and were supervised by a graduate student in tutor-directed decoding practice (including systematic, explicit phonics instruction), in decodable text reading practice accompanied by pre-reading and discussion of vocabulary, and in worksheet-guided reading comprehension exercises.
<b>Comparison</b>	The control group received the same regular English education curriculum as the treatment group but did not receive the <i>Read Well</i> program.
<b>Primary outcomes and measurement</b>	Study measures in the reading achievement domain included a researcher-developed oral reading assessment <sup>3</sup> and three scales from the Woodcock Reading Mastery Tests—Revised: Word Identification, Word Attack, and Reading Comprehension. (See Appendix A2 for a more detailed description of outcome measures.)
<b>Teacher training</b>	Twenty-three undergraduate students studying special education who were enrolled in a class for teaching students with reading difficulties served as tutors. Tutors received training and were supervised by a graduate student experienced in <i>Read Well</i> .

1. Students were assigned to one of the two interventions, *Read Well* or *Read Naturally*, based on their pretest scores on the Word Attack subtest of the Woodcock Reading Master Tests-Revised (WRMT-R). Students who scored below the first-grade equivalency (< 1.0) were assigned to *Read Well*, and students whose grade equivalency score was higher than first grade (≥ 1.0) were assigned to *Read Naturally*. Of the 93 students involved in the study, 33 participated in *Read Well* and 60 participated in *Read Naturally*.
2. The differential attrition within this study exceeds the WWC evidence standards for randomized controlled trial studies, so this study was judged to meet WWC evidence standards with reservations (see the [WWC Intervention Rating Scheme](#) for details).
3. Data from the researcher-developed oral reading assessment were not included in the Denton et al., 2004 study. Denton and colleagues (2004) stated that “logistical problems, some instances of potentially unreliable administration, and missing data points resulted in data that is invalid for analyses” (p. 296).

## Appendix A2 Outcome measures for the reading achievement domain

Outcome measure	Description
<b>Woodcock Reading Mastery Tests—Revised (WRMT-R): Word Identification</b>	The Word Identification subtest assesses basic reading skills by having participants read words presented in a list (as cited by Denton et al., 2004).
<b>WRMT-R: Word Attack</b>	The Word Attack subtest assesses phonemic decoding by having participants read aloud a list of nonsense words (as cited by Denton et al., 2004).
<b>WRMT-R: Passage Comprehension</b>	The Reading Comprehension subtest uses a cloze format that requires participants to read a passage that has an omitted word. The participants are asked to supply an appropriate word to complete the passage that they are reading (as cited by Denton et al., 2004).

## Appendix A3 Summary of study findings included in the rating for the reading achievement domain<sup>1</sup>

Outcome measure	Study sample	Sample size (students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> ( <i>Read Well</i> – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			<i>Read Well</i> group	Comparison group				
Denton et al., 2004 (randomized controlled trial with differential attrition)								
WRMT-R: Word Identification	Grades 2–5	33	4.06 (10.79)	.21 (7.06)	3.85	0.40	ns	+16
WRMT-R: Word Attack	Grades 2–5	33	5.16 (8.69)	2.35 (6.15)	2.81	0.35	ns	+14
WRMT-R: Reading Comprehension	Grades 2–5	33	1.58 (11.69)	1.57 (9.7)	0.01	0.00	ns	0
<b>Domain average<sup>7</sup> for reading achievement</b>						0.25	ns	+10

ns = not statistically significant

1. This appendix reports overall findings considered for the effectiveness rating and the improvement index.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are; a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Intervention and control group pretest to posttest change scores were used in the study authors' analyses and in the WWC calculations. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was calculated by the WWC and, where necessary, corrects for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of *Read Well*, no corrections for clustering or multiple comparisons were needed.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results.
7. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A4 *Read Well* rating for the reading achievement domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.<sup>1</sup>

For the outcome domain of reading achievement, the WWC rated *Read Well* as having potentially positive effects. It did not meet the criteria for positive effects because it had only one study that illustrated significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered, because *Read Well* was assigned the highest applicable rating.

### Rating received

**Potentially positive effects:** Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

**Met.** One study reviewed by the WWC reported an average effect size that is substantively important.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect. Fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

**Met.** The WWC analysis found no indeterminate, statistically significant negative, or substantively important negative effects in this domain.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

**Not met.** *Read Well* has only one study meeting WWC evidence standards with reservations. Although the effect was substantively positive, the study lacked a strong design.

- Criterion 2: No studies showing statistically significant or substantially important *negative* effects.

**Met.** The WWC analysis found no statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive effects. See the [WWC Intervention Rating Scheme](#) for a complete description.