



WWC Single Study Review

A review of the design and summary of findings for an individual study



September 2016

WWC Review of the Report “A Randomized Control Trial of a Statewide Voluntary Prekindergarten Program on Children’s Skills and Behaviors Through Third Grade”¹

The findings from this review do not reflect the full body of research evidence on the *Tennessee Voluntary Prekindergarten* program.

What is this study about?

The study authors examined whether children who attended at least 20 days of the *Tennessee Voluntary Prekindergarten* program (*TN-VPK*) performed better on tests of academic skills and social and emotional development than other children. *TN-VPK* is a public full-day program for 4-year-old children run by participating school districts.

The study described in this report is part of a larger evaluation of *TN-VPK*, a project in which researchers identified *TN-VPK* programs with more applicants than could be enrolled, and randomly offered enrollment to children who had applied. This was done across two cohorts of children, in the 2009–10 and 2010–11 school years. From this larger randomly assigned sample, the study authors obtained parental consent for data collection and had outcome data on 1,076 children (36% of the randomized sample). In the study described in this report, the authors used a quasi-experimental design (referred to by the study authors as the “Intensive Substudy”) with this subsample of children. The study authors used a quasi-experimental design to compare children who had attended *TN-VPK* for 20 or more days during the school year to children who did not attend *TN-VPK* or attended fewer than 20 days. The manuscript examined in this WWC report only reports findings for this subsample of the evaluation.

The study authors examined the impact of *TN-VPK* on academic achievement using a standardized assessment (Woodcock-Johnson III Tests of Achievement) that measured students’ skills in four areas: mathematics, alphabetics, comprehension, and language development. Separately, as a measure of cognition, the study examined a composite of the six subtests in these four areas. The study also examined impacts on students’ social-emotional development as measured by teacher ratings of the children’s work-related skills and social behavior. The study authors collected baseline data on academic outcomes in the fall of prekindergarten, and outcomes were assessed yearly, through the end of third grade. The analyses of outcomes included between 864 students (analysis of social-emotional development at the end of second grade) and 1,043 students (analysis of mathematics scores at the end of prekindergarten).^{2,3}

What did the study find?

The study authors reported that, at the end of prekindergarten, children who attended at least 20 days of *TN-VPK* had higher composite scores than other students on the standardized assessment and higher scores on outcomes related to cognition, mathematics, alphabetics, and comprehension. The authors also found that, at the start of kindergarten, children

WWC Rating

The research described in this report meets WWC group design standards with reservations^{1,2}

This study used a quasi-experimental design that compared outcomes for children based on whether they had attended at least 20 days of *TN-VPK*. The authors demonstrated baseline equivalence of the intervention group (attended more than 20 days) and comparison group (did not attend the program or attended fewer than 20 days), using observed data for the academic and social-emotional outcomes presented in this report. Baseline equivalence was demonstrated for the analytic samples used to examine outcomes from prekindergarten through third grade.

Features of the *Tennessee Voluntary Prekindergarten program (TN-VPK)*

TN-VPK is an optional state-funded full-day program for 4-year-olds who are expecting to enter kindergarten in the next school year. Instructional standards for *TN-VPK* are set by the State Board of Education. Each *TN-VPK* classroom is required to have a teacher with a license in early childhood development and education, an aide, at least one adult per 10 students, 20 students or fewer in the entire class, and an approved curriculum. Children from low-income families who apply to *TN-VPK* are admitted first, followed by other at-risk children, including students with disabilities or limited English proficiency, followed by children who are not considered to be at risk.

who attended at least 20 days of *TN-VPK* were rated by teachers as having better work-related skills and social behavior. All of these differences were statistically significant, and the WWC confirmed these findings after adjusting for multiple comparisons.

Furthermore, the study authors found that students who attended at least 20 days of *TN-VPK* had significantly lower scores than other students on quantitative concepts (mathematics) at the end of second grade and applied problems (mathematics) at the end of third grade. No other differences were statistically significant. After adjusting for multiple comparisons, the WWC confirmed the statistical significance for mathematics at the end of second grade but did not confirm the statistical significance for mathematics at the end of third grade.

Appendix A: Study details

Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade (Research report)*. Nashville, TN: Vanderbilt University, Peabody Research Institute. <http://files.eric.ed.gov/fulltext/ED566664.pdf>.

- Setting** The study took place within 21 school districts and 58 elementary schools in the state of Tennessee. Nineteen schools were in cities, 12 were in towns, 11 schools were in suburban areas, and 16 schools were in rural areas. Prekindergarten programs were operated through the school district, but actual services may have been provided through non-profit or for-profit child care providers, Head Start programs, or public schools.
- Study sample** As part of a larger evaluation of the *Tennessee Voluntary Prekindergarten program (TN-VPK)*, the authors randomly offered enrollment in 76 *TN-VPK* classes with more applicants than they could enroll in the 2009–10 and 2010–11 school years. In this study (referred to as the “Intensive Substudy” by the study authors), the sample was limited to children from the larger evaluation who had parental consent for data collection and had outcome data at the end of the prekindergarten year (1,076 of the 3,025 children who were randomized). Using the Intensive Substudy sample, the authors employed a quasi-experimental design to compare children who had attended *TN-VPK* for 20 or more days to children who did not attend *TN-VPK* or attended fewer than 20 days. In the Intensive Substudy sample, children were 51.8 months (4 years, 4 months) of age on average at baseline; 48% were male; 23% were African American, and 19% were Hispanic; and 79% were native English speakers. The sample sizes varied by outcome and wave of measurement.
- Intervention group** The intervention condition attended *TN-VPK* for 20 or more days. *TN-VPK* is a full-day program for 4-year-olds that operates on the same school-year calendar as the public school system in Tennessee. Enrollment is prioritized for students who are deemed as being at risk (i.e., eligible for free/reduced-price lunch, English language learners, children with disabilities, or “otherwise at risk”). The program must meet the State Board of Education standards regarding teacher licensing, adult:student ratio, class size, and approved curriculum. The study authors stated that the 20-day threshold for attendance corresponds to a state enrollment standard.
- Comparison group** The comparison group children did not attend *TN-VPK* or attended fewer than 20 days of *TN-VPK*. These children could have been cared for at home, received private childcare, attended Head Start, or attended another center-based preschool program.

Outcomes and measurement

The study examines multiple outcomes in prekindergarten, kindergarten, and grades 1–3. There are six direct assessment outcomes from the Woodcock-Johnson III (WJ-III) Tests of Achievement that are discussed in this single study review: Letter-Word Identification and Spelling (reviewed in the alphabets outcome domain); Oral Comprehension (under the language development domain); Picture Vocabulary (under the comprehension domain); and Applied Problems and Quantitative Concepts (under the mathematics domain). In addition, the authors averaged the six subtest scores to form a WJ-III Composite⁶ index, which is reviewed in the cognition domain. The study also assessed two outcomes using teacher ratings from the Cooper-Farran Behavior Rating Scales (CF): Work-Related Skills and Social Behavior (under the social-emotional development domain). For a more detailed description of these outcome measures, see Appendix B.

Three additional measures from the WJ-III (Calculation, Passage Comprehension, and a composite based on all eight subscale scores from the WJ-III) and four measures from a separate teacher rating instrument, the Academic Classroom and Behavior Record (Preparedness for Grade, Peer Relations, Behavior Problems, Feelings About School) were used by the study but are not discussed in this single study review. The analyses of these outcomes did not meet WWC group design standards because baseline equivalence was not demonstrated with observed data for the analytic samples.

In accordance with the WWC Evidence Review Protocol for Early Childhood Education Interventions (version 3.0), this review identifies the primary outcomes as those measured at the most recent follow-up period after the end of the intervention (spring of prekindergarten; Appendix C). Findings for students in kindergarten and grades 1, 2, and 3 are presented in Appendices D.1–D.4, respectively.

Support for implementation

The State Board of Education standards for *TN-VPK* specify a licensed classroom teacher, at least one adult per 10 children, maximum class size of 20, and use of an approved curriculum. Non-school settings (e.g., private child care or Head Start) needed to meet these standards and hold the highest ratings from the Tennessee Department of Human Services licensing system. No additional information is provided about implementation of the program.

Reason for review

This study was identified for review by receiving media attention.

Appendix B: Outcome measures for each domain

| Mathematics | |
|---|--|
| <i>Woodcock-Johnson III (WJ-III) Tests of Achievement, Applied Problems subtest</i> | The WJ-III, Applied Problems subtest is a standardized measure of children's abilities to solve mathematical problems presented orally with accompanying pictures of objects. The study authors used longitudinally-scaled W-scores in the analysis. W-scores use information on item difficulty to measure fixed increases in the probability that an individual can complete a task successfully. An increase of 10 in the W score indicates that the individual can perform tasks with 75% success that he or she previously performed with 50% success. |
| <i>WJ-III, Quantitative Concepts subtest</i> | The WJ-III, Quantitative Concepts subtest is a standardized measure of children's abilities to count, identify shapes, patterns, numbers, and series, as well as their knowledge of mathematical concepts and terms. The study authors used longitudinally-scaled W-scores in the analysis. W-scores use information on item difficulty to measure fixed increases in the probability that an individual can complete a task successfully. An increase of 10 in the W score indicates that the individual can perform tasks with 75% success that he or she previously performed with 50% success. |
| Alphabetic | |
| <i>WJ-III, Letter-Word Identification subtest</i> | The WJ-III, Letter-Word Identification subtest is a standardized measure of children's abilities to identify and pronounce individual letters of the alphabet and read words, ranging from commonly used words to less familiar words of the English language. The study authors used longitudinally-scaled W-scores in the analysis. W-scores use information on item difficulty to measure fixed increases in the probability that an individual can complete a task successfully. An increase of 10 in the W score indicates that the individual can perform tasks with 75% success that he or she previously performed with 50% success. |
| <i>WJ-III, Spelling subtest</i> | The WJ-III, Spelling subtest is a standardized measure of children's abilities to draw simple shapes and write orally-presented letters and words. The study authors used longitudinally-scaled W-scores in the analysis. W-scores use information on item difficulty to measure fixed increases in the probability that an individual can complete a task successfully. An increase of 10 in the W score indicates that the individual can perform tasks with 75% success that he or she previously performed with 50% success. |
| Comprehension | |
| <i>WJ-III, Picture Vocabulary subtest</i> | The WJ-III, Picture Vocabulary subtest is a standardized, expressive vocabulary test that asks students to name objects presented in pictures and choose the picture that corresponds to a word. The study authors used longitudinally-scaled W-scores in the analysis. W-scores use information on item difficulty to measure fixed increases in the probability that an individual can complete a task successfully. An increase of 10 in the W score indicates that the individual can perform tasks with 75% success that he or she previously performed with 50% success. |
| Language development | |
| <i>WJ-III, Oral Comprehension subtest</i> | The WJ-III, Oral Comprehension Subtest is a standardized measure of children's abilities to comprehend a short orally-presented passage. In the assessment, children are asked to fill in missing words using syntactic and semantic information in the passage. The study authors used longitudinally-scaled W-scores in the analysis. W-scores use information on item difficulty to measure fixed increases in the probability that an individual can complete a task successfully. An increase of 10 in the W score indicates that the individual can perform tasks with 75% success that he or she previously performed with 50% success. |
| Cognition | |
| <i>WJ-III, Composite6</i> | The WJ-III, Composite6 is a composite of six WJ-III subtests: Applied Problems, Quantitative Concepts, Letter-Word Identification, Spelling, Picture Vocabulary, and Oral Comprehension. A WJ-III, Composite6 score is the simple mean across the six WJ-III subtests. The study authors used longitudinally-scaled W-scores in the analysis. W-scores use information on item difficulty to measure fixed increases in the probability that an individual can complete a task successfully. An increase of 10 in the W score indicates that the individual can perform tasks with 75% success that he or she previously performed with 50% success. |

Social-emotional development

Cooper-Farran Behavioral Rating Scale (CF), Work-Related Skills subtest

This measure is part of a teacher rating form of individual children. The work-related skills subscale includes items about working independently, listening to the teacher, remembering and complying with instructions, and completing games and activities in general and whether they were completed within specified times for a given task. The scale consists of 16 items rated on a 1 to 7 scale. Based on a study by McClellan and Scalzo (2006), the CF interrater reliability was .79 for work-related skills.

CF, Social Behavior subtest

This measure is part of a teacher rating form of individual children. The social behavior subscale consists of 21 items rated on a 1 to 7 scale. Items assess whether the child's behavior is appropriate during the classroom's group activities, play, and outdoor time; how they express ideas and emotions during group activities; and how they respond to others' problems. Based on a study by McClellan and Scalzo (2006), the CF interrater reliability was .78 for the social behavior scale.

Table Notes: More information on W scores is available in Jaffe (2009).⁴ Three additional measures from the WJ-III (Calculation, Passage Comprehension, and Composite8) and four measures from a separate teacher rating instrument, the Academic Classroom and Behavior Record (Preparedness for Grade, Peer Relations, Behavior Problems, Feelings About School) were used by the study but are not included in the single study review. The analyses of these outcomes did not meet WWC group design standards because baseline equivalence was not demonstrated with observed data for the analytic samples.

Appendix C: Study findings for each domain, from the spring of prekindergarten and fall of kindergarten

| Domain and outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|--|---------------------------|----------------|---------------------------|------------------|------------------|-------------|-------------------|--------------------------------------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Mathematics | | | | | | | | |
| <i>Woodcock-Johnson-III (WJ-III), Applied Problems</i> | Spring of prekindergarten | 1,043 students | 409.05 (20.75) | 404.95 (29.31) | 4.10 | 0.18 | +7 | < .01 |
| <i>WJ-III, Quantitative Concepts</i> | Spring of prekindergarten | 1,043 students | 422.21 (15.58) | 417.72 (16.36) | 4.49 | 0.28 | +11 | < .01 |
| Domain average for mathematics | | | | | | 0.23 | +9 | Statistically significant |
| Alphabetics | | | | | | | | |
| <i>WJ-III, Letter-Word Identification</i> | Spring of prekindergarten | 1,043 students | 344.25 (24.51) | 333.12 (30.68) | 11.13 | 0.42 | +16 | < .01 |
| <i>WJ-III, Spelling</i> | Spring of prekindergarten | 1,043 students | 376.59 (24.66) | 369.52 (26.00) | 7.07 | 0.28 | +11 | < .01 |
| Domain average for alphabetics | | | | | | 0.35 | +14 | Statistically significant |
| Comprehension | | | | | | | | |
| <i>WJ-III, Picture Vocabulary</i> | Spring of prekindergarten | 1,043 students | 462.89 (14.97) | 459.04 (24.12) | 3.85 | 0.21 | +8 | < .01 |
| Domain average for comprehension | | | | | | 0.21 | +8 | Statistically significant |
| Language development | | | | | | | | |
| <i>WJ-III, Oral Comprehension</i> | Spring of prekindergarten | 1,043 students | 452.30 (16.34) | 450.84 (18.02) | 1.46 | 0.09 | +3 | .10 |
| Domain average for language development | | | | | | 0.09 | +3 | Not statistically significant |
| Cognition | | | | | | | | |
| <i>WJ-III, Composite6</i> | Spring of prekindergarten | 1,043 students | 411.23 (15.22) | 405.85 (19.31) | 5.39 | 0.33 | +13 | < .01 |
| Domain average for cognition | | | | | | 0.33 | +13 | Statistically significant |

| Social-emotional development | | | | | | | | |
|--|----------------------|--------------|----------------|----------------|------|-------------|-----------|----------------------------------|
| <i>Cooper-Farran Behavioral Rating Scale (CF), Work-Related Skills</i> | Fall of kindergarten | 900 students | 5.04 (1.14) | 4.82 (1.13) | 0.22 | 0.19 | +8 | .02 |
| <i>CF, Social Behavior</i> | Fall of kindergarten | 900 students | 5.80 (0.91) | 5.61 (0.99) | 0.19 | 0.20 | +8 | .03 |
| Domain average for social-emotional development | | | | | | 0.20 | +8 | Statistically significant |

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding.

Study Notes: Means included in the report used a method to account for missing data that is not accepted by the WWC for quasi-experimental designs. The unimputed, model-adjusted means, unadjusted standard deviations, and *p*-values displayed in this table were provided by the authors. Corrections for multiple comparisons were needed in the mathematics domain, alphabets domain, and social-emotional development domain but did not affect whether any of the contrasts were found to be statistically significant. This study is characterized as having a statistically significant positive effect in the mathematics domain, alphabets domain, comprehension domain, language development domain, and social-emotional development domain because the effect for at least one measure in each domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons if applicable. This study is characterized as having an indeterminate effect in the language development domain, because the estimated effect for the outcome in this domain is neither statistically significant nor substantively important. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D1: Supplemental findings from spring of kindergarten for each domain

| Domain and outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|---|------------------------|----------------|---------------------------|------------------|------------------|-------------|-------------------|---------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Mathematics | | | | | | | | |
| <i>WJ-III, Applied Problems</i> | Spring of kindergarten | 1,015 students | 436.67 (15.65) | 435.76 (16.36) | 0.92 | 0.06 | +2 | .44 |
| <i>WJ-III, Quantitative Concepts</i> | Spring of kindergarten | 1,015 students | 448.39 (13.56) | 449.62 (12.81) | -1.24 | -0.09 | -4 | .22 |
| Alphabetics | | | | | | | | |
| <i>WJ-III, Letter-Word Identification</i> | Spring of kindergarten | 1,015 students | 398.63 (27.09) | 399.19 (27.33) | -0.57 | -0.02 | -1 | .78 |
| <i>WJ-III, Spelling</i> | Spring of kindergarten | 1,015 students | 423.53 (20.90) | 424.55 (20.77) | -1.01 | -0.05 | -2 | .54 |
| Comprehension | | | | | | | | |
| <i>WJ-III, Picture Vocabulary</i> | Spring of kindergarten | 1,015 students | 471.89 (10.77) | 471.04 (12.52) | 0.85 | 0.08 | +3 | .22 |
| Language development | | | | | | | | |
| <i>WJ-III, Oral Comprehension</i> | Spring of kindergarten | 1,015 students | 466.02 (14.30) | 464.87 (17.12) | 1.15 | 0.08 | +3 | .26 |
| Cognition | | | | | | | | |
| <i>WJ-III, Composite6</i> | Spring of kindergarten | 1,015 students | 440.98 (13.46) | 440.98 (13.87) | 0.00 | 0.00 | 0 | .99 |

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards with reservations, but do not factor into the determination of the study rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

Study Notes: Means included in the report used a method to account for missing data that is not accepted by the WWC for quasi-experimental designs. The unimputed, model-adjusted means, unadjusted standard deviations, and p-values displayed in this table were provided by the authors. The WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences for findings displayed in this table. This supplemental analysis is characterized as having an indeterminate effect in the mathematics domain and the alphabetics domain because the mean effect reported in each of these domains is neither statistically significant nor substantively important. This supplemental analysis is characterized as having an indeterminate effect in the comprehension domain, language development domain, and cognition domain because the estimated effect for the outcome in each of these domains is neither statistically significant nor substantively important. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

Appendix D2: Supplemental findings from spring of first grade for each domain

| Domain and outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|---|-----------------------|--------------|---------------------------|------------------|------------------|-------------|-------------------|---------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Mathematics | | | | | | | | |
| <i>WJ-III, Applied Problems</i> | Spring of first grade | 997 students | 457.77 (15.43) | 457.87 (15.82) | -0.10 | -0.01 | 0 | .94 |
| <i>WJ-III, Quantitative Concepts</i> | Spring of first grade | 997 students | 464.33 (13.31) | 466.11 (13.17) | -1.78 | -0.13 | -5 | .09 |
| Alphabetics | | | | | | | | |
| <i>WJ-III, Letter-Word Identification</i> | Spring of first grade | 997 students | 446.40 (28.27) | 448.18 (29.16) | -1.77 | -0.06 | -2 | .45 |
| <i>WJ-III, Spelling</i> | Spring of first grade | 997 students | 459.87 (20.75) | 460.41 (20.00) | -0.55 | -0.03 | -1 | .75 |
| Comprehension | | | | | | | | |
| <i>WJ-III, Picture Vocabulary</i> | Spring of first grade | 997 students | 478.85 (11.18) | 478.28 (11.35) | 0.57 | 0.05 | +2 | .45 |
| Language development | | | | | | | | |
| <i>WJ-III, Oral Comprehension</i> | Spring of first grade | 997 students | 477.04 (13.51) | 478.44 (13.53) | -1.40 | -0.10 | -4 | .15 |
| Cognition | | | | | | | | |
| <i>WJ-III, Composite6</i> | Spring of first grade | 997 students | 464.11 (13.85) | 464.97 (13.84) | -0.86 | -0.06 | -2 | .36 |
| Social-emotional development | | | | | | | | |
| <i>CF, Work-Related Skills</i> | Spring of first grade | 886 students | 5.04 (1.20) | 5.20 (1.14) | -0.16 | -0.14 | -5 | .12 |
| <i>CF, Social Behavior</i> | Spring of first grade | 886 students | 5.80 (0.96) | 5.86 (0.99) | -0.06 | -0.06 | -2 | .50 |

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards with reservations, but do not factor into the determination of the study rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The means displayed in this table reflect adjusted means obtained through a response to author query. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

Study Notes: Means included in the report used a method to account for missing data that is not accepted by the WWC for quasi-experimental designs. The unimputed, model-adjusted means, unadjusted standard deviations, and p-values displayed in this table were provided by the authors. The WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences for findings displayed in this table. This supplemental analysis is characterized as having an indeterminate effect in the mathematics domain, alphabetics domain, and social-emotional development domain because the mean effect reported in each domain is neither statistically significant nor substantively important. This supplemental analysis is characterized as having an indeterminate effect in the comprehension domain, language development domain, and cognition domain because the estimated effect for the outcome in each of these domains is neither statistically significant nor substantively important. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

Appendix D3: Supplemental findings from spring of second grade for each domain

| Domain and outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|---|------------------------|--------------|---------------------------|------------------|------------------|-------------|-------------------|---------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Mathematics | | | | | | | | |
| <i>WJ-III, Applied Problems</i> | Spring of second grade | 985 students | 473.63 (16.45) | 475.35 (16.78) | -1.72 | -0.10 | -4 | .20 |
| <i>WJ-III, Quantitative Concepts</i> | Spring of second grade | 985 students | 476.35 (14.00) | 479.19 (13.78) | -2.84 | -0.20 | -8 | .01 |
| Alphabetics | | | | | | | | |
| <i>WJ-III, Letter-Word Identification</i> | Spring of second grade | 985 students | 470.46 (24.86) | 472.92 (26.57) | -2.46 | -0.10 | -4 | .26 |
| <i>WJ-III, Spelling</i> | Spring of second grade | 985 students | 476.91 (20.90) | 478.81 (19.47) | -1.90 | -0.09 | -4 | .28 |
| Comprehension | | | | | | | | |
| <i>WJ-III, Picture Vocabulary</i> | Spring of second grade | 985 students | 484.84 (11.45) | 485.30 (11.43) | -0.47 | -0.04 | -2 | .57 |
| Language development | | | | | | | | |
| <i>WJ-III, Oral Comprehension</i> | Spring of second grade | 985 students | 485.26 (13.24) | 486.84 (12.81) | -1.58 | -0.12 | -5 | .11 |
| Cognition | | | | | | | | |
| <i>WJ-III, Composite6</i> | Spring of second grade | 985 students | 478.00 (13.69) | 479.89 (13.49) | -1.89 | -0.14 | -6 | .05 |
| Social-emotional development | | | | | | | | |
| <i>CF, Work-Related Skills</i> | Spring of second grade | 864 students | 5.09 (1.18) | 5.10 (1.11) | -0.01 | -0.01 | 0 | .94 |
| <i>CF, Social Behavior</i> | Spring of second grade | 864 students | 5.88 (0.93) | 5.87 (0.95) | 0.01 | 0.01 | 0 | .95 |

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards with reservations, but do not factor into the determination of the study rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

Study Notes: Means included in the report used a method to account for missing data that is not accepted by the WWC for quasi-experimental designs. The unimputed, model-adjusted means, unadjusted standard deviations, and p-values displayed in this table were provided by the authors. A correction for multiple comparisons was needed for the mathematics domain but did not affect whether any of the contrasts were found to be statistically significant. This supplemental analysis is characterized as having a statistically significant negative effect in the mathematics domain because at least one measure is negative and statistically significant and no effects are positive and statistically significant, accounting for multiple comparisons. The WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences for findings in other domains displayed in this table. This supplemental analysis is characterized as having an indeterminate effect in the alphabetics domain, comprehension domain, language development domain, cognition domain, and social-emotional development domain because the mean effect reported in each of these domains is neither statistically significant nor substantively important. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

Appendix D4: Supplemental findings for spring of third grade from each domain

| Domain and outcome measure | Study sample | Sample size | Mean (standard deviation) | | WWC calculations | | | p-value |
|---|-----------------------|--------------|---------------------------|------------------|------------------|-------------|-------------------|---------|
| | | | Intervention group | Comparison group | Mean difference | Effect size | Improvement index | |
| Mathematics | | | | | | | | |
| <i>WJ-III, Applied Problems</i> | Spring of third grade | 962 students | 484.75 (17.83) | 487.81 (17.87) | -3.05 | -0.17 | -7 | .04 |
| <i>WJ-III, Quantitative Concepts</i> | Spring of third grade | 962 students | 489.23 (13.58) | 490.56 (13.07) | -1.33 | -0.10 | -4 | .25 |
| Alphabetics | | | | | | | | |
| <i>WJ-III, Letter-Word Identification</i> | Spring of third grade | 962 students | 487.95 (23.56) | 490.47 (25.61) | -2.52 | -0.10 | -4 | .25 |
| <i>WJ-III, Spelling</i> | Spring of third grade | 962 students | 489.45 (19.30) | 490.59 (19.75) | -1.14 | -0.06 | -2 | .51 |
| Comprehension | | | | | | | | |
| <i>WJ-III, Picture Vocabulary</i> | Spring of third grade | 962 students | 491.32 (11.48) | 490.25 (11.46) | 1.07 | 0.09 | +4 | .22 |
| Language development | | | | | | | | |
| <i>WJ-III, Oral Comprehension</i> | Spring of third grade | 962 students | 494.31 (13.60) | 494.45 (12.87) | -0.14 | -0.01 | 0 | .90 |
| Cognition | | | | | | | | |
| <i>WJ-III, Composite6</i> | Spring of third grade | 962 students | 489.56 (13.46) | 490.80 (13.37) | -1.24 | -0.09 | -4 | .23 |
| Social-emotional development | | | | | | | | |
| <i>CF, Work-Related Skills</i> | Spring of third grade | 875 students | 5.13 (1.18) | 5.06 (1.15) | 0.07 | 0.06 | +2 | .51 |
| <i>CF, Social Behavior</i> | Spring of third grade | 875 students | 5.91 (0.93) | 5.86 (0.88) | 0.05 | 0.05 | +2 | .59 |

Table Notes: The supplemental findings presented in this table are additional findings that meet WWC design standards with reservations, but do not factor into the determination of the study rating. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on individual outcomes, representing the average change expected for all individuals who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average individual's percentile rank that can be expected if the individual is given the intervention. Some statistics may not sum as expected due to rounding.

Study Notes: Means included in the report used a method to account for missing data that is not accepted by the WWC for quasi-experimental designs. The unimputed, model-adjusted means, unadjusted standard deviations, and p-values displayed in this table were provided by the authors. A correction for multiple comparisons was needed for the mathematics domain and resulted in a WWC-computed critical p-value of .03 that was higher than the p-value of .04 for the WJ-III, Applied Problems measure; therefore, the WWC does not find the result to be statistically significant. The analysis is characterized as having an indeterminate effect in the mathematics domain because the mean effect reported is neither statistically significant nor substantively important. The WWC did not need to make corrections for clustering, multiple comparisons, or to adjust for baseline differences for findings for other domains displayed in this table. This supplemental analysis is characterized as having an indeterminate effect in the alphabetics domain, comprehension domain, language development domain, cognition domain, and social-emotional development domain because the mean effect reported in each of these domains is neither statistically significant nor substantively important. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information.

Endnotes

¹ Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the authors) to assess whether the study design meets WWC group design standards. The review reports the WWC's assessment of whether the study meets WWC group design standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol (version 2.0). The review also referred to the Review Protocol for Early Childhood Education Interventions (version 3.0) to separate outcomes into Appendices C and D.1–D.4. A quick review of this study was released in January 2016, and this report is the follow-up review that replaces that initial assessment. The study rating in this report reflects additional information provided by the authors. The WWC rating applies only to the study outcomes that were eligible for review under the single study review protocol (version 2.0). The reported analyses in this single study review are only for those eligible outcomes that either met WWC group design standards without reservations or met WWC group design standards with reservations, and do not necessarily apply to all results presented in the study.

² In the study, the authors present results of statistical tests based on analyses that use an imputation method that is not accepted by the WWC for quasi-experimental designs. The authors provided the WWC with sample sizes, adjusted means, standard deviations, and *p*-values for the analytic samples of children with observed data. This information was provided across all follow-up periods, for most of the outcomes presented in the study. These analyses demonstrated baseline equivalence and are presented in this report because they meet WWC group design standards with reservations. Because only the findings obtained from the authors meet WWC group design standards, this report does not present findings from the manuscript.

³ There were seven outcomes included in the study that are not described in this WWC report. See the table notes in Appendix B for more information.

⁴ Jaffe, L. E. (2009). *Development, interpretation, and application of the W score and the relative proficiency index (Woodcock-Johnson III Assessment Service Bulletin No. 11)*. Rolling Meadows, IL: Riverside Publishing.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, September). *WWC review of the report: A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade*. Retrieved from <http://whatworks.ed.gov>

Glossary of Terms

| | |
|--|---|
| Attrition | Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study. |
| Clustering adjustment | If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary. |
| Confounding factor | A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor. |
| Design | The design of a study is the method by which intervention and comparison groups were assigned. |
| Domain | A domain is a group of closely related outcomes. |
| Effect size | The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes. |
| Eligibility | A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design. |
| Equivalence | A demonstration that the analytic sample groups are similar on observed characteristics defined in the review area protocol. |
| Improvement index | Along a percentile distribution of individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from -50 to +50. |
| Multiple comparison adjustment | When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary. |
| Quasi-experimental design (QED) | A quasi-experimental design (QED) is a research design in which study participants are assigned to intervention and comparison groups through a process that is not random. |
| Randomized controlled trial (RCT) | A randomized controlled trial (RCT) is an experiment in which eligible study participants are randomly assigned to intervention and comparison groups. |
| Single-case design (SCD) | A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention. |
| Standard deviation | The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample are spread out over a large range of values. |
| Statistical significance | Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < .05$). |
| Substantively important | A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance. |

Please see the [WWC Procedures and Standards Handbook \(version 3.0\)](#) for additional details.



Intervention
Report



Practice
Guide



Quick
Review



Single Study
Review

A **single study review** of an individual study includes the WWC's assessment of the quality of the research design and technical details about the study's design and findings.

This single study review was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.