

# What Works Clearinghouse



## Bright Beginnings

### Program Description<sup>1</sup>

*Bright Beginnings* is an early childhood curriculum, based in part on *High/Scope*<sup>®</sup> and *Creative Curriculum*<sup>®</sup>, with an additional emphasis on literacy skills. The curriculum consists of nine thematic units designed to enhance children’s cognitive, social, emotional, and physical development, and each unit includes

concept maps, literacy lessons, center activities, and home activities. Special emphasis is placed on the development of early language and literacy skills, and parent involvement is a key component of the program.

### Research<sup>2</sup>

No studies of *Bright Beginnings* meet What Works Clearinghouse (WWC) evidence standards, but one study meets WWC evidence standards with reservations. The one study included 198 students from 14 public preschool classrooms in Tennessee.<sup>3</sup>

print knowledge, phonological processing, and math. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Bright Beginnings* in the early reading and writing or the cognition domains.

Based on this one study, the WWC considers the extent of evidence for *Bright Beginnings* to be small for oral language,

### Effectiveness

*Bright Beginnings* was found to have potentially positive effects on print knowledge and no discernible effects on oral language, phonological processing, and math.

|                                | Oral language  | Print knowledge  | Phonological processing             | Early reading and writing | Cognition | Math   |
|--------------------------------|--|--|-------------------------------------|---------------------------|-----------|--|
| Rating of effectiveness        | No discernible effects   | Potentially positive effects   | No discernible effects              | na                        | na        | No discernible effects   |
| Improvement index <sup>4</sup> | Average: +4 percentile points<br>Range: +4 to +5 percentile points | Average: +12 percentile points<br>Range: +7 to +15 percentile points | Average: -3 percentile points<br>na | na<br>na                  | na<br>na  | Average: +4 percentile points<br>Range: -1 to +6 percentile points |

na = not applicable

1. The descriptive information for this program was obtained from publicly available sources: the program’s website (<http://www.cms.k12.nc.us/cmsdepartments/ci/pre-kservices/Pages/default.aspx>, downloaded August 2008) and the literature reviewed for this report. The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
2. The studies in this report were reviewed using WWC Evidence Standards, Version 1.0 (see the WWC Standards).
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
4. These numbers show the average and range of student-level improvement indices for all findings across the study (studies).

## Absence of conflict of interest

The PCER Consortium (2008) study summarized in this intervention report had numerous contributors, including staff of Mathematica Policy Research, Inc. (MPR). Because the principal

investigator for the WWC review of Early Childhood Education is also a MPR staff member, the study was rated by Chesapeake Research Associates, who also prepared the intervention report.

## Additional program information

### Developer and contact

Developed by former superintendent of Charlotte-Mecklenburg Schools, Eric Smith, in conjunction with district staff and local businesses. *Bright Beginnings* Office: Family Application Center, 700 Marsh Road, Charlotte, NC 28209. Email: prek@cms.k12.nc.us. Web: <http://www.cms.k12.nc.us/cmsdepartments/ci/pre-k-services/Pages/default.aspx>. Telephone: (980) 343-3797.

### Scope of use

*Bright Beginnings* was funded by the Charlotte-Mecklenburg district's Title I grant, and has been used primarily in the Charlotte-Mecklenburg school system, in five prekindergarten centers and 14 elementary schools. Recently, the program has been implemented in other districts as well.

### Teaching

The *Bright Beginnings* program is based on the *High/Scope*<sup>®</sup> and *Creative Curriculum*<sup>®</sup> models, with an additional emphasis on the development of early literacy skills. The curriculum is designed to create a child-centered, literacy-focused program that is relevant to the developmental needs of young children

and addresses their cognitive, social, emotional, and physical development. *Bright Beginnings* consists of nine units: language and literacy, mathematics, social and personal development, healthful living, scientific thinking, social studies, creative arts, physical development, and technology. Active exploration and interaction with other students, adults, and materials are important components of the *Bright Beginnings* model. As children participate in a variety of *Bright Beginnings* activities, they are continually monitored by teachers to assess their progress. Another feature of the *Bright Beginnings* program is an effort to engage parents; they are required to sign a parent-school partnership agreement affirming their active participation in their children's education.

### Cost

District cost estimates, based on use in the Charlotte-Mecklenburg school system, range from \$9.8 million for 1,672 *Bright Beginnings* students in the first cohort (1997–98) to \$16.7 million for 3,020 *Bright Beginnings* students in the sixth cohort (2002–03), resulting in a cost of approximately \$5,500 to \$5,800 per student.

## Research

Five studies reviewed by the WWC investigated the effects of *Bright Beginnings*. One study (PCER Consortium, 2008) is a randomized controlled trial design that had non-random allocations after random assignment, but the analytic groups were shown to be equivalent, so the study meets WWC evidence standards with reservations. No studies are randomized controlled trials or quasi-experimental designs that meet WWC evidence standards. The remaining four studies do not meet either WWC evidence standards or eligibility screens.

### Meets evidence standards with reservations

One study reviewed by the WWC (PCER Consortium, 2008) assessed the effectiveness of *Bright Beginnings* as part of the Preschool Curriculum Evaluation Research effort.<sup>5</sup> The PCER Consortium (2008) randomly assigned 36 classrooms to three conditions (*Bright Beginnings*, *Creative Curriculum*<sup>®</sup>, and control), randomly selecting 21 classrooms for the national evaluation. After the pilot year, 8 classrooms dropped out (2 *Bright Beginnings*, 3 *Creative Curriculum*<sup>®</sup>, and 3 control) and were

5. The Preschool Curriculum Evaluation Research Consortium (2008) evaluated a total of 14 preschool curricula, including *Bright Beginnings*, in comparison to respective control conditions.

## Research *(continued)*

replaced with classrooms from the original 36. For this study, 14 classrooms were used (including the original and replacement classrooms), and baseline equivalence on pretests was established for the *Bright Beginnings* and control children. Data were collected on 198 children (98 *Bright Beginnings* and 100 control). Just over half of the children (51%) were male, 82% were Caucasian, and 23% were reported to have a disability. Pretest data were collected in the fall and posttest data were collected in the spring of the preschool year. The study investigated effects on oral language, print knowledge, phonological processing, and math. The control condition varied across sites and included teacher-developed, nonspecific curricula with a focus on basic school readiness.

## Effectiveness Findings

The WWC review of interventions for early childhood education addresses child outcomes in six domains: oral language, print knowledge, phonological processing, early reading and writing, cognition, and math. The study included in this report covers four domains: oral language, print knowledge, phonological processing, and math. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Bright Beginnings* on children.<sup>7</sup>

*Oral Language.* The PCER Consortium (2008) analyzed the effectiveness of *Bright Beginnings* on oral language using the Peabody Picture Vocabulary Test (PPVT-III) and the Test of Language Development (TOLD). The authors report, and the WWC

## Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.<sup>6</sup>

The WWC considers the extent of evidence for *Bright Beginnings* to be small for oral language, print knowledge, phonological processing, and math. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Bright Beginnings* in the early reading and writing or the cognition domains.

confirms, that differences between the *Bright Beginnings* and control groups are not statistically significant or substantively important on any of these measures. According to WWC criteria, this study shows no discernible effects on oral language.

*Print Knowledge.* The PCER Consortium (2008) analyzed the effectiveness of *Bright Beginnings* on the Test of Early Reading Ability (TERA-3), the Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest, and the WJ-III Spelling subtest.<sup>8</sup> The study authors report a statistically significant positive effect of *Bright Beginnings* on the TERA-3, but according to WWC calculations, after correcting for multiple comparisons, the effect was not statistically significant. The study authors report, and the WWC confirms, there was no statistically significant effect on the WJ-III

6. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Bright Beginnings* is in Appendix A6.
7. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. No correction for clustering was needed for the study by the PCER Consortium (2008) because their analysis corrected for clustering by using hierarchical linear modeling (HLM). A correction for multiple comparisons was needed for the PCER Consortium (2008) study, so the significance levels in this report may differ from those reported in the original study.
8. By name, the TERA-3 sounds like it should be captured under the early reading and writing domain; however, the description of the measure identifies constructs that are pertinent to print knowledge, such as knowing the alphabet, understanding print conventions, and environmental print. More detailed explanations of the measures in each domain can be found in Appendices 2.1–2.4.

## Effectiveness *(continued)*

Letter-Word Identification subtest or on the WJ-III Spelling subtest. Although the effects are not statistically significant, the PCER Consortium found, and the WWC confirmed, that the effects on the TERA-3 and on the WJ-III Letter-Word Identification subtest are large enough to be considered substantively important according to WWC criteria (that is, at least 0.25). According to WWC criteria, the study shows potentially positive effects on print knowledge.

**Phonological Processing.** The PCER Consortium (2008) analyzed the effectiveness of *Bright Beginnings* on phonological processing using the Preschool Comprehensive Test of Phonological and Print Processing. The authors report, and the WWC confirms, that differences between the *Bright Beginnings* and control groups are not statistically significant or substantively important on this measure. According to WWC criteria, this study shows no discernible effects on phonological processing.

**Math.** The PCER Consortium (2008) analyzed the effectiveness of *Bright Beginnings* on math using the WJ-III Applied

Problems subtest, the Composite Score from the Child Math Assessment-Abbreviated, and Shape Composition. The authors report, and the WWC confirms, that differences between the *Bright Beginnings* and control groups are not statistically significant or substantively important on any of these measures. According to WWC criteria, this study shows no discernible effects on math.

## Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Procedures and Standards Handbook, Appendix E).

**The WWC found *Bright Beginnings* to have potentially positive effects on print knowledge and no discernible effects on oral language, phonological processing, and math**

## Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see WWC Procedures and Standards Handbook, Appendix F). The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.

Based on one study (PCER Consortium, 2008), the average improvement index of *Bright Beginnings* on two measures of oral language is +4 percentile points with a range from +4 to +5

percentile points across findings. For print knowledge, the average improvement index on three measures is +12 percentile points with a range from +7 to +15 percentile points across findings. For phonological processing, the improvement index on one outcome measure is -3 percentile points. For math, the average improvement index on three measures is +4 percentile points with a range from -1 to +6 percentile points.

## Summary

The WWC reviewed five studies of *Bright Beginnings*. One of these studies meets WWC evidence standards with reservations; the remaining four studies do not meet either WWC evidence standards or eligibility screens. Based on the one study, the WWC found potentially positive effects on print knowledge and no discernible effects on oral language, phonological processing, and math. The conclusions presented in this report may change as new research emerges.

## References **Meets WWC evidence standards with reservations**

Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). *Bright Beginnings and Creative Curriculum*: Vanderbilt University. In *Effects of preschool curriculum programs on school readiness* (pp. 41–54). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

## **Studies that fall outside the Early Childhood Education review protocol or do not meet WWC evidence standards**

Bucci, A. F. (2000). Using Title I and local funds to build quality preschool programs in Charlotte-Mecklenburg: A “*Bright Beginning*”. In *Current state and local initiatives to support student learning: Early childhood programs and innovative programs to better address the needs of youth* (pp. 12–17). Selected presentations from an “Ensuring Student Success through Collaboration Network” conference (September 12–15, 1999), Louisville, KY. The study is ineligible for review because it does not use a comparison group.

Charlotte-Mecklenburg Schools. (2004). *Bright Beginnings program cost-benefit analysis project report*. The study is ineligible for review because it does not include an outcome within a domain specified in the protocol.

Charlotte-Mecklenburg Schools. *Comparison study of 1997–98 Bright Beginning participants percent at or above grade level*

*on end-of-year assessment*. Retrieved from <http://www.cms.k12.nc.us/cmsdepartments/ci/pre-k/services/Pages/ComparisonStudy.aspx>. The study is ineligible for review because it does not provide enough information about its design to assess whether it meets standards.

Smith, E. J., Pellin, B. J., & Agruso, S. A. (2003). *Bright Beginnings: An effective literacy-focused PreK program for educationally disadvantaged four-year-old children*. Arlington, VA: Educational Research Service. The study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

### **Additional sources:**

McClure, M. (2000). For K–12 success, North Carolina district starts early. *Curriculum Administrator*, 36(8), 24.

Neuman, S. B. (2009). Changing the odds through high-quality early care and education. In *Changing the odds for at risk children: Seven essential principles of educational programs that break the cycle of poverty* (pp. 97–126). Westport, CT: Praeger.

Smith, E. J. (2007). Weaving the gifted into the full fabric. *School Administrator*, 64(2), 10.

Tough, P. (2008). The conveyor belt. In *Whatever it takes: Geoffrey Canada’s quest to change Harlem and America* (pp. 188–212). New York: Houghton Mifflin Harcourt.

# Appendix

## Appendix A1 Study characteristics: Preschool Curriculum Evaluation Research Consortium, 2008 (randomized controlled trial)

| Characteristic                          | Description   |
|---|---|
| <b>Study citation</b>                   | Preschool Curriculum Evaluation Research (PCER) Consortium (2008). <i>Bright Beginnings and Creative Curriculum</i> : Vanderbilt University. In <i>Effects of Preschool Curriculum Programs on School Readiness</i> (pp. 41–54). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.   |
| <b>Participants</b>                     | The study, conducted during the 2003/04 and 2004/05 school years, included three intervention groups: <i>Bright Beginnings</i> , <i>Creative Curriculum</i> <sup>®</sup> , and a control group. Thirty-six full-day prekindergarten classrooms in 28 public preschool programs in Tennessee were recruited and blocked into groups of three by matching them on composite factors for demographic characteristics (urban/rural, percent of races other than White) and achievement (percent free lunch, reading, language, mathematics, and science achievement scores). Within each block, one preschool was randomly assigned to <i>Bright Beginnings</i> , one to <i>Creative Curriculum</i> <sup>®</sup> , and one to the control group. The manuscript notes that the researchers randomly assigned the classrooms to three conditions; however, all classrooms within a preschool were assigned to the same condition. Subsequent to randomization, 21 of the 36 classrooms (7 from each of the three groups) were randomly selected to participate in the national PCER study of <i>Bright Beginnings</i> and <i>Creative Curriculum</i> <sup>®</sup> . All 36 classrooms participated in the local investigator’s pilot study during the first year. Following the pilot year, and prior to starting the national PCER study, 8 of the 21 originally assigned classrooms dropped out of the study, leaving 5 <i>Bright Beginnings</i> , 4 <i>Creative Curriculum</i> <sup>®</sup> , and 4 control classrooms. The 8 dropout classrooms were replaced by randomly selecting 8 from the 15 classrooms that had not been selected to participate in the national PCER study, including 2 <i>Bright Beginnings</i> , 3 <i>Creative Curriculum</i> <sup>®</sup> , and 3 control classrooms, restoring the sample of classrooms to 7 in each of the three intervention groups. The evaluation of <i>Bright Beginnings</i> included 14 classrooms (7 <i>Bright Beginnings</i> and 7 control) and a total of 208 children at baseline (103 <i>Bright Beginnings</i> and 105 control), while the analysis sample included 98 <i>Bright Beginnings</i> children and 100 control children. Pretest differences between the treatment and comparison groups were not statistically significant. At baseline, children in the study averaged 4.5 years of age; 51% were male; and 11% were Hispanic, and 82% were White. A higher percentage of parents in the control group reported that their child had an Individualized Education plan relative to those assigned to <i>Bright Beginnings</i> (33 percent vs. 13 percent), a difference that was statistically significant, but did not exceed the 25% upper limit on acceptable baseline differences between groups that is indicated in the WWC Early Childhood Education protocol. |
| <b>Setting</b>                          | The <i>Bright Beginnings</i> study was conducted in prekindergarten classrooms in 14 public schools (7 <i>Bright Beginnings</i> and 7 control) from 7 county school districts in Tennessee.   |
| <b>Intervention</b>                     | <i>Bright Beginnings</i> is an integrated curriculum with a focus on language and early literacy, based in part on the <i>High/Scope</i> <sup>®</sup> and <i>Creative Curriculum</i> <sup>®</sup> models, with an added focus on skills designed to promote school literacy. <i>Bright Beginnings</i> includes nine curriculum units: language and literacy, mathematics, social and personal development, healthful living, scientific thinking, social studies, creative arts, physical development, and technology. In the PCER study, each classroom’s fidelity to the curriculum was rated on a four-point scale, ranging from “not at all” (0) to “high” (3). The average score for the <i>Bright Beginnings</i> classrooms was 1.88 on the measure.  |
| <b>Comparison</b>                       | Control teachers used teacher-developed, nonspecific curricula with a focus on basic school readiness. Their classrooms were rated with the same fidelity measure used in the <i>Bright Beginnings</i> classrooms, which ranged from 0 to 3. The average score for the control classrooms was 2.0.  |
| <b>Primary outcomes and measurement</b> | The outcome domains assessed were children’s oral language, print knowledge, phonological processing, and math. Oral language was assessed with the Peabody Picture Vocabulary Test-III (PPVT-III) and the Test of Language Development-Primary III (TOLD-P:3) Grammatical Understanding subtest. Print knowledge was assessed with the Test of Early Reading Ability-III (TERA-3), the Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest, and the WJ-III Spelling subtest. Phonological processing was assessed with the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) Elision subtest. Math was assessed with the WJ-III Applied Problems subtest, the Child Math Assessment-Abbreviated (CMA-A), and the Shape Composition task. For a more detailed description of these outcome measures, see Appendices A2.1–2.4.   |
| <b>Staff/teacher training</b>           | <i>Bright Beginnings</i> teachers received 2.5 days of curriculum training prior to the start of the prekindergarten year. Onsite consultation to teachers was provided four times during the school year: twice by trained Tennessee staff members and twice by curriculum trainers. Consultation visits typically included a classroom observation, an opportunity for teachers to ask questions about the curriculum, and implementation feedback from the trainer.  |

## Appendix A2.1 Outcome measures for the oral language domain

| Outcome measure  | Description   |
|--|---|
| <b>Peabody Picture Vocabulary Test–3rd Edition (PPVT-III)</b>                                | A standardized measure of children’s receptive vocabulary in which children demonstrate understanding of a spoken word by pointing to a picture that best represents the meaning (as cited in PCER Consortium, 2008). |
| <b>Test of Language Development-Primary III (TOLD-P:3) Grammatical Understanding subtest</b> | A standardized measure of children’s ability to comprehend the meaning of sentences by selecting pictures that most accurately represent the sentence (as cited in PCER Consortium, 2008).                            |

## Appendix A2.2 Outcome measures for the print knowledge domain

| Outcome measure   | Description  |
|---|--|
| <b>Test of Early Reading Ability III (TERA-3)</b>                       | A standardized measure of children’s developing reading skills with three subtests: alphabet, conventions, and meaning (as cited in PCER Consortium, 2008). <sup>1</sup> |
| <b>Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest</b> | A standardized measure of identification of letters and reading of words (as cited in PCER Consortium, 2008).  |
| <b>Woodcock-Johnson III (WJ-III) Spelling subtest</b>                   | A standardized measure that assesses children’s prewriting skills, such as drawing lines, tracing, and writing letters (as cited in PCER Consortium, 2008).              |

1. By name, this measure sounds like it should be captured under the early reading and writing domain; however, the description of the measure identifies constructs that are pertinent to print knowledge, such as knowing the alphabet, understanding print conventions, and environmental print.

## Appendix A2.3 Outcome measures for the phonological processing domain

| Outcome measure  | Description  |
|--|--|
| <b>Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest</b> | A measure of children’s ability to identify and manipulate sounds in spoken words, using word prompts and picture plates for the first nine items and word prompts only for later items (as cited in PCER Consortium, 2008). |

## Appendix A2.4 Outcome measures for the math domain

| Outcome measure  | Description   |
|--|---|
| <b>Woodcock-Johnson III (WJ-III) Applied Problems subtest</b>    | A standardized measure of children's ability to solve numerical and spatial problems, presented verbally with accompanying pictures of objects (as cited in PCER Consortium, 2008).   |
| <b>Child Math Assessment-Abbreviated (CMA-A) Composite Score</b> | The average of four subscales: (1) solving addition and subtraction problems using visible objects, (2) constructing a set of objects equal in number to a given set, (3) recognizing shapes, and (4) copying a pattern using objects that vary in color and identity from the model pattern (as cited in PCER Consortium, 2008). |
| <b>Building Blocks, Shape Composition task</b>                   | Modified for PCER from the Building Blocks assessment tools. Children use blocks to fill in a puzzle and are assessed on whether they fill the puzzle without gaps or hangovers (as cited in PCER Consortium, 2008).  |



## Appendix A3.1 Summary of study findings included in the rating for the oral language domain<sup>1</sup>

| Outcome measure  | Study sample | Sample size (classrooms/ children)   | Authors' findings from the study               |                  | WWC calculations   |                          |   |                                |
|--|--------------|--------------------------------------|--|------------------|--|--------------------------|---|--------------------------------|
|  |              |                                      | Mean outcome (standard deviation) <sup>2</sup> |                  | Mean difference <sup>4</sup> (Bright Beginnings– comparison) | Effect size <sup>5</sup> | Statistical significance <sup>6</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>7</sup> |
|  |              | Bright Beginnings group <sup>3</sup> | Comparison group                               |                  |  |                          |   |                                |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>8</sup></b> |              |                                      |  |                  |  |                          |   |                                |
| PPVT-III   | Preschoolers | 14/195                               | 96.31<br>(14.71)                               | 93.93<br>(15.37) | 2.38   | 0.13                     | ns  | +5                             |
| TOLD-P:3 Grammatic Understanding subtest                                     | Preschoolers | 14/197                               | 9.60<br>(2.95)                                 | 9.11<br>(2.73)   | 0.49   | 0.09                     | ns  | +4                             |
| <b>Domain average for oral language<sup>9</sup></b>                          |              |                                      |  |                  |  | <b>0.11</b>              | <b>na</b>   | <b>+4</b>                      |

ns = not statistically significant

na = not applicable

PPVT-III = Peabody Picture Vocabulary Test-III

TOLD-P:3 = Test of Language Development Primary, Third Edition

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.1.
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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. 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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), and no impacts were statistically significant.
9. This row provides the study average, which in this instance 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 A3.2 Summary of study findings included in the rating for the print knowledge domain<sup>1</sup>

| Outcome measure  | Study sample | Sample size (classrooms/ children)   | Authors' findings from the study               |                  | WWC calculations   |                          |   |                                |
|--|--------------|--------------------------------------|--|------------------|--|--------------------------|---|--------------------------------|
|  |              |                                      | Mean outcome (standard deviation) <sup>2</sup> |                  | Mean difference <sup>4</sup> (Bright Beginnings– comparison) | Effect size <sup>5</sup> | Statistical significance <sup>6</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>7</sup> |
|  |              | Bright Beginnings group <sup>3</sup> | Comparison group                               |                  |  |                          |   |                                |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>8</sup></b> |              |                                      |  |                  |  |                          |   |                                |
| TERA-3   | Preschoolers | 14/198                               | 91.41<br>(15.91)                               | 87.98<br>(14.71) | 3.43   | 0.39                     | ns  | +15                            |
| WJ-III Letter-Word Identification subtest                                    | Preschoolers | 14/198                               | 106.06<br>(14.97)                              | 97.21<br>(13.03) | 8.85   | 0.35                     | ns  | +14                            |
| WJ-III Spelling subtest  | Preschoolers | 14/198                               | 95.75<br>(12.46)                               | 90.94<br>(12.98) | 4.81   | 0.18                     | ns  | +7                             |
| <b>Domain average for print knowledge<sup>9</sup></b>                        |              |                                      |  |                  |  | <b>0.31</b>              | <b>na</b>   | <b>+12</b>                     |

ns = not statistically significant

na = not applicable

TERA-3 = Test of Early Reading Ability III

WJ-III = Woodcock-Johnson III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.2.
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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. 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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no correction for clustering was needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), but a correction for multiple comparisons was necessary.
9. This row provides the study average, which in this instance 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 A3.3 Summary of study findings included in the rating for the phonological processing domain<sup>1</sup>

| Outcome measure  | Study sample | Sample size (classrooms/ children) | Authors' findings from the study               |                  | WWC calculations   |                          |   |                                |
|--|--------------|------------------------------------|--|------------------|--|--------------------------|---|--------------------------------|
|  |              |                                    | Mean outcome (standard deviation) <sup>2</sup> |                  | Mean difference <sup>4</sup> (Bright Beginnings– comparison) | Effect size <sup>5</sup> | Statistical significance <sup>6</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>7</sup> |
|  |              |                                    | Bright Beginnings group <sup>3</sup>           | Comparison group |  |                          |   |                                |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>8</sup></b> |              |                                    |  |                  |  |                          |   |                                |
| Pre-CTOPPP Elision subtest   | Preschoolers | 14/198                             | 10.02<br>(4.50)                                | 10.38<br>(4.78)  | –0.36  | –0.07                    | ns  | –3                             |
| <b>Domain average for phonological processing<sup>9</sup></b>                |              |                                    |  |                  |  | <b>–0.07</b>             | <b>na</b>   | <b>–3</b>                      |

ns = not statistically significant

na = not applicable

Pre-CTOPPP = Preschool Comprehensive Test of Phonological and Print Processing

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the phonological processing domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.3.
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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. 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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), and no impacts were statistically significant.
9. This row provides the study average, which in this instance 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 A3.4 Summary of study findings included in the rating for the math domain<sup>1</sup>

| Outcome measure  | Study sample | Sample size (classrooms/ children)   | Authors' findings from the study               |                  | WWC calculations   |                          |   |                                |
|--|--------------|--------------------------------------|--|------------------|--|--------------------------|---|--------------------------------|
|  |              |                                      | Mean outcome (standard deviation) <sup>2</sup> |                  | Mean difference <sup>4</sup> (Bright Beginnings– comparison) | Effect size <sup>5</sup> | Statistical significance <sup>6</sup> (at $\alpha = 0.05$ ) | Improvement index <sup>7</sup> |
|  |              | Bright Beginnings group <sup>3</sup> | Comparison group                               |                  |  |                          |   |                                |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>8</sup></b> |              |                                      |  |                  |  |                          |   |                                |
| WJ-III Applied Problems subtest  | Preschoolers | 14/198                               | 100.69<br>(14.68)                              | 96.48<br>(16.69) | 4.21   | 0.16                     | ns  | +6                             |
| CMA-A Composite  | Preschoolers | 14/198                               | 0.57<br>(0.25)                                 | 0.53<br>(0.27)   | 0.04   | 0.14                     | ns  | +6                             |
| Shape Composition  | Preschoolers | 14/198                               | 1.82<br>(0.93)                                 | 1.85<br>(0.91)   | –0.03  | –0.03                    | ns  | –1                             |
| <b>Domain average for math<sup>9</sup></b>                                   |              |                                      |  |                  |  | <b>0.09</b>              | <b>na</b>   | <b>+4</b>                      |

ns = not statistically significant

na = not applicable

WJ-III = Woodcock-Johnson III

CMA-A = Child Math Assessment - Abbreviated

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math domain. Follow-up findings from PCER Consortium (2008) are not included in these ratings but are reported in Appendix A4.4.
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. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's  $d$  based on a repeated measures analysis).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. 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 for the intervention group.
8. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), and no impacts were statistically significant.
9. This row provides the study average, which in this instance 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 A4.1 Summary of follow-up findings for the oral language domain<sup>1</sup>

| Outcome measure  | Study sample    | Sample size <sup>3</sup><br>(classrooms/<br>children) | Authors' findings from the study                   |                     | WWC calculations   |                             |   |                                   |
|--|-----------------|---|--|---------------------|--|-----------------------------|---|-----------------------------------|
|  |                 |   | Mean outcome<br>(standard deviation) <sup>2</sup>  |                     | Mean<br>difference <sup>5</sup><br>( <i>Bright<br/>Beginnings</i> –<br>comparison) | Effect<br>size <sup>6</sup> | Statistical<br>significance <sup>7</sup><br>(at $\alpha = 0.05$ ) | Improvement<br>index <sup>8</sup> |
|  |                 |   | <i>Bright<br/>Beginnings</i><br>group <sup>4</sup> | Comparison<br>group |  |                             |   |                                   |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>9</sup></b> |                 |   |  |                     |  |                             |   |                                   |
| PPVT-III   | Kindergarteners | nr/203  | 98.43<br>(10.83)                                   | 97.21<br>(13.74)    | 1.22   | 0.07                        | ns  | +3                                |
| TOLD-P:3 Grammatic<br>Understanding subtest                                  | Kindergarteners | nr/203  | 10.73<br>(2.91)                                    | 9.91<br>(2.93)      | 0.82   | 0.16                        | ns  | +6                                |

ns = not statistically significant

nr = not reported

PPVT-III = Peabody Picture Vocabulary Test-III

TOLD-P:3 = Test of Language Development Primary, Third Edition

1. This appendix presents follow-up findings considered for measures that fall in the oral language domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.1.
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. The PCER Consortium (2008) study included 134 kindergarten classrooms across all three conditions in this study (*Bright Beginnings*, *Creative Curriculum*<sup>®</sup>, and control). The number of classrooms for *Bright Beginnings* and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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 for the intervention group.
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), and no impacts were statistically significant.

## Appendix A4.2 Summary of follow-up findings for the print knowledge domain<sup>1</sup>

| Outcome measure  | Study sample    | Sample size <sup>3</sup><br>(classrooms/<br>children) | Authors' findings from the study                   |                     | WWC calculations   |                             |   |                                   |
|--|-----------------|---|--|---------------------|--|-----------------------------|---|-----------------------------------|
|  |                 |   | Mean outcome<br>(standard deviation) <sup>2</sup>  |                     | Mean<br>difference <sup>5</sup><br>( <i>Bright<br/>Beginnings</i> –<br>comparison) | Effect<br>size <sup>6</sup> | Statistical<br>significance <sup>7</sup><br>(at $\alpha = 0.05$ ) | Improvement<br>index <sup>8</sup> |
|  |                 |   | <i>Bright<br/>Beginnings</i><br>group <sup>4</sup> | Comparison<br>group |  |                             |   |                                   |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>9</sup></b> |                 |   |  |                     |  |                             |   |                                   |
| TERA-3   | Kindergarteners | nr/203  | 93.35<br>(16.02)                                   | 93.99<br>(17.75)    | –0.64  | –0.07                       | ns  | –3                                |
| WJ-III Letter-Word Identification subtest                                    | Kindergarteners | nr/204  | 106.12<br>(10.67)                                  | 103.96<br>(13.41)   | 2.16   | 0.09                        | ns  | +4                                |
| WJ-III Spelling subtest  | Kindergarteners | nr/204  | 102.12<br>(12.09)                                  | 100.57<br>(15.15)   | 1.55   | 0.06                        | ns  | +2                                |

ns = not statistically significant

nr = not reported

TERA-3 = Test of Early Reading Ability III

WJ-III = Woodcock-Johnson III

1. This appendix presents follow-up findings for measures that fall in the print knowledge domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.2.
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. The PCER Consortium (2008) study included 134 kindergarten classrooms across all three conditions in this study (*Bright Beginnings*, *Creative Curriculum*<sup>®</sup>, and control). The number of classrooms for *Bright Beginnings* and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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 for the intervention group.
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), and no impacts were statistically significant.

## Appendix A4.3 Summary of follow-up findings for the phonological processing domain<sup>1</sup>

| Outcome measure  | Study sample    | Sample size <sup>3</sup><br>(classrooms/<br>children) | Authors' findings from the study                   |                     | WWC calculations   |                             |   |                                   |
|--|-----------------|---|--|---------------------|--|-----------------------------|---|-----------------------------------|
|  |                 |   | Mean outcome<br>(standard deviation) <sup>2</sup>  |                     | Mean<br>difference <sup>5</sup><br>( <i>Bright<br/>Beginnings</i> –<br>comparison) | Effect<br>size <sup>6</sup> | Statistical<br>significance <sup>7</sup><br>(at $\alpha = 0.05$ ) | Improvement<br>index <sup>8</sup> |
|  |                 |   | <i>Bright<br/>Beginnings</i><br>group <sup>4</sup> | Comparison<br>group |  |                             |   |                                   |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>9</sup></b> |                 |   |  |                     |  |                             |   |                                   |
| CTOPP Elision subtest  | Kindergarteners | nr/203  | 4.34<br>(2.76)                                     | 4.30<br>(3.27)      | 0.04   | 0.01                        | ns  | 0                                 |

ns = not statistically significant

nr = not reported

CTOPP = Comprehensive Test of Phonological Processing

1. This appendix presents follow-up findings for measures that fall in the phonological processing domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.3.
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. The PCER Consortium (2008) study included 134 kindergarten classrooms across all three conditions in this study (*Bright Beginnings*, *Creative Curriculum*<sup>®</sup>, and control). The number of classrooms for *Bright Beginnings* and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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 for the intervention group.
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), and no impacts were statistically significant.

## Appendix A4.4 Summary of follow-up findings for the math domain<sup>1</sup>

| Outcome measure  | Study sample    | Sample size <sup>3</sup><br>(classrooms/<br>children) | Authors' findings from the study                   |                     | WWC calculations   |                             |   |                                   |
|--|-----------------|---|--|---------------------|--|-----------------------------|---|-----------------------------------|
|  |                 |   | Mean outcome<br>(standard deviation) <sup>2</sup>  |                     | Mean<br>difference <sup>5</sup><br>( <i>Bright<br/>Beginnings</i> –<br>comparison) | Effect<br>size <sup>6</sup> | Statistical<br>significance <sup>7</sup><br>(at $\alpha = 0.05$ ) | Improvement<br>index <sup>8</sup> |
|  |                 |   | <i>Bright<br/>Beginnings</i><br>group <sup>4</sup> | Comparison<br>group |  |                             |   |                                   |
| <b>PCER Consortium, 2008 (meets standards with reservations)<sup>9</sup></b> |                 |   |  |                     |  |                             |   |                                   |
| WJ-III Applied Problems subtest  | Kindergarteners | nr/204  | 103.21<br>(12.77)                                  | 99.88<br>(16.18)    | 3.33   | 0.13                        | ns  | +5                                |
| CMA-A Composite  | Kindergarteners | nr/203  | 0.71<br>(0.17)                                     | 0.69<br>(0.18)      | 0.02   | 0.07                        | ns  | +3                                |
| Shape Composition  | Kindergarteners | nr/204  | 2.49<br>(0.72)                                     | 2.36<br>(0.89)      | 0.13   | 0.15                        | ns  | +6                                |

ns = not statistically significant

nr = not reported

WJ III = Woodcock-Johnson III

CMA-A = Child Math Assessment - Abbreviated

1. This appendix presents follow-up findings for measures that fall in the math domain. End-of-preschool scores were used for rating purposes and are presented in Appendix A3.4.
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. The PCER Consortium (2008) study included 134 kindergarten classrooms across all three conditions in this study (*Bright Beginnings*, *Creative Curriculum*<sup>®</sup>, and control). The number of classrooms for *Bright Beginnings* and the control group is likely about two-thirds of the total.
4. In PCER Consortium (2008), the treatment group mean equals the sum of the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. In the case of PCER Consortium (2008), the mean differences are covariate-adjusted.
6. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B. In the case of PCER Consortium (2008), the WWC used the effect sizes reported by the study authors (Cohen's *d* based on a repeated measures analysis).
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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 for the intervention group.
9. The level of statistical significance was reported by the study authors or, when necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using hierarchical linear modeling (HLM), and no impacts were statistically significant.



## Appendix A5.1 *Bright Beginnings* rating for the oral language 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 oral language, the WWC rated *Bright Beginnings* as having no discernible effects.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important effects, either positive or negative, on oral language.

### 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.** Only one study of *Bright Beginnings* was included in this review and it showed no statistically significant or substantively important positive effects on oral language.

### AND

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

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on oral language.

**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.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on oral language.

### AND

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

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on oral language. No studies showed indeterminate effects and no studies showed statistically significant or substantively important positive effects.

**Mixed effects:** Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive or negative effects on oral language.

### OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive or negative effects on oral language.

(continued)

## Appendix A5.1 *Bright Beginnings* rating for the oral language domain (continued)

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

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

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on oral language.

**AND**

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on oral language.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

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

**Not met.** One study of *Bright Beginnings* showed no statistically significant negative effects on oral language.

**AND**

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on oral language.

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 or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A5.2 *Bright Beginnings* rating for the print knowledge 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 print knowledge, the WWC rated *Bright Beginnings* as having potentially positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *Bright Beginnings* 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 of *Bright Beginnings* showed a substantively important positive effect on print knowledge.

### AND

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

**Met.** No study of *Bright Beginnings* showed a statistically significant or substantively important negative effect on print knowledge.

### 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.** Only one study of *Bright Beginnings* was included in this review.

### AND

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

**Met.** No studies of *Bright Beginnings* showed statistically significant or substantively important negative effects on print knowledge.

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 or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

### Appendix A5.3 *Bright Beginnings* rating for the phonological processing 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 phonological processing, the WWC rated *Bright Beginnings* as having no discernible effects.

#### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important effects, either positive or negative, on phonological processing.

#### 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.** There was only one study of *Bright Beginnings* and it showed no statistically significant or substantively important positive effects on phonological processing.

#### AND

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

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on phonological processing.

**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.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on phonological processing.

#### AND

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

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on phonological processing. No studies showed indeterminate effects and no studies showed statistically significant or substantively important positive effects.

(continued)

### Appendix A5.3 *Bright Beginnings* rating for the phonological processing domain (continued)

**Mixed effects:** Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive or negative effects on phonological processing.

**OR**

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive or negative effects on phonological processing.

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

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

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on phonological processing.

**AND**

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on phonological processing.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

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

**Not met.** One study of *Bright Beginnings* showed no statistically significant negative effects on phonological processing.

**AND**

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on phonological processing.

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 or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A5.4 *Bright Beginnings* rating for the math 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 math, the WWC rated *Bright Beginnings* as having no discernible effects.

### Rating received

**No discernible effects:** No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important effects, either positive or negative, on math.

### 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.** There was only one study of *Bright Beginnings* and it showed no statistically significant or substantively important positive effects on math.

### AND

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

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on math.

**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.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on math.

### AND

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

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on math. No studies showed indeterminate effects and no studies showed statistically significant or substantively important positive effects.

(continued)

## Appendix A5.4 *Bright Beginnings* rating for the math domain (continued)

**Mixed effects:** Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on math.

**OR**

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on math.

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

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

**Not met.** One study of *Bright Beginnings* showed no statistically significant or substantively important negative effects on math.

**AND**

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on math.

**Negative effects:** Strong evidence of a negative effect with no overriding contrary evidence.

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

**Not met.** One study of *Bright Beginnings* showed no statistically significant negative effects on math.

**AND**

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

**Met.** One study of *Bright Beginnings* showed no statistically significant or substantively important positive effects on math.

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 or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A6 Extent of evidence by domain

| Outcome domain            | Number of studies | Sample size          |          | Extent of evidence <sup>1</sup> |
|---------------------------|-------------------|----------------------|----------|---------------------------------|
|                           |                   | Preschool classrooms | Students |                                 |
| Oral language             | 1                 | 14                   | 197      | Small                           |
| Print knowledge           | 1                 | 14                   | 198      | Small                           |
| Phonological processing   | 1                 | 14                   | 198      | Small                           |
| Early reading and writing | 0                 | na                   | na       | na                              |
| Cognition                 | 0                 | na                   | na       | na                              |
| Math                      | 1                 | 14                   | 198      | Small                           |

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.