

Appendix

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

Characteristic	Description
Study citation	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.
Participants	The study, conducted during the 2003/04 and 2004/05 school years, included three intervention groups: <i>Bright Beginnings</i> , <i>Creative Curriculum</i> [®] , 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> [®] , 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> [®] . 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> [®] , 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> [®] , 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.
Setting	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.
Intervention	<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> [®] and <i>Creative Curriculum</i> [®] 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.
Comparison	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.
Primary outcomes and measurement	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.
Staff/teacher training	<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
Peabody Picture Vocabulary Test–3rd Edition (PPVT-III)	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).
Test of Language Development-Primary III (TOLD-P:3) Grammatical Understanding subtest	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
Test of Early Reading Ability III (TERA-3)	A standardized measure of children’s developing reading skills with three subtests: alphabet, conventions, and meaning (as cited in PCER Consortium, 2008). ¹
Woodcock-Johnson III (WJ-III) Letter-Word Identification subtest	A standardized measure of identification of letters and reading of words (as cited in PCER Consortium, 2008).
Woodcock-Johnson III (WJ-III) Spelling subtest	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
Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP), Elision subtest	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
Woodcock-Johnson III (WJ-III) Applied Problems subtest	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).
Child Math Assessment-Abbreviated (CMA-A) Composite Score	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).
Building Blocks, Shape Composition task	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 domain¹

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (Bright Beginnings– comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
		Bright Beginnings group ³	Comparison group					
PCER Consortium, 2008 (meets standards with reservations)⁸								
PPVT-III	Preschoolers	14/195	96.31 (14.71)	93.93 (15.37)	2.38	0.13	ns	+5
TOLD-P:3 Grammatic Understanding subtest	Preschoolers	14/197	9.60 (2.95)	9.11 (2.73)	0.49	0.09	ns	+4
Domain average for oral language⁹						0.11	na	+4

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 domain¹

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (Bright Beginnings– comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
		Bright Beginnings group ³	Comparison group					
PCER Consortium, 2008 (meets standards with reservations)⁸								
TERA-3	Preschoolers	14/198	91.41 (15.91)	87.98 (14.71)	3.43	0.39	ns	+15
WJ-III Letter-Word Identification subtest	Preschoolers	14/198	106.06 (14.97)	97.21 (13.03)	8.85	0.35	ns	+14
WJ-III Spelling subtest	Preschoolers	14/198	95.75 (12.46)	90.94 (12.98)	4.81	0.18	ns	+7
Domain average for print knowledge⁹						0.31	na	+12

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 domain¹

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (<i>Bright Beginnings</i> –comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			<i>Bright Beginnings</i> group ³	Comparison group				
PCER Consortium, 2008 (meets standards with reservations)⁸								
Pre-CTOPPP Elision subtest	Preschoolers	14/198	10.02 (4.50)	10.38 (4.78)	–0.36	–0.07	ns	–3
Domain average for phonological processing⁹						–0.07	na	–3

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 domain¹

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (Bright Beginnings– comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
		Bright Beginnings group ³	Comparison group					
PCER Consortium, 2008 (meets standards with reservations)⁸								
WJ-III Applied Problems subtest	Preschoolers	14/198	100.69 (14.68)	96.48 (16.69)	4.21	0.16	ns	+6
CMA-A Composite	Preschoolers	14/198	0.57 (0.25)	0.53 (0.27)	0.04	0.14	ns	+6
Shape Composition	Preschoolers	14/198	1.82 (0.93)	1.85 (0.91)	–0.03	–0.03	ns	–1
Domain average for math⁹						0.09	na	+4

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 domain¹

Outcome measure	Study sample	Sample size ³ (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁵ (<i>Bright Beginnings</i> – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Bright Beginnings</i> group ⁴	Comparison group				
PCER Consortium, 2008 (meets standards with reservations)⁹								
PPVT-III	Kindergarteners	nr/203	98.43 (10.83)	97.21 (13.74)	1.22	0.07	ns	+3
TOLD-P:3 Grammatic Understanding subtest	Kindergarteners	nr/203	10.73 (2.91)	9.91 (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*[®], 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 domain¹

Outcome measure	Study sample	Sample size ³ (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁵ (<i>Bright Beginnings</i> – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Bright Beginnings</i> group ⁴	Comparison group				
PCER Consortium, 2008 (meets standards with reservations)⁹								
TERA-3	Kindergarteners	nr/203	93.35 (16.02)	93.99 (17.75)	–0.64	–0.07	ns	–3
WJ-III Letter-Word Identification subtest	Kindergarteners	nr/204	106.12 (10.67)	103.96 (13.41)	2.16	0.09	ns	+4
WJ-III Spelling subtest	Kindergarteners	nr/204	102.12 (12.09)	100.57 (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*[®], 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 domain¹

Outcome measure	Study sample	Sample size ³ (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁵ (<i>Bright Beginnings</i> – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Bright Beginnings</i> group ⁴	Comparison group				
PCER Consortium, 2008 (meets standards with reservations)⁹								
CTOPP Elision subtest	Kindergarteners	nr/203	4.34 (2.76)	4.30 (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*[®], 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 domain¹

Outcome measure	Study sample	Sample size ³ (classrooms/ children)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁵ (<i>Bright Beginnings</i> – comparison)	Effect size ⁶	Statistical significance ⁷ (at $\alpha = 0.05$)	Improvement index ⁸
			<i>Bright Beginnings</i> group ⁴	Comparison group				
PCER Consortium, 2008 (meets standards with reservations)⁹								
WJ-III Applied Problems subtest	Kindergarteners	nr/204	103.21 (12.77)	99.88 (16.18)	3.33	0.13	ns	+5
CMA-A Composite	Kindergarteners	nr/203	0.71 (0.17)	0.69 (0.18)	0.02	0.07	ns	+3
Shape Composition	Kindergarteners	nr/204	2.49 (0.72)	2.36 (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*[®], 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.¹ 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.¹

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.¹ 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.¹ 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 ¹
		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.