Appendix

Appendix A1.1 Study characteristics: Preschool Curriculum Evaluation Research Consortium, 2008 (randomized controlled trial)

Characteristic	Description
Study citation	Preschool Curriculum Evaluation Research (PCER) Consortium. (2008). *Creative Curriculum*: University of North Carolina at Charlotte. In *Effects of preschool curriculum programs on school readiness* (pp. 55–64). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
Participants	This randomized controlled study, conducted during the 2003/04 and 2004/05 school years, included an intervention group that implemented *The Creative Curriculum*® and a control group that continued using the teacher-developed, nonspecific curriculum. Both teachers and children were randomized within the centers. During the pilot year, teachers were blocked on education and teacher certification status, then randomly assigned within blocks to treatment or control groups. Thus, each of the five participating Head Start centers included both *The Creative Curriculum*® and control classrooms. A total of 20 classrooms (10 in North Carolina and 10 in Georgia) were randomly assigned in 2002/03, the pilot year. In the following year, which was the year of the PCER study, two North Carolina classrooms were dropped because they participated in the state's More at Four program, had degreed teachers, and had excessive teacher attrition (10% attrition at the assignment level). Children within a center were sorted into blocks based on gender, disability status, and ethnicity. They were then randomly assigned to *The Creative Curriculum*® or control classrooms. Participants included 18 classrooms (9 *Creative Curriculum*® and 9 control) and 190 children at baseline (95 *Creative Curriculum*® and 95 control). The spring follow-up data collection included 171 children (90 *Creative Curriculum*® and 81 control). Overall attrition at follow-up was 10.0%. At baseline, children in the study were 4.5 years of age on average; 46% were boys; and 85% were African-American, 8% were Hispanic, and 3% were white. Additional findings reflecting students' outcomes at the end of kindergarten can be found in Appendices A4.1–A4.4.
Setting	*The Creative Curriculum*® study was conducted in a total of 18 full-day Head Start preschool classrooms in five Head Start centers (three centers with 8 classrooms in North Carolina and two centers with 10 classrooms in Georgia).
Intervention	*The Creative Curriculum*® is a comprehensive preschool curriculum for children ages 3–5. The curriculum addresses four areas of development: social/emotional, physical, cognitive, and language. *The Creative Curriculum*® requires the physical space of the classroom to be structured into 10 interest areas (blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, and computers). Curriculum content includes literacy, mathematics, science, social studies, the arts, technology, and skills such as observing, exploring, and problem solving. Teachers conduct ongoing child assessments employing a Developmental Checklist. 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 Creative Curriculum*® classrooms was 2.11 on this measure.
Comparison	Business-as-usual using teacher-developed, nonspecific curricula. Control teachers' classrooms were rated with the same fidelity measure used in *The Creative Curriculum*® classrooms, which ranged from 0 to 3. The average score for the control classrooms using this measure was 1.5.
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) Grammatic 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–A2.4.
Staff/teacher training	Teachers in *The Creative Curriculum*® treatment group were in their second year of implementing the program at the time of the evaluation. The research team provided refresher training to the treatment group teachers. Four (North Carolina) or five (Georgia) training periods were provided to teachers. Training was delivered in one half-day or one full-day session (both NC and GA teachers received the same training in total). Training topics included choosing and planning in-depth topics of study; providing materials and interactions for content learning in literacy, math, science, social studies, the arts, and technology; and observation-based assessment of children's learning. Training included a mix of lecture, small group projects, video viewing, and hands-on practical applications. Technical assistance was provided to teachers throughout the school year.

Appendix A1.2 Study characteristics: Preschool Curriculum Evaluation Research Consortium, 2008 (randomized controlled trial)

Characteristic	Description
Study citation	Preschool Curriculum Evaluation Research (PCER) Consortium (2008). *Bright Beginnings* and *Creative Curriculum:* Vanderbilt University. In *Effects of preschool curriculum programs on school readiness* (ch. 2, pp. 41–54). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
Participants	This randomized controlled study, conducted during the 2003/04 and 2004/05 school years, included three intervention groups: *The Creative Curriculum*®, Bright Beginnings, and a control group. Thirty-six full-day prekindergarten classrooms in 28 public schools were recruited and blocked into groups of three by matching them on composite factors for demographic characteristics (urban/rural, percentages of races other than white) and achievement (percentage receiving free lunch and reading, language, mathematics, and science achievement scores). Within each block, one preschool was randomly assigned to *The Creative Curriculum*®, one to Bright Beginnings, and one to the control group. The manuscript notes that the researchers randomly assigned the classrooms to three conditions; however, all classrooms in a preschool were assigned to the same study 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 *The Creative Curriculum*®, Bright Beginnings, and a control group. All 36 classrooms participated in the local investigator's pilot-year study during the first year. Following the pilot year, and prior to starting the national PCER study, 8 of the 21 PCER classrooms dropped out of the study, leaving 4 *Creative Curriculum*®, 5 Bright Beginnings, and 4 control classrooms (attrition of 43%, 29%, and 43% respectively). 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 Bright Beginnings, 3 *Creative Curriculum*®, and 3 control classrooms, restoring the sample of classrooms to 7 in each of the three intervention groups. The study demonstrated the baseline equivalence of the analytic sample of children in the intervention and control groups. At baseline, children in the study averaged 4.5 years of age; 52% were male; and 11% were Hispanic, 80% were white, and 7% were African-American. Child-level attrition was 6.7% overall; 8.6% in *The Creative Curriculum*® classrooms and 5% in the comparison group. The analysis sample included 93 children in 7 *Creative Curriculum*® classrooms and 100 children in 7 control classrooms. Additional findings reflecting students' outcomes at the end of kindergarten can be found in Appendices A4.1–A4.4.
Setting	*The Creative Curriculum*® study was conducted in prekindergarten classes in 14 public schools (7 *Creative Curriculum*® and 7 control) from seven county school districts in Tennessee.
Intervention	*The Creative Curriculum*® is a comprehensive preschool curriculum for children ages 3–5. The curriculum addresses four areas of development: social/emotional, physical, cognitive, and language. *The Creative Curriculum*® requires the physical space of the classroom to be structured into 10 interest areas (blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, and computers). Curriculum content includes literacy, mathematics, science, social studies, the arts, technology, and skills such as observing, exploring, and problem solving. Teachers conduct ongoing child assessments employing a Developmental Checklist. 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 Creative Curriculum*® classrooms was 2.11 on this measure.
Comparison	Business-as-usual using teacher-developed, nonspecific curricula with a focus on basic school readiness. Control teachers' classrooms were rated with the same fidelity measure used in *The Creative Curriculum*® classrooms, which ranged from 0 to 3. The average score for the control classrooms using this measure 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) Grammatic 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–A2.4.
Staff/teacher training	*The Creative Curriculum*® was implemented in treatment schools in fall 2002 (pilot-study year) and in fall 2003 for additional teachers participating in the intervention year. Treatment group teachers received 2.5 full days of training and had access to ongoing curriculum implementation throughout the school 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. No specific additional professional development activities for control group teachers are described.

Appendix A1.3 Study characteristics: Henry, Ponder, Rickman, Mashburn, Henderson, & Gordon, 2004 (quasi-experimental design)

Characteristic	Description
Study citation	Henry, G. T., Ponder, B. D., Rickman, D. K., Mashburn, A. J., Henderson, L. W., & Gordon, C. S. (2004). *An evaluation of the implementation of Georgia's pre-K program: Report of the findings from the Georgia early childhood study (2002–03).* Atlanta, GA: Georgia State University, Andrew Young School of Policy Studies.
Participants	The authors used a probability sample of children who attended prekindergarten in Georgia. To obtain a representative sample of classrooms and children, they used a four-stage sampling approach by (1) sampling counties stratified by the number of 4-year-olds; (2) sampling Georgia pre-K, Head Start, and private preschool sites within selected counties; (3) sampling classes within sites; and (4) selecting children within classes. A total of 135 sites were selected, and 126 agreed to participate. Within selected and participating classrooms, 75% of the families of children selected for the study gave consent for their children to participate. At the end of the preschool year, 482 children had both fall and spring assessments.[1] The average age of children in the sample was 4.5 years; 52% were boys; and 33% were African-American, 4% were Hispanic, and 58% were white. The analysis sample included 120 children in 18 *Creative Curriculum®* classrooms and 362 children in 51 control classrooms.
Setting	This study took place in a total of 69 full-day state preschool, Head Start, and private preschool classrooms in 69 centers or schools across Georgia.
Intervention	*The Creative Curriculum®* is a comprehensive preschool curriculum for children ages 3–5. The curriculum addresses four areas of development: social/emotional, physical, cognitive, and language. *The Creative Curriculum®* requires the physical space of the classroom to be structured into 10 interest areas (blocks, dramatic play, toys and games, art, library, discovery, sand and water, music and movement, cooking, and computers). Curriculum content includes literacy, mathematics, science, social studies, the arts, technology, and skills such as observing, exploring, and problem solving. Teachers conduct ongoing child assessments employing a Developmental Checklist. Fidelity to the curriculum was not measured in this study.
Comparison	Classrooms using High/Scope, High Reach, and a variety of other curricula were used as the comparison group. Fidelity to either *The Creative Curriculum®* or the other curricula was not measured in this study.
Primary outcomes and measurement	The outcome domains assessed at the end of preschool were children's oral language, print knowledge, and math. Oral language was assessed with the Peabody Picture Vocabulary Test–III (PPVT-III) and the Oral and Written Language Scale (OWLS) Oral Expression subtest. Print knowledge was assessed with the Woodcock-Johnson–III (WJ-III) Letter-Word Identification subtest. Math was assessed with the WJ-III Applied Problems subtest. For a more detailed description of these outcome measures, see Appendices A2.1–A2.4.
Staff/teacher training	Teachers were already using particular curricula when the study began, so they had already been trained to use them. The study provides no information on the amount of training or technical assistance teachers received in implementing particular curricula.

1. This sample size was obtained through an author query and includes children from the Georgia prekindergarten program, Head Start, and private preschools (for a discussion of this sample see Henry et al., 2003). This sample differs from that included in Henry et al. (2004), which focused solely on children from the Georgia prekindergarten program (sample size of 326 children).

Appendix A2.1 Outcome measures for the oral language domain

Outcome measure	Description
Peabody Picture Vocabulary Test–III (PPVT-III)	A standardized measure of children's receptive vocabulary in which children show understanding of a spoken word by pointing to a picture that best represents the meaning (as cited in PCER Consortium, 2008, and Henry et al., 2004).
Test of Language Development–Primary: III (TOLD-P:3) Grammatic 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).
Oral and Written Language Scales (OWLS) Oral Expression subscale	A standardized measure of children's expressive language that requires the child to answer questions and finish sentences (as cited in Henry et al., 2004).

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, and Henry et al., 2004).
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 as if 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 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, and Henry et al., 2004).
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 domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group[4]	WWC calculations: Mean difference[5] (*The Creative Curriculum®*–comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
PCER Consortium [Chapter 3], 2008[9]								
PPVT-III	Preschoolers	18/165	86.64 (14.43)	85.42 (13.40)	1.22	0.08	ns	+3
TOLD-P:3 Grammatic Understanding subtest	Preschoolers	18/169	7.70 (2.58)	8.44 (2.68)	–0.74	–0.16	ns	–6
Average for oral language (PCER Consortium [Chapter 3], 2008)[10]						**–0.04**	**na**	**–2**
PCER Consortium [Chapter 2], 2008[9]								
PPVT-III	Preschoolers	14/192	98.06 (13.27)	93.93 (15.37)	4.13	0.23	ns	+9
TOLD-P:3 Grammatic Understanding subtest	Preschoolers	14/193	9.44 (2.55)	9.11 (2.73)	0.33	0.07	ns	+3
Average for oral language (PCER Consortium [Chapter 2], 2008)[10]						**0.15**	**na**	**+6**
Henry, Ponder, Rickman, Mashburn, Henderson, and Gordon, 2004[9]								
PPVT-III	Preschoolers	69/482	97.67 (14.17)	95.95 (13.78)	1.72	0.12	ns	+5
OWLS Oral Expression subtest	Preschoolers	69/482	94.11 (13.96)	92.83 (13.57)	1.28	0.09	ns	+4
Average for oral language (Henry et al., 2004)[10]						**0.11**	**na**	**+4**
Domain average for oral language across all studies[9]						**0.07**	**na**	**+3**

ns = not statistically significant
na = not applicable
PPVT-III = Peabody Picture Vocabulary Test–III
TOLD-P:3 = Test of Language Development–Primary-III
OWLS = Oral and Written Language Scales

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language domain. Kindergarten 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.

(continued)

Appendix A3.1 Summary of study findings included in the rating for the oral language domain[1] *(continued)*

3. In PCER Consortium (2008), the treatment group mean equals the unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. In Henry et al. (2004), the comparison group mean equals the mean across all three alternative curriculum groups (High/Scope, High Reach, and others).
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 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 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 HLM and no impacts were statistically significant. In the case of Henry et al. (2004), the WWC corrected the mean comparisons for clustering.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group[4]	WWC calculations: Mean difference[5] (*The Creative Curriculum®*–comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
PCER Consortium [Chapter 3], 2008[9]								
TERA-3	Preschoolers	18/170	85.81 (13.97)	86.39 (13.88)	–0.58	–0.08	ns	–3
WJ-III Letter-Word Identification subtest	Preschoolers	18/169	99.87 (12.11)	101.74 (13.08)	–1.87	–0.08	ns	–3
WJ-III Spelling subtest	Preschoolers	18/169	87.39 (14.38)	91.95 (13.23)	–4.56	–0.18	ns	–7
Domain average for print knowledge (PCER Consortium [Chapter 3], 2008)[10]						**–0.11**	**na**	**-4**
PCER Consortium [Chapter 2], 2008[9]								
TERA-3	Preschoolers	14/193	88.12 (12.06)	87.98 (14.71)	0.14	0.02	ns	+1
WJ-III Letter-Word Identification subtest	Preschoolers	14/193	100.80 (11.06)	97.21 (13.03)	3.59	0.16	ns	+6
WJ-III Spelling subtest	Preschoolers	14/193	95.39 (11.07)	90.94 (12.98)	4.45	0.19	ns	+8
Domain average for print knowledge (PCER Consortium [Chapter 2], 2008)[10]						**0.12**	**na**	**+5**
Henry, Ponder, Rickman, Mashburn, Henderson, and Gordon, 2004[9]								
WJ-III Letter-Word Identification subtest	Preschoolers	69/482	104.95 (14.25)	102.46 (12.85)	2.49	0.19	ns	+7
Domain average for print knowledge (Henry et al., 2004)[11]						**0.19**	**na**	**+7**
Domain average for print knowledge across all studies[10]						**0.07**	**na**	**+3**

ns = not statistically significant
na = not applicable
TERA-3 = Test of Early Reading Ability–III
WJ-III = Woodcock-Johnson–III
WJ-R = Woodcock-Johnson–Revised

(continued)

Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain[1] *(continued)*

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge domain. Kindergarten 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 unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. In Henry et al. (2004), the comparison group mean equals the mean across all three alternative curriculum groups (High/Scope, High Reach, and others).
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 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 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 HLM and no impacts were statistically significant. In the case of Henry et al. (2004), the WWC corrected the mean comparisons for clustering.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.
11. 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[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] *(The Creative Curriculum®–comparison)*	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium [Chapter 3], 2008								
Pre-CTOPPP Elision subtest	Preschoolers	18/171	8.38 (4.08)	8.19 (4.03)	0.19	0.02	ns	+1
Domain average for phonological processing (PCER Consortium [Chapter 3], 2008)[8]						**0.02**	**na**	**+1**
PCER Consortium [Chapter 2], 2008[9]								
Pre-CTOPPP Elision subtest	Preschoolers	14/193	10.34 (3.60)	10.38 (4.78)	–0.04	–0.10	ns	–4
Domain average for phonological processing (PCER Consortium [Chapter 2], 2008)						**–0.10**	**na**	**–4**
Domain average for phonological processing across all studies[8]						**–0.04**	**na**	**–2**

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. Kindergarten 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 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. 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.
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 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 HLM and no impacts were statistically significant.

Appendix A3.4 Summary of study findings included in the rating for the math domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group[4]	WWC calculations: Mean difference[5] (*The Creative Curriculum®*–comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
PCER Consortium [Chapter 3], 2008[9]								
WJ-III Applied Problems subtest	Preschoolers	18/169	94.07 (12.26)	89.45 (13.75)	4.62	0.20	ns	+8
CMA-A composite	Preschoolers	18/170	0.42 (0.27)	0.44 (0.29)	–0.02	–0.10	ns	–4
Shape Composition	Preschoolers	18/169	1.42 (0.89)	1.25 (0.83)	0.17	0.19	ns	+8
Domain average for math (PCER Consortium [Chapter 3], 2008)[10]						**0.10**	**na**	**+4**
PCER Consortium [Chapter 2], 2008[9]								
WJ-III Applied Problems subtest	Preschoolers	14/193	100.45 (12.03)	96.48 (16.69)	3.97	0.17	ns	+7
CMA-A Composite	Preschoolers	14/193	0.55 (0.23)	0.53 (0.27)	0.02	0.10	ns	+4
Shape Composition	Preschoolers	14/193	1.74 (0.95)	1.85 (0.91)	–0.11	–0.12	ns	–5
Domain average for math (PCER Consortium [Chapter 2], 2008)[10]						**0.05**	**na**	**+2**
Henry, Ponder, Rickman, Mashburn, Henderson, and Gordon, 2004[9]								
WJ-III Applied Problems subtest	Preschoolers	69/482	99.48 (14.73)	96.94 (12.68)	2.54	0.19	ns	+8
Domain average for math (Henry et al., 2004)[11]						**0.19**	**na**	**+8**
Domain average for math across all studies[10]						**0.11**	**na**	**+4**

ns = not statistically significant
na = not applicable
WJ-III = Woodcock-Johnson–III
CMA-A = Child Math Assessment–Abbreviated

(continued)

Appendix A3.4 Summary of study findings included in the rating for the math domain[1] *(continued)*

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math domain. Kindergarten 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 unadjusted control group mean and the covariate-adjusted mean difference. Standard deviations are unadjusted.
4. In Henry et al. (2004), the comparison group mean equals the mean across all three alternative curriculum groups (High/Scope, High Reach, and others).
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 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 HLM and no impacts were statistically significant. In the case of Henry et al. (2004), the WWC corrected the mean comparisons for clustering.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.
11. 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[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum*®[3] group	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*The Creative Curriculum*®–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium [Chapter 2], 2008[8]								
PPVT-III	Kindergarten	nr/199	99.29 (10.82)	97.21 (13.74)	2.08	0.12	ns	+5
TOLD-P:3 Grammatic Understanding subtest	Kindergarten	nr/199	10.45 (2.24)	9.91 (2.93)	0.54	0.11	ns	+4
PCER Consortium [Chapter 3], 2008[8]								
PPVT-III	Kindergarten	nr/160	90.44 (11.94)	88.09 (13.60)	2.35	0.15	ns	–7
TOLD-P:3 Grammatic Understanding subtest	Kindergarten	nr/161	8.81 (2.67)	9.63 (2.88)	–0.82	–0.17	ns	–7

ns = not statistically significant
nr = not reported
PPVT-III = Peabody Picture Vocabulary Test–III
TOLD-P:3 = Test of Language Development Primary–III

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. In PCER Consortium (2008), the treatment group mean equals 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 (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 statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A4.2 Summary of follow-up findings for the print knowledge domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*The Creative Curriculum®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium [Chapter 2], 2008[8]								
TERA-3	Kindergarten	nr/199	94.73 (15.33)	93.99 (17.75)	0.74	0.10	ns	+4
WJ-III Letter-Word Identification subtest	Kindergarten	nr/200	112.35 (11.92)	103.96 (13.41)	8.39	0.38	ns	+15
WJ-III Spelling subtest	Kindergarten	nr/200	106.55 (11.62)	100.57 (15.15)	5.98	0.25	ns	+10
PCER Consortium [Chapter 3], 2008[8]								
TERA-3	Kindergarten	nr/161	92.21 (17.62)	92.51 (15.30)	–0.30	–0.04	ns	–2
WJ-III Letter-Word Identification subtest	Kindergarten	nr/161	105.21 (15.25)	105.28 (12.95)	–0.07	0.00	ns	0
WJ-III Spelling subtest	Kindergarten	nr/161	100.99 (17.90)	102.28 (16.25)	–1.29	–0.05	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. In PCER Consortium (2008), the treatment group mean equals 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 (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 statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A4.3 Summary of follow-up findings for the phonological processing domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*The Creative Curriculum®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium [Chapter 2], 2008[8]								
CTOPP Elision subtest	Kindergarten	nr/199	4.50 (3.41)	4.30 (3.27)	0.20	0.06	ns	+2
PCER Consortium [Chapter 3], 2008[8]								
CTOPP Elision subtest	Kindergarten	nr/161	2.68 (3.03)	2.51 (2.83)	0.17	0.06	ns	+2

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. In PCER Consortium (2008), the treatment group mean equals 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 ANCOVA).
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 (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 statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A4.4 Summary of follow-up findings for the math domain[1]

Outcome measure	Study sample	Sample size (classrooms/ children)	Authors' findings from the study: Mean outcome (standard deviation)[2] — *The Creative Curriculum®* group[3]	Authors' findings from the study: Mean outcome (standard deviation)[2] — Comparison group	WWC calculations: Mean difference[4] (*The Creative Curriculum®*–comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PCER Consortium [Chapter 2], 2008[8]								
WJ- III Applied Problems subtest	Kindergarten	nr/200	103.79 (9.60)	99.88 (16.18)	3.91	0.17	ns	+7
CMA-A Composite	Kindergarten	nr/199	0.70 (0.17)	0.69 (0.18)	0.01	0.05	ns	+2
Shape Composition	Kindergarten	nr/200	2.36 (0.70)	2.36 (0.89)	–0.00	0.00	ns	0
PCER Consortium [Chapter 3], 2008[8]								
WJ- III Applied Problems subtest	Kindergarten	nr/161	95.58 (14.29)	93.46 (13.21)	2.12	0.09	ns	+4
CMA-A Composite	Kindergarten	nr/161	0.66 (0.18)	0.63 (0.20)	0.03	0.14	ns	+6
Shape Composition	Kindergarten	nr/161	2.05 (0.80)	2.05 (0.92)	–0.00	–0.01	ns	0

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. In PCER Consortium (2008), the treatment group mean equals 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 (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 statistical significance, see WWC Procedures and Standards Handbook, Appendix C. In the case of PCER Consortium (2008), no corrections were needed because the analysis corrected for clustering by using HLM and no impacts were statistically significant.

Appendix A5.1 *The Creative Curriculum*® rating for the oral language domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of oral knowledge, the WWC rated *The Creative Curriculum*® 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. None of the three studies of *The Creative Curriculum*® that measured oral language showed a statistically significant or substantively important effect, either positive or negative.

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. None of the three studies of *The Creative Curriculum*® that measured oral language showed a statistically significant positive effect.

AND

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

  Met. None of the three studies of *The Creative Curriculum*® that measured oral language showed a statistically significant or substantively important negative effect.

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. None of the three studies of *The Creative Curriculum*® that measured oral language showed a statistically significant or substantively important positive effect.

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. None of the three studies of *The Creative Curriculum*® that measured oral language showed a statistically significant or substantively important negative effect. None of the studies showed a statistically significant or substantively important positive effect 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.

(continued)

Appendix A5.1 *The Creative Curriculum®* rating for the oral language 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. None of the three studies of *The Creative Curriculum®* that measured oral language showed a statistically significant or substantively important positive effect.

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. None of the three studies of *The Creative Curriculum®* that measured oral language showed a statistically significant or substantively important positive or negative effect.

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. None of the three studies of *The Creative Curriculum®* that measured oral language showed a statistically significant or substantively important negative effect.

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. None of the three studies of *The Creative Curriculum®* that measured oral language showed a statistically significant or substantively important positive effect. None of the three studies showed a statistically significant or substantively important negative effect 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. None of the three studies of *The Creative Curriculum®* that measured oral language showed a statistically significant negative effect.

AND

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

  Met. None of the three studies of *The Creative Curriculum®* that measured oral language showed a statistically significant or substantively important positive effect.

Appendix A5.2 *The Creative Curriculum*® rating for the print knowledge domain

The WWC rates an intervention's effects in 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 *The Creative Curriculum*® 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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important effect, either positive or negative.

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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant positive effect.

AND

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

  Met. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important negative effect.

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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important positive effect.

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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important negative effect. None of the studies showed a statistically significant or substantively important positive effect 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.

(continued)

Appendix A5.2 *The Creative Curriculum*® rating for the print knowledge 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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important positive effect.

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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important positive or negative effect.

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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important negative effect.

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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important positive effect. None of the studies showed a statistically significant or substantively important negative effect on print knowledge.

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. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant negative effect.

AND

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

  Met. None of the three studies of *The Creative Curriculum*® that measured print knowledge showed a statistically significant or substantively important positive effect.

Appendix A5.3 *The Creative Curriculum*® rating for the phonological processing domain

The WWC rates an intervention's effects in 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 *The Creative Curriculum*® 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important effect, either positive or negative. No other studies measured 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant positive effect. No other studies measured phonololgical processing.

AND

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

  Met. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important negative effect. No other studies measured phonololgical 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important positive effect. No other studies measured phonololgical 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.

  Not met. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important negative effect. No other studies measured 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.

(continued)

Appendix A5.3 *The Creative Curriculum*® 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important positive or negative effect. No other studies measured 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important positive or negative effect. No other studies measured 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important negative effect. No other studies measured 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important positive effect. No other studies measured 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. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant negative effect. No other studies measured phonological processing.

AND

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

  Met. Neither of the two studies of *The Creative Curriculum*® that measured phonological processing showed a statistically significant or substantively important positive effect. No other studies measured phonological processing.

Appendix A5.4 *The Creative Curriculum*® rating for the math domain

The WWC rates an intervention's effects in 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 *The Creative Curriculum*® 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. None of the three studies of *The Creative Curriculum*® that measured math showed a statistically significant or substantively important effect, either positive or negative.

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. None of the three studies of *The Creative Curriculum*® that measured math showed a statistically significant positive effect.

AND

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

  Met. None of the three studies of *The Creative Curriculum*® that measured math showed statistically significant or substantively important negative effects.

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. None of the three studies of *The Creative Curriculum*® that measured math showed a statistically significant or substantively important positive effect.

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. None of the three studies of *The Creative Curriculum*® that measured math showed a statistically significant or substantively important negative effect.

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.

(continued)

Appendix A5.4 *The Creative Curriculum®* 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. None of the three studies of *The Creative Curriculum®* that measured math showed a statistically significant or substantively important positive or negative effect.

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. None of the three studies of *The Creative Curriculum®* that measured math showed a statistically significant or substantively important positive or negative effect.

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. None of the three studies of *The Creative Curriculum®* that measured math showed a statistically significant or substantively important negative effect.

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. None of the three studies of *The Creative Curriculum®* that measured math showed a statistically significant or substantively important positive effect.

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. None of the three studies of *The Creative Curriculum®* that measured math showed a statistically significant negative effect.

AND

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

  Met. None of the three studies of *The Creative Curriculum®* that measured math showed a statistically significant or substantively important positive effect.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[2]
		Schools	Students[1]	
Oral language	3	101	839	Medium to large
Print knowledge	3	101	844	Medium to large
Phonological processing	2	32	364	Medium to large
Early reading or writing	0	0	0	na
Cognition	0	0	0	na
Math	3	101	844	Medium to large

na = not applicable/not studied

1. The sample size of students shown in this table is based on the smallest number of children with valid posttest measurements within a domain. Posttest responses for the PCER [Chapter 2] (2008) study ranged from 192 to 193. Posttest responses for the PCER [Chapter 3] (2008) study ranged from 165 to 171. Posttest responses for the Henry et al. (2004) study totaled 482 children.
2. 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.