Appendix

Appendix A1 Study characteristics: Cole, Dale, Mills, & Jenkins, 1993 (randomized controlled trial with attrition problems)

Characteristic	Description
Study citation	Cole, K. N., Dale, P. S., Mills, P. E., & Jenkins, J. R. (1993). Interaction between early intervention curricula and student characteristics. *Exceptional Children, 60*(1), 17–28.
Participants	This study included 164 preschool and kindergarten children from a larger study.[1] They had no previous preschool experience. Eighty-one children were in the intervention group and 83 children were in the comparison group.[2] Individual children were first randomly assigned to a *Direct Instruction* intervention group or a Mediated Learning comparison group; then children in each condition were randomly assigned to classrooms (six preschool classes and two kindergarten classes). All the children had disabilities, including language delay (80%), cognitive delay (50%), fine motor delay (60%), gross motor delay (60%), and socio-emotional delay (60%). The mean age of the children in the sample was 4.75 years, their mean IQ was 76.03, and 32% were female. Sixty-two percent of the children were European-American, 29% were African-American, and 10% were Hispanic, Pacific Islander, Asian, Native American, or another race/ethnicity.
Setting	The study took place at the Experimental Education Unit of the University of Washington's Child Development and Mental Retardation Center. Preschool classes were two hours a day, five days a week for 180 days, and kindergarten classes were five and a half hours a day, five days a week for 180 days.
Intervention	The WWC designated the *Direct Instruction* condition as the intervention for this review. The intervention condition included the *Direct Instruction* programs *DISTAR Arithmetic, DISTAR Reading,* and *DISTAR Language,* which are academically-oriented programs characterized by instruction that is fast-paced, teacher-directed, prescribed, and explicit, with all children receiving instruction on a pre-specified sequence of activities at the same time. Although there is information on the length of each school program, there is no information about the duration or frequency of *Direct Instruction.*
Comparison	The WWC designated the Mediated Learning condition as the comparison condition for this review. The Mediated Learning curriculum included units that were two to three weeks long and focused on topics such as identifying patterns, making comparisons, identifying feelings, and planning ahead. Mediated Learning is a cognitively-oriented, non-academic program that emphasizes processes related to input, elaboration, and output. As opposed to the direct approach used by teachers of *Direct Instruction,* teachers using Mediated Learning scaffold children's learning around cognitive processes such as classification and sequencing. Although there is information on the length of each school program, there is no information about the duration or frequency of Mediated Learning.
Primary outcomes and measurement[3]	The primary outcome domains assessed were children's oral language, print knowledge, cognition, and math. Oral language was assessed with five standardized measures and one non-standardized measure. The standardized measures include: the McCarthy Scales of Children's Abilities (MSCA) Verbal Scale; the Peabody Picture Vocabulary Test-Revised (PPVT-R);[4] the Test of Early Language Development (TELD);[4] the Preschool Language Assessment Inventory (PLAI); and the Basic Language Concepts Test-Errors (BLCT; criterion referenced measure). The non-standardized measure is the Mean Length of Utterance (MLU) derived from language samples. Print knowledge was assessed with one standardized measure, the Test of Early Reading Ability (TERA).[4] Cognition was assessed with three standardized measures, the MSCA Composite Scale (General Cognitive Index), the MSCA Perceptual Scale, and the MSCA Memory Scale. Math was assessed with one standardized measure, the MSCA Quantitative Scale (see Appendices A2.1–2.4 for detailed descriptions of outcome measures).[5] Data analyses were based on the children's scores after the first year of participation in the four-year study period.

(continued)

Appendix A1 Study characteristics: Cole, Dale, Mills, & Jenkins, 1993 (randomized controlled trial with attrition problems) *(continued)*

Characteristic	Description
Teacher training	Each class was staffed by a head teacher and an assistant teacher. Other staff included related services personnel and practicum students. All head teachers had Master's degrees in special education. Of the eight *Direct Instruction* head teachers, three received their degrees from a University of Oregon program that emphasized *Direct Instruction,* four received in-service training from that program, and one was trained on-site. Two of the Mediated Learning teachers were trained in Mediated Learning at Vanderbilt University and two additional teachers received training from Dr. Feuerstein (an Israeli psychologist whose work is the basis for Mediated Learning) and consultation on curriculum and procedures. The five additional teachers in the Mediated Learning program received ongoing professional development from the teachers who were trained at Vanderbilt.

1. This sample of children without previous preschool experience is a sub-sample drawn from a related study (Dale & Cole, 1988) that did not meet WWC evidence screens because effect sizes could not be calculated. In consultation with the study authors, the sub-sample used in Cole et al. (1993) was determined to be an acceptable study with which to determine the effectiveness of *Direct Instruction* in comparison with Mediated Learning. The ECE topic includes studies with preschool and kindergarten children when the majority (60% or more) of children in the sample are in preschool. In this study, there were six preschool classes and two kindergarten classes per year, and the mean age of the children was under five, indicating that the study meets this criterion for inclusion.
2. The study was classified as "meets evidence standards with reservations" due to severe overall attrition. Based on the number of classes and children in the original study, the sample size at assignment was 368 children with disabilities [Cole et al. (1993) stated that the full sample included just 206 children]. However, the analysis sample was 164 children. Based upon the inconsistency between the figures at assignment, the study was downgraded for severe overall attrition.
3. The BLCT, MLU, and PLAI are not included in the overall rating of effectiveness because they were not administered to the full sample of children in the study and cannot be used to determine the overall effectiveness of the intervention. Additionally, they are not included in the later appendices for this study because the WWC was unable to obtain the analysis sample sizes for each of these measures. The WWC did not include the MSCA Composite Scale in the WWC intervention report because the WWC includes the individual measures used to develop the composite. There was one additional outcome measure, the MSCA Motor Scale; however, this measure is not included in this report because it is not relevant to the WWC review.
4. The authors reported raw scores and standardized scores for the PPVT-R, the TELD, and the TERA. The WWC report includes the standardized scores only because they are adjusted for age.
5. For further details about the outcomes included in the early childhood education topic review, please see the Early Childhood Education Protocol.

Appendix A2.1 Outcome measures in the oral language domain

Characteristic	Description
McCarthy Scales of Children's Abilities (MSCA) Verbal subtest	A subtest from a standardized measure that assesses children's receptive and expressive language (as cited in Cole et al., 1993, and Cole et al., 1991).
Peabody Picture Vocabulary Test-Revised (PPVT-R)	A standardized measure of children's receptive vocabulary that requires them to identify pictures that correspond to spoken words (as cited in Cole et al., 1993, and Cole et al., 1991).
Test of Early Language Development (TELD)	A standardized measure of children's semantic and syntactic skills in receptive and expressive language (as cited in Cole et al., 1993, and Cole et al., 1991).
Basic Language Concepts Test (BLCT)	A criterion-referenced measure of children's receptive and expressive language that focuses on common nouns, adjectives, plural, and tense. This test was developed as part of the *Direct Instruction* program and measures the number of errors (as cited in Cole et al., 1993, and Cole et al., 1991) and a lower score reflects a better outcome.
Preschool Language Assessment Inventory (PLAI)	A standardized measure that assesses children's ability to respond to increasingly abstract language (as cited in Cole et al., 1993, and Cole et al., 1991).
Mean Length of Utterance (MLU)	A non-standardized measure of children's expressive language based on 20-minute language samples (as cited in Cole et al., 1993, and Cole et al., 1991).

Appendix A2.2 Outcome measure in the print knowledge domain

Characteristic	Description
Test of Early Reading Ability (TERA)[1]	A standardized measure of young children's early reading skills that captures the following constructs: awareness of print in environmental contexts, vocabulary, listening, comprehension, knowledge of alphabet, and concepts about printed language (as cited in Cole et al., 1993).

1. By name, this measure sounds like it should be captured under the early reading/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 in the cognition domain

Characteristic	Description
MSCA General Cognitive Index[1]	A standardized measure of children's general intellectual functioning that is based on their scores of the Perceptual, Verbal, and Quantitative measures (as cited in Cole et al., 1993 and Cole et al., 1991).
MSCA Perceptual Scale subtest	A subtest from a standardized measure that assesses children's ability to conceptualize and reason (as cited in Cole et al., 1993, and Cole et al., 1991).
MSCA Memory Scale subtest	A subtest from a standardized measure that assesses children's short-term recall of words, pictures, numbers, and sound sequences (as cited in Cole et al., 1993, and Cole et al., 1991).

1. The WWC did not include the MSCA General Cognitive Index in the overall rating of effectiveness. However, the results for this measure are reported in Appendix A4.2.

Appendix A2.4 Outcome measure in the math domain

Characteristic	Description
MSCA Quantitative subtest	A subtest from a standardized measure that assesses children's mathematical ability (as cited in Cole et al., 1993).

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

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Direct Instruction* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Mediated Learning group[3]	WWC calculations: Mean difference[4] (*Direct Instruction* – Mediated Learning)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Cole et al., 1993 (randomized controlled trial with attrition problems)[8]								
MSCA Verbal	3–7 year olds	164	39.20[9] (10.80)	37.70 (9.80)	1.50	0.14	ns	+6
PPVT-R Scale Score	3–7 year olds	164	81.70 (16.20)	84.60 (14.70)	–2.90	–0.19	ns	–7
TELD Language Quotient	3–7 year olds	164	89.50 (13.70)	88.20 (11.80)	1.30	0.10	ns	+4
Domain average[10] for oral language						0.02	ns	+1

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Composite scores and subgroup findings from the same study are not included in these ratings, but are reported in Appendices A4.1 and 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. The *Direct Instruction* group mean equals the Mediated Learning group mean plus the mean difference.
4. Positive differences and effect sizes favor the *Direct Instruction* group; negative differences and effect sizes favor the Mediated Learning group. The mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the *Direct Instruction* group had lower pretest scores than the Mediated Learning group and underestimate the intervention's effect when the *Direct Instruction* group had higher pretest scores than the Mediated Learning group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Direct Instruction* condition versus the percentile rank of the average student in the Mediated Learning condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Direct Instruction* group.
8. The level of statistical significance was reported by the study authors or, where 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Cole et al. (1993), no corrections for clustering or multiple comparisons were needed.
9. For all means and standard deviations in Appendix A3.1 the study authors reported data to a single decimal place. For formatting purposes, the WWC added zero in the second decimal place.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Direct Instruction* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Mediated Learning group[3]	WWC calculations: Mean difference[4] (*Direct Instruction* – Mediated Learning)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Cole et al., 1993 (randomized controlled trial with attrition problems)[8]								
TERA Reading Quotient	3–7 year olds	164	78.00[9] (13.80)	79.10 (11.10)	–1.10	–0.09	ns	–3
Domain average[10] for print knowledge						–0.09	ns	–3

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Composite scores and subgroup findings from the same study are not included in these ratings, but are reported in Appendices A4.1 and 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. The *Direct Instruction* group mean equals the Mediated Learning group mean plus the mean difference.
4. Positive differences and effect sizes favor the *Direct Instruction* group; negative differences and effect sizes favor the Mediated Learning group. The mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the *Direct Instruction* group had lower pretest scores than the Mediated Learning group and underestimate the intervention's effect when the *Direct Instruction* group had higher pretest scores than the Mediated Learning group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Direct Instruction* condition versus the percentile rank of the average student in the Mediated Learning condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Direct Instruction* group.
8. The level of statistical significance was reported by the study authors or, where 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Cole et al. (1993), no corrections for clustering or multiple comparisons were needed.
9. For all means and standard deviations in Appendix A3.2 the study authors reported data to a single decimal place. For formatting purposes, the WWC added zero in the second decimal place.
10. 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 cognition domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Direct Instruction* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Mediated Learning group[3]	WWC calculations: Mean difference[4] (*Direct Instruction* – Mediated Learning)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Cole et al., 1993 (randomized controlled trial with attrition problems)[8]								
MSCA Perceptual	3–7 year olds	164	38.20[9] (10.80)	38.50 (10.80)	–0.30	–0.03	ns	–1
MSCA Memory	3–7 year olds	164	36.90 (10.70)	37.30 (10.70)	–0.40	–0.04	ns	–1
Domain average[10] for cognition						–0.03	ns	–1

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Composite scores and subgroup findings from the same study are not included in these ratings, but are reported in Appendices A4.1 and 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. The *Direct Instruction* group mean equals the Mediated Learning group mean plus the mean difference.
4. Positive differences and effect sizes favor the *Direct Instruction* group; negative differences and effect sizes favor the Mediated Learning group. The mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the *Direct Instruction* group had lower pretest scores than the Mediated Learning group and underestimate the intervention's effect when the *Direct Instruction* group had higher pretest scores than the Mediated Learning group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Direct Instruction* condition versus the percentile rank of the average student in the Mediated Learning condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Direct Instruction* group.
8. The level of statistical significance was reported by the study authors or, where 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Cole et al. (1993), no corrections for clustering or multiple comparisons were needed.
9. For all means and standard deviations in Appendix A3.3 the study authors reported data to a single decimal place. For formatting purposes, the WWC added zero in the second decimal place.
10. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Direct Instruction* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Mediated Learning group[3]	WWC calculations: Mean difference[4] (*Direct Instruction* – Mediated Learning)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Cole et al., 1993 (randomized controlled trial with attrition problems)[8]								
MSCA Quantitative	3–7 year olds	164	38.70[9] (9.70)	39.60 (9.00)	1.80	0.19	ns	+8
Domain average[10] for math						0.19	ns	+8

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Composite scores and subgroup findings from the same study are not included in these ratings, but are reported in Appendices A4.1 and 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. The *Direct Instruction* group mean equals the Mediated Learning group mean plus the mean difference.
4. Positive differences and effect sizes favor the *Direct Instruction* group; negative differences and effect sizes favor the Mediated Learning group. The mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the *Direct Instruction* group had lower pretest scores than the Mediated Learning group and underestimate the intervention's effect when the *Direct Instruction* group had higher pretest scores than the Mediated Learning group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Direct Instruction* condition versus the percentile rank of the average student in the Mediated Learning condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Direct Instruction* group.
8. The level of statistical significance was reported by the study authors or, where 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Cole et al. (1993), no corrections for clustering or multiple comparisons were needed.
9. For all means and standard deviations in Appendix A3.4 the study authors reported data to a single decimal place. For formatting purposes, the WWC added zero in the second decimal place.
10. 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 subgroup findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Direct Instruction* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Mediated Learning group[3]	WWC calculations: Mean difference[4] (*Direct Instruction* – Mediated Learning)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Cole et al., 1991 (randomized controlled trial with attrition problems; children with language delays subgroup)[8]								
BLCT	3–7 year olds	58	38.40[9] (15.10)	37.80 (17.40)	–0.60	–0.04	ns	–1
PLAI	3–7 year olds	52	25.10 (11.30)	25.30 (9.00)	–0.20	–0.02	ns	–1
MLU	3–7 year olds	59	3.80 (1.10)	3.70 (1.10)	0.10	0.09	ns	+4
PPVT-R Standard Score	3–7 year olds	103	79.80 (14.40)	84.00 (12.90)	–4.20	–0.30	ns	–12
TELD Quotient	3–7 year olds	92	84.00 (10.10)	84.90 (10.20)	–0.90	–0.09	ns	–4

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the oral language domain for a related study. Cole et al. (1991) is not a pure subgroup analysis of Cole et al. (1993) (both studies are subsamples of a study that did not meet WWC evidence screens); however, Cole et al. (1991) compares the effects of *Direct Instruction* with the effects of Mediated Learning for children with language delays using a related subsample.
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 *Direct Instruction* group mean equals the Mediated Learning group mean plus the mean difference.
4. Positive differences and effect sizes favor the *Direct Instruction* group and negative differences and effect sizes favor the Mediated Learning group. The mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the *Direct Instruction* group had lower pretest scores than the Mediated Learning group and underestimate the intervention's effect when the *Direct Instruction* group had higher pretest scores than the Mediated Learning group. The BLCT is a measure of the number of errors, but the mean difference is calculated so that a positive effect indicates that the *Direct Instruction* group performed better (had fewer errors) than the Mediated Learning group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Direct Instruction* condition versus the percentile rank of the average student in the Mediated Learning condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Direct Instruction* group.
8. The level of statistical significance was reported by the study authors or, where 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Cole et al. (1991; 1993), no corrections for clustering were needed.
9. For all means and standard deviations in Appendix A4.1 the study authors reported data to a single decimal place. For formatting purposes, the WWC added zero in the second decimal place.

Appendix A4.2 Summary of subgroup findings for the cognition domain[1]

Outcome measure	Study sample	Sample size (children)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Direct Instruction* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Mediated Learning group[3]	WWC calculations: Mean difference[4] (*Direct Instruction* – Mediated Learning)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Cole et al., 1993 (randomized controlled trial with attrition problems)[8]								
MSCA GCI	3–7 year olds	164	77.90[9] (18.20)	76.60 (16.90)	1.30	0.07	ns	+3
Cole et al., 1991 (randomized controlled trial with attrition problems; children with language delays subgroup)[8]								
MSCA GCI	3–7 year olds	105	77.00 (14.00)	78.90 (14.40)	–1.90	–0.13	ns	–5

ns = not statistically significant

1. This appendix presents composite subgroup findings for measures that fall in the cognition domain. Total group scores were used for rating purposes and are presented in Appendix A3.3. As noted in Appendix A2.3, the WWC does not include the MSCA GCI in the overall rating of effectiveness because the WWC includes the individual measures used to develop the composite. Cole et al. (1991) is not a pure subgroup analysis of Cole et al. (1993) (both studies are subsamples of a study that did not meet WWC evidence screens); however, Cole et al. (1991) compares the effects of *Direct Instruction* with the effects of Mediated Learning for children with language delays.
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 *Direct Instruction* group mean equals the Mediated Learning group mean plus the mean difference.
4. Positive differences and effect sizes favor the *Direct Instruction* group; negative differences and effect sizes favor the Mediated Learning group. The mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the *Direct Instruction* group had lower pretest scores than the Mediated Learning group and underestimate the intervention's effect when the *Direct Instruction* group had higher pretest scores than the Mediated Learning group.
5. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
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 *Direct Instruction* condition versus the percentile rank of the average student in the Mediated Learning condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the *Direct Instruction* group.
8. The level of statistical significance was reported by the study authors or, where 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. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Cole et al. (1991; 1993), no corrections for clustering were needed.
9. For all means and standard deviations in Appendix A4.2 the study authors reported data to a single decimal place. For formatting purposes, the WWC added zero in the second decimal place.

Appendix A5.1 *Direct Instruction* 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 language, the WWC rated *Direct Instruction* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects, as no studies showed statistically significant or substantively important effects, either positive or negative.

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. The study did not show statistically significant or substantively important effects, 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. Only one study examined effects on oral language.

AND

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

  Met. The study did not show 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. The one study did not show statistically significant or substantively important positive effects.

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. The single study showed indeterminate 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. The study did not show statistically significant or substantively important effects, either positive or negative.

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. The single study showed indeterminate effects.

(continued)

Appendix A5.1 *Direct Instruction* 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. The study did not show statistically significant or substantively important negative effects.

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. The study did not show statistically significant or substantively important positive effects.

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. Only one study examined effects on oral language.

AND

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

  Met. The study did not show statistically significant or substantively important positive effects.

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. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *Direct Instruction* 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 *Direct Instruction* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects, as no studies showed statistically significant or substantively important effects, either positive or negative.

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. The study did not show statistically significant or substantively important effects, 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. Only one study examined effects on print knowledge.

AND

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

  Met. The study did not show 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. The one study did not show statistically significant or substantively important positive effects.

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. The single study showed indeterminate 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. The study did not show statistically significant or substantively important effects, either positive or negative.

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. The single study showed indeterminate effects.

(continued)

Appendix A5.2 *Direct Instruction* rating for the print knowledge 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. The study did not show statistically significant or substantively important negative effects.

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. The study did not show statistically significant or substantively important positive effects.

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. Only one study examined effects on print knowledge.

AND

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

  Met. The study did not show statistically significant or substantively important positive effects.

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. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.3 *Direct Instruction* rating for the cognition 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 cognition, the WWC rated *Direct Instruction* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects, as no studies showed statistically significant or substantively important effects, either positive or negative.

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. The study did not show statistically significant or substantively important effects, 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. Only one study examined effects on cognition.

AND

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

  Met. The study did not show 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. The one study did not show statistically significant or substantively important positive effects.

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. The single study showed indeterminate 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. The study did not show statistically significant or substantively important effects, either positive or negative.

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. The single study showed indeterminate effects.

(continued)

Appendix A5.3 *Direct Instruction* rating for the cognition 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. The study did not show statistically significant or substantively important negative effects.

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. The study did not show statistically significant or substantively important positive effects.

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. Only one study examined effects on cognition.

AND

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

  Met. The study did not show statistically significant or substantively important positive effects.

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. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.4 *Direct Instruction* 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 *Direct Instruction* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects, as no studies showed statistically significant or substantively important effects, either positive or negative.

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. The study did not show statistically significant or substantively important effects, 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. Only one study examined effects on math.

AND

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

  Met. The study did not show 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. The one study did not show statistically significant or substantively important positive effects.

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. The single study showed indeterminate 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. The study did not show statistically significant or substantively important effects, either positive or negative.

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. The single study showed indeterminate effects.

(continued)

Appendix A5.4 *Direct Instruction* rating for the math 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. The study did not show statistically significant or substantively important negative effects.

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. The study did not show statistically significant or substantively important positive effects.

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. Only one study examined effects on math.

AND

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

  Met. The study did not show statistically significant or substantively important positive effects.

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. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Centers	Children	
Oral language	1	1	164	Small
Print knowledge	1	1	164	Small
Phonological processing	0	0	0	na
Early reading/writing	0	0	0	na
Cognition	1	1	164	Small
Math	1	1	164	Small

na = not applicable/not studied

1. A rating of "moderate 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."