

What Works Clearinghouse



Tools of the Mind

Program description¹ *Tools of the Mind* is an early childhood curriculum for preschool and kindergarten children, based on the ideas of Russian psychologist Lev Vygotsky. The curriculum is designed to foster children’s executive function, which involves developing self-regulation, working memory, and cognitive flexibility. Many activities emphasize both executive functioning and academic skills.

Research One study of *Tools of the Mind* meets the What Works Clearinghouse (WWC) evidence standards. The study included more than 200 three- to four-year-old children attending preschool in a low-income, urban school district.² The WWC considers the extent of evidence for *Tools of the Mind* to be small for oral language, print knowledge, cognition, and math. No studies that meet the WWC evidence standards with or without reservations addressed phonological processing or early reading/writing.

Effectiveness *Tools of the Mind* was found to have no discernible effects on oral language, print knowledge, cognition, and math.

	Oral Language	Print knowledge	Cognition	Math	Phonological processing	Early reading/writing
Rating of effectiveness	No discernible effects	No discernible effects	No discernible effects	No discernible effects	na	na
Improvement index³	Average: +6 percentile points	Average: 0 percentile points	+2 percentile points	+7 percentile points	na	na
	Range: +4 to +8 percentile points	Range: -1 to +1 percentile points	na	na	na	na

na = not applicable

- The descriptive information for this program was obtained from publicly available sources: the program website (http://www.msced.edu/extendedcampus/tools_ofthemind/index.shtml, retrieved July 2008) and the literature reviewed for this report. The WWC asks developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
- The study was conducted in one school, with a full-day Abbott preschool education program, in which both the intervention and comparison group children participated. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
- These numbers show the average and range of student-level improvement indices for all findings across the study. For cognition and math, the improvement index is based on a single finding.

Additional program information

Developer and contact

Developed by Deborah J. Leong and Elena Bodrova, *Tools of the Mind* is distributed by Metropolitan State College of Denver, Center for Improving Early Learning. Address: 5660 Greenwood Plaza Blvd., Suite 100, Greenwood Village, CO 80111. Email: leongd@mscd.edu. Web: <http://www.mscd.edu/extendedcampus/toolsofthemind/index.shtml>. Telephone: (303) 721-1313.

Scope of use

Tools of the Mind was first implemented in preschool classrooms in 1993. During the 2008/09 school year *Tools of the Mind* will be active in more than 450 full- and half-day classrooms in Colorado, Maine, Massachusetts, New Jersey, New Mexico, and Oregon. *Tools of the Mind* is used in Head Start centers, public preschool programs, and child care centers in various settings. The curriculum is appropriate for use with typically developing children, as well as English language learners, and has been used in both inclusion and special education classrooms.

Teaching

Tools of the Mind can be implemented in a variety of early childhood settings. The curriculum focuses on 40 activities designed to develop children's executive function, including child-directed, teacher-supported, and cooperative peer activities. Instruction is individualized through teacher scaffolding. Dramatic play is a main component of the curriculum. With intentional planning by the children and support from the teacher, this component exposes children to a range of experiences that foster self-regulation skills. For example, children are encouraged to write or draw a representation of their plan for a pretend play activity. Self-regulation is viewed as a necessary prerequisite to school readiness and is embedded in activities throughout the day.

Thus, activities are designed for children to simultaneously practice self-regulation and cognitive skills, such as "Buddy Reading," during which children explore concepts of print but also practice staying in the role of "reader" and "listener." Professional development for teachers, paraprofessionals, and program coaches are provided by *Tools of the Mind* staff during the first two years of implementation. In the first year the trainers offer four workshops and conduct at least four site visits, depending on the program's size. The program coaches receive specialized training, a coaching manual, pacing guides, and a fidelity checklist.

Cost

Tools of the Mind is typically implemented over a two-year period. During this time the developer provides intensive professional development to facilitate implementation. The first year of implementation costs about \$3,000 per classroom, excluding travel and depending on the program's size. The price includes training for most staff that work with the students, such as paraprofessionals and supervisors—although special education staff are trained separately at additional cost. The cost and number of site visits provided vary depending on the number of classrooms in the program. The curriculum guides cost an additional \$100. The developer and adopters negotiate the cost of the second year of professional development services, typically about \$1,500. Although the developers of *Tools of the Mind* do not sell classroom materials, they provide a list of recommended materials that programs can purchase from other vendors.

Research Four studies reviewed by the WWC investigated the effects of *Tools of the Mind* on preschool children’s cognitive and language competencies and their school readiness. One study (Barnett et al., 2008) was a randomized controlled trial that meets WWC evidence standards. The remaining three studies did not meet WWC evidence standards.

Seven other studies did not meet WWC eligibility screens. Five did not investigate the effects of *Tools of the Mind* on children’s outcomes, one did not focus on preschool-age children, and one did not provide enough information to assess its study design.

Barnett et al. (2008) conducted a randomized controlled trial of teachers and students to investigate the effects of the program. In an urban school, teachers and their assistants were randomly assigned to classrooms using a stratified random assignment procedure. Three- to four-year-old children attending preschools were then randomly assigned to either *Tools of the Mind* or comparison group classrooms. In all, 85 children in 7 classrooms used *Tools of the Mind*, and 117 children in the 11 comparison group classrooms used their regular district

curriculum. According to the study authors, the district curriculum covered much of the same academic content and topics as *Tools of the Mind*, but there was greater emphasis on teacher-imposed control and less on children’s self-regulation. The study reported students’ outcomes after the first year of program implementation.

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the [What Works Clearinghouse Extent of Evidence Categorization Scheme](#)). The extent of evidence takes into account the number of studies and the total sample size across the studies that meet WWC evidence standards with or without reservations.⁴

The WWC considers the extent of evidence for *Tools of the Mind* to be small for oral language, print knowledge, cognition, and math. No studies that meet the WWC evidence standards with or without reservations addressed phonological processing or early reading/writing.

Effectiveness Findings

The WWC review of interventions for Early Childhood Education addresses student outcomes in six domains: oral language, print knowledge, cognition, math, phonological processing, and early reading/writing. The study included in this report covers four domains: oral language, print knowledge, cognition, and math. The findings below present the authors’ estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Tools of the Mind* on children.⁵

Oral language. Barnett et al. (2008) reported results separately for regression and hierarchical linear model (HLM) analyses. For regression analysis, the authors found a statistically significant positive effect of *Tools of the Mind* on the Peabody Picture Vocabulary Test (PPVT-III). For hierarchical linear model analysis, which accounted for clustering of children within classrooms, the effect was not statistically significant. The study authors did not find statistically significant effects of *Tools of the Mind* on the second oral language measure: Expressive One-Word Picture

4. The Extent of Evidence Categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as the students’ demographics and the types of settings in which studies took place, are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Tools of the Mind* is in Appendix A6.
5. 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, see the [WWC Tutorial on Mismatch](#). For the formulas the WWC used to calculate the statistical significance, see [Technical Details of WWC-Conducted Computations](#). For the *Tools of the Mind* study summarized here, no corrections for clustering and multiple comparisons were needed.

Effectiveness *(continued)*

Vocabulary Test-Revised (EOWPVT-R). The WWC found that the average effect size across the two outcomes was neither statistically significant nor large enough to be considered substantively important (an effect size at least 0.25) according to WWC criteria.

Print knowledge. The study authors did not find statistically significant effects of *Tools of the Mind* on either measure of print knowledge: Woodcock-Johnson-Revised Letter-Word Identification subtest or Get Ready to Read! assessment. The average effect size across the two outcomes was not large enough to be considered substantively important according to WWC criteria.

Cognition. Barnett et al. (2008) did not find a statistically significant effect of the *Tools of the Mind* curriculum on the Animal Pegs Subtest of the Wechsler Preschool Primary Scale of Intelligence, and the effect was not large enough to be considered substantively important according to WWC criteria.

The WWC found *Tools of the Mind* to have no discernible effects for oral language, print knowledge, cognition, or math.

Improvement index

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

The average improvement index for oral language is +6 percentile points in the one study, with a range of +4 to +8

Math. Barnett et al. (2008) did not find a statistically significant effect of the *Tools of the Mind* curriculum on the Woodcock-Johnson-Revised Applied Problems subtest, and the effect was not large enough to be considered substantively important according to WWC criteria.

Rating of effectiveness

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

percentile points across findings. The average improvement index for print knowledge is 0 percentile points in the one study, with a range of -1 to +1 percentile points across findings. The improvement index for cognition is +2 percentile points for a single finding of the study. The improvement index for math is +7 percentile points for a single finding of the study.

Summary

The WWC reviewed four studies on *Tools of the Mind*. One study meets WWC evidence standards and three studies did not meet WWC evidence standards; seven other studies did not meet eligibility screens. Based on the one study, the WWC found no discernible effects in oral language, print knowledge, cognition, or math. The evidence presented in this report may change as new research emerges.

References **Meets WWC evidence standards**

Barnett, W., Jung, K., Yarosz, D., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the *Tools of the Mind* Curriculum: a randomized trial. *Early Childhood Research Quarterly, 23*(3), 299–313.

Did not meet WWC evidence standards

Bodrova, E., & Leong, D. J. (2002). *Tools of the Mind* research project: implementation of Vygotskian principles of development and learning in an early childhood literacy program. [Unpublished manuscript]. Denver, CO: Mid-continent Research for Education and Learning. This study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Bodrova, E., Leong, D. J., & Semenov, D. (1997). *Tools of the Mind end of the year report*. Denver, CO: Metropolitan State College of Denver, ECE project. This study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Diamond, A., Barnett, S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(30), 1387–1388. This study does not meet WWC evidence standards because the intervention and comparison groups are not shown to be equivalent at baseline.

Did not meet WWC eligibility screens for *Tools of the Mind*

Bodrova, E., & Leong, D. J. (2001). *Tools of the Mind: a case study of implementing the Vygotskian approach in American early childhood and primary classrooms* (Innodata monographs 7). Geneva: International Bureau of Education. Retrieved from <http://www.ibe.unesco.org>. The study is ineligible for review because it does not provide enough information to assess whether it meets standards.

Bodrova, E., Leong, D. J., & Semenov, D. (1997). *Tools of the Mind end of the year report, Adams School District 50*. Denver, CO: Metropolitan State College of Denver. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Copple, C. (2003). Fostering young children's representation, planning, and reflection: a focus in three current early childhood models. *Journal of Applied Developmental Psychology, 24*(6), 763. This study is ineligible for review because it does not examine the effectiveness of an intervention.

Grigorenko, E. L. (1998). Mastering tools of the mind in school (trying out Vygotsky's ideas in classrooms). In R. J. Sternberg, & W. M. Williams (Eds.), *Intelligence, Instruction, and Assessment: Theory into Practice* (pp. 201–231). Mahwah, NJ: Lawrence Erlbaum Associates. This study is ineligible for review because it does not examine the effectiveness of an intervention.

Hyson, M. (2008). *Enthusiastic and engaged learners: approaches to learning in the Early Childhood Classroom*. New York: Teacher College Press and Washington, DC: NAEYC. This study is ineligible for review because it does not examine the effectiveness of an intervention.

Hyson, M., Copple, C., & Jones, J. (2006). Early childhood development and education. In K. A. Renninger, I. E. Sigel, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (pp. 3–47). Hoboken, NJ: John Wiley & Sons, Inc. This study is ineligible for review because it does not examine the effectiveness of an intervention.

Zigler, E. F., & Bishop-Josef, S. J. (1996). The cognitive child vs. the whole child: lessons from 40 years of Head Start. In D. G. Singer, R. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play = learning: how play motivates and enhances children's cognitive and social-emotional growth*. New York: Oxford University Press. This study is ineligible for review because it does not examine the effectiveness of an intervention.

For more information about specific studies and WWC calculations, please see the [WWC Tools of the Mind Technical Appendices](#).

Appendix

Appendix A1.1 Study Characteristics: Barnett, Jung, Yarosz, Thomas, Hornbeck, Stechuk, & Burns, 2008 (randomized controlled trial)

Characteristic	Description
Study citation	Barnett, W., Jung, K., Yarosz, D., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the <i>Tools of the Mind</i> curriculum: a randomized trial. <i>Early Childhood Research Quarterly</i> , 23(3), 299–313.
Participants	In one school selected for the study, 7 classrooms on one floor were available for <i>Tools of the Mind</i> implementation and 11 classrooms on another floor were available for the control condition. Teachers and assistants were randomly assigned to classrooms using a stratified assignment procedure, and then three- and four-year-old children were randomly assigned to either <i>Tools of the Mind</i> curriculum classrooms or district curriculum classrooms. Poverty level, achievement, and minority status were similar across intervention and comparison groups. Among the children sampled, 93 percent are Hispanic, and about 70 percent consider Spanish their primary home language. Although the overall student attrition rate was more than 25 percent, and student consent after random assignment led to differential attrition, the post-attrition intervention and comparison samples were equivalent on achievement pretests. After one year, 85 <i>Tools of the Mind</i> students and 117 comparison students remained in the sample.
Setting	This study was conducted in 18 classrooms in a low-income urban school with state-financed Abbott full-day preschool education.
Intervention	<i>Tools of the Mind</i> aims to aid learning and development while emphasizing emergent literacy and self-regulation. The two main goals of the curriculum are to develop underlying cognitive skills (such as self-regulation, deliberate memory, and focused attention) and to develop specific academic skills (such as symbolic thought, literacy, and an understanding of math). Play is the leading activity for developing such skills and the curriculum emphasizes the teacher's role in supporting the development of mature intentional dramatic play. The study was conducted during the first year of program implementation of <i>Tools of the Mind</i> .
Comparison	Control classrooms implemented the standard district-created curriculum, which was described as a full-day PreK balanced literacy curriculum with themes. In structured observations of the control group, frequently observed activities were art projects that correlated with the "letter of the week," free play, large group movement and/or music, and such large group activities as story time. According to the study authors, although the control curriculum covered much of the same academic content and topics as <i>Tools of the Mind</i> , there was greater emphasis on teacher-imposed control and less on children's self-regulation.
Primary outcomes and measurement	For both pre- and post-tests, the authors administered Peabody Picture Vocabulary Test-III, Expressive One-Word Picture Vocabulary Test-Revised, Animal Pegs Subtest of the Wechsler Preschool Primary Scale of Intelligence, and two subtests of the Woodcock-Johnson-Revised test (Applied Problems and Letter-Word Identification). Get Ready to Read! screening tool was used only at post-test assessment. IDEA Oral Language Proficiency Test was administered for the subsample of Spanish-speaking children. Problem Behaviors Scale of the Social Skills Rating System was also used in the study, but not included in this report because it was outside the scope of the Early Childhood Education review. For a more detailed description of these outcome measures, see Appendix A2.1–2.4.
Teacher and staff training	Teachers assigned to the <i>Tools of the Mind</i> group received four full days of curriculum training before the start of the school year. During the school year, they received 30-minute classroom visits approximately once a week from a <i>Tools of the Mind</i> trainer to address any difficulties they were having with the curriculum. In addition, <i>Tools of the Mind</i> teachers received 1 half-day workshop and 5 one-hour lunchtime meetings to discuss aspects of the curriculum. Control group teachers received similar amounts of training. They attended workshops on the already established district curriculum given by the district for the same amount of time.

Appendix A2.1 Outcome measures for the oral language domain

Outcome measure	Description
Peabody Picture Vocabulary Test-III	A standardized measure of children's receptive vocabulary that requires children to identify pictures that correspond to spoken words (as cited in Barnett et al., 2008).
Expressive One-Word Picture Vocabulary Test-Revised	A standardized measure of children's expressive vocabulary that requires them to name pictures of common objects, actions, and concepts (as cited in Barnett et al., 2008).
IDEA Oral Language Proficiency Test	This test assesses the receptive and expressive Spanish language skills of Spanish-speaking children (as cited in Barnett et al., 2008).

Appendix A2.2 Outcome measures for the print knowledge domain

Outcome measure	Description
Woodcock-Johnson-Revised: Letter-Word Identification subtest	A standardized measure of children's ability to name printed letters and words (as cited in Barnett et al., 2008).
Get Ready to Read!	A nonstandardized measure of readiness for reading instruction focusing on three core domains (print knowledge, emergent writing skills, and linguistic awareness) across 20 items to which children indicate their response by pointing (as cited in Barnett et al., 2008).

Appendix A2.3 Outcome measures for the cognition domain

Outcome measure	Description
Wechsler Preschool Primary Scale of Intelligence: Animal Pegs subtest	A subset from a standardized measure that assesses a child's nonverbal problem-solving and visual-motor proficiency as they place pegs of correct colors in a series of holes under pictures of animals (as cited in Barnett et al., 2008).

Appendix A2.4 Outcome measures for the math domain

Outcome measure	Description
Woodcock-Johnson-Revised: Applied Problems subtest	A subtest from a standardized measure that assesses children's math skills by asking children to count small sets and to solve simple addition and subtraction questions using pictures (as cited in Barnett et al., 2008).

Appendix A3.1 Summary of study findings included in the rating for the oral language domain¹

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (<i>Tools of the Mind</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Tools of the Mind</i> group	Comparison group				
Barnett et al. (2008) (randomized controlled trial)⁷								
Peabody Picture Vocabulary Test-III (PPVT-III)	Three- to four-year-olds	18/198	nr (19.19)	nr (15.90)	3.59	0.21	ns	+8
Expressive One-Word Picture Vocabulary Test-Revised (EOWPVT-R)	Three- to four-year-olds	18/193	nr (14.06)	nr (12.22)	1.19	0.09	ns	+4
Average for oral language⁸						0.15	ns	+6

ns = not statistically significant

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language domain. Subgroup findings for children who consider Spanish their primary language 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. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. 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 results favorable to the intervention group.
7. 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](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
8. This row provides the study average, which is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain¹

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Tools of the Mind</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Tools of the Mind</i> group	Comparison group				
Barnett et al. (2008) (randomized controlled trial)⁷								
Woodcock-Johnson-Revised Letter-Word Identification subtest	Three- to four-year-olds	18/202	nr (12.87)	nr (11.92)	–0.45	–0.04	ns	–1
Get Ready to Read!	Three- to four-year-olds	18/220	nr (3.90)	nr (3.91)	0.13	0.03	ns	+1
Average for print knowledge⁸						0.00	ns	0

ns = not statistically significant

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge domain.
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. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. 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 results favorable to the intervention group.
7. 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](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
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.

Appendix A3.3 Summary of study findings included in the rating for the cognition domain¹

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Tools of the Mind</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Tools of the Mind</i> group	Comparison group				
Barnett et al. (2008) (randomized controlled trial)⁷								
Wechsler Preschool Primary Scale of Intelligence (WPPSI) Animal Pegs Subtest	Three- to four-year-olds	18/200	nr (15.22)	nr (16.35)	0.84	0.05	ns	+2
Average for cognition⁸						0.05	ns	+2

ns = not statistically significant
nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the cognition domain.
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. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. 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 results favorable to the intervention group.
7. 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](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
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.

Appendix A3.4 Summary of study findings included in the rating for the math domain¹

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Tools of the Mind</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Tools of the Mind</i> group	Comparison group				
Barnett et al. (2008) (randomized controlled trial)⁷								
Woodcock-Johnson-Revised Applied Problems	Three- to four-year-olds	18/202	nr (16.19)	nr (18.86)	3.07	0.17	ns	+7
Average for math⁸						0.17	ns	+7

ns = not statistically significant

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math domain.
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. The standard deviations were provided by the author at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. 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 results favorable to the intervention group.
7. 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](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering or multiple comparisons were needed because the study reported findings were based on hierarchical linear model (HLM) analyses and were not statistically significant.
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.

Appendix A4.1 Summary of subgroup findings by age for the oral language domain¹

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Tools of the Mind</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Tools of the Mind</i> group	Comparison group				
Barnett et al. (2008) (randomized controlled trial)⁷								
IDEA Oral Language Proficiency Test in Spanish	Three- to four-year-olds	18/132	nr (8.49)	nr (6.82)	2.36	0.31	ns	+12

ns = not statistically significant

nr = not reported

1. This appendix presents subgroup findings for measures that fall in the oral language domain. The Oral Language Proficiency Test in Spanish was administered to children who considered Spanish their primary language (approximately 70 percent of the sample) to assess their Spanish language development. Total group 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. The standard deviations were provided by the authors at the WWC request.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean difference is the hierarchical linear model (HLM) coefficient for the intervention's effect provided by the author at the WWC request.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. 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 results favorable to the intervention group.
7. 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](#). For the formulas the WWC used to calculate statistical significance, see [Technical Details of WWC-Conducted Computations](#). For Barnett et al. (2008), no corrections for clustering were needed because the study reported findings were based on hierarchical linear model (HLM) analyses.

Appendix A5.1 *Tools of the Mind* rating for the oral language domain

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

For the outcome domain of oral language, the WWC rated *Tools of the Mind* 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 because 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. One study examined effects on oral language and 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. No study showed statistically significant positive effects.

AND

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

Met. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through EITHER of the following.

- 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. No study showed a statistically significant or substantively important effect, 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. No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

Appendix A5.1 *Tools of the Mind* 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. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect.

AND

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

Met. No study showed 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. For a complete description, see the [WWC Intervention Rating Scheme](#).

Appendix A5.2 *Tools of the Mind* rating for the print knowledge domain

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

For the outcome domain of print knowledge, the WWC rated *Tools of the Mind* 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 because 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. One study examined effects on print knowledge and 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. No study showed statistically significant positive effects.

AND

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

Met. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through EITHER of the following.

- 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. No study showed a statistically significant or substantively important effect, 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. No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

Appendix A5.2 *Tools of the Mind* 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. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect.

AND

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

Met. No study showed 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. For a complete description, see the [WWC Intervention Rating Scheme](#).

Appendix A5.3 *Tools of the Mind* rating for the cognition domain

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

For the outcome domain of cognition, the WWC rated *Tools of the Mind* 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 because 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. One study examined effects on cognition and 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. No study showed statistically significant positive effects.

AND

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

Met. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through EITHER of the following.

- 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. No study showed a statistically significant or substantively important effect, 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. No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

Appendix A5.3 *Tools of the Mind* 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. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect.

AND

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

Met. No study showed 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. For a complete description, see the [WWC Intervention Rating Scheme](#).

Appendix A5.4 *Tools of the Mind* rating for the math domain

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

For the outcome domain of math, the WWC rated *Tools of the Mind* 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 because 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. One study examined effects on math and 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. No study showed statistically significant positive effects.

AND

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

Met. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through EITHER of the following.

- 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. No study showed a statistically significant or substantively important effect, 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. No study showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

(continued)

Appendix A5.4 *Tools of the Mind* 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. No study 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. No study 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. No study showed a statistically significant or substantively important negative effect.

AND

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

Met. No study showed 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. For a complete description, see the [WWC Intervention Rating Scheme](#).

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Oral language	1	1	198	Small
Print knowledge	1	1	220	Small
Cognition	1	1	200	Small
Math skills	1	1	202	Small
Phonological processing	0	na	na	na
Early reading/writing	0	na	na	na

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”