

What Works Clearinghouse



Arthur

- Program description** *Arthur*, a book-based educational television program designed for children ages 4–8, is popular among preschool and kindergarten students. The program is based on the storybooks, by Marc Brown, about Arthur, an 8-year-old aardvark. Each show is 30 minutes in length and includes two stories involving characters dealing with moral issues. The show has been used as a listening comprehension and language development intervention for English language learning students.
- Research** One study of *Arthur* met the What Works Clearinghouse (WWC) evidence standards. This study, which included 108 kindergarten Spanish-speaking English language learners from six schools in a large urban school district on the East Coast, assessed students based on narrative skill in English—the ability to talk about events in a coherent fashion.¹
- Effectiveness** *Arthur* was found to have potentially positive effects on English language development.

	<i>Reading achievement</i>	<i>Mathematics achievement</i>	<i>English language development</i>
Rating of effectiveness	Not reported	Not reported	Potentially positive effects
Improvement index²	Not reported	Not reported	Average: +11 percentile points Range: –5 to +17 percentile points

1. The evidence presented in this report is based on the available research. Findings and conclusions may change as new research becomes available.
 2. These numbers show the average and the range of improvement indices for all findings across the study.

Additional program information

Developer and contact

WGBH Boston. P.O. Box 200, Boston, MA 02134. Cookie Jar Education, Inc., Carson-Dellosa Publishing Co., Inc., 7027 Albert Pick Road, Greensboro, NC 27409. Web: www.wgbh.org; www.cinar.com. Email: tvdistribution@thecookiejarcompany.com. Telephone: 617-300-5400 (WGBH Boston); 336-632-0084 (Cookie Jar Education, Inc.). A direct link to a description of *Arthur* and related materials is available at <http://pbskids.org/arthur/index.html>.

Scope of use

Public Broadcasting Service (PBS) stations throughout the United States broadcast *Arthur* daily, Monday through Friday. The PBS Kids website provides a number of lesson plans and activities for parents and teachers. According to PBS, there is flexibility in how to use these lessons, and parents and teachers may choose whether to use them. *Arthur* is not specifically

designed for English language learners but, according to the developer, can be used with these students.

Teaching

Arthur, an animated children's series based on a storybook, is intended for use with young children. Each 30-minute episode consists of two stories—each with a plot, conflict, and resolution. The website also offers a link for teachers to send queries about classroom activities.

Cost

All materials for teaching are available as free online downloads from the PBS Kids website.³ The program is broadcast on PBS stations at no cost to viewers. Schools using the program would need access to televisions. *Arthur* videos can be found in bookstores, video stores, or public libraries.

Research

One study (Uchikoshi, 2005) reviewed by the WWC investigated the effects of *Arthur* on English language learners. The study was a randomized controlled trial that met WWC evidence standards.

Participants in the study were 108 English language learning kindergarten students randomly assigned to either the intervention group or a comparison group. Intervention group students were assigned to watch three episodes of *Arthur* a week from

October to May (a total of 54 episodes), while comparison group students were assigned to watch an alternative educational program, *Between the Lions*. *Between the Lions* is a 30-minute, book-based program aired by PBS that focuses on phonics and reading skills but does not have the listening comprehension or language development emphasis of *Arthur*.⁴ To maintain consistency across classrooms, and because of limited classroom time, teachers were directed not to use follow-up activities.

Effectiveness

Findings

The WWC review of English language learners addresses student outcomes in three domains: reading achievement, mathematics achievement, and English language development.

English language development. Uchikoshi reported that

students watching *Arthur* showed greater improvement in narrative skill development than students in the comparison group. Although the individual and average effects (as calculated by the WWC) were not statistically significant, the average effect was large enough to be considered substantively important.⁵

3. Public Broadcasting Service (March, 2006). Lesson plans. Retrieved March 21, 2006 from <http://pbskids.org/arthur/parentsteachers/lesson/index.html>.

4. *Arthur* focuses on narrative skills. Uchikoshi (2005) defines a narrative as at least two sequential independent clauses describing a single past event, and he states that the ability to produce a narrative demonstrates a child's ability to talk about the world (p. 465). *Arthur* purportedly presents a well formed story structure (plot, conflict, and resolution), and the study author investigated whether narrative development is enhanced by watching *Arthur*.

5. The level of statistical significance was calculated by the WWC and corrects for multiple comparisons. For an explanation see the [WWC Tutorial on Mismatch](#). See the [WWC Intervention Rating Scheme](#) for the formulas the WWC used to calculate statistical significance.

Effectiveness *(continued)*

Rating of effectiveness

The WWC rates interventions 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 (as calculated by the WWC), the size of the differences

between participants in the intervention condition and the comparison condition, and the consistency of the findings across studies (see the [WWC Intervention Rating Scheme](#)).

The WWC found *Arthur* to have potentially positive effects for English language development

Improvement index

For each outcome domain, the WWC computed an improvement index based on the effect size (see the [WWC Improvement Index Technical Paper](#)). 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 entirely based on the size of the effect, regardless of the statistical significance of the effect, study design, or analysis. The improvement

index can take on values between –50 and +50, with positive numbers denoting favorable results. The average improvement index is +11 percentile points, with a range of –5 to +17 percentile points across findings, for the English language development domain.

Summary

The WWC reviewed one study on *Arthur*, which met WWC evidence standards. The WWC rated the program as having potentially positive effects on English language development.

References

Met WWC evidence standards

Uchikoshi, Y. (2005). Narrative development in bilingual kindergartners: Can *Arthur* help? *Developmental Psychology*, 41(3), 464–478.

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

Appendix

Appendix A1 Study characteristics: Uchikoshi, 2005 (randomized controlled trial)

Characteristic	Description
Study citation	Uchikoshi, Y. (2005). Narrative development in bilingual kindergarteners: Can <i>Arthur</i> help? <i>Developmental Psychology</i> , 41(3), 464–478.
Participants	The study involved 108 kindergarten students (47 girls and 61 boys). Fifty-one children were assigned to watch <i>Arthur</i> ; 57 were assigned to watch <i>Between the Lions</i> . ¹ Picture Vocabulary Test scores indicated that, at the beginning of the intervention, participants' average English vocabulary was at the three-year two-month age level of a monolingual English child. The Spanish version of this measure indicated that their native language vocabulary was at the five-year level; the average age of the children at the beginning of the study was 5 years, 7 months (boys) and 5 years, 6 months (girls). At least 80% of the students in the study qualified for free lunch. The time their families lived in the United States ranged from three months to seven years. According to parent survey responses, only 22% of the children in the sample were born outside of the country. These surveys also indicated that, on average, there were 21 books (in both English and Spanish) in the home, although there was wide variation on this number, ranging from zero to 300.
Setting	The study was conducted in six schools in a large urban district on the East Coast. Spanish-English classrooms (classrooms providing instruction in both languages) were selected, and all teachers were fluent in both languages. All children came from primarily Spanish-speaking homes and neighborhoods with heavy concentrations of Spanish-speaking people.
Intervention	The intervention group watched a 30-minute episode of <i>Arthur</i> at school, three times a week between October and May of one school year, for a total of 54 episodes. Although follow-up activities are available at the PBS website, teachers were directed only to show the videos.
Comparison	The comparison group watched the same number of episodes of <i>Between the Lions</i> over the same time period. <i>Between the Lions</i> is an educational television program with a focus on phonics and reading skills. <i>Arthur</i> focuses on narrative structure. As with the intervention group, none of the follow-up activities associated with the show were used. Each program in this show entails a story that a family of lions read together, focusing on phonological skills and the alphabet.
Primary outcomes and measurement	The outcome measure in the study was an instrument used to assess children's ability to tell a coherent story narrative, total number of words uttered by students, and the average length of the clauses used when describing a story.
Teacher training	Little information about teacher training was provided, other than they were bilingual.

1. Students were assigned within the six classrooms and matched as closely as possible on gender and pretest scores. Each classroom was presumably selected from one school.

Appendix A2 Outcome measures in the English language development domain

Outcome measure	Description
Combined narrative measure	Students were assessed by asking them to tell a “Bear Story” in English; three pictures of a family of teddy bears served as story prompts. The measure was taken from the School-Home Early Language and Literacy assessment developed by Catherine Snow and colleagues, as cited in Uchikoshi (2005). The measure assesses a student’s ability to develop a coherent narrative in English. Five dimensions are assessed: story structure coding, events coding, evaluation coding, temporality and reference, and storybook language. Coding entailed searching for whether any given dimension is present in the story. So stories were reviewed for an introduction, problem, and resolution (story structure); whether events related to the characters and plot (events coding); whether the children’s perspective were captured in the story (evaluation); presence of temporality and the presence of quotes; and use of adverbs and conjoined noun/verb phrases (storybook language) (Uchikoshi, 2005, p. 468). Children’s narratives were transcribed by trained assessors, and stories were read back to children to ensure they were accurately recorded. Although Spanish outcomes are available, these fall outside the parameters of this review.
Total number of words	The total number of words uttered by students offers a measure of story length.
Mean clause length	The complexity of clauses is thought to be associated with narrative skill development, and the length of clauses served as a proxy. Mean clause length was determined by total number of words (the above measure) divided by the number of clauses.

Appendix A3 Summary of study findings included in the rating for the English language development domain¹

Outcome measure	Study sample	Sample size (students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (column 1–column 2)	Effect size ⁴	Significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Arthur group (column 1)	Control group (column 2)				
Uchikoshi, 2005 (randomized controlled trial)								
Combined narrative measure	K	102	4.13 (4.35)	2.34 (3.75)	1.79	0.44	ns	+17
Total number of words	K	102	20.74 (39.4)	10.88 (24.7)	9.86	0.30	ns	+12
Mean clause length	K	102	0.54 (1.67)	0.76 (2.08)	–0.22	–0.11	ns	–5
Domain average⁷ for English language development (Uchikoshi, 2005)						0.29	ns	+11

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
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. Change scores on each measure were calculated (the difference between measures taken in October and May/June) and represent mean scores for the intervention and comparison groups. This difference focuses on how much higher the intervention group scored relative to the comparison condition. This differs from the study author's focus, which was based on how much faster intervention students learned relative to the comparison students. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the [WWC Technical Working Paper on Effect Size](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The level of statistical significance was calculated by the WWC and corrects for multiple comparisons. For an explanation see the [WWC Tutorial on Mismatch](#). See the [WWC Intervention Rating Scheme](#) for the formulas the WWC used to calculate statistical significance. These significance levels differ from those in the original study report, because the author presented a growth curve model, which is meant to examine the rate of change among individuals over time. The effect size estimations presented here focus on comparing the rate of change between groups while considering the multiple outcomes (total number of words, mean clause length, and combined narrative measure), thus impacting estimates of whether the groups have statistically significant differences. Note that the study tested outcomes at three time points (October, February, and May/June of the same school year). The WWC analysis used the October and May/June tests as the pre and posttests.
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 favorable results.
7. This row provides the study average, which is also the domain average in this case. 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 Rating for the English language development domain

The WWC rates interventions as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of English language development, the WWC rated *Arthur* as having potentially positive effects. It did not meet the criteria for positive effects, because it only had one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Arthur* was assigned the highest applicable rating.

Rating received

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

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, thus qualifying as a *positive* effect.

Met. In the one study on *Arthur* that examined English language development, the average effect size was substantively important.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect. The number of studies showing *indeterminate* effects is not greater than the number showing statistically significant or substantively important *positive* effects.

Met. The WWC analysis found no statistically significant or substantively important negative effects or indeterminate effects in this domain.

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. *Arthur* had only one study meeting WWC evidence standards. Although the effect was substantively important, the study lacked a strong design.

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

Met. The WWC analysis found no statistically significant or substantively important negative effects in this domain.

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 effects. See the [WWC Intervention Rating Scheme](#) for a complete description.