# What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Beginning Reading June 11, 2007

# Start Making a Reader Today® (SMART®)

## Program description

*Start Making a Reader Today® (SMART®)* is a volunteer tutoring program widely implemented in Oregon for students in grades K-2 who are at risk of reading failure. The program is designed to be a low-cost, easy-to-implement intervention. Volunteer tutors go into schools where at least 40% of students are eligible for free or reduced-price lunch and read one-on-one with students twice a week for half an hour. Typically, one volunteer works with two children on four types of activities: reading to the child, reading with the child, re-reading with the child, and asking the child questions about what has been read. The program also gives each student two new books a month to encourage families to read together.

## Research

One study of *SMART®* met the What Works Clearinghouse (WWC) evidence standards. The one study included more than 125 students in first grade in six schools across four school districts in Oregon.[1] The WWC considers the extent of evidence for *SMART®* to be small for alphabetics, fluency, and comprehension. No studies that met WWC evidence standards with or without reservations addressed general reading achievement.

## Effectiveness

*Start Making a Reader Today®* was found to have potentially positive effects on alphabetics, fluency, and comprehension.

| | *Alphabetics* | *Fluency* | *Comprehension* | *General reading achievement* |
|---|---|---|---|---|
| **Rating of effectiveness** | Potentially positive effects | Potentially positive effects | Potentially positive effects | na |
| **Improvement index[2]** | +16 percentile points | Average: +17 percentile points<br>Range: +16 to +18 percentile points | Average: +14 percentile points<br>Range: +11 to +16 percentile points | na |

na = not applicable

1. The evidence presented in this report is based on available research.
2. These numbers show the average and range of improvement indices for all findings in the study. The range is provided only if more than one outcome was measured within a domain.

## Additional program information

### Developer and contact

Developed by the Oregon Children's Foundation, *Start Making a Reader Today*® is self-distributed. Address: 219 NW 12th Ave, Suite 203, Portland, OR 97209. Email: smart@getsmartoregon.org. Web: www.getsmartoregon.org. Telephone: (503) 937-4800 or (877) 598-4633.

### Scope of use

Since its start in 1992, the program reports serving 100,000 children in the state of Oregon through more than 2.3 million volunteer hours. It has also given students more than 1.4 million books. The goal for the 2006–07 school year is to serve 12,000 students in 280 schools in 32 of Oregon's 36 counties.

### Teaching

*SMART*® accepts applications from schools where at least 40% of students are eligible for free or reduced-price lunch. The *SMART*® organization hires a part-time school coordinator for each participating school who works under the direction of a regional manager. The coordinator recruits and trains volunteers, is present at the school during all program hours, schedules reading sessions, and serves as primary contact for school personnel. In rural areas, *SMART*® offers the *SMART*® *Kit* as an alternative delivery model. The kit assists a school and its surrounding community to implement the program themselves without a regional manager. It includes instructions for setting up the program, organizing classrooms, recruiting volunteers, scheduling the intervention into classrooms, and coordinating the overall program.

Once the program is in place, the *SMART*® organization assists the school with materials, books, volunteer training, and technical assistance. *SMART*® staff facilitate the creation of local Leadership Councils, made up of school and community members, which assist in local fundraising and serve as local advocates for *SMART*®. In *SMART*® Kit communities, this group is known as a Leadership Committee and takes on primary responsibility for program operation.

Volunteers, who range from high school students to senior citizens, undergo a 1–2 hour long training that provides an introduction to the program and to reading strategies instruction. Volunteers are trained by the *SMART*® coordinator where the program is housed. They draw on the handbook that outlines the four *SMART*® reading strategies: reading to students, reading with students, re-reading, and asking comprehension questions. Using these strategies, volunteers tutor students one-on-one for 30 minutes twice a week throughout the school year.

### Cost

The *SMART*® program is funded through a wide range of statewide and local activities involving businesses, foundations, and individuals. There is no cost to a school participating through the standard delivery model. Local fundraising pays the salary of program coordinators, though some program coordinators volunteer and thus do not incur this cost.

For communities using the *SMART*® *Kit* delivery model, local groups raise money to cover the cost of *SMART*® licensing and for the salary of a school coordinator. Overall program cost runs approximately $300 a year per child. This cost is largely covered through donations to the *SMART*® parent organization.

## Research

One study (Baker, Gersten, & Keating, 2000) reviewed by the WWC investigated the effects of *SMART*®. This study was a randomized controlled trial that met WWC evidence standards.

Baker, Gersten, & Keating (2000) randomly assigned low-performing first-grade students in 24 classrooms from six Title I schools to the intervention or the comparison group within each classroom and assessed reading outcomes at the end of first and second grades. Students in the intervention group received the *SMART*® program as a supplement to the regular reading curriculum during first and second grades. Students in the comparison group did not receive the *SMART*® program, but received the same classroom instruction as students in the intervention group.

## Research *(continued)*

The study also included an average-achieving comparison but the WWC did not include this portion of the study in its review.

### Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate 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 met WWC evidence standards with or without reservations.[3]

The WWC considers the extent of evidence for *SMART*® to be small for alphabetics, fluency, and comprehension. No studies that met WWC evidence standards with or without reservations addressed general reading achievement.

## Effectiveness

### Findings

The WWC review of interventions for beginning reading addresses student outcomes in four domains: alphabetics, fluency, comprehension, and general reading achievement.[4] The Baker, Gersten, & Keating (2000) study reported outcomes in the first three domains. The findings below report outcomes assessed at the end of second grade.[5]

*Alphabetics.* The Baker, Gersten, & Keating (2000) study reported a statistically significant positive effect of *SMART*® on the Woodcock Reading Mastery Tests–Revised (WRMT-R) word identification subtest. This result was confirmed by the WWC.

*Fluency.* The Baker, Gersten, & Keating (2000) study reported, and the WWC confirmed, statistically significant positive effects of *SMART*® on the Oral Reading Fluency test, first- and second-grade passages (both administered to students at the end of second grade).

*Comprehension.* The Baker, Gersten, & Keating (2000) study reported a statistically significant positive effect of *SMART*® on the word comprehension subtest of the WRMT-R, and no statistically significant effect on the passage comprehension subtest of the WRMT-R. The WWC did not find that either of these effects was statistically significant. The average effect size across the two outcomes, however, was large enough to be considered substantively important according to WWC criteria (that is, an effect size of at least 0.25).

### 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 effects. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings,[6] 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).

3. 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 sizes of studies. Additional factors associated with a related concept, external validity, such as students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.
4. For definitions of the domains, see the Beginning Reading Protocol.
5. Outcomes assessed at the end of second grade are shown in Appendix A3 and outcomes assessed at the end of first grade are shown in Appendix A4.
6. 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. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of the *Start Making a Reader Today*®, a correction for multiple comparisons was needed for some domains.

**The WWC found *Start Making a Reader Today®* to have potentially positive effects for behavior**

### 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.

In the one study of *SMART®,* the improvement index for the single outcome in alphabetics is +16 percentile points; the average improvement index for fluency is +17 percentile points, with a range of +16 to +18 percentile points across findings. The average improvement index for comprehension is +14 percentile points, with a range of +11 to +16 percentile points across findings.

### Summary

The WWC reviewed one study on *SMART®* and this study met WWC evidence standards. Based on this study, the WWC found potentially positive effects for alphabetics, fluency, and comprehension. The evidence presented in this report may change as new research emerges.

## References

### Met WWC evidence standards

Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A two-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly, 35*(4), 494–519.

**For more information about specific studies and WWC calculations, please see the WWC *Start Making a Reader Today®* Technical Appendices.**

# Appendix

## Appendix A1.1 Study characteristics: Baker, Gersten, & Keating, 2000 (randomized controlled trial)

| Characteristic | Description |
|---|---|
| **Study citation** | Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A two-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly, 35*(4), 494–519. |
| **Participants** | Participants were 127 first-grade students from 24 classrooms in six Title I schools in four districts. Participants were nominated by their teachers as needing supplemental reading assistance based on two criteria: low reading skills and relatively little reading experience with adults or others at home. The students were randomly assigned to intervention and comparison conditions within classrooms after being matched on the Rapid Letter Naming pretest. The study presented findings after the intervention students completed two years of the program. At the end of second grade, 84 students of the original sample remained (43 students in the intervention and 41 students in the comparison group).[1] The study included an additional comparison group of 36 average-achieving readers from the same schools. Analysis involving these comparison groups was not eligible for WWC review because the WWC considers only comparisons of students with similar achievement backgrounds in assessing the effectiveness of *SMART®*. Student ethnicity was 47% European-American, 30% African-American, 10% American Indian, 6% Asian-American, and 6% Latino. |
| **Setting** | The study took place in two large counties in western Oregon. The schools represented a diverse range of communities, from low income/large city to working class/moderate size-city to rural settings. |
| **Intervention** | Students received one-to-one tutoring for six months each year while they were in first and second grade. The program consisted of two 30-minute sessions a week. Students could also take home two books a month. The number of sessions per student ranged from 49 to 98 with a mean of 73 sessions. |
| **Comparison** | Students in the comparison group received the same regular classroom reading instruction as students in the intervention group, but did not receive the tutoring program. |
| **Primary outcomes and measurement** | The Woodcock Reading Mastery Tests–Revised (WRMT-R) word identification subtest was used to test students' knowledge of alphabetics. First- and second-grade passages from the Oral Reading Fluency were used to test fluency. The WRMT-R passage comprehension subtest was used to test comprehension. Authors also looked at referral rates for special education; however this is not an outcome specified for the beginning reading topic (see Appendices A2.1–2.3 for more detailed descriptions of outcome measures). |
| **Teacher training** | The *SMART®* program intentionally places minimal demands on volunteer tutors and classroom teachers. Volunteer tutors are given 1-2 hours of training, preferably before the school year begins, but occasionally in an "on the job" setting. The training focuses as much on the logistics of tutoring as it does on reading instruction techniques. A key resource for the volunteers is a volunteer handbook, which describes four reading strategies that they can use with students: reading to the child, reading with the child, re-reading with the child, and asking the child questions about what has been read. Volunteers rely on their own judgment for any other needs. |

1. The beginning reading team does not have a set cut-off point for attrition but rather examines the pretest comparability of intervention and comparison groups after attrition. In this case, the WWC examined the baseline scores of the remaining students and found the two groups were comparable on the pretest measure.

## Appendix A2.1 Outcome measures in the alphabetics domain

| Outcome measure | Description |
|---|---|
| **Woodcock Reading Mastery Tests–Revised (WRMT-R) Word Identification subtest** | The word identification subtest is a standardized test of decoding skills. It requires the student to read aloud isolated real words that vary in frequency and difficulty. It includes 51 items (as cited in Baker, Gersten, & Keating, 2000). |

## Appendix A2.2 Outcome measures in the fluency domain

| Outcome measure | Description |
|---|---|
| **Oral Reading Fluency–First- and Second-Grade Passages** | Each student reads aloud a story from a first- or second-grade basal reader. The passages have been used in numerous other studies in the past. The number of words read correctly in one minute was used as the outcome measure (as cited in Baker, Gersten, & Keating, 2000). |

## Appendix A2.3 Outcome measures in the comprehension domain

| Outcome measure | Description |
|---|---|
| **WRMT-R Word Comprehension subtest** | This standardized measure assesses students' vocabulary through antonyms, synonyms, and analogies (as cited in Baker, Gersten, & Keating, 2000). |
| **WRMT-R Passage Comprehension subtest** | This standardized test assesses reading comprehension by having students read a text silently and fill in missing words in a short paragraph (as cited in Baker, Gersten, & Keating, 2000). |

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

| Outcome measure | Study sample[3] | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation[2]) — *SMART* group | Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group | WWC calculations: Mean difference[4] (*SMART* – comparison) | WWC calculations: Effect size[5] | WWC calculations: Statistical significance[6] (at $\alpha = 0.05$) | WWC calculations: Improvement index[7] |
|---|---|---|---|---|---|---|---|---|
| **Baker, Gersten, & Keating, 2000 (randomized controlled trial)—Two years of intervention[8]** | | | | | | | | |
| Construct: Phonics | | | | | | | | |
| Woodcock Reading Mastery Tests–Revised: Word Identification subtest | Grade 1 | 84 | 449.4 (30.2) | 437.9 (25.9) | 11.5 | 0.40 | Statistically significant | +16 |
| **Domain average[9] for alphabetics** | | | | | | 0.40 | Statistically significant | +16 |

1. This appendix reports findings considered for the effectiveness rating. Interim findings (end of first grade after one year of intervention) from the same study are not included in these ratings, but are reported in Appendix A4.1.
2. The means in the Baker, Gersten, & Keating (2000) study were adjusted for student pretest scores on two measures: the Phonemic Segmentation test and the word identification subtest of the WRMT-R. 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 sample at the beginning of the study consisted of students in first grade. Results in this table are based on outcomes assessed at the end of second grade.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison 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 intervention condition versus the percentile rank 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.
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 or multiple outcomes within one domain. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. In the case of Baker, Gersten, & Keating (2000), no corrections were needed for this domain.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

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

| Outcome measure | Study sample[3] | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation[2]) *SMART* group | Authors' findings from the study: Mean outcome (standard deviation[2]) Comparison group | WWC calculations: Mean difference[4] (*SMART* – comparison) | WWC calculations: Effect size[5] | WWC calculations: Statistical significance[6] (at $\alpha = 0.05$) | WWC calculations: Improvement index[7] |
|---|---|---|---|---|---|---|---|---|
| **Baker, Gersten, & Keating, 2000 (randomized controlled trial)—Two years of intervention[8]** | | | | | | | | |
| Oral Reading Fluency First-Grade Passage | Grade 1 | 84 | 71.3 (35.2) | 55.9 (32.1) | 15.4 | 0.45 | Statistically significant | +17 |
| Oral Reading Fluency Second-Grade Passage | Grade 1 | 84 | 61.5 (35.5) | 45.9 (29.5) | 15.6 | 0.47 | Statistically significant | +18 |
| **Domain average[9] for fluency** | | | | | | 0.46 | Statistically significant | +17 |

1. This appendix reports findings considered for the effectiveness rating. Interim findings (end of first grade after one year of intervention) from the same study are not included in these ratings, but are reported in Appendix A4.2.
2. The means in the Baker, Gersten, & Keating (2000) study were adjusted for student pretest scores on two measures: the Phonemic Segmentation test and the word identification subtest of the WRMT-R. 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 sample at the beginning of the study consisted of students in first grade. Results in this table are based on outcomes assessed at the end of second grade.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison 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 intervention condition versus the percentile rank 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.
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 or multiple outcomes within one domain. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. In the case of Baker, Gersten, & Keating (2000), corrections for multiple comparisons were needed for this domain.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A3.3 Summary of study findings included in the rating for the comprehension domain[1]

| Outcome measure | Study sample[3] | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation[2]) — *SMART* group | Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group | WWC calculations: Mean difference[4] (*SMART* – comparison) | WWC calculations: Effect size[5] | WWC calculations: Statistical significance[6] (at $\alpha = 0.05$) | WWC calculations: Improvement index[7] |
|---|---|---|---|---|---|---|---|---|
| **Baker, Gersten, & Keating, 2000 (randomized controlled trial)—Two years of intervention[8]** | | | | | | | | |
| Construct: Vocabulary development | | | | | | | | |
| Woodcock Reading Mastery Test–Revised: Word Comprehension subtest | Grade 1 | 84 | 472.30 (17.3) | 456.4 (16.2) | 6.90 | 0.41 | ns | +16 |
| Construct: Reading comprehension | | | | | | | | |
| Woodcock Reading Mastery Test-Revised: Passage Comprehension subtest | Grade 1 | 84 | 468.90 (16.0) | 464.70 (13.1) | 4.20 | 0.28 | ns | +11 |
| **Domain average[9] for comprehension** | | | | | | 0.35 | ns | +14 |

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating. Interim findings (end of first grade after one year of intervention) from the same study are not included in these ratings, but are reported in Appendix A4.3.
2. The means in the Baker, Gersten, & Keating (2000) study were adjusted for student pretest scores on two measures: the Phonemic Segmentation test and the word identification subtest of the WRMT-R. 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 sample at the beginning of the study consisted of students in first grade. Results in this table are based on outcomes assessed at the end of second grade.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison 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 intervention condition versus the percentile rank 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.
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 or multiple outcomes within one domain. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. In the case of Baker, Gersten, & Keating (2000), corrections for multiple comparisons were needed for this domain.
9. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

## Appendix A4.1 Summary of findings at the end of first grade for the alphabetics domain[1]

| Outcome measure | Study sample | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation[2]) — *SMART* group | Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group | WWC calculations: Mean difference[3] (*SMART* – comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| **Baker, Gersten, & Keating, 2000 (randomized controlled trial)—One year of intervention[7]** | | | | | | | | |
| Woodcock Reading Mastery Tests–Revised: Word Identification subtest | Grade 1 | 84 | 409.20 (29.70) | 398.90 (24.40) | 10.30 | 0.37 | ns | +15 |

ns = not statistically significant

1. This appendix presents interim findings for measures that fall in the alphabetics domain. First-grade scores, which reflect student outcomes after one year of the intervention, are reported here. Second-grade scores (after two years of the intervention) were used for rating purposes and are reported in Appendix A3.1.
2. The means in the Baker, Gersten, & Keating (2000) study were adjusted for student pretest scores on two measures: the Phonemic Segmentation test and the word identification subtest of the WRMT-R. 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of 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 the percentile rank 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 or multiple outcomes within one domain. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. In the case of Baker, Gersten, & Keating (2000), no corrections were needed for this domain.

## Appendix A4.2 Summary of findings at the end of first grade for the fluency domain[1]

| Outcome measure | Study sample | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation[2]) — *SMART* group | Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group | WWC calculations: Mean difference[3] (*SMART* – comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| **Baker, Gersten, & Keating, 2000 (randomized controlled trial)—One year of intervention[7]** | | | | | | | | |
| Oral Reading Fluency First-Grade Passage | Grade 1 | 84 | 27.80 (22.80) | 18.70 (17.30) | 9.10 | 0.44 | Statistically significant | +17 |

1. This appendix presents interim findings for measures that fall in the fluency domain. First-grade scores, which reflect student outcomes after one year of the intervention, are reported here. Second-grade scores (after two years of the intervention) were used for rating purposes and are reported in Appendix A3.2.
2. The means in the Baker, Gersten, & Keating (2000) study were adjusted for student pretest scores on two measures: the Phonemic Segmentation test and the word identification subtest of the WRMT-R. 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. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of 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 versus the percentile rank 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 or multiple outcomes within one domain. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. In the case of Baker, Gersten, & Keating (2000), no corrections were needed for this domain.

## Appendix A4.3 Summary of findings at the end of first grade for the comprehension domain[1]

| Outcome measure | Study sample | Sample size (students) | Authors' findings from the study: Mean outcome (standard deviation[2]) — *SMART* group | Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group | WWC calculations: Mean difference[3] (*SMART* – comparison) | WWC calculations: Effect size[4] | WWC calculations: Statistical significance[5] (at $\alpha = 0.05$) | WWC calculations: Improvement index[6] |
|---|---|---|---|---|---|---|---|---|
| **Baker, Gersten, & Keating, 2000 (randomized controlled trial)—One year of intervention[7]** | | | | | | | | |
| Woodcock Reading Mastery Test–Revised (WRMT-R): Passage Comprehension subtest | Grade 1 | 84 | 449.30 (24.40) | 443.20 (14.20) | 6.10 | 0.30 | ns | +12 |

ns = not statistically significant

1. This appendix presents interim findings for measures that fall in the comprehension domain. First-grade scores, which reflect student outcomes after one year of the intervention, are reported here. Second-grade scores (after two years of the intervention) were used for rating purposes and are reported in Appendix A3.3.
2. The means for the Baker, Gersten, & Keating (2000) study were adjusted for student pretest scores on two measures: the Phonemic Segmentation test and the word identification subtest of the WRMT-R. 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. 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, 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 versus the percentile rank 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 or multiple outcomes within one domain. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. In the case of Baker, Gersten, & Keating (2000), no corrections were needed for this domain.

## Appendix A5.1 *Start Making a Reader Today*® rating for the alphabetics 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 alphabetics, the WWC rated *Start Making a Reader Today*® as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC evidence standards. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because the intervention was assigned the highest applicable rating.

### Rating received

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

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

  **Met.** The single study of *SMART*® showed statistically significant 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.

  **Met.** No studies of *SMART*® showed statistically significant or substantively important negative effects, and no studies showed indeterminate effects.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  **Not met.** Only one study of *SMART*® showed statistically significant positive effects.

***and***

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

  **Met.** The single study of *SMART*® did not show statistically significant or substantively important negative 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 *Start Making a Reader Today*® rating for the fluency 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 fluency, the WWC rated *Start Making a Reader Today*® as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC evidence standards. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because the intervention was assigned the highest applicable rating.

### Rating received

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

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

  **Met.** The single study of *SMART*® showed statistically significant 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.

  **Met.** No studies of *SMART*® showed statistically significant or substantively important negative effects, and no studies showed indeterminate effects.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  **Not met.** Only one study of *SMART*® showed statistically significant positive effects.

***and***

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

  **Met.** The single study of *SMART*® did not show statistically significant or substantively important negative 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 *Start Making a Reader Today*® rating for the comprehension 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 comprehension, the WWC rated *Start Making a Reader Today*® as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC evidence standards. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because the intervention was assigned the highest applicable rating.

### Rating received

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

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

  **Met.** The single study of *SMART*® showed statistically significant 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.

  **Met.** No studies of *SMART*® showed statistically significant or substantively important negative effects, and no studies showed indeterminate effects.

### Other ratings considered

**Positive effects:** Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  **Not met.** Only one study of *SMART*® showed statistically significant positive effects.

***and***

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

  **Met.** The single study of *SMART*® did not show statistically significant or substantively important negative 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] |
|---|---|---|---|---|
| | | Schools | Students | |
| Alphabetics | 1 | 6 | 84 | Small |
| Comprehension | 1 | 6 | 84 | Small |
| Fluency | 1 | 6 | 84 | Small |
| General reading achievement | 0 | 0 | 0 | na |

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."