What Works Clearinghouse™

ies INSTITUTE OF EDUCATION SCIENCES

Science January 2013

Great Explorations in Math and Science® (GEMS®) The Real Reasons for Seasons

Report Contents

Program Description[1]

Great Explorations in Math and Science® (GEMS®) The Real Reasons for Seasons is a curriculum unit for grades 6–8 that focuses on the connections between the Sun and the Earth to teach students the scientific concepts behind the seasons. The unit utilizes models, hands-on investigations, peer-to-peer discussions, reflection, and informational student readings to help students understand science content and develop scientific investigation skills.

Research[2]

The What Works Clearinghouse (WWC) identified one study of *GEMS® The Real Reasons for Seasons* that both falls within the scope of the Science topic area and meets WWC evidence standards. This one study meets standards without reservations and included 4,777 seventh-grade students in 10 middle schools in Maryland.

The WWC considers the extent of evidence for *GEMS® The Real Reasons for Seasons* on the science performance of middle school students to be small for one outcome domain: general science achievement. (See the Effectiveness Summary on p. 4 for further description of the domain.)

Effectiveness

GEMS® The Real Reasons for Seasons was found to have potentially negative effects on general science achievement for middle school students.

Table 1. Summary of findings[3]

		Improvement index *(percentile points)*				
Outcome domain	Rating of effectiveness	Average	Range	Number of studies	Number of students	Extent of evidence
General science achievement	Potentially negative effects	–10	–14 to –6	1	4,777	Small

Program Information

Background

GEMS® The Real Reasons for Seasons was developed at the Lawrence Hall of Science, the public science and mathematics curriculum development and educational research center of the University of California, Berkeley. The unit is available from the *GEMS®* distributor, Carolina Curriculum. Address: Carolina Biological Supply Company, 2700 York Road, Burlington, NC 27215–3398. Email: curriculum@carolina.com. Web: http://www.carolinacurriculum.com/GEMS/. Telephone: (800) 334-5551.

Program details

GEMS® The Real Reasons for Seasons is a three-week curriculum unit designed for grades 6–8 that consists of eight hands-on activities focusing on Sun–Earth connections. Each of the eight activities requires 30–90 minutes of class time and builds on key concepts in earth and space science. The unit comes with a teacher's guide, a materials kit, and master copies for duplication or electronic presentation. Working in small groups, students explore the role of models and evidence in science. During the class sessions, students take a "Trip to the Sun" to determine the real shape of the Earth's orbit, evaluate data on world temperature and hours of sunlight in different locations, and model how the angle at which the Sun's rays strike a surface affect their concentration. These activities target core science concepts and common misconceptions that students might have about them. Students are encouraged to evaluate alternative explanations of concepts, use evidence to support them, and critique the merits of an explanation.

Cost

GEMS® The Real Reasons for Seasons teacher's guide costs $28 (rate effective January 2012). The guide includes an assessment system and a CD-ROM, which offers a collection of resources, software programs, and web links. Cost information for other *GEMS®* products is available from the program distributor, Carolina Curriculum. *GEMS®* network sites and centers also provide ongoing training and support for teachers on how to use *GEMS®* within their larger curriculum.

Research Summary

The WWC identified one study on the effects of *GEMS® The Real Reasons for Seasons* on the science achievement of middle school students.

The WWC reviewed this study against group design evidence standards. The study (Pyke, Lynch, Kuipers, Szesze, & Watson, 2004) is a randomized controlled trial that meets WWC evidence standards without reservations. This study is summarized in this report. The citation for this study is in the References section, which begins on p. 5.

Table 2. Scope of reviewed research

Grade	7
Delivery method	Small group/Whole class
Program type	Curriculum
Studies reviewed	1 study
Group design studies that meet WWC evidence standards	
• **without reservations**	1 study
• **with reservations**	0 studies

Summary of studies meeting WWC evidence standards without reservations

Pyke et al. (2004) examined the effects of *GEMS® The Real Reasons for Seasons* on seventh-grade students' knowledge and understanding of earth and space science. The authors studied two cohorts of students in 10 Maryland schools.[4]

The authors used a three-step process to randomly assign schools to the intervention and comparison groups. First, Pyke et al. (2004) grouped all district schools into five school profile categories, each having similar demographic and achievement characteristics. Next, the authors selected one pair of schools from each of the five school profile categories. Finally, one school from each pair was randomly assigned either to implement *GEMS® The Real Reasons for Seasons* or to serve as a comparison school and use the regular science curriculum. Through this process, 10 schools were identified for participation in the study. In the first year of the study, seventh-grade students in these 10 schools were referred to as Cohort 1. In the second year, seventh-grade students in the same 10 schools were referred to as Cohort 2.[5]

Cohort 1 was formed in the 2003–04 school year and consisted of 1,318 seventh-grade students who received *GEMS® The Real Reasons for Seasons* and 1,051 seventh-grade students in the comparison group. Cohort 2 was formed in the 2004–05 school year and consisted of 1,287 seventh-grade students who received *GEMS® The Real Reasons for Seasons* and 1,121 seventh-grade students in the comparison group. The total study sample (Cohorts 1 and 2) included 4,777 seventh-grade students.

Summary of studies meeting WWC evidence standards with reservations

No studies of *GEMS® The Real Reasons for Seasons* meet WWC evidence standards with reservations.

Effectiveness Summary

The WWC review of *GEMS® The Real Reasons for Seasons* for the Science topic includes student outcomes in one domain: general science achievement. The domain includes three outcome constructs: life science, earth/space science, and physical science. The one study of *GEMS® The Real Reasons for Seasons* that meets WWC evidence standards reported findings that cover one construct: earth/space science. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *GEMS® The Real Reasons for Seasons* on the science achievement of middle school students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 12.

Summary of effectiveness for the general science achievement domain

One study reported findings in the general science achievement domain.

Pyke et al. (2004) reported statistically significant negative effects of *GEMS® The Real Reasons for Seasons* on concept assessments for both Cohort 1 and Cohort 2 seventh-grade students. According to WWC calculations, the effects were not statistically significant (when adjusted for clustering), but the average effect across the two cohorts was large enough to be considered substantively important according to WWC criteria (i.e., an effect size of at least 0.25).

Thus, for the general science achievement domain, one study showed substantively important negative effects. This results in a rating of potentially negative effects, with a small extent of evidence.

Table 3. Rating of effectiveness and extent of evidence for the general science achievement domain

Rating of effectiveness	Criteria met
Potentially negative effects *Evidence of a negative effect with no overriding contrary evidence.*	In the one study that reported findings, the estimated impact of the intervention on outcomes in the *general science achievement domain* was negative and substantively important.
Extent of evidence	**Criteria met**
Small	One study that included 4,777 students in 10 schools reported evidence of effectiveness in the *general science achievement domain*.

References

Study that meets WWC evidence standards without reservations

Pyke, C., Lynch, S., Kuipers, J., Szesze, M., & Watson, W. (2004). *Implementation study of The Real Reasons for Seasons (2003–2004): SCALE-uP Report No. 4*. Washington, DC: George Washington University, SCALE-uP.

Additional sources:

Pyke, C., Lynch, S., Kuipers, J., Szesze, M., & Watson, W. (2005). *Implementation study of The Real Reasons for Seasons (2004–2005): SCALE-uP Report No. 7*. Washington, DC: George Washington University, SCALE-uP.

Rethinam, V., Pyke, C., & Lynch, S. (2008). Using multi-level analyses to study the effectiveness of science curriculum materials. *Evaluation & Research in Education, 21*(1), 18–42.

Studies that meet WWC evidence standards with reservations

None

Studies that do not meet WWC evidence standards

None

Studies that are ineligible for review using the Science Evidence Review Protocol

None

Appendix A: Research details for Pyke et al. (2004)

Pyke C., Lynch, S., Kuipers, J., Szesze, M., & Watson, W. (2004). *Implementation study of The Real Reasons for Seasons (2003–2004): SCALE-uP Report No. 4.* Washington, DC: George Washington University, SCALE-uP.

Table A. Summary of findings

Meets WWC evidence standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
General science achievement	10 schools/4,777 students	–10	No

Setting

The study took place in 10 schools in Maryland's Montgomery County School District. The student population of this large suburban district was 43% White, 22% African American, 14% Asian American, and 20% Hispanic. The study was part of a multiyear research project called "Scaling up Curriculum for Achievement, Learning, and Equity Project" (SCALE-uP).[6]

Study sample

In this randomized controlled trial, researchers created a sampling frame consisting of five school profile categories, with approximately seven schools in each category. The sampling frame was based on achievement and demographic factors. Each school category had a similar profile determined by: the percentage of students eligible for free and reduced-price meals, math and reading achievement scores, ethnicity, eligibility for English for Speakers of Other Languages (ESOL) services, and eligibility for special education (SPED) services. Two schools were randomly selected from each category to participate in the study. In each category, one school of the matched pair was then randomly chosen to implement the intervention, and the other served as the comparison school. The study school sample consisted of five schools implementing *GEMS® The Real Reasons for Season*s and five schools not implementing it.

The analysis is based on two cohorts of seventh-grade students that attended the study schools during two consecutive school years. Cohorts 1 and 2 consisted of seventh-grade students in the 2003–04 and 2004–05 school years, respectively. The Cohort 1 analysis sample included 1,318 seventh-grade students who received *GEMS® The Real Reasons for Season*s and 1,051 seventh-grade students who received the regular science curriculum. Cohort 2 included 1,287 seventh-grade students who received *GEMS® The Real Reasons for Season*s and 1,121 seventh-grade students who received the regular science curriculum. Overall and differential attrition rates of students were low for Cohort 1 (9% and 3%, respectively) and Cohort 2 (13% and 6%, respectively). The study reported student outcomes for the two cohorts after the completion of the unit; these findings are included in the WWC effectiveness rating and can be found in Appendix C. Additional findings for Cohort 1 subgroups by gender, race/ethnicity, eligibility for ESOL services, and eligibility for SPED services are considered supplemental findings by the WWC and can be found in Appendix D.

Intervention group

The intervention teachers implemented the eight activities of the *GEMS® The Real Reasons for Seasons* curriculum unit over a period of three weeks. Each activity required about 30–90 minutes of class time. The curriculum unit addressed common misconceptions about seasons and was designed to either validate students' accurate ideas about seasons or to address common problems students experienced when learning about seasons. The curriculum came with a teacher's guide, student lab materials, and master copies for duplication or electronic presentation. Montgomery County Public Schools purchased and distributed to teachers all student lab materials needed for use with the unit. The *GEMS® The Real Reasons for Seasons* curriculum was embedded in a larger astronomy unit using the district-approved curriculum.

Comparison group

Comparison group teachers used regular curriculum materials normally available to Montgomery County Public Schools' teachers. The district materials addressed the same instructional benchmarks as the *GEMS® The Real Reasons for Seasons* curriculum unit.

Outcomes and measurement

Students took a concept assessment test for both the pretest and posttest. For Cohort 1, the authors used the Reasons for the Seasons Assessment (RSA). For Cohort 2, the authors used the Causes for the Seasons Assessment (CSA). Although named differently, essentially the same concept test was used for data analysis for both cohorts of students. For a more detailed description of this outcome measure, see Appendix B. Study authors also assessed each student's personal orientation toward learning using the Science Learning Orientation and Engagement for Students Questionnaire. This outcome measure is outside the scope of the Science review protocol and this review.

Support for implementation

The study did not describe any information about training provided to teachers or staff.

Appendix B: Outcome measures for each domain

General science achievement	
Earth/space science construct	
Concept Assessment–Reasons for the Seasons Assessment (RSA) score	The RSA Concept Assessment consists of 15 items (10 constructed response and five selected response) that require understanding of the reasons for seasons. Each item relates to one of the following: sun and shadows, sun and water temperature, the Earth's rotation, concentration of sunlight, and spatial representation of the Earth's orbit and tilt. Student responses to the constructed items were judged by trained raters. For the selected response items, students were presented with a set of responses from which they chose the best answer. Three assessment items were excluded from data analysis. The excluded items were either redundant with the other test items or showed a very low coefficient for discrimination. Cronbach's alpha for the remaining 12 items of the RSA was 0.72 (as cited in Pyke et al., 2004).
Concept Assessment–Causes for the Seasons Assessment (CSA) score	The CSA Concept Assessment consists of 12 items (eight constructed response and four selected response) that require understanding of the reasons for seasons. The CSA excluded at the outset three items that performed poorly in the RSA. Each item relates to one of the following: sun and shadows, sun and water temperature, the Earth's rotation, concentration of sunlight, and spatial representation of the Earth's orbit and tilt. Student responses to the constructed items were judged by trained raters. For the selected response items, students were presented with a set of responses from which they chose the best answer. Internal consistency of Cronbach's alpha for the CSA was 0.73 (as cited in Pyke et al., 2005).

Appendix C: Findings included in the rating for the general science achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Pyke et al., 2004[a]								
Concept Assessment–RSA score	Grade 7 Cohort 1	10 schools/ 2,369 students	27.80 (20.00)	35.39 (21.86)	–7.59	–0.36	–14	< 0.05
Concept Assessment–CSA score	Grade 7 Cohort 2	10 schools/ 2,408 students	38.54 (23.59)	42.26 (22.88)	–3.72	–0.16	–6	< 0.05
Domain average for general science achievement (Pyke et al., 2004)						**–0.26**	**–10**	**Not statistically significant**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC. Outcomes for each cohort were provided in two separate reports. The report for Cohort 2 (Pyke et al., 2005) is included as an additional source of information for the Pyke et al. (2004) study in the references section. RSA = Reasons for the Seasons Assessment. CSA = Causes for the Seasons Assessment.

[a] The *p*-values presented here were reported in the original study. For Pyke et al. (2004), a correction for clustering was needed and resulted in significance levels that differ from those in the original study. As a result of the clustering adjustment, the WWC does not find the results for the RSA and CSA scores to be statistically significant. The intervention group mean outcome values are the unadjusted comparison group posttest means plus the difference in mean gains between the intervention and comparison groups. Comparison group means are unadjusted. The data reported in the table for Cohort 1 and Cohort 2 were provided by the author to the WWC and are not the data included in the original reports. This study is characterized as having a substantively important negative effect, because no effects are statistically significant within the domain and the negative mean effect is at least 0.25.

Appendix D: Description of supplemental findings for the general science achievement domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Pyke et al., 2004[a]								
Grade 7, Cohort 1								
Concept Assessment–RSA score	Males	10 schools/ 1,241 students	29.72 (20.23)	36.00 (22.12)	–6.28	–0.30	–12	> 0.05
Concept Assessment–RSA score	Females	10 schools/ 1,128 students	26.76 (19.57)	34.76 (21.44)	–8.00	–0.39	–15	> 0.05
Concept Assessment–RSA score	White	10 schools/ 1,111 students	34.14 (19.40)	39.99 (20.33)	–5.85	–0.29	–12	> 0.05
Concept Assessment–RSA score	Never ESOL	10 schools/ 1,929 students	29.76 (19.97)	37.38 (21.58)	–7.62	–0.37	–14	> 0.05
Concept Assessment–RSA score	Prior ESOL	10 schools/ 314 students	21.22 (17.87)	29.28 (21.76)	–8.06	–0.41	–16	> 0.05
Concept Assessment–RSA score	Not SPED	10 schools/ 2,090 students	28.62 (20.06)	36.72 (21.64)	–8.10	–0.39	–15	< 0.05
Concept Assessment–RSA score	Current SPED	10 schools/ 279 students	22.82 (15.43)	24.06 (15.61)	–1.24	–0.08	–3	> 0.05

Table Notes: The supplemental findings presented in this table are additional subgroup findings for Cohort 1 students from Pyke et al. (2004) that do not factor into the determination of the intervention rating. Student subgroups include gender, ethnicity, students' status as English language learners (ESOL), and eligibility for special education (SPED) services. For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the average change expected for all students who are given the intervention (measured in standard deviations of the outcome measure). The improvement index is an alternate presentation of the effect size, reflecting the change in an average student's percentile rank that can be expected if the student is given the intervention. The WWC reports findings only for the subgroups of interest that are equivalent with regard to pretest scores: Males, Females, White, Never ESOL, Prior ESOL, Not SPED, Current SPED. Subgroup findings for Cohort 2 (Pyke et al., 2005) are excluded from this report because the authors imputed a missing pretest or posttest value for 675 cases. "Never ESOL" students are those whose primary language is English and who have never been classified as English language learners. "Prior ESOL" students are those who have previously been enrolled in the ESOL instructional program but are currently either in their first year of transition from the ESOL program to the general education program or have achieved proficiency in English and are no longer considered transition students. "Not SPED" students are those who are not currently eligible for special education services. "Current SPED" students are those who are currently eligible for special education services and who are taught science in mainstream classrooms. RSA = Reasons for the Seasons Assessment.

[a] The *p*-values presented for special education services' subgroups ("Not SPED" and "Current SPED") were reported in the original study. For other subgroups, *p*-values presented in Appendix D were not reported in the original study but were computed by the WWC. For Pyke et al. (2004), corrections for clustering and multiple comparisons were needed and resulted in significance levels that differ from those in the original study. When adjusted for clustering, the WWC-calculated effect on the RSA score for the "Not SPED" services' subgroup was not statistically significant ($p = 0.18$). For special education services' subgroups, the intervention and comparison group mean outcome values are ANCOVA-adjusted posttest scores, with pretest scores being treated as a covariate. For all other subgroups, the intervention group mean outcome values are the unadjusted comparison group posttest means plus the difference in mean gains between the intervention and comparison groups.

Endnotes

[1] The descriptive information for this program was obtained from publicly available sources: the developer's website (http://www.lawrencehallofscience.org/gems, downloaded June 2011) and the distributor's website (http://www.carolinacurriculum.com/GEMS/About+GEMS.asp, downloaded June 2011). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in August 2011, and we incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review. The literature search reflects documents publicly available by May 2012.

[2] The studies in this report were reviewed using the Evidence Standards from the WWC Procedures and Standards Handbook (version 2.1), along with those described in the Science review protocol (version 2.0). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

[3] For criteria used in the determination of the rating of effectiveness and extent of evidence, see the WWC Rating Criteria on p. 12. These improvement index numbers show the average and range of student-level improvement indices for all findings across the study.

[4] The study results are presented in separate research reports. Findings for Cohort 1 students are reported in Pyke et al. (2004). Findings for Cohort 2 students are reported in Pyke et al. (2005). Although both sources contribute to the effectiveness rating in this review, the WWC conventionally lists Pyke et al. (2004) as a primary reference for the whole study.

[5] The WWC review team has determined that this review should consider the analysis of the Cohort 2 students as providing evidence of the effect of *GEMS® The Real Reasons for Seasons* on student performance, despite the potential threat of validity associated with selection bias associated with student mobility occurring between the time of random assignment and the implementation of the program in Year 2. The Science topic area principal investigator has determined that a selection bias stemming from student mobility is unlikely to have affected the observed impacts at the end of Cohort 2, primarily because science was not an elective course.

[6] SCALE-uP is funded by the Interagency Education Research Initiative and administered by the National Science Foundation. In the first two years of the project (referred to by the authors as Project Years 0 and 1), authors reported the results of *Chemistry That Applies* (State of Michigan, 1993) on eighth-grade students in the Montgomery County Public Schools in Maryland. In Years 2 and 3, presented in this intervention report, authors reported the results of the first and second year of implementation of the curriculum unit *GEMS® The Real Reasons for Seasons* study (Lawrence Hall of Science, 2000) for seventh-grade students in the same district. In Years 2, 3, and 4, authors also reported the results of the first, second, and third year of implementation of the curriculum unit *Exploring Motion and Forces: Speed, Acceleration, and Friction* (Harvard-Smithsonian Center for Astrophysics, 2001) for sixth-grade students in the same district.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, January). Science intervention report: *Great Explorations in Math and Science® (GEMS®) The Real Reasons for Seasons*. Retrieved from http://whatworks.ed.gov.

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC evidence standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC evidence standards with reservations	A study that provides weaker evidence for an intervention's effectiveness, such as a QED or an RCT with high attrition that has established equivalence of the analytic samples.

Criteria used to determine the rating of effectiveness for an intervention

Rating of effectiveness	Criteria
Positive effects	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important negative effects.
Potentially positive effects	At least one study shows a statistically significant or substantively important positive effect, AND No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects	At least one study shows a statistically significant or substantively important positive effect AND at least one study shows 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, OR At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects	One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND No studies show statistically significant or substantively important positive effects.
No discernible effects	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Criteria used to determine the extent of evidence for an intervention

Extent of evidence	Criteria
Medium to large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Extent of evidence An indication of how much evidence supports the findings. The criteria for the extent of evidence levels are given in the WWC Rating Criteria on p. 12.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Rating of effectiveness The WWC rates the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance, and consistency in findings. The criteria for the ratings of effectiveness are given in the WWC Rating Criteria on p. 12.

Single-case design A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.