What Works Clearinghouse

ies INSTITUTE OF EDUCATION SCIENCES

Science February 2012

Chemistry That Applies

Program Description[1]

Chemistry That Applies is an instructional unit designed to help students in grades 8–10 understand the law of conservation of matter. It consists of 24 lessons organized in four clusters. Working in groups, students explore four chemical reactions: burning, rusting, the decomposition of water, and the reaction of baking soda and vinegar. As part of the unit, students conduct experiments in which they cause these reactions to happen, obtain and record data in individual notebooks, analyze the data, and use evidence-based arguments to explain the data. The instructional unit engages the students in a structured sequence of hands-on laboratory investigations interwoven with other forms of instruction.

Research[2]

One study of *Chemistry That Applies* that falls within the scope of the Science review protocol meets What Works Clearinghouse (WWC) evidence standards. The one study included more than 4,000 students in grade 8 in 10 middle schools in Maryland. Based on this study, the WWC considers the extent of evidence for *Chemistry That Applies* on middle school students to be small for the general science achievement domain.

Report Contents

Overview	p. 1
Program Information	p. 2
Research Summary	p. 3
Effectiveness Summary	p. 4
References	p. 5
Research Details for Each Study	p. 6
Outcomes Measures for Each Domain	p. 8
Findings Included in the Rating for Each Outcome Domain	p. 9
Supplemental Findings for Each Outcome Domain	p. 10
Endnotes	p. 12
Rating Criteria	p. 13
Glossary of Terms	p. 14

Effectiveness

Chemistry That Applies was found to have potentially positive effects on general science achievement for middle school students.

Table 1. Summary of findings[3]

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
General science achievement	Potentially positive effects	+11	+10 to +13	1	4,176	Small

Program Information

Background

Chemistry That Applies was developed by researchers[4] at the Michigan Department of Education. Address: Michigan Department of Education, Office of Education Options, Public School Academy Program, Hannah Building, 4th Floor, 608 W. Allegan St., P.O. Box 30008, Lansing, MI 48909. Web: http://www.michigan.gov/mde. Telephone: (517) 241-4715. Fax: (517) 241-0197.

Program details

Chemistry That Applies is a six- to ten-week instructional unit composed of 24 lessons organized in four clusters. Students explore the same four chemical reactions as the units advance in order to understand conservation of matter. In Cluster 1, students mix substances and describe the changes that occur. Students learn to write accurate descriptions of reactants and products and use their descriptions to find evidence of the formation of new substances. In Cluster 2, students predict weight changes during physical and chemical reactions and test their predictions with lab activities, which lead students to understand the law of conservation of matter. In Cluster 3, students build models of the observed chemical reactions, which demonstrate that atoms are conserved as new molecules are formed. In Cluster 4, students explore the energy changes that take place between reactants and products.

Throughout the unit, students learn to conduct their own research of a chemical substance, pose questions, search for solutions to problems, work with others, and value the need for evidence in making decisions. They apply the concepts learned in each cluster to their specific substance as they learn its chemical name, physical properties, history, uses, chemical composition, disposal method, and energy requirements. The activities within the investigation enable students to observe change in the states of matter shown by the chemical reactions (i.e., solid to liquid, liquid to solid, and liquid to gas). At the conclusion of the unit, students make presentations.

Cost

The curriculum materials consist of a teacher's guide, which describes lessons, optional activities, and readings. The guide is available for free download on the George Washington University website (http://www.gwu.edu/~scale-up/documents/CTA.pdf).

Research Summary

Two studies reviewed by the WWC Science Topic Area investigated the effects of *Chemistry That Applies* on middle school students. One study (Pyke, Lynch, Kuipers, Szesze, & Driver, 2004), summarized in this report, is a randomized controlled trial that meets WWC evidence standards. The remaining study does not meet WWC eligibility screens. (See references beginning on page 5 for citations for both studies.)

Table 2. Scope of reviewed research

Grade	8
Delivery method	Small group/Whole class
Program type	Curriculum
Studies reviewed	2
Meets WWC standards	1 study
Meets WWC standards with reservations	0 studies

Summary of a study meeting WWC evidence standards without reservations

Pyke et al. (2004) conducted a randomized controlled trial that examined the effects of *Chemistry That Applies* on eighth-grade students' knowledge and understanding of physical science. Two separate cohorts were formed, and the results were presented in three research reports.[5] These reports have been combined into a single study for this review. The total study sample included 4,176 eighth-grade students attending 10 middle schools in Maryland.

The study used a school-level random design that involved a three-step process. 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 the intervention or to serve as a control school. Through this process, 10 study schools were identified.

Two cohorts of eighth-grade students attended the study schools during two consecutive school years. Cohort 1 was formed in the 2001–02 school year and consisted of 1,087 eighth-grade students who received *Chemistry That Applies* and 809 eighth-grade students in the control group. Cohort 2 was formed in the 2002–03 school year and consisted of 1,121 eighth-grade students who received *Chemistry That Applies* and 1,159 eighth-grade students in the control group. All control group students were taught using their school's regular science curriculum. The study reported students' outcomes after approximately seven weeks of program implementation.

Summary of studies meeting WWC evidence standards with reservations

No studies of *Chemistry That Applies* meet WWC evidence standards with reservations.

Effectiveness Summary

The WWC review of interventions for Science addresses student outcomes in one domain: general science achievement. The domain includes three outcome constructs: life science, earth/space science, and physical science. The study that contributes the effectiveness rating of this report covers one construct: physical science. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Chemistry That Applies* on middle school students. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria later in this report.

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 positive effects of *Chemistry That Applies* on Conservation of Matter Assessment for both the Cohort 1 and Cohort 2 eighth-grade students. According to WWC calculations, the effects were not statistically significant (when adjusted for clustering), but were 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 positive effects. This results in a rating of potentially positive 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 positive effects *Evidence of a positive effect with no overriding contrary evidence.*	The review of *Chemistry That Applies* had one study showing a substantively important positive effect and no studies showing a statistically significant or substantively important negative effect or indeterminate effects.
Extent of evidence	**Criteria met**
Small	The review of *Chemistry That Applies* for the general science achievement domain was based on one study that included 10 schools and 4,176 students.

References

Study that meets WWC evidence standards without reservations

Pyke C., Lynch, S., Kuipers, J., Szesze, M., & Driver, H. (2004). *Implementation study of Chemistry That Applies (2002–2003): SCALE-uP Report No. 2*. Washington, DC: George Washington University and Montgomery County Public Schools.

Additional sources:

Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching, 42*(8), 912–946.

Lynch, S., Taymans, J., Watson, W. A., Ochsendorf, R. J., & Pyke, C. (2007). Effectiveness of a highly rated science curriculum unit for students with disabilities in general education classrooms. *Council for Exceptional Children, 73*(2), 202–223.

Study that is ineligible for review using the Science Evidence Review Protocol

Lynch, S., Pyke, C., & Jansen, J. (2003). Deepening understanding of science and mathematics education reform principles: Novice teachers design web-based units using Project 2061's curriculum analysis. *Journal of Science Teacher Education, 14*(3), 193–216. The study is ineligible for review because it does not use a sample aligned with the protocol—the sample is not within the specified age or grade range.

Appendix A: Research details for Pyke et al., 2004

Pyke C., Lynch, S., Kuipers, J., Szesze, M., & Driver, H. (2004). *Implementation study of Chemistry That Applies (2002–2003): SCALE-uP Report No. 2*. Washington, DC: George Washington University and Montgomery County Public Schools.

Table A. Summary of findings

Meets WWC standards

Outcome domain	Sample size	Study findings	
		Average improvement index *(percentile points)*	Statistically significant
General science achievement	10 schools/4,176 students	+ 11	No

Setting

The study took place in 10 schools in Montgomery County Public Schools, a large, suburban school district in Maryland. The study population has no ethnic majority and is among the highest performing in Maryland.

Study sample

In this randomized study,[6] researchers created a sampling frame consisting of five profile categories, with approximately seven schools in each category. Each school category has a similar demographic and achievement profile determined by 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 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 was the comparison school. The study school sample consisted of five schools implementing the intervention and five schools not implementing it. The analysis is based on two cohorts of eighth-grade students that attended the study schools during two consecutive school years. Cohort 1 was formed in the 2001–02 school year and consisted of 1,087 eighth-grade students who received *Chemistry That Applies* in the five intervention schools and 809 eighth-grade students in the five comparison schools who received a regular science curriculum. Cohort 2 was formed in the 2002–03 school year in the same schools and consisted of 1,121 eighth-grade students who received *Chemistry That Applies* in the five intervention schools and 1,159 eighth-grade students in the five comparison schools who received a regular science curriculum. Differential attrition rate of students was low for Cohort 2 (3%) and high for Cohort 1 (13%). Because of the high attrition[7] in Cohort 1, the WWC confirmed that baseline equivalence for Cohort 1 intervention and comparison groups was demonstrated.[8] The study reported student outcomes for the two cohorts after seven weeks of program implementation; these findings can be found in Appendix C. Additional findings for subgroups by gender, race/ethnicity, students in the Free and Reduced Price Lunch (FRPL) program, those in the ESOL program, and those eligible for Special Education (SPED) can be found in Appendix D.

Intervention group

The curriculum unit employed by the experimental group was *Chemistry That Applies* (State of Michigan, 1993). *Chemistry That Applies* is a middle school science curriculum that received an acceptable rating by Project 2061, a curriculum analysis project funded by the Interagency Educational Research Initiative of the National Science Foundation (NSF). *Chemistry That Applies* consists of 24 lessons. In this study, teachers were instructed to cover the first 18 lessons only because the topics covered in the last six lessons were not part of the district curriculum and hence not covered in the comparison group. *Chemistry That Applies* focuses on "guided inquiry" with hands-on, student-centered material. Working in large and small groups, students explore chemical reactions, collect data, and use evidence-based arguments to support their claims. Students keep individual science notebooks for analyzing results. *Chemistry That Applies* provides question prompts (called "Think and Write") that require students to use critical thinking skills. Complicated vocabulary is kept to a minimum. The unit is implemented over a period of approximately seven weeks.

Comparison group

Comparison group teachers used regular curriculum materials normally available to Montgomery County Public Schools teachers that addressed the same target benchmarks. The comparison group curriculum comes from a range of sources, including traditional textbooks, Prentice Hall, reform-based NSF-funded materials, and teacher-designed materials. All teachers were exposed to professional development and "reform-based" strategies.

Outcomes and measurement

For both the pretest and the posttest, students took the Conservation of Matter Assessment (COMA). For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

All intervention group eighth-grade science teachers participated in two days of professional development. They also were given a box of lab materials, instructions for implementation, and an unspecified number of follow-up meetings during the school year. All teachers had access to their regular professional development meetings.

Appendix B: Outcome measures for each domain

General science achievement	
Physical science construct	
Conservation of Matter Assessment (COMA)	The Conservation of Matter Assessment was created to align with the middle-grade science standards, as articulated in Benchmarks for Science Literacy (AAAS, 1993).[9] The concept assessment consists of 10 items (four constructed response and six selected response) that focus on four phenomena that require understanding of conservation of matter, such as "closed versus open systems," appearance or disappearance of substances, and chemical or physical changes. Inter-rater reliability for the four constructed items was based on a 2% sample of assessments and had kappa scores from 0.7 to 0.81. Cronbach's alpha estimated on the entire sample for the 10 items was 0.71. Ideas on the conservation of matter made up 60% of the exam; ideas about atoms made up 40%. Scores were then mapped into a 0 to 100 range: 0–2 to 0–23, 3–5 to 24–50, 6–8 to 51–70, and 9–10 to 71–100 (as cited in Pyke et al., 2004).

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]								
Conservation of Matter Assessment	Grade 8/ Cohort 1	10 schools/ 1,896 students	41.68 (29.55)	32.71 (25.84)	8.97	0.32	+13	< 0.05
Conservation of Matter Assessment	Grade 8/ Cohort 2	10 schools/ 2,280 students	50.22 (30.09)	42.73 (29.66)	7.49	0.25	+10	< 0.05
Domain average for general science achievement (Pyke et al., 2004)						**0.29**	**+11**	**ns**

Table Notes: Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if that student is given the intervention. 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. Findings for Cohort 1 students are reported in Lynch et al. (2005). Findings for Cohort 2 students are reported in Pyke et al. (2004) and Lynch et al. (2005). See the References section for more information. 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 a study's domain average was determined by the WWC. ns = not statistically significant.

[a] For Pyke et al. (2004), a correction for clustering was needed and resulted in significance levels that differ from those in the original study. The p-values presented here were reported in the original study. For Cohort 1 (Lynch et al., 2005), the group mean outcomes values are unadjusted posttest means. For Cohort 2 (Pyke et al., 2004), the intervention and control group mean outcome values are ANCOVA-adjusted posttest scores, with pretest scores being treated as a covariate.

Appendix D: Summary of subgroup 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 8, Cohort 1								
Conservation of Matter Assessment	Females	10 schools/ 648 students	42.86 (28.89)	33.89 (23.75)	8.97	0.33	+13	< 0.05
Conservation of Matter Assessment	African American	10 schools/ 395 students	36.07 (26.68)	25.17 (22.56)	10.90	0.43	+17	< 0.05
Conservation of Matter Assessment	White	10 schools/ 444 students	51.04 (28.55)	45.44 (27.57)	5.60	0.20	+8	< 0.05
Conservation of Matter Assessment	Prior FRPL	10 schools/ 247 students	39.63 (27.72)	28.50 (22.99)	11.13	0.43	+16	< 0.05
Conservation of Matter Assessment	Current FRPL	10 schools/ 349 students	33.17 (26.10)	21.98 (18.84)	11.19	0.46	+18	< 0.05
Conservation of Matter Assessment	Current ESOL	10 schools/ 54 students	19.49 (24.04)	21.53 (23.19)	–2.04	–0.08	–3	> 0.05
Grade 8, Cohort 2								
Conservation of Matter Assessment	Males	10 schools/ 1/163 students	49.56 (30.50)	42.30 (30.07)	7.26	0.24	+9	< 0.05
Conservation of Matter Assessment	Females	10 schools/ 1/119 students	50.87 (29.64)	43.14 (29.26)	7.73	0.26	+10	< 0.05
Conservation of Matter Assessment	African American	10 schools/ 692 students	42.78 (27.31)	35.26 (23.56)	7.52	0.29	+12	< 0.05
Conservation of Matter Assessment	Asian American	10 schools/ 268 students	54.56 (28.74)	50.63 (32.12)	3.93	0.13	+5	> 0.05
Conservation of Matter Assessment	Hispanic	10 schools/ 544 students	45.03 (28.70)	34.95 (24.97)	10.10	0.37	+15	< 0.05
Conservation of Matter Assessment	White	10 schools/ 778 students	59.53 (29.06)	51.49 (29.79)	8.04	0.27	+11	< 0.05
Conservation of Matter Assessment	Never FRPL	10 schools/ 1/128 students	55.53 (29.72)	50.38 (29.98)	5.15	0.17	+7	< 0.05
Conservation of Matter Assessment	Prior FRPL	10 schools/ 486 students	46.51 (28.51)	38.91 (27.72)	7.60	0.27	+11	< 0.05
Conservation of Matter Assessment	Current FRPL	10 schools/ 630 students	44.31 (27.87)	32.32 (22.21)	11.99	0.48	+18	< 0.05
Conservation of Matter Assessment	Never ESOL	10 schools/ 1/716 students	52.12 (30.19)	46.17 (29.99)	5.95	0.20	+8	< 0.05
Conservation of Matter Assessment	Prior ESOL	10 schools/ 388 students	46.53 (28.78)	34.14 (25.21)	12.39	0.46	+18	< 0.05
Conservation of Matter Assessment	Current ESOL	10 schools/ 140 students	39.04 (24.63)	27.52 (20.24)	11.52	0.51	+20	< 0.05

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Conservation of Matter Assessment	Not SPED	10 schools/ 2,042 students	51.34 (29.84)	43.89 (29.91)	7.45	0.25	+10	< 0.05
Conservation of Matter Assessment	Current SPED	10 schools/ 202 students	40.68 (29.66)	32.99 (29.93)	7.69	0.26	+10	< 0.05

Table Notes: The supplemental findings presented in this table are additional subgroup findings from the studies in this report that do not factor in the determination of the intervention rating. Student subgroups include gender, ethnicity, socioeconomic status as indicated by eligibility for the Free and Reduced Price Meals System (FRPL; abbreviated as FARMS in the original report), students' status as English language learners (ESOL), and eligibility for special education services (SPED). Total group scores were used for rating purposes and are presented in Appendix C. Positive results for mean difference, effect size, and improvement index favor the intervention group; negative results favor the comparison group. The effect size is a standardized measure of the effect of an intervention on student outcomes, representing the change (measured in standard deviations) in an average student's outcome that can be expected if that student is given the intervention. 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. Findings for Cohort 1 students are reported in Lynch et al. (2005). Findings for cohort 2 students are reported in Lynch et al. (2007). See the References section for more information.

"Never FRPL" students are those who have never been eligible for free or reduced-price meals during the time they have been students in Montgomery County Public Schools. "Prior FRPL" students are those who were previously eligible for free or reduced-price meals during the time they have been students in Montgomery County Public Schools but are not currently eligible. "Current FRPL" students are those who are currently eligible for free and reduced-price meals. "Never ESOL" students are those whose primary language is English and who have never been classified as an English language learner. "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. "Current ESOL" students are those who are currently enrolled in the ESOL instructional program. "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.

[a] 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. The p-values presented here were reported in the original study. For Cohort 1 (Lynch et al., 2005), the *Chemistry That Applies* group mean outcome values are the unadjusted control group posttest means plus the difference in mean gains between the intervention and control groups. Control group means are unadjusted. Attrition is high for all subgroups, and the analysis does not control analytically for the baseline scores. Therefore, only the subgroups that are equivalent at baseline with no required adjustment meet WWC evidence standards with reservations and are shown in this table. For Cohort 2 (Lynch et al., 2007), the intervention and control group mean outcome values are ANCOVA-adjusted posttest scores, with pretest scores being treated as a covariate.

Endnotes

[1] The descriptive information for this program was obtained from publicly available sources: the program's teacher guide (http://www.gwu.edu/~scale-up/documents/CTA.pdf, retrieved June 2011), and *America's Lab Report* (2005). The WWC requests developers to review the program description sections for accuracy from their perspective. The program description was provided to the developer in August 2011; however the WWC received no response. 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 June 2011.

[2] The studies in this report were reviewed using WWC Evidence Standards, Version 2.1, as described in the Science review protocol. 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 later in this report. These improvement index numbers show the average and range of student-level improvement indices for all findings across the study.

[4] Blakeslee, T., Bronstein, L., Chapin, M., Hesbitt, D., Peek, Y., Thiele, E., & Vellanti, J. (1993). *Chemistry That Applies*. Lansing, MI: Michigan Department of Education.

[5] Findings for Cohort 1 students are reported in Lynch et al. (2005). Findings for Cohort 2 students are reported in Pyke et al. (2004) and Lynch et al. (2007). The evidence rating for Cohort 1 is meets standards with reservations due to high student attrition and demonstrated baseline equivalence. The evidence rating for Cohort 2 is meets standards without reservations due to low attrition. Conventionally, the WWC gives a study the highest possible evidence rating achieved by one of its components. Hence, Pyke et al. (2004) meets standards without reservations based on Cohort 2's evidence rating.

[6] Pyke et al. (2004) is part of a multiyear research project, Scaling Up Curriculum for Achievement, Learning, and Equity Project (SCALE-uP), which is funded by the Interagency Education Research Initiative and administered by the National Science Foundation. In project Years 0 and 1, presented in this intervention report, 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 Year 2 and Year 3, authors reported the results of the first and second year of implementation of the curriculum unit *GEMS® Real Reasons for Seasons* study (Lawrence Hall of Science, 2000) for seventh-grade students in the same district. In Year 2, Year 3, and Year 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.

[7] The WWC considers both the overall sample attrition rate and the differential in sample attrition between the intervention and comparison groups, as both contribute to the potential bias of the estimated effect of an intervention. For Cohort 1, combination of overall (16%) and differential (13%) attrition rates exceeded the applicable threshold.

[8] The reported pretest sample was slightly larger than the analysis posttest sample, but the difference in samples was deemed too small to alter the demonstrated baseline equivalence. We sent out an author query but did not receive a response. As a result of the high attrition rate, Cohort 1 analyses received a lower evidence rating: meet WWC standards with reservations. Analyses for the four-month delayed posttest sample (Lynch et al., 2005; Cohort 1) do not meet WWC evidence standards because the intervention and comparison groups were not shown to be equivalent at baseline.

[9] American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2012, February). *Science intervention report: Chemistry That Applies*. Retrieved from http://whatworks.ed.gov.

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets evidence standards	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets 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 treatment 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 earlier in this report.

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 treatment 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 treatment 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 earlier in this report.

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.