



WWC Intervention Report

A summary of findings from a systematic review of the evidence



Teacher Training, Evaluation, and Compensation

June 2015*

MyTeachingPartner–Secondary

Program Description¹

MyTeachingPartner–Secondary (MTP-S) is a professional development program that aims to increase student learning and development through improved teacher–student interactions. Through the program, middle and high school teachers access a video library featuring examples of high-quality interactions and receive individualized, web-based coaching approximately twice per month during the school year. *MTP-S* uses the secondary school version of the *Classroom Assessment Scoring System®–Secondary (CLASS-S)* to define and observe effective teaching practices.

Research^{2,3}

The What Works Clearinghouse (WWC) identified one study of *MTP-S* that both falls within the scope of the Teacher Training, Evaluation, and Compensation topic area and meets WWC group design standards. This one study meets WWC group design standards with reservations. The study included 2,237 middle and high school students in 12 schools in Virginia.⁴

The WWC considers the extent of evidence for *MTP-S* on the academic achievement of middle and high school students to be small for one student outcome domain—general achievement. There were no studies that meet WWC design standards in the five other student outcome domains and six teacher outcome domains, so this intervention report does not report on the effectiveness of *MTP-S* for those domains. (See the Effectiveness Summary on p. 4 for more details of effectiveness by domain.)

Effectiveness

MTP-S was found to have potentially positive effects on general achievement for middle and high school students.

Table 1. Summary of findings^{5,6}

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
General achievement	Potentially positive	+6	+2 to +9	1	2,237	Small

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
Outcome Measures for Each Domain	p. 8
Findings Included in the Rating for Each Outcome Domain	p. 9
Endnotes	p. 10
Rating Criteria	p. 12
Glossary of Terms	p. 13

This intervention report presents findings from a systematic review of *MyTeaching Partner–Secondary* conducted using the WWC Procedures and Standards Handbook, version 3.0, and the Teacher Training, Evaluation, and Compensation review protocol, version 3.1.

Program Information

Background

The Center for Advanced Study of Teaching and Learning (CASTL) developed *MTP-S*. Robert Pianta, founding director of CASTL, and Bridget Hamre, associate director of CASTL, founded Teachstone to scale implementation of *MTP-S* and other CASTL-developed tools.

CASTL's address is 350 Old Ivy Way, Suite 202, Charlottesville, VA 22903. Web: www.mtpsecondary.net and <http://curry.virginia.edu/research/centers/castl/project/mtp-Secondary>. Telephone: (434) 243-2406. Email: chris.hafen@virginia.edu.

Teachstone's address is 105 Monticello Avenue, Suite 201, Charlottesville, VA 22902. Web: www.teachstone.com. Telephone: (866) 998-8352. Fax: (434) 293-4338. Email: contact@teachstone.com.

Program details

MTP-S has two components:

- **A video library of annotated examples of best practice.** The online video library contains more than 400 one- to two-minute video clips of teachers interacting with students. A description accompanying each video clip explains the behavior shown and why the interaction is effective. The videos, which are organized by *CLASS-S* dimensions, feature examples of a variety of instructional activities in different content areas with students from diverse socioeconomic backgrounds. The *CLASS-S* dimensions are organized into three domains: (1) Emotional Support, which includes positive climate, negative climate, teacher sensitivity, and regard for adolescent perspectives; (2) Classroom Organization, which includes behavior management, productivity, and instructional learning formats; and (3) Instructional Support, which includes content understanding, analysis and problem solving, and quality of feedback.
- **Web-based individualized coaching.** Each teacher participating in the *MTP-S* coaching program receives web-based support from a trained consultant throughout the school year. Twice a month, teachers submit videotaped sessions of their own instruction to a consultant, who identifies segments of the session for review and provides feedback linked to effective practices from the video library. Teachers review the identified segments for their own behaviors and for student reactions and then confer with the consultant, who recommends strategies to enhance teacher–student interactions.

MTP-S is currently available in two formats: a direct teacher support model and a coach support model. In the direct teacher support model, Teachstone delivers the *MTP-S* coaching program directly to teachers. The coach support model uses a “train-the-trainer” approach, in which Teachstone trains and supports school and district staff in serving as *MTP-S* coaches, with the level of Teachstone support decreasing over multiple years of implementation. The study reviewed in this report used the direct teacher support model.

Cost

MTP-S can be purchased by schools, districts, training and technical assistance centers, and institutes of higher education. The direct teacher support model currently costs \$5,000 per teacher. The cost of the coach support model varies from \$8,500 per coach in the start-up phase with biweekly support, to \$4,000 per coach in the final phase of semiannual support; an additional fee of \$325 per teacher in the coach support model covers teacher materials and equipment. Each coach can support up to 18 teachers under the coach support model. Specific pricing may vary based on factors such as the number of teachers and coaches.

Research Summary

The WWC identified three eligible studies that investigated the effects of *MTP-S* on the academic achievement of middle and high school students. An additional three studies were identified but do not meet WWC eligibility criteria for review in this topic area. Citations for all six studies are in the References section, which begins on p. 5.

The WWC reviewed three eligible studies against group design standards. One study (Allen, Pianta, Gregory, Mikami, & Lun, 2011) is a randomized controlled trial that meets WWC group design standards with reservations. The study is summarized in this report. Two studies do not meet WWC group design standards.

Table 2. Scope of reviewed research

Grades	6–12
Delivery method	Whole class
Program type	Teacher level

Summary of studies meeting WWC group design standards without reservations

No studies of *MTP-S* met WWC group design standards without reservations.

Summary of study meeting WWC group design standards with reservations

Allen et al. (2011) examined the effectiveness of *MTP-S* using a randomized controlled trial conducted in 12 middle and high schools in Virginia. Participating teachers in each school were grouped by the subject (i.e., math/science, language arts/social studies) they would teach in the evaluation. Within each subject in each school, participating teachers were randomly assigned to either the *MTP-S* intervention or “business-as-usual” professional development. Teachers were then asked to select their “focal class” for inclusion in the evaluation—the class they expected would be most academically challenging and had end-of-year achievement tests. Although initial selection of focal classes occurred before teachers learned their research condition, some teachers later changed their focal classes.⁷ Parental consent and student assent to participate in the evaluation also occurred after randomization. Because the teachers’ ultimate selection of focal classes, parental consent, and student assent could have been affected by knowledge of the teacher’s research condition, the study was determined by the WWC to have a compromised random assignment process. The study demonstrated baseline equivalence of the analytic samples and, therefore, meets WWC group design standards with reservations.

Student achievement was assessed using end-of-year scores from Commonwealth of Virginia Standards of Learning (SOL) tests taken in core subjects, with the SOL score on the most comparable course in the previous school year used as a pretest. Because the authors report a single SOL test score without distinguishing academic achievement in specific subject areas, the outcome falls in the general achievement domain.

The study authors report student achievement findings at two time points, each involving a different student cohort: (1) the end of the school year, during which *MTP-S* was delivered to intervention condition teachers (“intervention year”); and (2) the end of the following school year, during which participating teachers taught a different set of students, and teachers in the intervention condition no longer received *MTP-S* coaching (“post-intervention year”). The analytic sample for the intervention year included 76 teachers (39 *MTP-S* and 37 comparison) and 1,267 students (606 *MTP-S* and 661 comparison). The analytic sample for the post-intervention year included 61 teachers (27 *MTP-S* and 34 comparison) and 970 students (419 *MTP-S* and 551 comparison).

Effectiveness Summary

The WWC review of *MTP-S* for the Teacher Training, Evaluation, and Compensation topic area includes both student and teacher outcomes. The review encompasses student outcomes in six domains: English language arts achievement, mathematics achievement, science achievement, social studies achievement, general achievement, and student progression. The review includes teacher outcomes in six domains: teacher instruction, teacher attendance, student growth scores, teacher retention at the school, teacher retention in the school district, and teacher retention in the profession. The one study of *MTP-S* that meets WWC group design standards reported findings in one of the six domains for student outcomes: general achievement. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *MTP-S* on middle and high 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 achievement domain

One study that meets WWC group design standards with reservations reported findings in the general achievement domain.

Allen et al. (2011) examined one outcome in the general achievement domain: an end-of-year score from SOL tests taken in core subjects. For the intervention year, the authors reported, and the WWC confirmed, that the difference between the *MTP-S* group and the comparison group at the end of the school year was not statistically significant. For the post-intervention year, the authors reported, and the WWC confirmed, a positive and statistically significant difference between the *MTP-S* group and the comparison group at the end of the school year. The WWC-computed average effect across the two time points is positive and statistically significant. The WWC characterizes these study findings as statistically significant positive effects.

Thus, for the general achievement domain, one study showed statistically significant 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 achievement domain

Rating of effectiveness	Criteria met
Potentially positive effects <i>Evidence of a positive effect with no overriding contrary evidence.</i>	In the one study that reported findings, the estimated impact of the intervention on outcomes in the <i>general achievement</i> domain was positive and statistically significant.
Extent of evidence	Criteria met
Small	One study that included 2,237 students in 12 schools reported evidence of effectiveness in the <i>general achievement</i> domain.

References

Studies that meet WWC group design standards without reservations

None.

Study that meets WWC group design standards with reservations

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037.

Additional sources:

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher-student interactions in secondary school classrooms: Predicting student achievement with the Classroom Assessment Scoring System–Secondary. *School Psychology Review*, 42(1), 76–98.

Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., Hamre, B., & Pianta, R. C. (2012, November). *Improving teaching quality in secondary schools through professional development: Evidence from two RCT's of the My Teaching Partner program* [Study 1]. Paper presented at Association for Public Policy Analysis and Management Conference, Baltimore, MD. Retrieved from <https://appam.confex.com/appam/2012/webprogram/>

Hafen, C. A., Allen, J. P., Mikami, A. Y., Gregory, A., Hamre, B., & Pianta, R. C. (2012). The pivotal role of autonomy in secondary school classrooms. *Journal of Youth and Adolescence*, 41, 245–255.

Mikami, A. Y., Gregory, A., Allen, J. P., Pianta, R. C., & Lun, J. (2011). Effects of a teacher professional development intervention on peer relationships in secondary classrooms. *School Psychology Review*, 40(3), 367–385.

Studies that do not meet WWC group design standards

Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2014). Effects of a professional development program on behavioral engagement of students in middle and high school. *Psychology in the Schools*, 51(2), 143–163. The study does not meet WWC group design standards because the eligible outcomes do not meet WWC requirements.

Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., Hamre, B., & Pianta, R. C. (2012, November). *Improving teaching quality in secondary schools through professional development: Evidence from two RCT's of the My Teaching Partner program* [Study 2]. Paper presented at Association for Public Policy Analysis and Management Conference, Baltimore, MD. Retrieved from <https://appam.confex.com/appam/2012/webprogram/> The study does not meet WWC group design standards because the eligible outcomes do not meet WWC requirements.

Studies that are ineligible for review using the Teacher Training, Evaluation, and Compensation Evidence Review Protocol

Center for Advanced Study of Teaching and Learning. (n.d.). *Measuring and improving teacher-student interactions in PK-12 settings to enhance student learning*. Charlottesville, VA: Author. Retrieved from <http://curry.virginia.edu> The study is ineligible for review because it does not use an eligible design.

Gregory, A., Allen, J., Mikami, A., Hafen, C., & Pianta, R. (2013, January). *The promise of a teacher professional development program in reducing the racial disparity in classroom exclusionary discipline*. Paper presented at the Center for Civil Rights Remedies national conference, Closing the School to Research Gap: Research to Remedies conference, Washington, DC. Retrieved from <http://civilrightsproject.ucla.edu> The study is ineligible for review because it is out of the scope of the protocol.

Pianta, R. C. (2011). *Teaching children well: New evidence-based approaches to teacher professional development and training*. Washington, DC: Center for American Progress. Retrieved from <http://files.eric.ed.gov/fulltext/ED535637.pdf> The study is ineligible for review because it does not use an eligible design.

Appendix A: Research details for Allen et al., 2011⁸

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037.

Table A. Summary of findings **Meets WWC group design standards with reservations**

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General achievement	76 teachers/2,237 students	+6	Yes

Setting The study was conducted in 12 middle and high schools in Virginia.

Study sample The study recruited secondary school teachers who expected to be the primary instructor of a course that had end-of-course standardized achievement exams. Participating teachers in each school were grouped by the subject (i.e., math/science, language arts/social studies) they would teach in the evaluation. Within each subject in each school, participating teachers were randomly assigned to either the *MTP-S* intervention or “business-as-usual” professional development. Teachers were then asked to select their “focal class” for inclusion in the evaluation—the class that they expected to be the most academically challenging and that had end-of-year achievement tests. Although initial selection of focal classes occurred before teachers learned their research condition, some teachers later changed their focal classes. Written parental consent and student assent were obtained for participating students in the focal classes after randomization. Because the teachers’ ultimate selection of focal classes, parental consent, and student assent could have been affected by knowledge of the teacher’s research condition, the study was determined by the WWC to have a compromised random assignment process.

The study included 78 teachers (40 *MTP-S* and 38 comparison) from 12 schools.⁹ The analytic sample for the intervention year included 1,267 students (606 *MTP-S* and 661 comparison) taught by 76 teachers (39 *MTP-S* and 37 comparison). The analytic sample for the post-intervention year was composed of a new cohort of 970 students (419 *MTP-S* and 551 comparison) who had not participated in the study during the previous year. These post-intervention year students were taught by 61 teachers (27 *MTP-S* and 34 comparison) who had participated in the study during the intervention year; however, the *MTP-S* teachers did not receive *MTP-S* coaching during the post-intervention year.

The mean grade of the students across both cohorts was grade 8. Forty-six percent of students were female, and 29% had families with incomes at less than 200% of the poverty line. The racial/ethnic demographics were as follows: 72% were White, 22% were Black, 4% were Hispanic, 2% were Asian, and 1% were of other race/ethnicity.¹⁰

Intervention group

Intervention group teachers attended an initial workshop, where consultants defined the *MTP-S* principles and described the dimensions of high-quality teacher–student interactions from the *CLASS-S*. Approximately twice a month throughout the school year, each teacher submitted videotaped sessions from his or her focal class to a consultant, who identified brief segments of the session for review and discussion. The teacher reviewed the identified segments for his or her own behaviors and for student reactions and answered questions from the consultant about the connection between the behaviors and the reactions. The teacher then conferred with the consultant in a 20- to 30-minute discussion by phone in which the consultant recommended strategies to enhance teacher–student interactions. Consultants also directed teachers to view annotated video exemplars of high-quality teaching available on the *MTP-S* website. The year of coaching was followed by a brief “booster” workshop. The intervention involved a total of about 20 hours of in-service training over 13 months. The mean years of teaching experience was 7.6 during the intervention year. Sixty-nine percent of *MTP-S* teachers were female, and 59% held a master’s or higher level degree during the intervention year.

Comparison group

Comparison group teachers received regular in-service training. They videotaped six classroom segments throughout the school year at times that coincided with *MTP-S* teacher videotapings, but they did not receive feedback on the videotapes. Comparison teachers did not have access to the *MTP-S* library of video exemplars. The mean years of teaching experience was 10.1 during the intervention year. Fifty-seven percent of comparison teachers were female, and 70% held a master’s or higher level degree during the intervention year.

Outcomes and measurement

An outcome in the general achievement domain was reported. Student achievement was measured using end-of-year scores from tests taken in core subjects, with the score on the most comparable course in the previous school year used as a pretest. Because the authors report a single test score without distinguishing academic achievement in specific subject areas, the outcome falls in the general achievement domain. For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

Prior to the beginning of the school year, both *MTP-S* and comparison group teachers participated in a 3-hour workshop in which they received instruction in procedures for videotaping classes and submitting the videotapes (as well as procedures for study-related tasks, such as obtaining student assent/parent consent and collecting self-reported data). Consultants for the *MTP-S* teachers were master teachers trained in using the *CLASS-S* system. Two consultants led an initial workshop for intervention teachers that outlined *MTP-S* principles and discussed the *CLASS-S* dimensions of high-quality teacher–student interactions.

Appendix B: Outcome measure for the general achievement domain

General achievement

Commonwealth of Virginia Standards of Learning (SOL) tests

These standardized tests comprise the accreditation testing program for Virginia. The tests are administered to students at the end of core subject courses. Each test consists of between 45 and 63 multiple-choice questions and is standardized on a 200 to 600 point scale (as cited in Allen et al., 2011).

Appendix C: Findings included in the rating for the general 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	
Allen et al., 2011^a								
<i>Commonwealth of Virginia Standards of Learning (SOL) tests</i>	Intervention year	76 teachers/ 1,267 students	468.9 (74.4)	465.6 (78.0)	3.32	0.04	+2	> .05
<i>SOL tests</i>	Post-intervention year	61 teachers 970 students	498.0 (72.0)	482.2 (65.9)	15.77	0.23	+9	< .05
Domain average for general achievement (Allen et al., 2011)						0.14	+6	Statistically significant
Domain average for general achievement across all studies						0.14	+6	na

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 outcomes, representing the average change expected for all individuals 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 individual's percentile rank that can be expected if the individual is given the intervention. The statistical significance of the study's domain average was determined by the WWC. Some statistics may not sum as expected due to rounding. na = not applicable.

^a For Allen et al. (2011), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. The p-values presented here were reported in the original study. The WWC calculated the intervention group mean by adding the impact of the intervention (the estimated coefficient on the intervention group indicator from a hierarchical linear model) to the unadjusted comparison group posttest mean. This study is characterized as having a statistically significant positive effect because the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), page 26.

Endnotes

* On September 22, 2015, the WWC modified this report in response to clarification in the design and analysis. The study authors report student achievement findings at two time points, each involving a different student cohort: (1) the end of the school year, during which *MyTeachingPartner–Secondary (MTP-S)* was delivered to intervention condition teachers (“intervention year”); and (2) the end of the following school year, during which participating teachers taught a different set of students, and teachers in the intervention condition no longer received *MTP-S* coaching (“post-intervention year”). For the June 16, 2015 report, the WWC treated student achievement findings for the post-intervention year as a follow-up outcome and placed them in Appendix D as supplemental findings that do not factor into the intervention’s rating of effectiveness. Upon further review, WWC methodological and content experts determined it was more accurate to include findings from both time points in the effectiveness rating, as the analysis of the second cohort of students provides additional evidence of program effectiveness on a new set of students 1 year after their teachers had received *MTP-S*. As a result of these changes, the effectiveness rating for *MTP-S* in the general achievement domain is *potentially positive*, rather than *no discernible effects*. The WWC changed the Research, Effectiveness, Research Summary, Effectiveness Summary, Appendix A, and Appendix C sections, and removed Appendix D of this report. Studies have not been added to the evidence base, and the literature search has not been updated since the June 16, 2015 report. This revised report replaces the June 16, 2015 version of the intervention report.

¹ The descriptive information for this program was obtained from publicly available sources: the program’s websites (www.mtpsecondary.net and <http://curry.virginia.edu/research/centers/castl>, downloaded March 2014) and the research literature (Allen et al., 2011; Pianta, 2011). The WWC requests developers review the program description sections for accuracy from their perspective. The program description was provided to the developer in March 2014, and the WWC incorporated feedback from the developer. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.

² The literature search reflects documents publicly available by July 2014. A quick review of Allen et al. (2011) was released in February 2012. For the 2012 quick review, the study was reviewed according to version 2.1 of the WWC Procedures and Standards Handbook. The review for this intervention report was conducted under version 3.0 of the Standards. During the review for this intervention report, an author query was sent to clarify randomization as described in the supplemental materials related to the Allen et al. (2011) study. The authors’ response confirmed that the random assignment process was compromised. As a result, the rating of the analysis of the intervention year changed from *meets WWC group design standards without reservations* (as reported in the quick review) to *meets WWC group design standards with reservations* (as reported in this intervention report). The rating for the post-intervention year analysis did not change, as the quick review had rated the analysis as *meets WWC group design standards with reservations* due to high attrition.

³ The studies in this report were reviewed using the Standards from the WWC Procedures and Standards Handbook (version 3.0), along with those described in the Teacher Training, Evaluation, and Compensation review protocol (version 3.1). The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

⁴ The study sample included 78 teachers in 12 schools, but the analytic sample included only 76 teachers in the intervention year and 61 teachers in the post-intervention year. The authors did not report the number of schools included in the analytic samples; however, there is no indication that any schools were dropped.

⁵ 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 individual-level improvement indices for all findings across the studies.

⁶ The following domains were not examined by the one study that meets WWC group design standards: English language arts achievement, mathematics achievement, science achievement, social studies achievement, student progression, teacher instruction, teacher attendance, student growth scores, teacher retention at the school, teacher retention in the school district, and teacher retention in the profession.

⁷ Information about the timing of the selection of focal classes relative to randomization was provided by the study authors at the WWC’s request.

⁸ The WWC identified four additional sources related to Allen et al. (2011). These studies do not contribute unique information to Appendix A and are not listed here.

⁹ The teacher sample sizes by condition were provided by the study authors at the WWC’s request.

¹⁰ The WWC calculated these sample characteristics, which are the student-weighted averages of the characteristics that the authors reported separately by time point for the intervention and comparison groups. Intervention and comparison group percentages differed by less than four percentage points for each characteristic, with one exception—the post-intervention year difference in the percentage of students who were White was about seven percentage points. None of the differences was statistically significant.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2015, June). *Teacher Training, Evaluation, and Compensation intervention report: MyTeachingPartner–Secondary*. Retrieved from <http://whatworks.ed.gov>

WWC Rating Criteria

Criteria used to determine the rating of a study

Study rating	Criteria
Meets WWC group design standards without reservations	A study that provides strong evidence for an intervention's effectiveness, such as a well-implemented RCT.
Meets WWC group design 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 group design 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 group design 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 analytic 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 individuals, the improvement index represents the gain or loss of the average individual due to the intervention. As the average individual starts at the 50th percentile, the measure ranges from -50 to +50.
Intervention	An educational program, product, practice, or policy aimed at improving student outcomes.
Intervention report	A summary of the findings of the highest-quality research on a given program, product, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against design standards, and summarizes the findings of those that meet WWC design standards.
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 study participants 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 eligible study participants are randomly assigned to 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% ($p < .05$).
- Substantively important** A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.
- Systematic review** A review of existing literature on a topic that is identified and reviewed using explicit methods. A WWC systematic review has five steps: 1) developing a review protocol; 2) searching the literature; 3) reviewing studies, including screening studies for eligibility, reviewing the methodological quality of each study, and reporting on high-quality studies and their findings; 4) combining findings within and across studies; and, 5) summarizing the review.

Please see the WWC Procedures and Standards Handbook (version 3.0) for additional details.



An **intervention report** summarizes the findings of high-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.

This intervention report was prepared for the WWC by Mathematica Policy Research under contract ED-IES-13-C-0010.