



WWC Intervention Report

A summary of findings from a systematic review of the evidence



Transition to College

October 2016

ACT/SAT Test Preparation and Coaching Programs

Program Description¹

Most colleges and universities in the United States require students to take the SAT or ACT as part of the college application process. These tests are high stakes in at least three ways. First, most universities factor scores on these tests into admissions decisions. Second, higher scores can increase a student's chances of being admitted to selective schools, while lower scores can limit the number of institutions students have available to choose from. Finally, many colleges use admissions tests when determining eligibility for merit-based financial aid. Therefore, increasing scores on standardized college admissions tests is one way to help students access postsecondary education at the institution of their choice, while potentially helping them reduce the costs associated with college attendance.

Test preparation programs—sometimes referred to as test coaching programs—have been implemented with the goal of increasing student scores on college entrance tests. They generally (a) familiarize students with the format of the test; (b) introduce general test-taking strategies (e.g., get a good night's sleep); (c) introduce specific test-taking strategies (e.g., whether the test penalizes incorrect answers, and what this means for whether or not one should guess an answer if it is not known); and (d) specific drills (e.g., practice factoring polynomial expressions). The programs can be delivered in person or online, and in whole class settings, in small groups, and individually.

Research²

The What Works Clearinghouse (WWC) identified six studies of *ACT/SAT Test Preparation and Coaching Programs* that both fall within the scope of the Transition to College topic area and meet WWC group design standards. Three studies meet WWC group design standards without reservations, and three studies meet WWC group design standards with reservations. Together, these studies included 65,603 high school students across the United States.

The WWC considers the extent of evidence for *ACT/SAT Test Preparation and Coaching Programs* to be medium to large for one student outcome domain—general academic achievement (high school). There were no studies that meet WWC group design standards in the 11 other domains eligible for review in the Transition to College topic area, so this intervention report does not report on the effectiveness of *ACT/SAT Test Preparation and Coaching Programs* for those domains. (See the Effectiveness Summary on p. 6 for more details of effectiveness by domain.)

Report Contents

Overview	p. 1
Program Information	p. 3
Research Summary	p. 4
Effectiveness Summary	p. 6
References	p. 8
Research Details for Each Study	p. 12
Outcome Measures for Each Domain	p. 18
Findings Included in the Rating for Each Outcome Domain	p. 19
Supplemental Findings for Each Outcome Domain	p. 21
Endnotes	p. 22
Rating Criteria	p. 23
Glossary of Terms	p. 24

This intervention report presents findings from a systematic review of *ACT/SAT Test Preparation and Coaching Programs* conducted using the WWC Procedures and Standards Handbook, version 3.0, and the Transition to College review protocol, version 3.2.

Effectiveness

ACT/SAT Test Preparation and Coaching Programs were found to have positive effects on general academic achievement (high school) for high school students, with a medium to large extent of evidence.

Table 1. Summary of findings³

Outcome domain	Rating of effectiveness	Improvement index (percentile points)		Number of studies	Number of students	Extent of evidence
		Average	Range			
General academic achievement (high school)	Positive effects	+9	-3 to +19	6	65,603	Medium to large

Program Information

Background

Test preparation or coaching programs for the two common college admissions tests, the SAT and the ACT, have been created by test developers, educational organizations, and various businesses and can include software programs, workbooks, practice tests, class curricula, and a number of other resources. The SAT (formerly called the Scholastic Aptitude Test) was first introduced and administered in 1926, but did not become widely used for college admissions until the 1940's.⁴ Designed to measure students' likelihood of success in college, the test was introduced to simplify the college admission process, especially for those students applying to more than one college.⁵ The ACT was introduced about 30 years after the SAT. Originally referred to as the American College Test and billed as an alternative to the SAT, the American College Test is now referred to simply as the ACT. Both assessments have undergone many changes since their inceptions. Slightly more students take the ACT each year than the SAT, but both are widely used, with nearly 2 million students each taking one of the two tests annually.

Standardized tests are used to make important decisions for students attempting to access and enter college; as such, interest has increased in helping students better prepare to take and score well on these assessments. Preparation or coaching programs were introduced not long after the SAT was first developed.⁶ Currently, it has been estimated that nearly 50,000 students spend approximately \$10,000,000 annually on different forms of commercial test preparation and coaching for all standardized examinations.⁷

Program details

In the six studies that met WWC group design standards, students participated in a variety of test preparation programs. Four studies focused on SAT coaching, and two focused on ACT coaching. Two of the studies focused on computerized coaching, which involved students interacting individually with the coaching programs either in their classrooms or in a computer lab at their school, while the other four studies examined group classes with an instructor.

Cost

The cost of test preparation or coaching programs can vary based on the program or practice used. The responsibility for bearing the cost often depends on who initiates the program (i.e., school-bought course or individual program purchased).

Of the six studies that meet WWC group design standards, only one study reported the cost of the course, which was \$350 per student.

Research Summary

The WWC identified 26 eligible studies that investigated the effects of *ACT/SAT Test Preparation and Coaching Programs* for high school students. An additional 14 studies were identified but do not meet WWC eligibility criteria for review in this topic area. Citations for all 40 studies are in the References section, which begins on p. 8.

The WWC reviewed 26 eligible studies against group design standards.

Three studies are randomized controlled trials that meet WWC group design standards without reservations, and three studies are randomized controlled trials or quasi-experimental design studies that meet WWC group design standards with reservations. Those six studies are summarized in this report. Twenty studies do not meet WWC group design standards.

Table 2. Scope of reviewed research

Grades	10–12
Delivery method	Individual, Small group, Whole class
Program type	Practice, Curriculum

Summary of studies meeting WWC group design standards without reservations

Holmes and Keffer (1995) conducted a randomized controlled trial of an intervention focused on helping students improve their vocabulary and their verbal scores on the SAT via the study of Greek and Latin root words. The study took place at a high school in rural Georgia. One hundred and fifteen students in college-preparatory English classes volunteered to participate and were randomly assigned to one of four groups. Two groups were assigned to receive the intervention and were allowed two 45-minute sessions with the computer program per week for 6 weeks. The program used a flash-card style interface in which students matched root words to their definitions; upon mastery of the root words, students were given a similar matching task with the English derivatives. The two groups of comparison students did not have access to the computer program.

McClain (1999) conducted a randomized controlled trial in which two computerized coaching programs for the SAT were tested against a comparison group that received no coaching. The sample included 60 high school seniors from a public high school in Maryland who had previously taken the SAT and who were randomly assigned to one of two coaching programs or to a comparison group. *The Stanford Study Guide for the SAT* covers both the mathematics and verbal portions of the SAT, has a large number of drill items, provides specific test-taking strategies for each topic area on the test, and has a diagnostic component with hints for many items. *Your Personal Trainer for the SAT* includes mathematics and verbal drill questions and has a diagnostic component in which students take a single full-length pretest and receive a personalized training plan based on their performance. The students in the comparison group received no test preparation.

McMann (1994) conducted a randomized controlled trial to determine whether student ACT mathematics scores could be improved by embedding general test taking strategies and ACT practice items into high school algebra. Students in the sample came from eight different sections of an algebra class offered at a public high school located in southeastern Michigan. Students were in grades 10 and 11. One hundred ninety-six students were randomly assigned to either the intervention or the comparison group (99 to intervention and 97 to comparison). Content of the course included general test-taking strategies, test practice, and review of practice test items that were embedded within the regular mathematics curriculum. The course lasted 10 weeks. Comparison group students did not receive the curriculum with integrated test-taking strategies or review practice test items, but continued to attend their usual algebra classes.

Summary of studies meeting WWC group design standards with reservations

Domingue and Briggs (2009) used a quasi-experimental design to examine the effects of participating in commercial SAT preparation courses on student test scores. Drawn from a national sample, 353 students who self-reported participating in a commercial SAT preparation course on the Educational Longitudinal Survey of 2002 who also had

high school transcript data and PSAT and SAT scores were matched to a group of 353 students who did not report participating in a preparation course on socioeconomic status, demographics, prior achievement, and several motivational factors. No information about the specific preparation courses was available.

Filizola (2008) used a quasi-experimental design to examine the impact of a SAT preparation course on student test scores at a Texas high school. A total of 17 students enrolled in the preparation course and paid a \$350 fee. Each intervention student was then matched with a non-participating student who had similar characteristics at the same school. Students were matched according to their gender, ethnicity, and grade point average (GPA) in English and Mathematics. The course included a full-length practice SAT at the beginning and end of the course plus content instruction and practice on the mathematics, verbal, and writing skills sections on the test.

Scholes and Lain (1997) [Experiment 2]⁸ used a large sample of students who self-reported participating in a test preparation course prior to taking the ACT. The authors used a national database of ACT takers to identify students who had taken the ACT more than once between October 1, 1994 and September 20, 1995. Students who self-reported taking a test preparation course served as the intervention group (n=3,071). Students who did not report any test preparation activities served as the comparison group (n=61,496).

Effectiveness Summary

The WWC review of *ACT/SAT Test Preparation and Coaching Programs* for the Transition to College topic area includes outcomes in 12 domains: (a) general academic achievement (high school), (b) attendance, (c) progressing in high school, (d) staying in high school, (e) completing high school, (f) college readiness, (g) college access and enrollment, (h) college attendance, (i) credit accumulation, (j) general academic achievement (college), (k) degree attainment, and (l) labor market. The six studies of *ACT/SAT Test Preparation and Coaching Programs* that meet WWC group design standards reported findings in one of the 12 domains: general academic achievement (high school). The findings below present the authors’ estimates and WWC-calculated estimates of the size and statistical significance of the effects of *ACT/SAT Test Preparation and Coaching Programs* for high school students. Additional comparisons are presented as supplemental findings in Appendix D. Supplemental finding does not factor into the intervention’s rating of effectiveness. For a more detailed description of the rating of effectiveness and extent of evidence criteria, see the WWC Rating Criteria on p. 23.

Summary of effectiveness for the general academic achievement (high school) domain

Table 3. Rating of effectiveness and extent of evidence for the general academic achievement (high school) domain

Rating of effectiveness	Criteria met
<p>Positive effects <i>Strong evidence of a positive effect with no overriding contrary evidence.</i></p>	<p>In the six studies that reported findings, the estimated impact of the intervention on outcomes in the <i>general academic achievement (high school)</i> domain was positive because three studies show statistically significant positive effects and no studies show statistically significant or substantively negative effects.</p>
Extent of evidence	Criteria met
<p>Medium to large</p>	<p>Six studies that included 64,897 students in <i>multiple</i> schools reported evidence of effectiveness in the <i>general academic achievement (high school)</i> domain.</p>

Three studies that meet WWC group design standards without reservations and three studies that meet WWC group design standards with reservations reported findings in the general academic achievement (high school) domain.

Holmes and Keffer (1995) compared intervention students’ scores on the verbal portion of the SAT verbal following completion of the computerized intervention to the scores achieved by the students in the comparison group. The authors reported, and the WWC confirmed, that there was a statistically significant difference between students who participated in the test preparation program and comparison participants on verbal SAT scores. The WWC characterizes this finding as a statistically significant positive effect.

McClain (1999) reported on students’ final SAT scores following completion of the intervention. Composite SAT scores for students who received one of the two intervention test preparation programs were compared to those for students in the comparison group. The author reported, and the WWC confirmed, that there was no statistically significant difference between students who participated in one of the two intervention programs and comparison students on SAT scores. The WWC characterizes this finding as an indeterminate effect.

McMann (1994) reported on students’ posttest scores on a practice ACT test in mathematics. Students who received the test preparation curriculum were compared to students who did not receive the program who participated in the usual mathematics classes. The author reported, and the WWC confirmed, that there was a statistically significant difference between intervention and comparison group students on their ACT test scores in mathematics. The WWC characterizes this finding as a statistically significant positive effect.

Domingue and Briggs (2009) examined the effects of participating in commercial test preparation courses on the math and verbal portions of the SAT using propensity score matched groups taken from the Educational Longitudinal Survey of 2002. The authors reported, and the WWC confirmed, that there was no statistically significant difference between students who participated in test preparation courses and those who did not on SAT scores. The WWC characterizes this finding as an indeterminate effect.

Filizola (2008) reported students' scores on the reading, writing, and mathematics subtests of the SAT. Students who participated in the SAT preparation course were compared to a matched group of students who did not participate in the course. The author reported, and the WWC confirmed, that there was a statistically significant positive difference between the intervention students and the comparison students on SAT scores. The WWC characterizes this finding as a statistically significant positive effect.

Scholes and Lane (1997) [Experiment 2] compared students' ACT composite scores for a group of students who self-reported participating in test preparation activities to a group of students who did not report participating in such activities. The authors reported, and the WWC confirmed, that there was no significant difference in ACT composite scores between students in the intervention group and students in the comparison group. The WWC characterizes this finding as an indeterminate effect.

Thus, for the general academic achievement (high school) domain, three studies showed statistically significant positive effects, and three studies showed indeterminate effects. This results in a rating of positive effects, with a medium to large extent of evidence.

References

Studies that meet WWC group design standards without reservations

- Holmes, C. T., & Keffer, R. L. (1995). A computerized method to teach Latin and Greek root words: Effect on Verbal SAT scores. *Journal of Educational Research*, 89, 47–50.
- McClain, T. B. (1999). *The impact of computer-assisted coaching on the elevation of twelfth-grade students' SAT scores* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9945906)
- McMann, P. K. (1994). *The effects of teaching practice review items and test-taking strategies on the ACT mathematics scores of second-year algebra students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9423737)

Studies that meet WWC group design standards with reservations

- Domingue, B., & Briggs, D. C. (2009). *Using linear regression and propensity score matching to estimate the effect of coaching on the SAT*. Boulder: University of Colorado.
- Additional source:**
- Domingue, B., & Briggs, D. C. (2009). Using linear regression and propensity score matching to estimate the effect of coaching on the SAT. *Multiple Linear Regression Viewpoints*, 35(1), 12–29.
- Filizola, E. (2008). *The effect of a test preparation course on the SAT scores of students at Saint Joseph Academy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3309546)
- Scholes, R. J., & Lain, M. M. (1997, March). *The effects of test preparation activities on ACT assessment scores* [Experiment 2]. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. <http://files.eric.ed.gov/fulltext/ED409341.pdf>.

Studies that do not meet WWC group design standards

- Briggs, D. C. (2001). The effect of admissions test preparation: Evidence from NELS: 88. *Chance*, 14(1), 10–18. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.
- Additional source:**
- Briggs, D. C. (2002). *SAT coaching, bias, and causal inference* (Unpublished doctoral dissertation). University of California, Berkeley.
- Butler, J. D. (2002). *The effect of participation in a Scholastic Assessment Test preparation program on student performance on the Scholastic Assessment Test* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3042431) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.
- Dale, P. A. (2013). *The effect of ACT preparatory classes on secondary school students' academic achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3587517) The study does not meet WWC group sign standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Donen, T. (2012). *Formal individualized teaching, a differentiated instructional model, effects on student learning in a high school ACT preparation course* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3519038) The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Hargis, A. (2015). *The impact of a preparatory ACT course on student achievement connected to the ACT* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3706827) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.

- Holthaus, R. A. (2008). *The effects of a high school coaching program on SAT-Mathematics test scores* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3291445) The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Joseph, M. W. (2004). *A detailed and comprehensive operationalization of SAT coaching and an analysis of coaching efficacy* (Doctoral dissertation). Available from ProQuest Information & Learning. (UMI No. 3128547) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.
- Justus, L. C. (2010). *Impact of a school-based test preparation course on ACT scores with consideration of cultural associations* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3443004) The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Kaplan, J. (2005). The effectiveness of SAT coaching on math SAT scores. *Chance*, 18(2), 25–34. The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Additional source:**
- Kaplan, J. (2002). An SAT coaching program that works. *Chance*, 15(1), 12–22.
- Keltner, K. R. (2004). *A study to determine the effectiveness of the administrative development and selection of an ACT test preparation course* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3154270) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.
- Lane, K. L., Kalberg, J. R., Mofield, E., Wehby, J. H., & Parks, R. J. (2009). Preparing students for college entrance exams: Findings of a secondary intervention conducted within a three-tiered model of support. *Remedial and Special Education*, 30(1), 3–18. doi:10.1177/0741932507314022 The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Moss, G. L., Chippendale, E. K., Mershon, C. W., & Carney, T. (2012). Effects of a coaching class on the ACT scores of students at a large Midwest high school. *Journal of College Admission*, 217, 16–23. <http://files.eric.ed.gov/fulltext/EJ992994.pdf>. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.
- Mudambi, A. (2013). *The use of computer games for SAT vocabulary acquisition* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3579010) The study does not meet WWC group design standards because the eligible outcomes do not meet WWC requirements.
- Murphy, A. M. (2003). *The relationship between a structured semester SAT-I coaching course and SAT-I results* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3102917) The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Parrott, T. N. (2012). *ACT test preparation course and its impact on students' college- and career-readiness* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3552840) The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Petrey, T. L. (2002). *The efficacy of remediation on the Scholastic Aptitude Test with multicultural high school students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3129169) The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.

- Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. *Journal of Educational Measurement*, 36(2), 93–118. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.
- Additional source:**
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467), 609–618.
- Ringer, V. M. (2014). *The effects of the Cambridge Test Preparation Program on Scholastic Assessment Test scores* (Unpublished doctoral dissertation). Walden University, Minneapolis, MN. The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.
- Scholes, R. J., & Lain, M. M. (1997, March). *The effects of test preparation activities on ACT assessment scores* [Experiment 1]. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. <http://files.eric.ed.gov/fulltext/ED409341.pdf>. The study does not meet WWC group design standards because equivalence of the analytic intervention and comparison groups prior to the intervention was necessary and not demonstrated.
- Wrinkle, G. W. (1996). *A Scholastic Assessment Test preparation class and its effect on Scholastic Assessment Test scores* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9626761) The study does not meet WWC group design standards because the measure of effectiveness cannot be attributed solely to the intervention.

Studies that are ineligible for review using the Transition to College Evidence Review Protocol

- Allalouf, A., & Ben-Shakhar, G. (1998). The effect of coaching on the predictive validity of scholastic aptitude tests. *Journal of Educational Measurement*, 35(1), 31–47. The study is ineligible for review because it does not have a sample aligned with the protocol.
- Byun, S. Y., & Park, H. (2012). The academic success of East Asian American youth: The role of shadow education. *Sociology of Education*, 85(1), 40–60. The study is ineligible for review because it does not have an intervention in line with the protocol.
- Din, F. S., & Soldan, C. (2001). *The effects of practice reading of ACT science passages*. Retrieved from <http://files.eric.ed.gov/fulltext/ED460865.pdf>. The study is ineligible for review because it does not use an eligible design.
- Doe, C. D. (2006). *The role of preparation in test takers' strategy use and test performance* (Doctoral dissertation). Retrieved from the Library and Archives Canada. (ISBN: 978-0-494-18257-4) The study is ineligible for review because it does not use an eligible design.
- Gaye, Z. Z. (2001). *The ACT Prep and its effects on students who participated in the Multicultural Educational Enrichment (MEE) program at Tennessee State University* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3007576) The study is ineligible for review because it does not use an eligible design.
- Hamilton, M. (2015). *The effect of lab based instruction on ACT science scores* (Masters thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1587194) The study is ineligible for review because it does not use an eligible design.
- Harris Badgett, T. L. (2015). *Exploring an ACT preparation course as an intervention method for African American students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3732101) The study is ineligible for review because it does not use an eligible design.
- Jones, P. L. (1997). *The effects of academic team participation on SAT scores and self-esteem of high school students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9726931) The study is ineligible for review because it does not have an intervention in line with the protocol.
- McTeer, D. E., Jr. (2004). *An evaluation of the South Carolina laptop program to improve SAT scores* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3130471) The study is ineligible for review because it does not have an intervention in line with the protocol.

- Montgomery, P., & Lilly, J. (2012). Systematic reviews of the effects of preparatory courses on university entrance examinations in high school age students. *International Journal of Social Welfare*, 21(1), 3–12. doi:10.1111/j.1468-2397.2011.00812.x The study is ineligible for review because it is not a primary analysis of the effect of an intervention.
- Price, M., & Finkelstein, A. (1994). Pictures improve memory of SAT vocabulary words. *Journal of Reading*, 38(2), 134–135. The study is ineligible for review because it is not a primary analysis of the effect of an intervention.
- Urbina-Lilback, R. N. (2003). *The effect of incorporating the development of metacognitive skills into test coaching for the math component of an SAT preparatory elective in a private urban high school* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 1414501) The study is ineligible for review because it does not include an outcome within the domain specified in the protocol.
- Warch, K. L. (2001). The impact of computer-assisted coaching on high school students' SAT I scores (Masters thesis). Available from ProQuest Dissertations & Theses database. (UMI No. 1402424) The study is ineligible for review because it does not use an eligible design.
- Wilson, D. F. (2003). *The effects of a test preparation guide in problem-solving in improving students' achievement on the New York Math 'A' Regents Examinations* (Doctoral dissertation). Available from ProQuest Dissertation and Theses database. (UMI No. 3091307) The study is ineligible for review because it does not include an outcome within the domain specified in the protocol.

Appendix A.1: Research details for Holmes and Keffer (1995)

Holmes, C. T., & Keffer, R. L. (1995). A computerized method to teach Latin and Greek root words: Effect on Verbal SAT scores. *Journal of Educational Research*, 89, 47–50.

Table A1. Summary of findings

Meets WWC group design standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General academic achievement (high school)	70 students	+19	Yes

Setting The study took place at a high school in rural northeast Georgia with students in college-preparatory-level English classes. The high school population was comprised of about 15% Black students, and about 16% were in the free/reduced-price lunch program. Overall, 59% of the students in this high school typically enroll in college. The average SAT scores at the school are below the national average.

Study sample The sample demographics in the study were not representative of the school population. Four students in the study sample were Black, and none participated in the free/reduced-price lunch program. Nineteen (56%) of the 34 students in the intervention group were female. Twenty-eight (78%) of the 36 students in the comparison group were female. The average age of both groups was about 15 and a half years.

Intervention group The intervention in this study was a computerized program designed to help students improve their vocabulary scores on the SAT through the study of Latin and Greek root words. The program focused on a list of 90 common Latin root words and 11 common Greek root words. About 800 English words and derivatives have these 101 roots. Participants in the intervention group were allowed two 45-minute periods per week to use the program. Times were available both before and after school. The program employed a flash card-style interface in which students matched definitions to root words. Once students mastered the root words, they were then given a similar matching task with the English derivatives. The intervention period lasted 6 weeks.

Comparison group Students in the comparison group were not offered the computerized coaching program. They were recruited from the same college-preparatory English classes as the intervention students. No information about any alternative services received by the comparison students was provided in the study.

Outcomes and measurement The study reported intervention effects on one eligible outcome that meets review requirements. The outcome was the verbal portion of the SAT. This outcome falls in the domain of general academic achievement (high school). For a more detailed description of these outcome measures, see Appendix B.

Support for implementation No information was provided regarding support for implementation.

Appendix A.2: Research details for McClain (1999)

McClain, T. B. (1999). *The impact of computer-assisted coaching on the elevation of twelfth-grade students' SAT scores* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9945906)

Table A2. Summary of findings

Meets WWC group design standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General academic achievement (high school)	40 students	-2	No

Setting This study took place at one suburban high school in Maryland, near the Washington, DC area. The majority of students in the high school were Black. The school is located in a suburban county with a population of approximately 764,000. The author reported that residents in this county have an Effective Buying Income (EBI) that is 17% higher than the average US EBI.

Study sample All 60 students in the sample were high school seniors. The sample of students included 26 males (43%) and 34 females (57%), all of whom were Black. No other demographic characteristics specific to the study sample were reported.

Intervention group Students were randomly assigned to participate in either *The Stanford Study Guide for the SAT* or Davidson's *Your Personal Trainer for the SAT*. Intervention group students were excused from their regular classroom three times per week for 1 hour to use the test preparation programs. During this time, students would go to the computer lab, where they worked with one of the two computer coaching programs. Students receiving the intervention spent 26 hours with the program over the course of 9 weeks.

Comparison group Students in the comparison group were not offered the computerized coaching programs and continued with their curriculum in their regular classrooms. These students took the SAT at the same time as the intervention students, both at pretest and posttest.

Outcomes and measurement The outcome reported in this study is student SAT scores. This outcome falls under the general academic achievement (high school) domain. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation No information was provided regarding support for implementation.

Appendix A.3: Research details for McMann (1994)

McMann, P. K. (1994). *The effects of teaching practice review items and test-taking strategies on the ACT mathematics scores of second-year algebra students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9423737)

Table A3. Summary of findings

Meets WWC group design standards without reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General academic achievement (high school)	196 students	+13	Yes

Setting This study took place at one suburban high school located in southeastern Michigan. The author described the location as a predominantly blue-collar community. The racial make-up of the school is predominantly White (97%). The high school’s total enrollment is 1,410 students.

Study sample The sample in this study consisted of tenth- and eleventh-grade students across eight different second-year algebra course sections. Four of the sections were intervention sections and were randomly assigned 99 students. The comparison group also had four class sections and were randomly assigned 97 students. There were a total of four teachers, with two instructing intervention sections and two instructing comparison sections. The author reports that there were 45 (45%) males and 54 (55%) females in the intervention group and 51 males (53%) and 46 females (47%) in the comparison group. No other demographic characteristics were provided for the sample.

Intervention group The intervention lasted 10 weeks. Students took the ACT pretest prior to the implementation of the intervention. Students then participated in their normal second year algebra course using the Algebra II and Trigonometry textbooks. Test-taking strategies and practice ACT items were reviewed during the course along with the regular curriculum. These materials came from suggested items from the ACT or were written by the researcher. Once the intervention was complete, students took the ACT posttest.

Comparison group Students in the comparison group also took the ACT pretest and posttest following implementation of the intervention. The comparison students received the regular curriculum of the second year algebra course, using the same Algebra II and Trigonometry textbooks as the intervention group. Comparison group students did not learn additional test-taking strategies or review practice test items.

Outcomes and measurement The study reports on one eligible outcome: student scores on the math subtest of an official ACT practice test. This outcome is eligible under the general academic achievement (high school) domain. Student scores on this outcome were measured at the beginning of the study (pretest) and at the end of the study (posttest) For a more detailed description of this outcome measure, see Appendix B.

Support for implementation No information was provided regarding support for implementation.

Appendix A.4: Research details for Domingue and Briggs (2009)

Domingue, B., & Briggs, D. C. (2009). *Using linear regression and propensity score matching to estimate the effect of coaching on the SAT*. Boulder: University of Colorado.

Additional source:

Domingue, B., & Briggs, D. C. (2009). Using linear regression and propensity score matching to estimate the effect of coaching on the SAT. *Multiple Linear Regression Viewpoints*, 35(1), 12–29.

Table A4. Summary of findings

Meets WWC group design standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General academic achievement (high school)	706 students	+4	No

Setting Students in this study were high school students in the United States who participated in the Educational Longitudinal Survey of 2002 (ELS:02). Students who reported participating in a commercial SAT preparation course were selected for the intervention group. Information about the setting for each prep course included in the analysis was not available.

Study sample Sample characteristics were reported on the full sample (n=1552). In the intervention group, the average age at posttest (twelfth grade) was 17.8 years; 53% of the sample was female; 24% were Asian, 12% Black, 6% Hispanic, 3% Native American; 23% were taking English as a Second Language (ESL), 7% had enrolled in a remedial English course; 8% had enrolled in a remedial math course; and 58% had taken an Advanced Placement (AP) course. In the comparison group, the average age at posttest was 17.9 years; 56% of the sample was female; 12% were Asian, 8% Black, 9% Hispanic, 3% Native American; 13% were taking ESL, 6% had enrolled in a remedial English course; 7% had enrolled in a remedial math course; and 52% had taken an AP course. The authors also report mean SES indices for each group (0.68 intervention and 0.38 comparison), but additional information about the calculation of these indices was not provided.

Intervention group The intervention group consisted of students who reported participating in any commercial test preparation course. The authors did not limit this to any one specific SAT preparation course; participation in any commercial SAT prep course was sufficient. In the ELS:02 survey, students were asked a number of questions regarding how (and whether) they prepared for the SAT. Students who reported participating in a commercial preparation course are considered to have been “coached” (note that students who only prepared using tutoring or self-prep materials are not considered “coached” under these criteria). Students who participated in a commercial preparation course were eligible for inclusion in the intervention group. Additional information about the content of the SAT preparation courses was not available.

Comparison group The comparison group consisted of students who reported that they did not take a commercial SAT preparation course.

Outcomes and measurement All data used in this study were taken from the National Center for Educational Statistics' Educational Longitudinal Survey of 2002 (ELS:02) dataset. The authors report SAT-Math and SAT-Verbal scores. Both SAT-Math and SAT-Verbal scores are eligible under the review protocol and can be categorized into the general academic achievement (high school) domain. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation No information was provided regarding support for implementation.

Appendix A.5: Research details for Filizola (2008)

Filizola, E. (2008). *The effect of a test preparation course on the SAT scores of students at Saint Joseph Academy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3309546)

Table A5. Summary of findings

Meets WWC group design standards with reservations

Outcome domain	Sample size	Study findings	
		Average improvement index (percentile points)	Statistically significant
General academic achievement (high school)	24 students	+11	Yes

Setting This study was conducted with high school students at Saint Joseph Academy, a parochial school in Brownsville, Texas. About 40% of the Academy's students live in Mexico.

Study sample The intervention group consisted of 17 students who enrolled in an SAT preparation course. The students in the intervention group consisted of seven male and 10 female students. Of these students, nine lived in Brownsville, Texas, and eight lived in Matamoros, Tamaulipas, Mexico. Fifteen of these students were Hispanic, one was Black, and one was "of Anglo ethnicity" (p. 23).

Intervention group Students in the intervention group registered to participate in the SAT preparation class, which consisted of eight 4-hour sessions, two of which were used for the administration of the pre- and posttest, a practice SAT. The remaining six sessions were split between math and reading/writing instruction, for a total of 12 hours of instruction in each content area. The verbal sessions focused instruction and practice on the essay portion of the SAT, as well as review, instruction, and practice for the multiple choice questions. The math sessions included strategies, practice items, and practice tests.

Comparison group Students in the comparison group did not participate in the SAT preparation class. The students did participate in the administration of the pre- and posttest. The author does not report any additional information.

Outcomes and measurement

The study reported on the following eligible outcomes: 1) Reading scores on the SAT, 2) Writing scores on the SAT, and 3) Mathematics scores on the SAT. All of these outcomes are in the general academic achievement (high school) domain. They were reported using pretest and posttest means and standard deviations. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

No information was provided regarding support for implementation.

Appendix A.6: Research details for Scholes and Lain (1997)

Scholes, R. J., & Lain, M. M. (1997, March). *The effects of test preparation activities on ACT assessment scores [Experiment 2]*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. <http://files.eric.gov/fulltext/ED409341.pdf>.

Table A6. Summary of findings

Meets WWC group design standards with reservations

Outcome domain	Sample size ^a	Study findings	
		Average improvement index (percentile points)	Statistically significant
General academic achievement (high school)	64,567 students	+1	No

Setting

The students in this study were high school juniors and seniors who lived in the United States. They were selected from students who had taken the ACT assessment more than once between October 1, 1994 and September 20, 1995.

Study sample

The total sample included 64,567 students, 3,071 in the intervention group and 61,496 in the comparison group. There were 36% males, 64% females, 46% high school juniors, and 52% high school seniors in the intervention group, and 43% males, 57% females, and 37% high school juniors in the comparison group. Of those in the intervention group, 74% were White, 13% were Black, and 10% had a family income of less than \$18,000. Of those in the comparison group, 77% were White, 10% were Black, and 12% had a family income of less than \$18,000.

Intervention group

The intervention group participated in a test preparation course. Test preparation consisted of activities that include components such as drills with feedback, familiarization with the test, test-taking strategies, and subject matter review.

Comparison group

The comparison group reported that they did not participate in any test preparation courses or any type of test preparation.

Outcomes and measurement

The outcome addressed in this study was ACT composite scores. This outcome falls in the domain of general academic achievement (high school). For a more detailed description of this outcome measure, see Appendix B.

Support for implementation

No information was provided regarding support for implementation.

Appendix B: Outcome measures for the general academic achievement (high school) domain

General academic achievement (high school)	
<i>ACT composite scores</i>	General academic achievement (high school) was assessed using the students' final scores received on the ACT following completion of the intervention (as cited in Scholes & Lain, 1997).
<i>Final SAT scores</i>	General academic achievement (high school) was assessed using the students' final scores received on the SAT following completion of the intervention (as cited in McClain, 1999).
<i>Practice ACT mathematics exam</i>	General academic achievement (high school) was assessed using scores on a practice ACT mathematics exam administered to all students following completion of the intervention (as cited in McMann, 1994).
<i>SAT scores in math, reading, verbal, and writing</i>	General academic achievement (high school) was assessed using students' scores on the Math, Reading, and Writing aptitude SAT tests administered following completion of the intervention (as cited in Domingue & Briggs, 2009; Filizola, 2008; and Holmes & Keffer, 1995).

Appendix C: Findings included in the rating for the general academic achievement (high school) domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Holmes & Keffer (1995)^a								
<i>SAT Verbal</i>	High school students	70	402.94 (88.44)	361.39 (81.00)	41.55	0.49	+19	.03
Domain average for general academic achievement (high school) (Holmes & Keffer, 1999)						0.49	+19	Statistically significant
McClain, 1999^b								
<i>Final SAT scores</i>	High school students, Stanford intervention	40	756.08 (110)	765.50 (110)	-9.42	-0.08	-3	.79
<i>Final SAT scores</i>	High school students, Davidson intervention	40	765.44 (110)	765.50 (110)	-0.06	-0.00	0	.99
Domain average for general academic achievement (high school) (McClain, 1999)						-0.04	-2	Not statistically significant
McMann, 1994^c								
<i>Practice ACT mathematics exam</i>	High school students	196	26.89 (8.24)	24.01 (8.69)	2.88	0.34	+13	.02
Domain average for general academic achievement (high school) (McMann, 1994)						0.34	+13	Statistically significant
Domingue & Briggs (2009)^d								
<i>SAT Math</i>	High school students	706	nr	nr	nr	0.13	+5	.08
<i>SAT verbal</i>	High school students	706	nr	nr	nr	0.08	+3	.31
Domain average for general academic achievement (high school) (Domingue & Briggs, 2009)						0.10	+4	Not statistically significant
Filizola, 2008^e								
<i>SAT Math</i>	High school students	24	498.16 (96.24)	460.17 (56.48)	37.99	0.67	+18	.008
<i>SAT Reading</i>	High school students	24	458.95 (75.09)	437.71 (83.49)	21.24	0.25	+10	.42
<i>SAT Writing</i>	High school students	24	485.81 (78.21)	475.02 (81.22)	10.79	0.13	+5	.55
Domain average for general academic achievement (high school) (Filizola, 2008)						0.28	+11	Statistically significant

Scholes & Lain, 1997 ^f								
<i>ACT composite</i>	High school students	64,567	21.30 (4.50)	21.20 (4.50)	0.10	0.02	+1	.23
Domain average for general academic achievement (high school) (Scholes & Lain, 1997)						0.02	+1	Not statistically significant
Domain average for general academic achievement (high school) across all studies							+9	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. na = not applicable; nr = not reported.

^a For Holmes & Keffer (1995), no corrections for clustering or multiple comparisons and no difference-in-differences adjustment were needed. The *p*-value presented here was computed by the authors. This study is characterized as having a statistically significant positive effect. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

^b For McClain (1999), a correction for multiple comparisons was needed but did not affect whether any of the contrasts were found to be statistically significant. Additionally, a difference-in-differences adjustment was used. The WWC calculated the program group mean using a difference-in-differences approach by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. The *p*-values presented were computed by the WWC. This study is characterized as having indeterminate effects because the reported effect size for all measures within the domain is neither statistically significant nor substantively important, accounting for multiple comparisons. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

^c For McMann (1994), no corrections for clustering or multiple comparisons were needed. However, the WWC calculated the program group mean using a difference-in-differences approach by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. The *p*-value presented here was computed by the WWC. 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), p. 26.

^d For Domingue and Briggs (2009), 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 computed by the WWC. Effect sizes for this study were computed using the ordinary least squares regression coefficient reported in the study and publicly available unadjusted standard deviations for the SAT. This study is characterized as having indeterminate effects because the reported effect size for all measures within the domain is neither statistically significant nor substantively important, accounting for multiple comparisons. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

^e For Filizola (2008), 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. 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), p. 26.

^f For Scholes and Lain (1997), no corrections for clustering or multiple comparisons were needed. However, the WWC calculated the program group mean using a difference-in-differences approach by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. The *p*-value presented here was computed by the WWC. This study is characterized as having indeterminate effects because the reported effect size for all measures within the domain is neither statistically significant nor substantively important, accounting for multiple comparisons. For more information, please refer to the WWC Procedures and Standards Handbook (version 3.0), p. 26.

Appendix D: Description of supplemental findings for the general academic achievement (high school) domain

Outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			p-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
McMann, 1994^a								
<i>Practice ACT mathematics exam</i>	Female high school students	100	27.17 (6.76)	24.26 (8.63)	2.91	0.38	+15	.06
<i>Practice ACT mathematics exam</i>	Male high school students	96	26.76 (8.36)	23.78 (8.68)	2.98	0.35	+14	.09

Table Notes: The supplemental findings presented in this table are additional findings from studies in this report that meet WWC design standards with or without reservations, but do not factor into the determination of the intervention rating. 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. Some statistics may not sum as expected due to rounding.

^a For McMann (1994), no corrections for clustering or multiple comparisons were needed. However, the WWC calculated the program group mean using a difference-in-differences approach by adding the impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the WWC Procedures and Standards Handbook (version 3.0) for more information. The p-values presented here were computed by the WWC.

Endnotes

¹ The descriptive information for this program was obtained from McClain (1999), McMann (1994), and Filizola (2008). The WWC requests developers review the program description sections for accuracy from their perspective. 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 February 2016. 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 Review Protocol for Studies of Interventions to Support the Transition to College (version 3.2). 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. 23. These improvement index numbers show the average and range of individual-level improvement indices for all findings across the studies. The outcome domains of attendance, progressing in high school, staying in high school, completing high school, college readiness, college access and enrollment, college attendance, credit accumulation, general academic achievement (college), degree attainment (college), and labor market were not included in the table because they did not have any reported findings.

⁴ Braswell, J. S. (1992). Changes in the SAT in 1994. *The Mathematics Teachers*, 85(1), 16–21.

⁵ Filizola, E. (2008). *The effect of a test preparation course on the SAT scores of students at Saint Joseph Academy* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3309546)

⁶ McMann, P. K. (1994). *The effects of teaching practice review items and test-taking strategies on the ACT mathematics scores of second-year algebra students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9423737)

⁷ <http://www.pbs.org/wgbh/pages/frontline/shows/sats/test/history.html>

⁸ This report described two separate experiments with two distinct samples. Only Experiment 2 met WWC group design standards.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, October). *Transition to College intervention report: ACT/SAT Test Preparation and Coaching Programs*. 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 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. 23.
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. 23.
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.

Glossary of Terms

- 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.



Practice Guide



Quick Review



Single Study Review

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 Development Services Group under contract ED-IES-12-C-0084.