

REVIEW PROTOCOL FOR TEACHER TRAINING, EVALUATION, AND COMPENSATION VERSION 3.1 (MAY 2015)

This review-specific protocol guides the review of research that informs the What Works Clearinghouse (WWC) intervention reports in the Teacher Training, Evaluation, and Compensation topic area. The review-specific protocol is used in conjunction with the [*WWC Procedures and Standards Handbook \(version 3.0\)*](#).

PURPOSE STATEMENT

This review focuses on interventions aimed at making teachers more effective at improving the academic achievement of students in grades PK–12. These interventions are intended for use by adults in a variety of teacher development settings who are considering teaching, undergoing teacher preparation, or already employed in the teaching profession. The primary focus is on student achievement and those teacher-level outcomes that have been shown to be related to student achievement (e.g., teacher retention).

The following research questions guide this review:

- Which teacher-focused interventions improve the academic achievement of students in grades PK–12?
- Which teacher-focused interventions improve those teacher outcomes that are associated with student achievement?

KEY DEFINITIONS

Categories of relevant research. The review team identified and defined five categories of research examining interventions aimed at making teachers more effective at improving the academic achievement of their students. Those categories are:

- ***Teacher preparation:*** Studies of programs that train individuals to serve as teachers, including both traditional programs based in college or university schools of education and alternative route programs.
- ***Teacher induction:*** Studies of programs that provide specialized training targeting the needs of novice teachers (i.e., those in their first 3 years of service).
- ***Teacher evaluation:*** Studies that examine processes or systems used to determine teacher performance for the purposes of identifying professional development needs and providing formative feedback to teachers, setting compensation, and making other personnel decisions.
- ***Teacher compensation:*** Studies of programs or systems that provide monetary incentives to teachers for improving student academic achievement, teaching performance, or related

school outcomes; taking on additional professional responsibilities; demonstrating particular knowledge or skills; or filling hard-to-staff teaching positions. Such systems might include bonuses or salary structures that differ from a traditional salary schedule.

- ***Teacher professional development:*** Studies of programs that provide training addressing the needs of in-service teachers (i.e., those presently employed as teachers of record by a school or school district).

Teacher. An adult who is employed by a school or school district to provide instruction to students in grades PK–12.

In cases where teachers within a study are described in more detail, the decision rules for classifying people as teachers are:

Teachers

- Those who provide students with more than 50% of instruction on a subject (i.e., teachers of record), regardless of whether those students are general education students, students with special needs, or a combination of general and special education students
- Long-term substitutes (i.e., individuals who fill in for particular teachers for more than half the period between pretest and posttest)

Non-teachers

- Those who provide instruction outside of school hours (i.e., tutors)
- Those whose primary role is administrative (e.g., principal, dean, superintendent)
- Those providing instruction to individuals outside of grades PK–12 (e.g., college lecturers)
- Those who provide non-instructional support to students (e.g., nurse, school psychologist, speech language pathologist)
- Short-term substitutes (i.e., individuals who fill in for particular teachers for less than half the period between pretest and posttest)

Teacher candidates. Individuals who are participating in a teacher pre-service training program. These individuals may be in a traditional teacher preparation program at a college or university, or using an alternative route to certification. Individuals are no longer teacher candidates when they complete their training program and become fully certified as a teacher.

Pre-service training. Training of prospective teachers as part of obtaining teacher certification. Generally, this occurs before being hired by a school/district and becoming responsible for a classroom or becoming the teacher of record. Pre-service training can involve coursework completed toward certification, as well as student-teaching or other practice-based experiences.

PROCEDURES FOR CONDUCTING THE LITERATURE SEARCH

The *WWC Procedures and Standards Handbook* discusses the procedures for conducting a literature search described in Section II: Developing the Review Protocol and Identifying Relevant Literature (p. 4) and in Appendix B: Policies for Searching and Prioritizing Studies for Review.

Search Terms

The following table presents the search terms by category.

Category	Search Terms	
Study Design	<ul style="list-style-type: none"> • ABAB design* • Alternating treatment* • Assignment • Baseline • Causal • Changing criterion design* • Comparison group* • Control* • Control group* • Counterfactual* • Effect* • Efficacy • Evaluat* • Experiment* • Field trial* • Growth • Impact* • Intrasubject replication design* • Matched group* • Meta analys* • Metaanalys* • Meta-analys* • Multi element design* • Multi-element design* • Multiple baseline design* • Outcome* 	<ul style="list-style-type: none"> • Posttest • Post-test • Pretest • Pre-test • QED • Quasi experimental design • Quasi experimental study • Quasiexperiment* • Quasi-experiment* • Random* • Random* assign* • Random control* trial* • Randomized controlled trial* • RCT • RDD • Regression discontinuity • Reversal design* • SCD • Simultaneous treatment* • Single case design* • Single subject design* • Studies • Study • Treatment* • Withdrawal design*

Category	Search Terms	
Intervention (Broadly)	<ul style="list-style-type: none"> • Merit pay • MTP • My Teaching Partner • MyTeachingPartner • New Teacher Center • NTC • Pay for performance • Pay-for-performance • Performance based compensation • Performance-based compensation • Performance bonus • Performance pay 	<ul style="list-style-type: none"> • System for Teacher and Student Advancement • TAP • Teach for America • Teacher Advancement Program • Teacher compensation • Teacher NEAR bonus* • Teacher NEAR incentive* • Teacher NEAR pay reform • Teacher pay • Teacher salary* • TFA
Population	<ul style="list-style-type: none"> • Elementary NEAR teacher* • Elementary school* • High school* • Junior high • K-5 • K-6 • K-8 • K-12 • Middle school* • PK-5 • PK-6 • PK-8 • PK-12 • PreK-5 	<ul style="list-style-type: none"> • PreK-6 • PreK-8 • PreK-12 • Pre-K-5 • Pre-K-6 • Pre-K-8 • Pre-K-12 • Primary school* • School* • School-aged • School-based • Secondary NEAR teacher* • Secondary school*
Outcomes	<ul style="list-style-type: none"> • Achiev* • Instruction • Instructional practice • Student NEAR growth • Student NEAR outcome* • Teacher NEAR attrition • Teacher NEAR growth 	<ul style="list-style-type: none"> • Teacher NEAR knowledge • Teacher NEAR mobility • Teacher NEAR outcome* • Teacher NEAR practice • Teacher NEAR retention • Teacher NEAR stability • Teacher NEAR turnover

Additional Sources

In addition to those listed in the *WWC Procedures and Standards Handbook*, Appendix B, this review searched the following electronic database:

- ***Campbell Collaboration***. C2-SPECTR (Social, Psychological, Educational, and Criminological Trials Register) is a registry of over 10,000 randomized and possibly randomized trials in education, social work and welfare, and criminal justice.

In addition to those listed in the *WWC Procedures and Standards Handbook*, Appendix B, this review searched the following websites:

- Bill & Melinda Gates Foundation
- Center for Teaching Quality
- Center on Great Teachers and Leaders
- Consortium for Policy Research in Education
- Institute of Education Sciences
- National Center for Alternative Certification
- National Center for Analysis of Longitudinal Data in Education Research (CALDER)
- National Center on Performance Incentives
- National Council on Teacher Quality
- RTI International
- University of Chicago Consortium on Chicago School Research
- Westat
- WestEd
- Intervention-specific websites (e.g., TAP™: The System for Teacher and Student Advancement, Teach For America, Troops to Teachers)

ELIGIBILITY CRITERIA

Eligible Populations

In this review, the following populations are of interest:

- **Grade.** Teachers must provide instruction to students in grades PK–12.
- **Location.** Teachers must be employed by schools located within the United States, its territories, or tribal entities.

Potential subgroups of interest for this review include:

- Characteristics of teachers:
 - Experience
 - Certification status
 - Demographic characteristics
- Characteristics of students:
 - Special education status
 - English learner status
 - Economically disadvantaged (e.g., free or reduced-price lunch status)
 - Grade
- Characteristics of school/classroom settings:
 - Location of teaching setting (e.g., urban, suburban, rural)
 - School level (e.g., elementary, middle, high)
 - School governance (e.g., traditional public, charter, private)
 - School size
 - Economically disadvantaged (e.g., Title I status, percentage of students eligible for free or reduced-price lunch)

Eligible Interventions

Only interventions that are replicable (i.e., can be reproduced) are eligible for review. The following characteristics of an intervention must be known to reliably reproduce the intervention with different participants, in other settings, and at other times:

- Intervention description: skills being targeted, approach to enhancing the skill(s) (e.g., strategies, activities, and materials), unit of delivery of the intervention (e.g., whole group, individual), medium/media of delivery (e.g., teacher-led instruction or software), and targeted population
- Intervention duration and intensity
- Description of individuals delivering or administering the intervention

In this review, the following types of interventions may be included:

- **Practices.** A *practice* is typically an action taken by teachers as they plan, implement, or evaluate instruction. The practice must be clearly described and commonly understood in the field and literature. For example, instructional scaffolding is a teacher-enacted practice of providing strong support to students when introducing topics and concepts, but then gradually withdrawing that support so students can use/integrate the new concepts independently. An intervention that provides professional development to support such a practice is of interest to this review.
- **Policies.** A *policy* is a named condition, system, or set of formal rules that affect teachers. The policy must be commonly understood in the field and literature. Policies may be set by federal, state, or local governments or by the organization providing services. Policies may focus on changing teachers' behaviors or motivation. Examples of teacher-focused policies of interest include:
 - Financial incentives for effective teaching, or
 - Professional development requirements for renewal of certification.
- **Programs.** A *program* is a system of training supports that aim to improve the performance of teachers. Examples of interest include well-defined teacher preparation and professional development programs.

Both “branded” and “non-branded” interventions will be reviewed. Branded interventions are commercial or published programs and products that may possess any of the following characteristics:

- An external developer who provides technical assistance (e.g., instructions/guidance on the implementation of the intervention) or sells or distributes the intervention
- Trademark or copyright

Eligible Research

The *WWC Procedures and Standards Handbook* discusses the types of research reviewed by the WWC in Section II: Developing the Review Protocol and Identifying Relevant Literature (p. 4). In this review, the following additional parameters define the scope of research studies to be included:

- **Topic.** The study must examine the effect of teacher preparation, teacher induction, teacher evaluation, teacher compensation, or teacher professional development on student academic achievement or teacher behaviors that are linked to student academic achievement.
- **Time frame.** Studies must have been released or made public in 1993 or later and be obtained by the WWC for review prior to the drafting of the intervention report.
- **Sample.** The study sample must meet the requirements described in the “Eligible Populations” section above. In addition, if the study examines student outcomes, English learners must account for less than half of the student sample. If half or more

of the sample are English learners, the study may be eligible for review in the English Language Learner topic area.

- **Language.** The study must be available in English to be included in the review. Studies of interventions delivered in languages other than English will not be included in the review.
- **Location.** The study must include teachers working in the United States, its territories, or tribal entities.

Eligible Outcomes

This review includes outcomes in the following domains:

Student-level Outcome Domains

English language arts achievement. Includes outcomes in the following areas: foundational reading (word reading, fluency and/or accuracy in reading connected text, vocabulary, reading comprehension), general reading, measures of English language conventions (e.g., grammar), writing, and general English language arts achievement (i.e., a standardized test covering an array of language arts topics).

Mathematics achievement. Includes outcomes in the following areas: understanding of different subjects within mathematics, including algebra, arithmetic, calculus, geometry, probability, statistics, and trigonometry; understanding of concepts and procedures; understanding of word problems and applications; and general math achievement (i.e., a standardized test covering an array of mathematics topics).

Science achievement. Includes outcomes in any of the physical or life science disciplines, such as biology, chemistry, earth science, general science, and physics.

Social studies achievement. Includes outcomes in social studies subdisciplines, such as civics, economics, geography, history, and world cultures.

General achievement. Includes a general measure of student academic achievement, only to be documented if study authors do not distinguish students' achievement in specific areas (e.g., math, reading). Examples include composite scores from state assessments that represent a combination of reading and math scores.

Course grades, teacher reports of proficiency in the different subject areas, and assessments unrelated to academic achievement are not eligible outcome measures.

Student progression. Includes measures of students' progression in school. Constructs include:

- **Student promotion** (e.g., students' advancement to next grade level)
- **Student graduation** (e.g., students' completion of the PK–12 education system)

Outcomes measuring student behavior, health, nutrition, and other student outcomes that are not measures of academic achievement or progression in school are outside of the scope of this review.

Teacher-level Outcome Domains

Teacher instruction. Includes outcomes that reflect the quality of teachers' instruction. Examples of eligible assessments of the quality of teacher instruction include:¹

- Charlotte Danielson's Framework for Teaching (FFT)
- Classroom Assessment Scoring System (CLASS)
- Mathematical Quality of Instruction (MQI, predicting mathematics achievement)
- Protocol for Language Arts Teaching Observations (PLATO)
- UTeach Observation Protocol (UTOP)

Teacher attendance. Includes outcomes that indicate the number (percentage) of eligible work days for which the teacher is present.

Student growth scores. Includes measures of student achievement growth attached to a specific teacher or school. Examples of eligible measures include teacher or school scores from a value-added model or the Colorado Growth Model. Measures based on teacher reports of student achievement growth (e.g., scores derived from teacher reports of meeting a student learning objective) are not eligible outcomes.

Teacher retention

This review focuses on outcomes that assess whether or not teachers return to their school, their school district, or the teaching profession from year to year. More detailed mobility outcomes generally will not be reviewed because they either are captured by the key, commonly measured retention outcomes of interest (e.g., an indicator for moving to another high school in the school district would be captured by a broader outcome that measures whether a teacher returned to teach in the same school district), or may not be defined consistently across studies (e.g., an indicator for moving from an intervention school to a comparison school). Furthermore, the focus for the teacher retention outcome must be teachers' *actual* movement from a teaching position, not *expected* movement. For example, teacher ratings on whether they *expect* to return to their positions are not eligible for review.

Teacher retention at the school. Includes outcomes that measure the percentage of teachers who return to teach in the same school from year to year.

Teacher retention in the school district. Includes outcomes that measure the percentage of teachers who return to teach in the same school district from year to year.

Teacher retention in the profession. Includes outcomes that measure the percentage of teachers who return to teach from year to year, regardless of teaching location.

¹ All of these measures have been validated through the MET study (Kane & Staiger, 2012).

EVIDENCE STANDARDS

Eligible studies are assessed against WWC evidence standards, as described in the *WWC Procedures and Standards Handbook* Section III: Screening and Reviewing Studies (pp. 8–21).

Sample Attrition

The *WWC Procedures and Standards Handbook* discusses the sample attrition standards used by the WWC in Section III: Subsection B.2 Sample Attrition: Is the combination of overall and differential attrition high? (pp. 11–15).

This review uses the *liberal* boundary for attrition. This boundary was based on the assumption that most attrition in studies of teacher training, evaluation, and compensation was due to factors that were not strongly related to intervention status. For example, most attrition in these teacher-focused interventions results from exogenous factors, such as teachers' absence on the day of observations, reductions in force among teachers (especially new teachers), or parent mobility and students' absence on the days that assessments are conducted. The *WWC Procedures and Standards Handbook* contains a figure illustrating the attrition boundary and an associated table with attrition levels that define high and low attrition. Based on the choice of the boundary, the study review guide calculates attrition and whether it is high or low.

Baseline Equivalence

If the study design is a randomized controlled trial or regression discontinuity design with high levels of attrition or a quasi-experimental design, the study must demonstrate baseline equivalence of the intervention and comparison groups for the analytic sample. The onus for demonstrating equivalence in these studies rests with the authors. The *WWC Procedures and Standards Handbook* discusses how authors must demonstrate baseline equivalence in Section III: Subsection B.3 Baseline Equivalence: Is equivalence established at baseline for the groups in the analytic sample? (pp. 15–16).

Baseline equivalence must be demonstrated for the intervention and comparison groups in the analytic sample on the following pre-intervention (or baseline) characteristic:

- A pre-intervention measure of the outcome; or,
- If a pre-intervention measure is not available, an acceptable alternative pre-intervention measure, as explained below and summarized in Table 1.

Acceptable measures for student achievement outcomes. For outcomes in the student achievement domains, studies must show that the groups are equivalent on an acceptable pre-intervention measure of student achievement. A pretest measure in the same subject as the outcome is preferred; however, if a same-subject pretest is not available, a pretest measure of general achievement (e.g., a combined mathematics and reading score) would be acceptable. In addition, a pretest measure of mathematics achievement can be used to establish baseline equivalence for a science achievement outcome, and a pretest measure of reading achievement can be used for a social studies achievement outcome.

Acceptable measures for student progression outcomes. For outcomes in the student progression domain, studies must show that groups are equivalent on the following set of characteristics that are correlated with student progression.

- Grade level; **AND**
- One of the following measures of student academic performance: standardized test scores, whether behind in grade level (could be measured by age among students in the same grade), prevalence of school behavior or discipline issues, rate of school attendance, or grade point average (GPA); **AND**
- One of the following: student race/ethnicity or a measure of degree of disadvantage (i.e., free or reduced-price lunch status, poverty status, family income, English learner status, special education status, or disability status); **AND,**
- **If the unit of assignment is the school,** a school-level measure of the student progression outcome.

Acceptable measures for teacher instruction, teacher attendance, and student growth score outcomes. For outcomes in the teacher instruction, teacher attendance, and student growth scores domains, studies must show that the groups are equivalent on a pre-intervention measure of the outcome. Because no teacher or school characteristics have been demonstrated to be highly related to teacher instructional outcomes, attendance, or student growth scores, measures of teacher or student characteristics are not acceptable pre-intervention measures of teacher instruction, teacher attendance, or student growth score outcomes.

Acceptable measures for teacher retention outcomes. For outcomes in the teacher retention domains, studies must show that groups are equivalent on the following set of characteristics that are correlated with teacher retention.

- Teacher experience; **AND**
- One of the following measures of student academic performance: standardized test scores, whether behind in grade level (could be measured by age among students in the same grade), prevalence of school behavior or discipline issues, rate of school attendance, or GPA; **AND**
- One of the following: student race/ethnicity or a measure of degree of disadvantage (i.e., free or reduced-price lunch status, poverty status, family income, English learner status, special education status, or disability status); **AND,**
- **If the unit of assignment is the school,** a school-level measure of the outcome. For example, if the outcome is teacher retention in the profession and the unit of assignment is the school, equivalence must also be demonstrated on a baseline measure of the percentage of teachers in the school who returned to the teaching profession.

This review requires that, in a domain that requires statistical adjustments, the adjustment is made only for that outcome. For example, if A, B, and C are available as pre- and post-intervention measures, and the pre-intervention difference in B requires statistical adjustment, only the analysis of outcome B must adjust for B.

A review should clearly document if a study has a baseline difference in any of the following characteristics, since it could be evidence that the populations were drawn from different

settings and that the intervention and comparison groups are not sufficiently comparable for the purposes of this review:

- Socioeconomic status (SES)
- Racial/ethnic breakdown
- School location (e.g., urban, rural)

The provision of such information, however, is not a requirement of the review.

Table 1. Acceptable Pre-Intervention Measures by Outcome Domain

Outcome Domain	Acceptable Pre-Intervention Measures
Student achievement ^a	<p>Achievement in the same subject as the outcome OR Achievement in the general achievement domain OR If the outcome is science, mathematics achievement OR If the outcome is social studies, reading achievement</p>
Student progression	<p>Grade level AND Student academic performance^c AND One of the following: student race/ethnicity or degree of disadvantage^c AND, if the unit of assignment is the school, a school-level measure of the student progression outcome</p>
Teacher instruction Teacher attendance Student growth scores	<p>The same measure as the outcome</p>
Teacher retention ^b	<p>Teacher experience AND Student academic performance^c AND One of the following: student race/ethnicity or degree of disadvantage^c AND, if the unit of assignment is the school, a school-level measure of the teacher retention outcome</p>

^a Includes all five student achievement domains (English language arts, mathematics, science, social studies, and general).

^b Includes all three teacher retention domains (at the school, in the school district, and in the profession).

^c See text for examples of acceptable measures of student academic performance and degree of disadvantage.

Outcomes

The *WWC Procedures and Standards Handbook* discusses the types of outcomes, criteria the outcomes must meet, and how outcomes are reported by the WWC in Section III: Subsection B.4 (pp. 16–19). In this review, the requirements for outcome measures differ from guidance in the *WWC Procedures and Standards Handbook* in the following ways: This review follows more stringent guidance with respect to reliability for student growth scores and validity for teacher instructional outcomes. The use of a standardized test is not sufficient for establishing the reliability of a student growth score; rather, authors also must show that the student growth scores meet reliability standards. For teacher instructional outcomes to meet the validity standard for this topic area, a statistical relationship must be evident between the outcome and student achievement. The onus for demonstrating that student growth scores and teacher instructional outcomes meet reliability and validity standards rests with the authors.

For this review, measures obtained at the end of an intervention, as well as any time thereafter, are admissible. The Teacher Training, Evaluation, and Compensation review team prioritizes immediate post-intervention findings for developing intervention ratings and improvement indices because these findings are most prevalent in teacher training, evaluation, and compensation studies. Measures occurring several months or years after the intervention may provide strong evidence for an intervention’s effectiveness. Therefore, the Teacher Training, Evaluation, and Compensation review team will include follow-up findings, when available and appropriate, in supplemental appendices to the intervention report.

Statistical Adjustments

The *WWC Procedures and Standards Handbook* discusses the types of adjustments made by the WWC in Section IV: Subsection B Statistical Significance of Findings (p. 24).

Other Study Designs

Studies that use regression discontinuity or single-case designs are eligible for review using the appropriate pilot standards.

The *WWC Procedures and Standards Handbook* discusses the pilot standards for reviewing regression discontinuity design studies in Appendix D.

The *WWC Procedures and Standards Handbook* discusses the pilot standards for reviewing single-case design studies in Appendix E.