

REVIEW PROTOCOL FOR INDIVIDUAL STUDIES IN THE POSTSECONDARY EDUCATION TOPIC AREA VERSION 3.0

Topic Area Focus

This document outlines the processes to be used in the review of individual studies in the Postsecondary Education topic area. Postsecondary education is any form of schooling occurring after the secondary level (i.e., after high school). This area may include interventions that increase the success of students transitioning to postsecondary education, instructional programs that occur in-person or online as part of postsecondary schooling, out-of-classroom practices such as mentoring, adult education activities, and so on.

Types of Reports

The What Works Clearinghouse (WWC) reviews individual studies in the Postsecondary Education topic area in two formats: quick reviews and single study reviews. [For information about studies reviewed for intervention reports, please see the relevant topic area protocols.]

Quick reviews provide education practitioners and policymakers with timely, preliminary objective assessments of the quality of the research evidence from recently released research papers and reports that have received significant coverage in the media. They are brief study summaries that describe the study being reviewed, its findings, and the WWC's rating of the study (which may be provisional due to the nature of the review process).

Single study reviews are reviews of individual studies, describing the program, policy, or practice studied; indicating whether the study meets WWC evidence standards; and summarizing the study findings on effectiveness.

Quick reviews and single study reviews are carried out by a lead methodologist who is supported by a content expert and by WWC-certified coders. The lead methodologist and content expert play a central role in determining the content and quality of the final products. The lead methodologist makes technical decisions for the team and serves as the point of contact for study authors and the Institute of Education Sciences (IES). The content expert provides context-specific support and guidance (e.g., determining the specific subgroups and outcomes that the review should highlight).

Identifying Studies for Review

There are two distinct mechanisms by which a study will come to be reviewed under this protocol. First, a study could qualify for review by receiving significant media attention. In addition, a study could be reviewed at the request of IES.

To identify studies that have received significant media attention, media scans are run weekly, and the following scoring rubric is applied to studies receiving media mention:

1. Significant media mention:

5 points for a news item in *The New York Times*, *The Washington Post*, *The Chronicle of Higher Education*, *Education Week*, or *Inside Higher Ed*

3 points for a blog or opinion piece in *The New York Times* or *The Washington Post*

1 point for a blog or opinion piece in another higher education outlet

Based on the “significant media mention” criterion, studies not mentioned in these outlets would not be eligible for review.

2. Timeframe for study

To be eligible for review due to receiving significant media attention, the study must have been published or reported within the last 2 years.

3. Evidence of (or claims of) causal relations:

Studies must generate evidence of (or make claims about) causal relations. Assign 5 points for using a “key causality term” (e.g., “effects of,” “causes,” “impacted,” “created a change in”) or having a study design that claims to be a randomized controlled trial, “rigorous” quasi-experiment, single-subject, or regression discontinuity design.

Assign –10 points for studies that do not claim to be randomized controlled trials, rigorous quasi-experiments, single-subject, or regression discontinuity designs, or studies that are not described using one of the “key causality terms.” Operationally, this scoring system implies such studies would not be a priority for review (see scoring rubric below).

4. Study size:

Add 3 points for a multiple institution study.

5. “Buzz”:

Add 3 points for being in the most viewed or most emailed list for the higher education outlet.

Studies will be scored based on the following:

- Higher priority: 11+ points
- Lower priority: 1–9 points
- Not a priority: < 0 points

With IES approval, studies that meet the high priority threshold are reviewed using the outcomes identified in this review protocol or those identified in a relevant topic area protocol.

In most cases, studies identified for review through the media mentions mechanism will first be reviewed as a quick review, followed by a single study review. The results of study assessments

that are reviewed based on an IES request will be published only in a single study review.

ELIGIBILITY CRITERIA AND EVIDENCE STANDARDS

Studies must meet several criteria to be eligible for review. These relate to the population that was sampled, the study design that was used, the outcomes that were measured, and when the study was conducted. Each of these is discussed below.

Populations to be Included

To be eligible for review under this protocol, a study must utilize postsecondary students in the United States or Canada. Studies of other student populations (e.g., high school students) are eligible for review only if they assess an outcome relevant to postsecondary education (e.g., postsecondary enrollment; see below).

Many studies provide effect size estimates for subgroups of students. The content expert for the review is responsible for determining which subgroups should be reported for any given review; this determination is made independent of the study's results (e.g., before the study rating is known). In general, the WWC determines a study rating based on average intervention effects and will report subgroup analyses only for groups that are identified in the protocol as being of theoretical, policy, or practical interest. For studies reviewed under this protocol, these subgroups are students who are (a) first-generation college students, (b) racial/ethnic minorities, (c) academically underprepared, (d) students from low socioeconomic status backgrounds, e.g., Pell Grant students, and (e) community college students. In addition, the WWC will report subgroup effects for gender when they are available.

Types of Studies to be Reviewed

Following the current *WWC Procedures and Standards Handbook* (version 3.0), to be eligible for review a study must be a primary analysis of the effect of an intervention. If a study does not examine the effects of an intervention, or if it is not a primary analysis (e.g., if it is a meta-analysis or other literature review), then it is not eligible for review. Studies that do not examine the effectiveness of an intervention, but have been portrayed so in the media, may still be eligible for a quick review.

In addition, the study must have an eligible design. Eligible study designs include randomized controlled trials (RCTs), quasi-experimental designs (QEDs), and certain kinds of single-case designs. The category “quasi-experimental” is broad. The WWC currently has standards for reviewing studies that use matching or statistical control in an attempt to equate non-equivalent groups, and also pilot standards for simple versions of the regression discontinuity design (i.e., those that use a single forcing variable). The WWC currently does not have standards for other types of quasi-experimental designs, such as the instrumental variable approach and interrupted time series designs. Therefore, studies using these types of research designs are not eligible for review under this protocol.

Relevant Outcome Domains

To be eligible for review, a study must also assess a relevant outcome domain. The content expert, in consultation with the review team, identified the following domains to guide the review of studies in postsecondary education that do not have a corresponding protocol: (a) access and enrollment, (b) credit accumulation, (c) academic achievement, (d) attainment, and (e) the labor market. Measures of actual behavior are preferred to those that measure intentions and related constructs. When studies present both types of measures for an outcome (e.g., both intention to enroll and actual enrollment), the WWC will focus on the behavioral measure.

- ***Access and enrollment*** refers to the process of applying to, actually enrolling, and attending a postsecondary institution. Examples of ways that enrollment might be operationally defined in studies include (a) applied vs. didn't apply to college, (b) number of applications completed, (c) attended vs. did not attend a postsecondary institution, or enrolled vs. did not enroll in the first semester or year of college, (d) selectivity of enrollment institution, (e) full-time vs. part-time enrollment, and (f) 4-year vs. 2-year vs. non-enrollment.
- ***Credit accumulation*** and persistence refers to progress toward the completion of a degree, certificate, or program. Examples of ways that credit accumulation might be operationally defined in studies include: (a) number of credits earned toward degree completion, (b) proportion of degree bearing vs. non-degree bearing credits earned, and (c) ratio of credits earned to credits attempted. In addition, measures of persistence are also considered to be part of the credit accumulation domain. Examples of ways in which persistence might be operationally defined include (a) enrolled vs. did not enroll in the next semester or next year of college and (b) number of semesters or terms of continuous enrollment. Some studies may refer to persistence as retention, as in students are "retained" in college.
- ***Academic achievement*** assesses the extent to which students adequately complete expected coursework. Examples of ways that academic achievement might be operationally defined in studies include grade point average and the ratio of courses passed vs. failed.
- ***Attainment*** refers to the completion of a degree, certificate, or program. Examples of ways attainment might be operationally defined in a study include certificate completion rates and degree completion rates.
- ***Labor market*** refers to outcomes related to employment after the postsecondary experience. Examples of ways that labor market outcomes might be operationally defined in studies include (a) employed vs. not, (b) employed full-time vs. employed part-time, (c) employed in field vs. not, and (d) income earned.

Outcomes measured at different points in time. For most outcomes in the Postsecondary Education domain, the longest follow-up period available for a variable should be selected as primary; findings from any earlier time points should also be included in supplemental tables. In

the access and enrollment domain (defined below), the *first* measure of enrollment (e.g., enrolled vs. not enrolled) should be selected as primary. Measures of enrollment that occur *after* the first semester or year of college would fall under the credit accumulation domain and the longest follow-up period should be selected as the primary measure.

Timeframe for Studies

Studies must have been conducted within the last 20 years to be eligible for review under this protocol.

REVIEW OF STUDIES AGAINST WWC EVIDENCE STANDARDS

All studies will be reviewed against the WWC Evidence Standards, using the most current version of the *WWC Procedures and Standards Handbook*. Generally, these standards assess outcome reliability and validity, attrition, baseline equivalence, and similar methodological and statistical issues. This review determines the overall WWC study rating. Details related to sample attrition in RCTs and baseline equivalence in QEDs and high-attrition RCTs are further articulated in this protocol.

Sample Attrition

Reviews of studies that are governed by this protocol will use the *liberal* boundary for attrition. The selection of this boundary was based on the assumption that most attrition in studies of interventions focused on postsecondary students is due to factors that are not strongly related to intervention status. The WWC's postsecondary content expert can change the boundary to use if this assumption seems inappropriate for a given intervention.

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.

If demonstration of baseline equivalence is required for a study, the following pre-intervention (or baseline) characteristics should be used:

- A pre-intervention measure of the outcome (i.e., a pretest) or a close proxy. In the postsecondary literature, pretests on the outcomes are often not available. When pretests or a close proxy are not available, studies must demonstrate baseline equivalence on the following two domains:
 - A baseline measure of academic achievement (e.g., high school grade point average, SAT/ACT scores), and
 - A baseline measure of student socio-economic status (e.g., FAFSA expected family contribution, family income, free- or reduced-price lunch status, parent education levels, Pell Grant eligibility)

In cases where multiple baseline measures of SES and/or academic achievement are available, the content expert is responsible for selecting the variable(s) to be used in the baseline equivalence assessment prior to the equivalence assessment being performed. For example, if both math and verbal scores on a college entrance exam are available, and the primary outcome is whether or not students passed their first college-level math course, then the content expert may decide that the score on the math portion of the entrance exam is the only achievement measure on which baseline equivalence should be assessed. However, if the primary outcome is attainment, then the content expert might decide to assess balance on both the math subtest and the verbal subtest.

PROCEDURES FOR STATISTICAL ADJUSTMENT FOR STUDIES WITH BASELINE COVARIATE IMBALANCE

These procedures apply to all studies for which baseline equivalence must be demonstrated (i.e., RCTs with high attrition and QEDs).

If a pretest or close proxy is available for an outcome and the difference between conditions at baseline is shown to be within the range that requires statistical adjustment, the statistical adjustment is only needed for that outcome. For example, if vocabulary, reading comprehension, and reading fluency are available as pre- and post-intervention measures, and the pre-intervention difference in reading comprehension requires statistical adjustment but the others do not, only the analysis of reading comprehension must adjust for baseline differences in reading comprehension (no adjustments are required for the other outcomes).

For outcomes that do not have a pretest or close proxy, if the difference between conditions at baseline on one of the required covariates is shown to be within the range that requires statistical adjustment, then adjustment is required only for the covariate in the adjustment range. For example, if academic achievement is judged to be within the range that requires statistical adjustment and SES is very closely balanced (i.e., it is not in the adjustment range), then all outcomes without pretests must adjust for the measure of academic achievement, and adjustment for baseline SES is not required.