

REVIEW PROTOCOL FOR POSTSECONDARY CAREER AND TECHNICAL EDUCATION (CTE) INTERVENTIONS VERSION 4.0

This protocol guides the review of research that informs the What Works Clearinghouse (WWC) intervention reports in the postsecondary Career and Technical Education (CTE) topic area. The review-specific protocol is used in conjunction with the *WWC Procedures and Standards Handbook* (version 4.0).

PURPOSE STATEMENT

A better understanding of CTE interventions is critical because there has been a growing demand for workers with postsecondary education and training – and those with a high school diploma or less have struggled to find and keep jobs with middle-class wages (Acemoglu & Autor, 2011; Carnevale, Jayasundera, & Hanson, 2012; Carnevale, Jayasundera, & Gulish, 2016; Rosen, Visher, & Beal, 2018). Meanwhile, rising requirements for skilled service and blue-collar industry workers have resulted in employer demand for better alignment between technical education and industry needs (Carnevale, Cheah, Ridley, & Strohl, 2017). In this environment, attention from researchers and policymakers has led to significant investments to improve postsecondary CTE, which can potentially increase postsecondary attainment and earnings while also building the workforce’s skills needed in occupations currently in demand.

WWC reviews in this topic area focus on postsecondary CTE interventions that help students achieve technical skill proficiency and/or an industry-recognized credential, a certificate, a license, an associate degree, or in the longer term, a baccalaureate degree. The primary goal of these interventions is to improve labor market outcomes, including employment and earnings. Systematic reviews of evidence in this topic area address the following questions:

- Which CTE interventions are effective at helping students progress toward the completion of an industry-recognized credential, a license, a certificate, or a postsecondary degree?
- Which CTE interventions are effective at helping students obtain an associate or a baccalaureate degree?¹

¹ CTE interventions designed to support progression toward or completion of a graduate degree fall outside of the scope of such reviews. The reason is that graduate work will generally entail highly specialized training delivered to students who likely already have base credentials needed for middle-class employment and wages.

- Which CTE interventions are effective at helping students obtain an industry-recognized credential, a license, or a certificate? Which CTE interventions are effective at helping students achieve technical skill proficiency?
- Which CTE interventions are effective at helping students obtain and/or retain employment?
- Which CTE interventions are effective at increasing earnings?

KEY DEFINITIONS

Career and technical education (CTE) interventions. Postsecondary CTE interventions are programs that develop the skills and knowledge required for specific jobs or fields of work. CTE interventions may include components, such as coursework; experiences in a job setting, such as clinical experiences in training for healthcare professions, practica, apprenticeships, or on-the-job training; or general training, such as development of problem-solving skills, work attitudes, general employability skills, and technical skills.

Education outcome. CTE postsecondary interventions lead to a range of education outcomes: technical skill proficiency, industry-recognized credentials, certificates, licenses, and degrees.² Individuals demonstrate *technical skill proficiency* by passing technical skill assessments aligned with industry-recognized standards. *Certificates* and *associate degrees* are conferred by postsecondary institutions such as community and technical colleges, public technical schools, and private for-profit trade schools. Associate degrees typically require two years of coursework including the completion of general education requirements, whereas certificates typically require shorter courses of study.

Some CTE interventions are developed in partnership with local employers or joint labor-management partnerships and lead to industry-recognized credentials (*licenses* and *certifications*), though not occurring within a postsecondary institution such as a community college or technical school. Industry-recognized credentials are awarded in recognition of an individual's attainment of measurable technical or occupational skills necessary to obtain employment or advance within an occupation. These technical or occupational skills are generally based on standards developed or endorsed by employers or industry associations. General work skills certificates, such as work readiness or safety/hygiene, do not count as industry-recognized credentials even if required for employment (DOL/ETA, 2016).

Many CTE interventions prepare participants to obtain a certification or license, though they may not lead to another type of award. Licenses and certifications both are typically awarded through

² Carl D. Perkins Career and Technical Education Act of 2006, as Amended by the Strengthening Career and Technical Education for the 21st Century Act.

validation of a certain set of skills or competencies. Licenses give the holder a legal permission to perform specific regulated tasks or occupations; certifications indicate mastery of certain tasks, without granting legal permissions.

Labor market outcomes. Labor market outcomes include employment (defined as having a paid job); cumulative employment over time (defined as being employed a certain number of days, weeks, months, or years over a specified time period); hours worked; and earnings (defined as income received from work).

Institutions identified under Title II of WIOA. Title II of the 2014 Workforce Innovation and Opportunity Act (WIOA) identifies the types of institutions and organizations that are eligible to provide education and training.³ These include local education agencies, community-based or faith-based organizations, volunteer literacy organizations, higher education institutions, non-profit agencies, libraries, and public housing authorities.

Displaced workers. A displaced worker is one who has been permanently laid off due to a plant downsizing or closing or to the elimination of that position within the firm. These workers are of particular interest for many CTE interventions.

ELIGIBILITY CRITERIA

Eligible Population and Subgroups

To be eligible for review under this protocol, a study must include participants of a postsecondary CTE intervention in the United States who are age 16 years and older. Studies where at least half the sample consists of high school students, even those concurrently enrolled in postsecondary education (i.e., dual enrollment), are not eligible for review under this protocol.

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. Eligible subgroups of interest for this review include these:

- Students who do not have a high school diploma, GED®, or other alternative secondary credential;

³ The U.S. Department of Education's Office of Career, Technical, and Adult Education (OCTAE) is responsible for the administration of WIOA, Title II, the Adult Education and Family Literacy Act (AEFLA). AEFLA authorizes the expenditure of funds for programs integrating basic skills education and postsecondary occupational training and for dual enrollment in basic skills education and postsecondary career and technical education. Studies of these types of educational services would be candidates for review under this protocol.

<https://www2.ed.gov/about/offices/list/ovae/pi/AdultEd/wioa-reauthorization.html>

- Students with a high school diploma, GED®, or other alternative secondary credential;
- Students who are limited-English-proficient speakers;
- First-generation college students;
- Underemployed or unemployed workers;
- Displaced workers;
- Low-income or otherwise economically disadvantaged individuals;
- Single parents, including single pregnant women;
- Recipients of public assistance;
- Men or women in nontraditional occupations;
- Older students (e.g., age 25 or older); and
- Students with disabilities.

If a study presents findings separately for several groups of students without presenting an aggregate result, the WWC will query authors to see whether they conducted an analysis on the full sample of students. If the WWC is unable to obtain aggregate results from the author, the WWC will average findings across subgroups within a study to use as the primary result and will present the subgroup results as supplemental analyses.

Eligible Institutions

A postsecondary CTE intervention helps students achieve technical skill proficiency, an industry-recognized credential, a certificate, a license, or an associate or, in the longer term, baccalaureate degree. The primary goal is improving labor market outcomes, including employment and earnings. Eligible interventions must target services to individuals age 16 and older who are not enrolled in K-12 education.

For the purposes of this review, eligible CTE interventions may be developed for delivery by multiple types of entities. These entities could include these:

- Public or private non-profit institutions of higher education that offer CTE courses that lead to technical skill proficiency or a recognized postsecondary credential, including an industry-recognized credential, a certificate, a license, or a postsecondary degree;
- A local educational agency providing education at the postsecondary level;
- An area career and technical education school providing education at the postsecondary level;

- A postsecondary educational institution controlled by the Bureau of Indian Education or operated by or on behalf of any Indian Tribe;
- An educational service agency; or
- Other entities identified under the Higher Education Act, WIOA Title I and II, and the Perkins Act.⁴

Eligible Interventions

The interventions considered for inclusion will be determined after a search of the publicly available literature by the CTE Review Team, as well as after review of nominations submitted to the WWC through the Help Desk or through the topic area team content experts. The intervention must be a postsecondary CTE intervention with a primary focus on developing the technical skills and knowledge required for specific jobs or fields of work through formal coursework. Eligible CTE interventions may lead to a range of education outcomes, including technical skill proficiency, industry-recognized credentials, certificates, licenses, associate degrees, and in the longer term, baccalaureate degrees.

The intervention must operate in the United States, its territories, or tribal entities.⁵ Eligible CTE interventions include these:

- Certificate and associate degree programs;
- Training interventions;
- Sectoral training programs, which aim to actively match worker skills on the supply side of the labor market with what employers seek on the demand side of the market (Holzer, 2015);
- Career pathways programs, which are a series of structured and connected programs and support services that enable students to advance over time to higher levels of education and training (Career Ladders Project, 2013);
- Integrated education and training courses/curricula;
- Simultaneous enrollment in basic skills and career and technical education; and
- Industry-recognized apprenticeships that meet the U.S. Department of Labor's definition.⁶

⁴ Carl D. Perkins Career and Technical Education Act of 2006, as Amended by the Strengthening Career and Technical Education for the 21st Century Act, Sec. 3(19)(A).

⁵ The language of instruction does not need to be English.

Eligible CTE interventions may comprise a single component, such as a specific training course, or multiple components, such as a career pathways program that involves occupational training and case management to connect students to comprehensive supportive services. Both single- and multiple-component types of interventions are eligible for WWC review.

To be eligible, interventions must be able to be implemented by practitioners other than the developers of the approach. The following characteristics of an intervention will be documented by the WWC, so that others are able to determine whether they are interested in learning more about the intervention:

- Targeted population;
- Description of intervention provider or administrator;
- Description of the intervention, including details of the services provided and of other activities that are part of the intervention;
- Occupational field targeted by the intervention;
- Length of calendar time and number of hours required to implement the intervention;
- Type of organizations/institutions in which the intervention can be implemented;
- Cost, which may include staff salaries to participate in training or provide the intervention; expenses for space, materials, and equipment needed for training and/or providing the intervention; travel and per diem expenses for training; price charged for intervention participants; and other intervention inputs; and
- Source of funding (when available).

These criteria reflect that CTE interventions can target a range of students and be implemented in a range of institutional settings. Information about the types of students and institutions is needed to replicate the program appropriately.

Eligible Research

- ***Topic.*** The study must be focused on the receipt of postsecondary degrees, an industry-recognized credential, a license, or a certificate; credits earned; employment; or earnings.
- ***Time frame.*** The study must have been published in 1997 or later. Rigorous evaluations of interventions implemented in this time frame test versions of interventions most likely to be available today and under conditions most likely to be current.

⁶ The U.S. Department of Labor is in the process of using the definition adopted by the Task Force on Apprenticeship Expansion (2018), which specifies that an industry-recognized apprenticeship includes a paid-work component and an educational or instructional component, wherein an individual obtains workplace-relevant knowledge and skills. See <https://www.dol.gov/apprenticeship/docs/task-force-apprenticeship-expansion-report.pdf>

- **Sample.** The study sample must meet the requirements described in the “Eligible Population and Subgroups” section above.
- **Design.** The study must be empirical, using quantitative methods and inferential statistical analysis, and as described by the *WWC Procedures and Standards Handbook* (version 4.0), must take the form of a randomized controlled trial (RCT) or use a quasi-experimental design (QED), a regression discontinuity design (RDD), or a single-case design (SCD).
- **Language.** The study must be available in English.
- **Location.** The study must meet the requirements described in the “Eligible Population and Subgroups” section above.

Eligible Outcomes

To be eligible, research submitted for review for the CTE topic area must explore one or more of 10 primary outcome domains, organized by two categories (Education and Labor Market):⁷

- **Education Outcome Domains.** When a given education outcome is measured at different time points, the review prioritizes the measure with the shortest follow-up period.
 1. **Credit accumulation.** This domain includes the completion of “gateway” or requisite courses for CTE career pathways. Examples of ways that credit accumulation and course completion may be operationally defined in studies include the number of college-level credits earned toward a credential or degree or the number of non-credit-bearing courses completed toward a license or credential that are not “for credit.” Outcomes measuring graduate-level credits earned will not be considered.
 2. **Postsecondary degree attainment.** This domain refers to the completion of an associate or a baccalaureate degree. Outcomes pertaining to completing or progressing toward a graduate-level degree will not be included.
 3. **Industry-recognized credential, certificate, or license completion.** This domain refers to the completion of an industry-recognized credential, certificate, or license. Examples of ways completion might be operationally defined in a study include (a) certificate

⁷ The use of two categories of outcomes, where the first category captures the four education outcome domains and the second category captures the six labor market outcome domains, has relevance for tests of baseline equivalence. A study must establish baseline equivalence using an acceptable baseline measure within the category in which the outcome of interest is reported. In the section below on baseline equivalence, we provide further description of how these categories are used when testing for baseline equivalence for both (1) outcomes within the category of labor market domains and (2) outcomes within the category of education domains.

completion rates, (b) non-degree-award receipt rates, and (c) certifications from third-party licensing or credentialing bodies.

4. **Technical skill proficiency.**⁸ This domain refers to assessments that measure technical skills at the occupation level. These assessments are aligned with industry-recognized standards. An example of ways that technical skill proficiency might be operationally defined in studies is end-of-course assessments administered by the National Occupational Competency Testing Institute.⁹

- **Labor Market Outcome Domains.** For the labor market outcomes in the *employment* and *earnings* domains, the review considers short-, medium-, and long-term impacts as separate outcome domains. This yields a total of six domains for the Labor Market category. Short-term follow-up is defined as one to two years after earliest possible program completion;¹⁰ medium-term follow-up is defined as three to four years after earliest possible program completion; and long-term follow-up is defined as five or more years after earliest possible program completion. When a given labor market outcome is measured at different time points within the same outcome domain, the review prioritizes the measure with the shortest follow-up period.

The WWC review should clearly document the source of the outcome data (e.g., survey data, Unemployment Insurance data, etc.), as each source of data has limitations. For instance, self-reported employment and earnings data may have measurement error, whereas administrative data may have gaps.¹¹

- 5-7. **Short-, Medium-, and Long-Term Employment.** This domain refers to having a paid job. Examples of ways that employment might be operationally defined in studies include (a) indicator of any paid employment, (b) number of months or quarters employed over the follow-up period, or (c) number of hours worked in an average

⁸ The *technical skill proficiency* domain in this protocol is parallel to an *academic achievement* domain in other WWC protocols. Outcomes in this domain measure achievement in the content area covered by the intervention. In contrast, the *industry-recognized credential, certificate, or license completion* domain is parallel to an *education attainment* domain in other protocols. Although measures in these two domains may be related, they are distinct from each other in the same way that high school graduation is distinct from achievement measured by a high school exit exam.

⁹ <https://www.nocti.org/PDFs/BlueprintPageLinks/Assessments%20at%20a%20Glance.pdf>

¹⁰ The Principal Investigator has discretion to include outcomes with follow-up periods of less than a year as short-term follow-up.

¹¹ For instance, Unemployment Insurance data do not cover the self-employed, the military, some non-profit and church workers, student workers, agricultural workers, and those who work for the federal government. Review team leadership will have discretion to determine whether a given measure was gathered from an appropriate data source.

week over the previous month. Employment outcomes must be defined over a specific period (e.g., six months prior to data collection, the previous month).

8-10. *Short-, Medium-, and Long-Term Earnings.* This domain refers to income received from work. Examples of ways that earnings might be operationally defined in studies include (a) cumulative earnings over the previous six months or (b) earnings in a typical week in the previous month.¹²

For measures to be considered appropriate for this review, they must be defined in the same way for all study participants (including those who are not enrolled or not employed). For example, a study that assesses earnings and hours worked must do so for study participants who are not employed as well as for those who are. For this reason, this review does not examine hourly wage as an outcome, as it is not defined for study participants who are not employed.¹³

EVIDENCE STANDARDS

Eligible studies are assessed against WWC evidence standards, as described in the *WWC Procedures Handbook*, Section IV: Screening Studies and Section V: Reviewing Studies, as well as the *WWC Standards Handbook*.

Sample Attrition

The *WWC Standards Handbook* discusses the sample attrition standards used by the WWC in the following sections:

- Step 2 of the WWC review process for individual-level group design studies in Section II.A—“Sample Attrition: Is the combination of overall and differential attrition high?”
- Step 1 of the WWC review process for cluster-level group design studies in Section II.B—“Is the study a cluster RCT with low cluster-level attrition?”

¹² Total individual or household income is not an eligible outcome. A CTE program that successfully increases individual earnings might decrease public benefit receipt, with the result that the participant’s income might increase, decrease, or remain constant even though the intervention successfully increased earnings. Further, household income might include spousal earnings; a CTE program that increases a participant’s earnings might induce the spouse of the participant to reduce his/her hours worked, especially if the spouse was working an additional job to support the participant during training.

¹³ Individuals who are not employed are typically considered to have worked zero hours and had zero earnings. However, this approach cannot be used to define *hourly wage* because that measure is defined by dividing earnings by hours worked, and zero divided by zero is mathematically undefined. The econometric techniques that appropriately estimate impacts on hourly earnings involve multi-step estimation (see Heckman, 1979) and cannot be reviewed using current WWC standards.

- Step 3 of the WWC review process for cluster-level group design studies in Section II.B—“Is there a risk of bias due to non-response of individuals?”
- Section 3 of the WWC standards for reviewing complier average causal effect estimates in Section II.D—“Calculating attrition when rating CACE estimates”
- Standard 2 of the WWC standards for reviewing regression discontinuity designs in Section III.C

In the *WWC Standards Handbook*, Figure II.2 illustrates the attrition boundary and Table II.1 reports 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. For most studies this review will entail use of the *optimistic* boundary for attrition based on the assumption that most attrition in studies of postsecondary CTE would be due to factors that are not strongly related to intervention status. We assume that adult students in CTE can have a range of life events that lead them to have missing outcome data that are unrelated to intervention status.

However, the review team leadership has discretion to use the *cautious* boundary if they have reason to believe that much of the attrition is endogenous—that is, related to the intervention. For guidance on when to apply the more cautious boundary, consider the following examples:

- A training opportunity that is thought to have strong, public appeal in a community could lead toward demoralization among study participants assigned to a business-as-usual control condition. These discouraged control group members may be less likely to provide outcome data through a follow-up survey. Similarly, studies where the control group members receive no services may be less engaged with the study and, therefore, less likely to provide outcome data through a follow-up survey. This is in contrast to studies that use a sample of students in K-12, where students in the control condition have less choice about whether to provide outcome data through a post-test.
 - In these examples, studies that use administrative records to construct outcomes could obviate the concern that loss of outcome data is endogenous to the intervention and provide a basis for the review team leadership to continue to apply the optimistic boundary.
- Use of administrative data might in other cases not solve the issue but instead contribute to a concern. For example, some interventions may promote sample movement across states. If the study relies on administrative unemployment insurance records to capture outcomes, then outcome data may be missing for those who move across states if state-level unemployment data is not available (most states do not capture earnings data from workers who move across state lines). Therefore, incomplete unemployment insurance data could contribute to bias in impacts on labor market related outcomes.
- As another example, consider a study in which provision of an intervention credential makes it more likely that treated workers become self-employed, and the study relies on

administrative unemployment insurance data as an outcome measure. Such data do not typically capture self-employed workers, leading to differential attrition that may be caused by the intervention. This scenario may prompt reason to utilize a cautious attrition boundary.

However, these examples are not expected to be common; they instead provide guidance on for the study-specific attrition threshold that should be considered by the review team leadership. Study reviewers should remain vigilant about these issues and raise awareness among the review team leadership should there be a case where attrition might be related to the intervention. Note that changes to applications of the boundary will be documented and justified in the associated WWC reports.

Joiners in Cluster Randomized Controlled Trials (RCTs)

Clusters that might be seen in CTE studies include but are not limited to community colleges, Workforce Innovation Boards, and community agencies. According to the *WWC Standards Handbook* (page 23), a cluster RCT must limit the risk of bias due to individuals entering the cluster after the time of random assignment (joiners) in order to receive the highest rating. The WWC defines a **joiner** as any study participant (e.g., worker, community college student) who enters a cluster after the results of random assignment. A joiner might seek out access to a particular study condition. Or a joiner's placement in a study might be influenced by another individual (e.g., coworkers, job-training counselors). Since joiners are not a part of random assignment, their presence in an analytic sample has the potential to introduce bias into estimates of an intervention's effectiveness. Consider, for example, a situation where highly motivated displaced workers join a cluster such as a community agency randomly assigned to provide a new CTE intervention. Such workers might be prone to work hard to gain re-employment regardless of their CTE experience, and thereby bias an effect size estimate in a labor market category domain.

In some cases some joiners may enter clusters after random assignment, but before individuals involved with placement (the joiners themselves or other stakeholders) knew the randomly assigned conditions of the clusters. The WWC never considers these joiners to pose a risk of bias because the decisions that led these individuals to join clusters could not have been affected by the intervention. The burden for demonstrating that individuals could not have known about the intervention rests with the study authors. In some cases, joiners who enter clusters relatively early in the study period have less potential to introduce bias than those who enter later. Therefore, WWC review teams sometimes differentiate between **early** and **late** joiners.

Late joiners are those who enter clusters after some specified time period past initial intervention onset. For postsecondary CTE reviews, this period is set to end 6 weeks after the introduction of the intervention (often the start of the school year when an intervention is delivered in a college). Joiners who enter clusters after 6 weeks may be more likely to do so because of the intervention. A study that includes at least one late joiner in the analytic sample *does not limit the risk of bias*

from joiners, meaning associated study contrasts can at best meet WWC standards with reservations (see Figure II.4. Review Process for Cluster-Level Assignment Studies, page 22 of the WWC Standards Handbook).

With that background, the general default disposition in the CTE review is that ***all joiners in the analytic sample are expected to pose a risk of bias*** (there are exceptions outlined below). Therefore, a study that includes at least one such joiner in the analytic sample does not limit the risk of bias from joiners. This is because study samples will be comprised of adults who can presumably choose their preferred CTE experiences. It seems implausible that adults would not understand what CTE intervention they are going to experience and have the capacity to exercise choice when it comes to pursuing different types of CTE. The CTE Review Team Leadership does however have discretion to revise this general guidance since not all scenarios can be anticipated.

Key considerations when thinking about the risk of bias from joiners are the degree to which a CTE intervention was well-advertised, offers supports and benefits that are not made available in a comparison group, or is likely to be perceived by joiners as having some special benefit (e.g., industry-recognized apprenticeships).

One common exception to the general default rule that ***all joiners in the analytic sample pose a risk of bias*** is when the unit of assignment is a college, courses within colleges, or a group of colleges (such as a coordinated group of community colleges) and these conditions are in place:

- The intervention is not expected to directly affect enrollment or placement decisions. In this case, ***only late joiners pose a risk of bias***. One example of an intervention that should not directly affect enrollment or placement decisions is when treatment and comparison groups are offered different types of potentially useful services (e.g., career pathways program version 1 vs. career pathways program version 2), such that we would not expect that individuals would be more likely to go out of their way to join one version over the other.
- Another example of an intervention that does not directly affect enrollment is an intervention where treatment group members receive a low-profile basic skills approach that is integrated into curricula of their community college, where individuals may be unlikely to know about this add-on to the curricula even after the point of random assignment. This is consistent with aforementioned idea that joiners are likely to be unaware that a cluster is part of a study condition.

To reiterate the CTE default, if the intervention may affect enrollment or placement decisions (such as a highly publicized program for displaced workers), then ***all joiners pose a risk of bias***. A study of such an intervention that includes one or more joiners in the analytic sample *does not limit the risk of bias from joiners*.

Again, not all scenarios can be anticipated. When an intervention and unit of assignment in a cluster RCT do not fall into a category described above, the Review Team Leadership has discretion to make a decision about whether the joiners pose a risk of bias. Any time such discretion is exercised, the background and rationale of decisions will be documented in intervention reports.

Baseline Equivalence

If the study design is a randomized controlled trial or regression discontinuity design with high levels of attrition, or is a quasi-experimental design, the study must satisfy the baseline equivalence requirement for the analytic intervention and comparison groups. The onus for demonstrating equivalence in these studies rests with the authors. The *WWC Standards Handbook* discusses how authors must satisfy the baseline equivalence requirement in:

- Step 3 of the WWC review process for individual-level group design studies in Section II.A—“Baseline Equivalence: Is equivalence established at baseline for the groups in the analytic sample?”
- Steps 4 and 7 of the WWC review process for cluster-level group design studies in Section II.B—“Does the study establish equivalence of individuals at baseline for groups in the analytic sample?” and “Does the study establish equivalence of clusters at baseline for groups in the analytic sample?”, respectively
- Section 5 of the WWC standards for reviewing complier average causal effect estimates in Section II.D—“Procedures for Rating CACE Estimates when Attrition is High”
- Standard 3 of the WWC standards for reviewing regression discontinuity designs in Section III.C

1. Baseline equivalence of individuals

For studies that must satisfy baseline equivalence of individuals, including cluster-level assignment studies being reviewed for evidence of effects on individuals, the baseline equivalence requirement must be satisfied for the analytic intervention and comparison groups. Pre-intervention measures of the outcome used in the analysis will be acceptable (e.g., baseline knowledge of some technical skill) but in the postsecondary CTE literature this strategy is not

always feasible (e.g., it would be unusual to have a meaningful baseline measure of postsecondary degree attainment).¹⁴

As described in the section on eligible outcomes above, the use of two categories of outcomes, where the first category captures the four education outcome domains and the second category captures the six labor market outcome domains, has relevance for tests of baseline equivalence. In this section, we provide further description of how these categories are used when testing for baseline equivalence for both (1) outcomes within the category of labor market domains and (2) outcomes within the category of education domains.

Baseline equivalence must be established on a pre-intervention measure of the outcome or a close proxy (i.e. a pretest measure in the same outcome domain). In cases where pretests in the same domain are not feasible – that is, the outcome (e.g. enrollment in college) does not have a natural pre-test) – studies must demonstrate baseline equivalence on the following:

Education Outcomes

- (1) A measure within the education domain category. Baseline equivalence must be established on a minimum of **1 continuous education measure OR 2 dichotomous education measures**.¹⁵ Education domain measures include credit accumulation and course completion; postsecondary degree attainment; industry-recognized credential, certificate, or license completion; and technical skill proficiency (see page 9 for more information). Note that education measures of high school performance, including graduation, GPA, and college placement tests, generally qualify as education baseline measures, even though they are not acceptable outcomes.¹⁶

AND

- (2) A measure of one or more of the sociodemographic characteristics listed on page 4. Examples include students who are limited English-proficient speakers, single parents, students with disabilities, and older students (i.e., age).¹⁷

¹⁴ The WWC can perform its own difference-in-differences adjustment to allow the study to satisfy the statistical adjustment requirement when a baseline characteristic is the same as the outcome. A basic requirement is that the baseline characteristic must have a correlation of 0.6 or higher with the outcome.

¹⁵ The requirement for 2 dichotomous measures is motivated by the fact that these types of variables might be less sensitive to treatment and comparison group differences as compared to continuous measures.

¹⁶ Review Team Leadership should be consulted if there is use of high school data that are measured more than two years prior to the start of the intervention, which could be the case with older dislocated workers.

¹⁷ Review Team Leadership will have the discretion to determine whether sociodemographic characteristics provide a clear test of baseline equivalence. For example, if an intervention is targeted to single parents, Review Team Leadership will determine whether this sampling consideration fulfills the baseline equivalence requirement for

Labor Market Outcomes

- (1) A measure within the labor market domain category *OR* a measure within the education domain category. Baseline equivalence must be established on a minimum of **1 continuous measure OR 2 dichotomous measures**.

AND

- (2) A measure of one or more of the sociodemographic characteristics listed on page 4.

When categorical baseline measures are reported, the WWC reviewers will check for balance on each category that meets the threshold for balance described above. For example, a categorical measure of educational attainment with four categories – (1) less than 12th grade, (2) high school graduate/GED®, (3), some college, and (4) college degree – would be treated as 4 dichotomous variables. Each of the four categories could be treated as a separate dichotomous baseline measure.

2. Baseline equivalence of clusters

RCTs with high or unknown cluster-level attrition and quasi-experimental designs that entailed cluster assignment must satisfy the baseline equivalence requirement. If this requirement is not met, studies of this sort are not eligible for the rating: *Meets WWC Group Design Standards with Reservations*.

Assessing equivalence of clusters: characteristics to consider

There are no new cluster-level variables (e.g., selection criteria used by colleges, staff characteristics in a community agency) that must be considered in equating.¹⁸ Rather, assessing baseline equivalence of clusters across the intervention and comparison groups uses the same characteristics and domains described above for studies that use individual-level assignment. Recall that although high school data are not eligible to be used as outcomes, in some circumstances such data could be used for baseline equating. This remains true for some cluster-level assignment studies, as described below.

Additional CTE cluster equivalence parameters

Baseline equivalence for an outcome contrast can be established using data from an earlier assessment of individuals in the analytic sample. This is sometimes referred to as a “same cohort equating strategy.” However, use of any data for baseline equating purposes that were generated

sociodemographic characteristics or whether there is sufficient variability on other characteristics that require further demonstration of baseline equivalence.

¹⁸ Note that inequivalence of other cluster-level variables can render a *does not meet standards* rating.

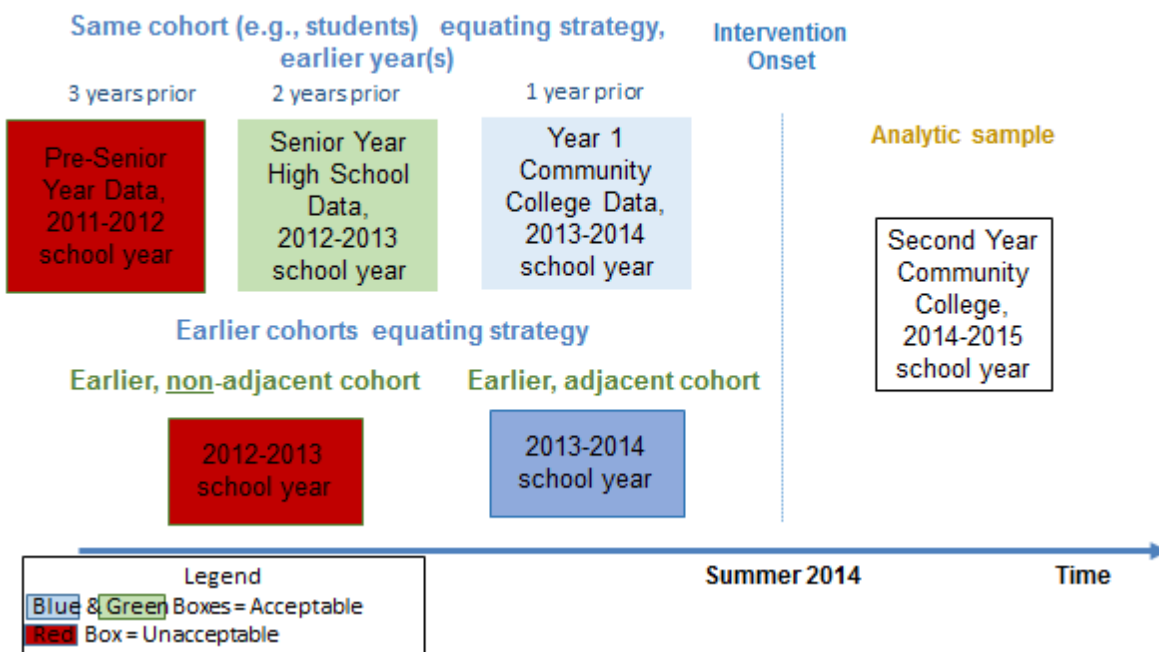
more than two years prior to intervention onset are ineligible. Recognizing that there are wide variation of clusters and samples covered by CTE (e.g., 18-year-old community college students to older displaced workers) exceptions to this general rule can be considered by the Review Team Leadership. Anytime an exception is enacted this will be documented in an intervention report.

It is possible that equating at the cluster-level could be accomplished using earlier data from the analytic sample that is gathered from outside of the clusters used in the study. For example, a study might establish baseline equivalence for an analytic sample in a community college cluster study by relying on the analytic sample's senior year high school data. Again, the default for this CTE review is that any equating data that are more than two years removed from intervention onset will yield a contrast that will not meet WWC standards. Note that high school data collected prior to the senior year to establish equating for a contrast will render a does not meet standards rating, even if the two-year criterion is met. See the top red box in Figure 1 (below).

Rather than relying on prior data on the analytic sample, the cluster baseline equivalence requirement could also be met using data from an earlier cohort from within the same clusters. This is sometimes called the "earlier cohorts equating strategy." This might occur when dealing with a college or some other organization that routinely processes cohorts (e.g., a community agency that provides annual CTE courses). Consider a study that entailed community college assignment and analyzed outcomes of second-year students at the end of the 2014-2015 school year. In this case, baseline equivalence of clusters can be assessed using the second-year students present in the colleges at the end of the 2013-14 school year. That is, an earlier, adjacent cohort can be used to establish baseline equivalence of clusters (see the bottom blue box in Figure 1). Use of an earlier *non*-adjacent cohort will, however, yield a contrast that does not meet evidence standards (see the bottom red box referencing the 2012-2013 school year). Note that when an earlier cohort strategy is used, the maximum elapsed time that is allowed between the generation of baseline data and intervention onset is one academic year.¹⁹

¹⁹ Consider years as "academic" timeframes and not strict calendar years. If for example outcome data are collected in the spring of 2015, and baseline data were generated in the fall of 2014, this spans more than one year but meets the adjacent cohort principle. In other words, this would be an acceptable approach to baseline equating.

Figure 1. Acceptable and Unacceptable Samples for Assessing Equivalence of Clusters



Outcome Measure Requirements

Overallignment of Outcome Measures. A study’s rating will be based only on those measures that are not overaligned. Overallignment occurs when outcome measures are more closely aligned to one of the research groups (intervention or comparison) than to the other group and so could bias a study’s results. For example, completion of a state certification for Licensed Practical Nurse or Vocational Nurse (LPN/LVN) could be interpreted as being overaligned in a study that gave students preferred access to LPN or LVN training. In this example, the intervention group might have higher completion rates as a function of the design.

Reliability of Outcome Measures. Measures of the outcome of interest should demonstrate adequate face validity. Acceptable outcome measures include administrative records of enrollment, term-to-term persistence, course taking, course completion, credit accumulation, and credential attainment. Administrative Unemployment Insurance records are an acceptable source of data on employment and earnings. Participant (e.g., employer) surveys are also an acceptable source of data for all outcome domains.²⁰ The Review Team Leadership will have discretion in

²⁰ Should outcomes be reported for both quarterly and annual Unemployment Insurance records, the former will be prioritized as it tends to yield more sensitive measurement, and findings based on annual records will be included in Appendix D.

determining the acceptability of outcome measures when their psychometric properties are not reported, as well as whether there appears to be a poor response rate to participant survey items.²¹

Measures of the *technical skill proficiency* domain, including third-party technical assessments, must demonstrate acceptable reliability. A measure is considered reliable if one the following minimum thresholds is met: (a) internal consistency (such as Cronbach’s alpha) of 0.50 or higher; (b) temporal stability/test-retest reliability of 0.40 or higher; or (c) inter-rater reliability (such as percentage agreement, correlation, or kappa) of 0.50 or higher.

Reporting Findings in Intervention Reports

This review follows the guidance in the *WWC Procedures Handbook* (in Chapter IV: Reporting on Findings) regarding reporting on findings from multiple analyses that use composite or subscale scores, or findings from subgroups.

For Outcomes in the Education Category. When a given education outcome is measured at different time points, the review prioritizes the measure with the shortest follow-up period. That is, the measure with the shortest follow-up period will be used to determine the WWC effectiveness rating and, therefore, will be reported as main findings in Appendix C of the intervention report. Later follow-up outcome measures will generally be reported in Appendix D as supplemental findings.

Findings can also be measured after different amounts of exposure to a CTE intervention’s implementation (for example, after 1, 2, or 3 years of intervention implementation). As a rule, the outcome measure that reflects the maximum exposure to the intervention will be used to determine the WWC effectiveness rating and will be reported as the main finding in Appendix C of the intervention report. The intermediate outcome measures will generally be reported as supplemental findings in Appendix D. For example, if a study evaluated the effectiveness of an intervention on a group of students after both one and two years, the outcome measure reflecting maximum exposure (i.e., after two years) will be considered as primary, while the intermediate outcome (i.e., after one year) will be considered as a supplemental finding.

While the above rules will guide how main and supplemental findings are identified, review team leadership has discretion to identify main and supplemental findings after considering additional factors about the findings under review, such as the prevalence of findings across implementation levels and the design of the intervention.

²¹ Comparisons of self-reported and administrative earnings records show that though measurement error indeed exists and that it is non-random (i.e., earnings are correlated with job tenure), self-reported earnings are in practice still an accurate representation of true earnings (Bound et al., 1994; Bound & Krueger, 1991).

For Outcomes in the Labor Market Category. As noted above, labor market outcomes (employment and earnings) are eligible for review if they measure short-, medium-, or long-term impacts, as defined above. When a given labor market outcome is measured at different time points within the same outcome domain, the review prioritizes the measure with the shortest follow-up period. For example, if two differed measures of long-term employment are reported – for example, an indicator for whether the subject was employed 5 years after enrollment in the study and an indicator for whether the subject was employed 7 years after enrollment in the study- the review would prioritize the measure with the shortest follow-up period. All other findings will generally be reported in Appendix D as supplemental findings.

Statistical Adjustments. The *WWC Procedures Handbook* discusses the types of adjustments made by the WWC in Section VI: Reporting on Findings. For “mismatched” analysis (i.e., when a study assigns units at the cluster-level but conducts analysis at the individual level), this topic area uses the intra-class correlation coefficient of 0.20 for all eligible education outcomes and uses the intra-class correlation coefficient of 0.10 for all eligible labor market outcomes unless a study-reported intra-class correlation coefficient is available.

Eligible Study Designs. Studies that use group designs (RCTs, QEDs, and regression discontinuity designs) or single-case designs are eligible for review using the appropriate WWC design standards.

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

PROCEDURES FOR CONDUCTING THE LITERATURE SEARCH

The *WWC Procedures Handbook, version 4.0*, discusses the procedures for conducting a literature search in Section III: Identifying Relevant Literature and Appendix B: Policies for Searching Studies for Review. This review will use a quick literature search process to identify research on a limited number of interventions that may be of most interest to decision makers, rather than using a broad keyword search on the full topic area to identify interventions. In the first step of this process, content experts identify and recommend interventions with a large body of causal evidence likely to be of interest to decision makers. This review will identify additional interventions that may be the focus of WWC-reviewed studies that are not already the subject of up-to-date WWC intervention reports.

After identifying these interventions, the second step of the process is to conduct intervention-specific literature searches, using the intervention name, to identify all publications on each intervention. This review may refine the potential scope of this search by including additional search terms.

In a third step, each citation gathered through this search process undergoes a screening process to determine whether the study meets the eligibility criteria established in the review protocol. This screening process is described in Chapter IV of the *WWC Procedures Handbook*. Finally, the interventions are prioritized for review based on the quantity and quality of eligible studies of the intervention. This prioritization process is described in Appendix A of the *WWC Procedures Handbook*.

Search Parameters for the I-BEST Intervention Report

The parameters below were used to conduct the literature search for the I-BEST Intervention Report. Searches conducted for other products in this topic area should have similar parameters, with the exception of the *Eligible Interventions and Components* terms.

Timeframe: Search for research studies for the past 20 years (1999–Present).

Bibliographic Databases (each database was searched separately)

- Academic Search Complete
- ProQuest Dissertations and Theses
- EconLit
- Education Research Complete
- ERIC (EBSCO version)
- JSTOR Journals
- PsycInfo
- MEDLINE
- Web of Science

Search Terms

The literature search will include four blocks of terms. Note that the blocks below contain our preliminary set of search terms. Our approach to the search will entail iterative searches, and the refinement of search terms as needed to target our search results to the extent possible.

Block A: Methods Terms

“control group” OR “control groups” OR “control condition” OR “control conditions” OR random OR randomly OR randomized OR randomization OR “comparison group” OR “comparison groups” OR “comparison condition” OR “comparison conditions” OR "regression discontinuity" OR “matched group” OR “matched groups” OR baseline OR treatment OR treatments OR experiment OR experiments OR experimental OR experimentally OR trial OR intervention OR interventions OR “intervention condition” OR “intervention conditions” OR empirical OR evaluation OR evaluations OR “research study” OR impact OR impacts OR “effect” OR effectiveness OR causal OR causally OR causality OR posttest OR “post-test” OR posttests OR “post-tests” OR “follow-up” or “follow up” OR pretest OR “pre-test” OR pretests OR “pre-tests” OR QED OR QEDs OR QES OR RCT OR RCTs OR “propensity score” OR “propensity scores” OR “quasi-experiment” OR “quasi-experimental” OR “quasi-experiments” OR “mixed method”

AND

Block B: Population or Setting Terms

"higher education" OR postsecondary OR "post-secondary" OR undergraduate OR undergraduates OR “student” OR “students” OR "tertiary education" OR “community college” OR “technical college” OR “technical school” OR “trade school” OR “career education” OR “technical education” OR “Workforce Investment Boards” OR “WIB” OR “dual enrollment” OR “adult education” OR “adult basic education” OR “nontraditional students”

AND

Block C: Eligible Interventions and Components

“Integrated Basic Education and Skills Training” OR “I-BEST” OR “Integrated education and training” OR “IET”

AND

Block D: Eligible Outcomes (with terms connected by OR)

- Assessment*
- Associate*
- Certificate
- Certification
- College-level
- Competen*
- Completion
- Course
- Credential
- Credit-bearing
- *Credit*
- Non-credit
- Degree
- Diploma
- Earnings
- Employ*
- Exam
- Hours
- Income
- Industry-recognized
- Job
- Labor market
- License
- Modular*
- National Occupational Competency Testing Institute
- NOCTI
- Paid
- Persist*
- Proficien*
- Salary
- Stackable
- Stacked
- Technical skill
- Test
- Wage

Additional Sources

Literature reviews for this topic area involve searching the electronic databases listed above as well as the following websites:

- Abt Associates
- American Institutes for Research (AIR)
- American Technical Education Association (ATEA)
- Annie E. Casey Foundation
- Aspen Institute
- California Center for Regional Leadership (CCRL)
- Center for Law and Social Policy (CLASP)
- Council for Adult and Experiential Learning (CAEL)
- ECMC Foundation
- Jobs for the Future (JFF)
- Gates Foundation
- Indiana Next Generation Manufacturing Competitiveness Center (N-MaC; Purdue University)
- John. J. Heldrich Center for Workforce Development (Rutgers University)
- Joyce Foundation
- JP Morgan
- Kresge Foundation
- Laura and John Arnold Foundation
- Mathematica Policy Research
- MDRC
- National Association of State Workforce Agencies (NASWA)
- National Association of Workforce Boards (NAWB)
- Oklahoma State University (OSU) Institute of Technology
- Social Policy Research Associates
- University of Illinois School of Labor and Employment Relations
- University of Wisconsin Center on for Education Research (WCER)
- Urban Institute
- U.S. Chamber of Commerce Foundation (USCCF)
- U.S. Department of Labor, Trade Adjustment Assistance Community College and Career Training Grant Program (TAACCCT)
- U.S. Department of Labor, Workforce Innovation Fund (WIF)
- Virginia Tech Office of Economic Development

References

- Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics* (Vol. 4, pp. 1043-1171). Elsevier.
- Belfield, C., and Bailey, T. (2017). The Labor Market Returns to Sub-Baccalaureate College: A Review. New York, NY: Center for the Analysis of Postsecondary Education and Employment. Accessible here: <https://capseecenter.org/labor-market-returns-sub-baccalaureate-college-review/>
- Bound, J., Brown, C., Duncan, G. J., & Rodgers, W. L. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12(3), 345-368.
- Bound, J., & Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of Labor Economics*, 9(1), 1-24.
- Carnevale, A. P., Cheah, B., Ridley, N., & Strohl, J. (2017). *Good jobs that pay without a BA*. Washington, DC: Georgetown University, Center on Education and the Workforce.
- Carnevale, A. P., Jayasundera, T., and Gulish, A. (2016). *America's Divided Recovery: College Haves and Have-Nots*. Washington, DC: Georgetown University, Center on Education and the Workforce.
- Carnevale, A. P., Jayasundera, T., & Hanson, A. R. (2012). *Career and Technical Education: Five Ways That Pay Along the Way to the B.A.* Washington, DC: Georgetown University Center on Education and the Workforce.
- Carnevale, A. P., Smith, N. & Strohl, J. (2013). *Recovery: Projections of Jobs and Education Requirements Through 2020*. Washington, DC: Georgetown University Center on Education and the Workforce.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47, 153-161.
- Holzer, H. J. (2015). Sector-based training strategies: The challenges of matching workers and their skills to well-paying jobs. *Future of Work Paper Series*, 48.
- Rosen, R., Visher, M., & Beal, K. (2018, September). *Career and Technical Education Current Policy, Prominent Programs, and Evidence*. Report. Retrieved from: <https://www.mdrc.org/publication/career-and-technical-education>.
- U.S. Department of Labor, Employment and Training Administration [DOL/ETA]. (2016). Performance accountability guidance for Workforce Innovation and Opportunity Act (WIOA) Title I, Title II, Title III, and Title IV core programs. Retrieved from https://wdr.doleta.gov/directives/attach/TEGL/TEGL_10-16.pdf