Improving Post-High School Outcomes for Transition-Age Students with Disabilities: An Evidence Review

R. Brian Cobb
Colorado State University

Stephen Lipscomb
Mathematica Policy Research

Jennifer Wolgemuth
University of South Florida

Theresa Schulte
Mathematica Policy Research

with

Abigail Veliquette
Colorado State University

Helen Holmquist-Johnson
Colorado State University

Morgen Alwell
University of Montana

Rebecca Orsi
Colorado State University

Keri Batchelder
Colorado State University

Laura Sample McMeeking
Colorado State University

Robert Bernard
Concordia University

Jun Wang
Colorado State University

Paul Hernandez
Colorado State University

Andrea Weinberg
Colorado State University

Project Officers

Yumiko Sekino

Marsha Silverberg

Institute of Education Sciences

U.S. Department of Education
August 2013

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

NLTS 2012
National Longitudinal Transition Study

Improving Post-High School Outcomes for Transition-Age Students with Disabilities: An Evidence Review

R. Brian Cobb
Colorado State University

Stephen Lipscomb
Mathematica Policy Research

Jennifer Wolgemuth
University of South Florida

Theresa Schulte
Mathematica Policy Research

with

Abigail Veliquette
Colorado State University

Morgen Alwell
University of Montana

Keri Batchelder
Colorado State University

Robert Bernard
Concordia University

Paul Hernandez
Colorado State University

Helen Holmquist-Johnson
Colorado State University

Rebecca Orsi
Colorado State University

Laura Sample McMeeking
Colorado State University

Jun Wang
Colorado State University

Andrea Weinberg
Colorado State University

Project Officers

Yumiko Sekino

Marsha Silverberg

Institute of Education Sciences

August 2013

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Ruth Curran Neild
Commissioner

#NCEE 2013–4011

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

August 2013

The report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-10-C-0073. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Cobb, R. B., Lipscomb, S., Wolgemuth, J., Schulte, T., Veliquette, A., Alwell, M., Batchelder, K., Bernard, R., Hernandez, P., Holmquist-Johnson, H., Orsi, R., Sample McMeeking, L., Wang, J., and Weinberg, A. (2013). *Improving Post-High School Outcomes for Transition-Age Students with Disabilities: An Evidence Review* (NCEE 2013-4011). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S.

This report is available on the National Center for Education Evaluation and Regional Assistance website at http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NCEE20134011.

Acknowledgments

This report is the product of collective efforts from staff at Colorado State University, Mathematica Policy Research, University of South Florida, University of Montana, and Concordia University, with support from the National Library of Education. We are grateful to John Burghardt and Bonnie O'Day for their suggestions and critical reviews of earlier versions of the report.

The opinions expressed herein are those of the authors and do not necessarily represent official positions or policies of the Institute of Education Sciences, the U.S. Department of Education, or the affiliations of the authors.

Disclosure of potential conflicts of interest

The research team for this review consists of key staff from Colorado State University, Mathematica Policy Research, University of South Florida, University of Montana, and Concordia University. The organizations and the key staff members do not have financial interests that could be affected by findings from the study.

Executive Summary

Nearly four decades have passed since the Individuals with Disabilities Education Act (IDEA) ensured access to public education for students with disabilities in the United States. During the years following its adoption, there was growing recognition that to lead productive and fulfilling lives as adults, many students need support in the transition from secondary school to post-high school environments. As a result, several reauthorizations of IDEA have emphasized transition planning in helping students with disabilities to obtain employment, pursue postsecondary education and training, and live more independently.

Despite the efforts of policymakers and practitioners, a gap remains between post-high school outcomes of students with disabilities and outcomes for other students. To help close that gap, this report reviews the research literature on programs (strategies, interventions, or sets of services) designed to help students with disabilities make transitions.[1] It deviates in the following ways from previous evidence reviews on this topic (for example, Cobb and Alwell 2009; Test et al. 2009):

- It updates earlier reviews by including studies publicly released between April 2008 and June 2011.
- It reviews studies using the standards and process developed by the Institute of Education Sciences (IES) What Works Clearinghouse (WWC).[2] Applying the WWC standards and procedures meant that we did not include some types of studies that were included in previous reviews.
- It focuses on direct measures of students' post-high school outcomes as evidence of a program's effectiveness.[3]

These criteria focus the evidence review on research results in which we are most confident, and that can best help us identify programs that are likely to improve the post-high school outcomes of students.

Review methods and results

The research team used the WWC's systematic procedures to guide this review (see Figure 1). Briefly, we first defined which types of studies the evidence review would include, then conducted an initial search for all studies that might be relevant to our topic and

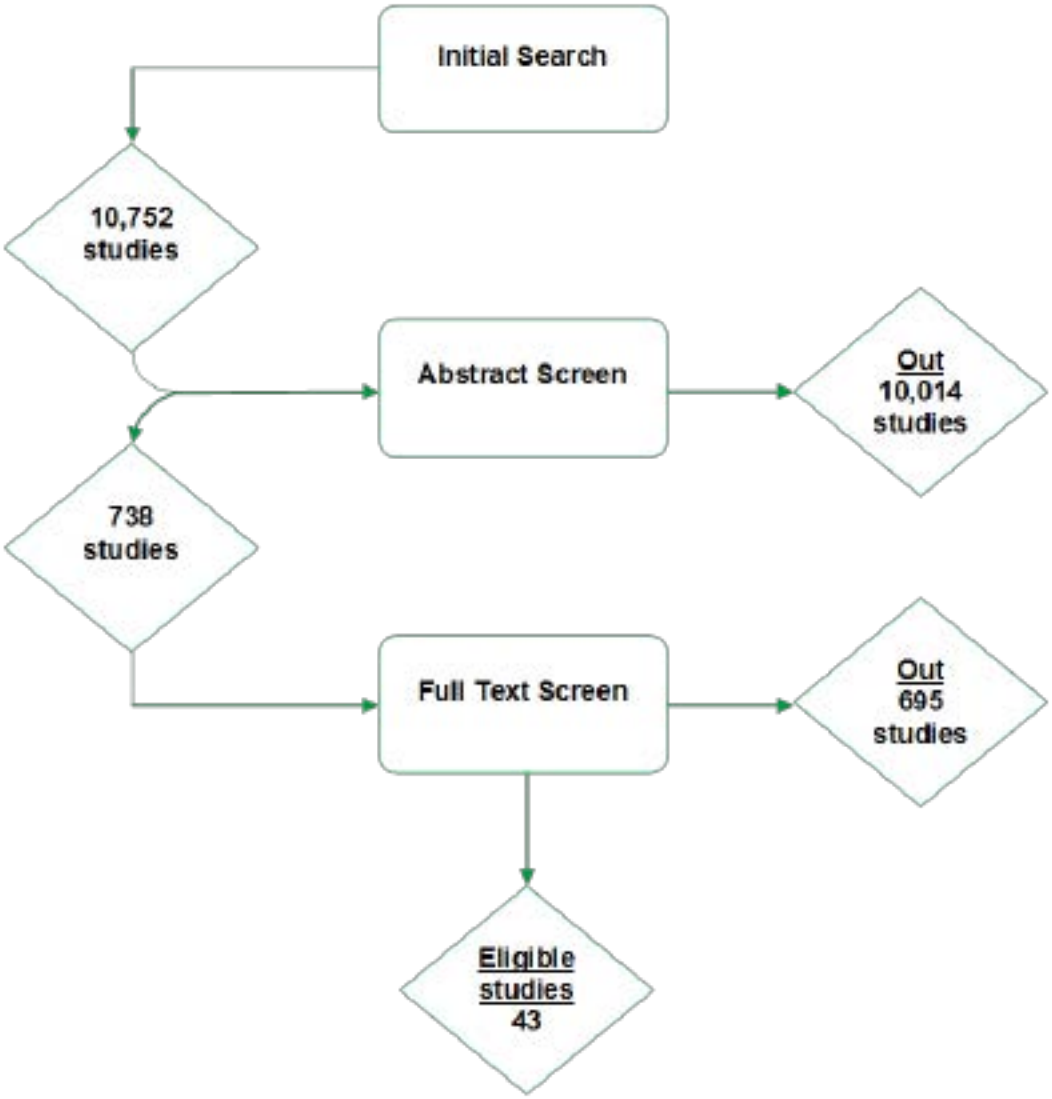
[1] The report is being prepared under ED's contract with Mathematica Policy Research to conduct the National Longitudinal Transition Study (NLTS) 2012. NLTS 2012 is the third in a series of data collections to track the characteristics, school experiences, and post-high school outcomes of a cohort of students with disabilities. See http://ies.ed.gov/ncee/nlts/ for more information.

[2] The research team needed to go beyond the WWC standards and procedures for rating the effectiveness of programs when the evidence included studies that use a type of research design called a single-case design. The WWC has pilot standards for reviewing single-case design studies but has not established a method for combining evidence across studies.

[3] Direct measures refer to explicit indicators of employment, postsecondary education and training, and independent living. This review does not include studies using indirect measures of these three outcome domains.

outcomes. Second, using short abstracts for each study, we screened out studies that were clearly ineligible. For the remaining studies, we read the full text of each article or report to determine whether it was eligible for the review. Only eligible studies that met the WWC evidence standards were to be included in the lessons we drew about promising programs.

Figure 1. Summary of the search and screening process



The 43 eligible studies were reviewed and assigned a WWC standards rating. Of this group, 16 studies met the WWC standards, but none at the highest level given to well-implemented studies that use the most rigorous research designs ("without reservations"):

- Of the 16 studies meeting the WWC standards, 13 were single-case design studies in which researchers compare how outcomes change for individual students in response to a program. A total of 42 students, all but 2 of whom had an intellectual disability[4], participated in these studies, which focused on examining independent living outcomes.
- Three studies used a quasi-experimental design, in which researchers compare outcomes for a group of students participating in a program with the outcomes of a comparison group that was selected to be as similar as possible to the participant group. Approximately 700 students participated in those three studies, which all measured employment outcomes; one of the studies (215 students) also examined postsecondary education outcomes.

Our review of transition research studies from the past two decades indicates that relatively few studies meet the WWC standards for credible evidence of effectiveness. Based on the 16 studies that met WWC standards, the review rated the effectiveness of programs designed to help students with disabilities make transitions to post-high school employment, postsecondary education, and independent living. Community-based work-experience programs were found to have mixed effects on employment outcomes based on a medium-to-large extent of evidence. These programs were found to have potentially positive effects on postsecondary education outcomes, although the extent of evidence was small. Functional life-skills development programs were found to have potentially positive effects on independent living outcomes. But, again, the extent of evidence was small. On the whole, evidence across eight program categories allowed us to review only two of the categories, providing little support from high-quality intervention research for identifying a range of programs to help students with disabilities make successful transitions to employment, postsecondary education and training, or independent living.

Hypotheses for future program development and research

The objective of this review was to draw some tentative lessons about how practitioners might develop successful programs and researchers might improve their investigations of transitions for students with disabilities. But the lack of studies that met the WWC's standards made achieving those objectives challenging. The research team decided to add to the 16 studies that met the WWC standards a set of 8 empirical studies that fell somewhat short of meeting the standards but that we felt could provide exploratory insights.

After examining the methods and results of these studies combined, we offer five hypotheses relevant to program implementation:

[4] Intellectual disability under IDEA [P.L. 108-448 and P.L. 111-256] refers to significantly sub-average general intellectual functioning, existing concurrently with deficits in adaptive behavior and manifested during the developmental period, that adversely affects a child's educational performance.

1. There are many more programs and bundles of strategies targeting different types of students, as well as greater experience with these programs, than when the earlier literature reviews were conducted.
2. The links between transition program offerings or components may be very important. Different findings for similar programs offered in different ways suggest that, for example, work-experience activities may need to be integrated with career and technical education classes or other aspects of students' educational programs to achieve desired benefits in post-high school outcomes.
3. Participation in career and technical education and or getting a job while in high school may be related to better employment outcomes for students with disabilities.
4. Inclusive education settings in high school may ease the path to postsecondary education.
5. Several strategies, such as computer-based instruction and prompting, may by increasing the functional skills of students with intellectual disabilities help them live more independently.

We suggest some caution in acting on these hypotheses, however, given the design of the studies we drew upon. It is likely that students in the participant groups of the studies were more interested and able to participate in program activities, and therefore more motivated than students in the comparison groups to become employed or enter college. This factor makes it difficult to determine whether the results of the studies reflect the effects of the programs, differences between the groups that existed even before they received the services studied, or some combination of both.

Because of these concerns about study design as well as the lack of studies that met WWC standards, we also offer five recommendations for researchers:

1. Pay greater attention to certain design elements, so that studies can improve their rigor (and therefore meet WWC standards). In particular, we strongly encourage researchers to randomly assign eligible participants into a program and comparison group and to include preprogram measures of all outcomes in the analysis.
2. Collect data on and control for students' "employment while in school" if the study focuses on post-high school employment outcomes. Correlational evidence suggests a positive relationship between having a job while in school and having one after leaving school. Studies that fail to account for in-school employment risk confusing the effects of that experience with the effects of other programs and services.
3. Control for type and severity of disability in the research analysis. As you might expect, these student attributes were the most consistently important in differentiating post-high school results. Although many of the studies we looked at described students using these attributes, the analyses rarely included these variables. Without ensuring that the program and comparison groups are equivalent on these attributes before a program is provided, it is difficult to know

whether differences after the program are due to the program or those pre-existing differences in attributes.

4. Measure post-high school outcomes directly. Many studies did not measure such outcomes. Moreover, the field has not convincingly shown a connection between interim or indirect measures deemed important (for example, self-determination and locus of control) for in-school programs and these later outcomes. We encourage the special education research community to extend its outcome measurement to include actual measures of employment, postsecondary education, or independent living. This approach will authenticate the effects of special school curricula and instruction. Fulfilling this goal is increasingly possible; post-high school outcomes are becoming easier and less costly to obtain. Researchers who convince state and local education officials of the importance of their analyses can access state longitudinal data systems supported by ED grants and perhaps tap other administrative databases (for example, the National Student Clearinghouse for postsecondary enrollment information). IES' National Center for Special Education Research runs various competitions to obtain grant funds for rigorous research including in the area of post-high school transitions.
5. Researchers studying follow-up effects of transition programs should include any information they can access on "treatment integrity" or "treatment fidelity." Such information adds precision to our understanding of why variations in outcome effects may occur across programs that appear to be similar but differ substantially in features or implementation processes.

Fortunately, several important evaluations on the horizon may provide useful information for policymakers and researchers who have an interest in this transition area. These studies of demonstration programs funded by the U.S. Social Security Administration were too early into data collection and analysis to provide information on the effects of the programs on post-high school outcomes. However, we mention them here for several reasons: (a) their designs provide examples of how to do the difficult work of conducting random assignment of students with disabilities and tracking their outcomes in secondary school and beyond; (b) early reports from these projects have also focused on issues of implementation fidelity; and (c) follow-up reports from these large-scale evaluations hold great promise for ascertaining the effects of the program implemented as well as the differential effects of varying components on students with different types of disabilities and in a number of implementation contexts. These types of next-generation research studies in special education transition may inform local policymakers about sequencing and intensity of program components and how these components vary in their effects on different transition outcome domains. These studies, coupled with the existing empirical research base, may also provide transition researchers and program designers with the tools to put together logic models that can clarify the strength and directions of program components and other inputs to different transition outcomes.

Contents

Figures

Tables

Chapter 1. Introduction

Nearly four decades have passed since the Individuals with Disabilities Education Act (IDEA) ensured access to public education for students with disabilities in the United States. During the years following its introduction, there was a growing recognition that helping students, especially students with disabilities, in their transitions from secondary school to post-high school environments is crucial to them leading productive and fulfilling lives as adults. Several reauthorizations of IDEA have emphasized the importance of transition planning in providing support for students with disabilities to obtain employment, pursue postsecondary education and training, and live more independently.

Despite the efforts of policymakers and practitioners, a gap in post-high school outcomes remains between students with disabilities and other students. Newman et al. (2011) reported that 60 percent of students with disabilities had ever taken a class from a postsecondary school within eight years of leaving high school[5], lower than the overall rate of 67 percent for same-age young adults in the general population. Students with disabilities were also less likely to be living independently as adults (45 percent versus 59 percent), be married (13 percent versus 19 percent), have a checking account (59 percent versus 74 percent), or have a credit card (41 percent versus 61 percent). Students with disabilities who had been out of high school for as many as eight years had lower rates of employment than same-age individuals in the general population, although employment differences were statistically significant only for students with more severe types of disabilities.[6]

To identify effective strategies for improving post-high school outcomes, this report reviews the research literature on programs (strategies, interventions, or sets of services) designed to help students with disabilities make transitions.[7] It deviates in several ways from previous evidence reviews (for example, Cobb and Alwell 2009; Test et al. 2009):

- It updates earlier reviews by including studies publicly released between April 2008 and June 2011.

[5] In the Newman et al. (2011) study, students who left high school included those who may not have graduated from high school but were no longer enrolled in school. The amount of time out of high school varied for the young adults with disabilities included in the Newman et al. (2011) report, ranging from one month or less to eight years post-high school.

[6] For instance, 39 percent of students with intellectual disabilities were employed at the eight-year follow-up. In contrast, 67 percent of students with learning disabilities were employed at the eight-year follow-up (see Table 19 of Newman et al. 2011).

[7] The report is being prepared under ED's contract with Mathematica Policy Research to conduct the National Longitudinal Transition Study (NLTS) 2012. NLTS 2012 is the third in a series of data collections to track the characteristics, school experiences, and post-high school outcomes of a cohort of students with disabilities. See http://ies.ed.gov/ncee/nlts/ for more information.

- It reviews studies using the standards and process developed by the Institute of Education Sciences (IES) What Works Clearinghouse (WWC).[8] Applying the WWC standards and procedures in this review meant that we did not include some types of studies that were included in previous reviews.
- It focuses on direct measures of students' post-high school outcomes as evidence of the program's effectiveness. By direct measures, we mean explicit indicators of employment, postsecondary education and training, and independent living.[9] This review does not include studies using indirect measures of these three outcome domains, such as vocational self-awareness or job satisfaction as a proxy for employment; ACT scores or high school GPA as a proxy for postsecondary education; or self-determination or social skills as a proxy for independent living. We recognize that limiting our review to studies that used direct outcome measures would dramatically reduce the number of studies we could include in the review. However, our intentions were to conduct a systematic review of the literature, guided by the WWC review procedures. As there is currently no WWC guidance on how to include in the review studies using indirect measures, we did not include them and focused the review on studies using direct measures.

These criteria focus the evidence review on research results in which we are most confident, that can best help us identify programs that are likely to improve the post-high school outcomes of students.

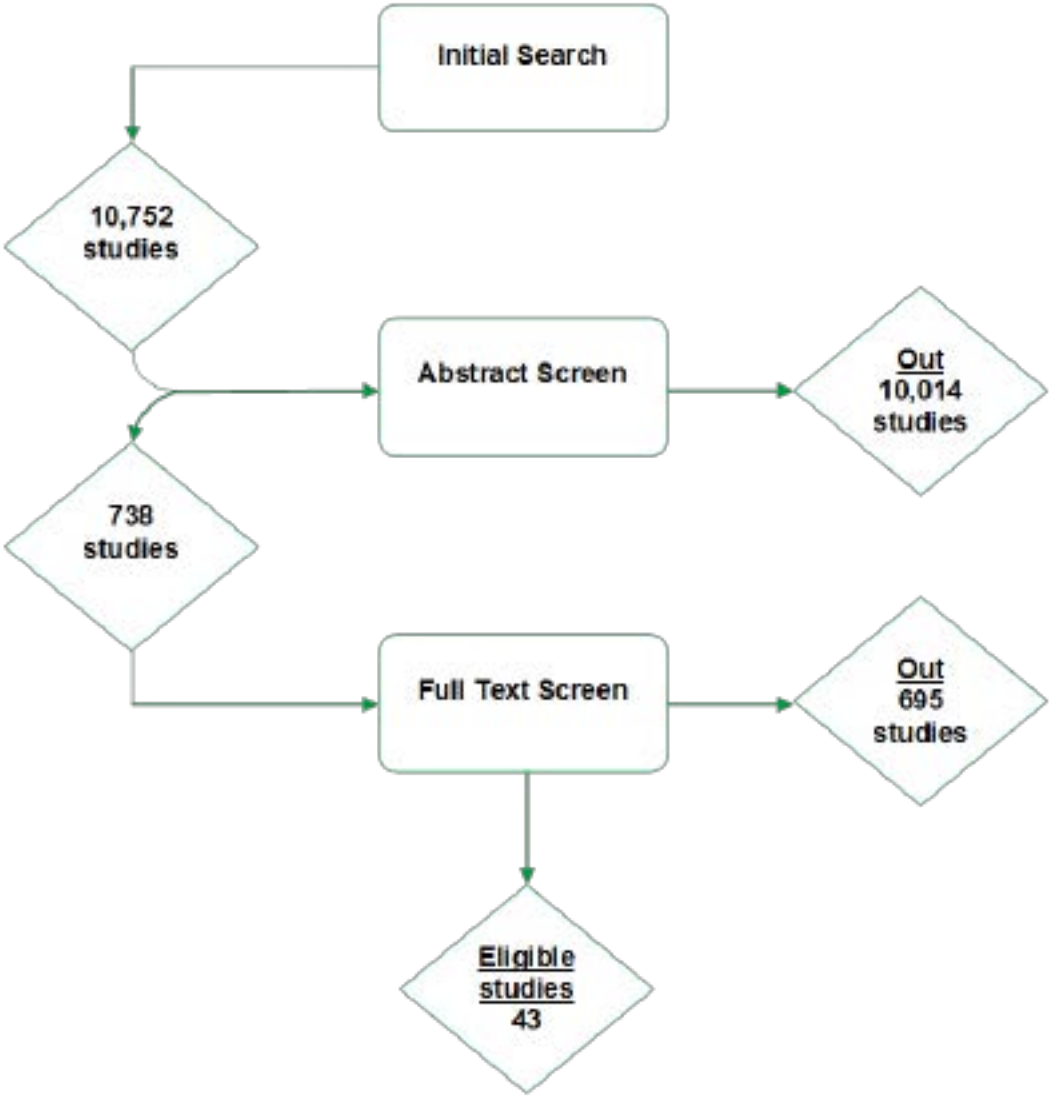
[8] Studies were reviewed using the *What Works Clearinghouse Procedures and Standards Handbook, Version 2.1* (U.S. Department of Education, 2011). The handbook can be accessed at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19.

[9] The independent living outcomes measure skills that are directly related to independent living post high school, although the timing of measurement is not necessarily after students have left school.

Chapter 2. Review and reporting methods

The research team used the WWC's systematic procedures to guide this review (see Figure 1). Briefly, we first defined which types of studies would and would not be included in the evidence review, then conducted an initial search for all studies that might be relevant to our topic and outcomes. Second, using short abstracts for each study, we screened out those that were clearly ineligible for our review (for example, citations that were not original research). Finally, we screened the remaining studies by reading the full text of each article or report to determine whether it was eligible for the review. Only eligible studies that met the WWC evidence standards were to be included in the lessons we drew about promising programs. The sections below provide more information on the search and screening process.

Figure 1. Summary of the search and screening process



Evidence review protocol

The research team developed a protocol to guide our review work that established rules for including studies in the review and for how the research literature would be searched (see Appendix A). To be considered for the review, studies needed to include primarily students ages 13 to 21 who are eligible for special education services under IDEA. Studies must have been publicly released or published within the last 20 years; must have examined a program whose primary goal is improving skills targeting post-high school employment, postsecondary education, or independent living; and must have reported one or more postsecondary education, employment, or independent living student outcomes.[10]

Employment outcomes could include involvement in competitive employment without support, supported employment (typically, employment in competitive work settings with support by professionals), self-employment, or features associated with these placements, such as earnings, hourly wages, levels of benefits, or stability of employment.

Postsecondary education outcomes could include participation or matriculation in discrete classes or degree programs in community colleges, tribal colleges, or four-year colleges or universities; or participation or enrollment in adult education classes, career and technical training programs, or private certification postsecondary institutions.

Independent living outcomes could include self-care or daily living skills (for example, cooking, grooming, housekeeping), community participation outcomes (for example, voting, shopping, using community services), or other types of outcomes associated with independent living, such as home ownership or home maintenance.

Eligible research designs

The WWC standards, as well as most expert research methodologists, consider only certain study designs as capable of generating credible evidence of effectiveness of a program. For our review, studies evaluating a program's effectiveness through one of the following broad categories of empirical research approaches were eligible:

- A randomized control trial (RCT)
- Certain well-documented quasi-experimental designs (QED)

[10] These outcome domains are cited most frequently as the ultimate source for determining whether an intervention is successful, regardless of the interim measures, for transition from school to post-high school environments for students with disabilities. For example, see Halpern (1994) for guiding conceptual work that became the adopted position of the Division on Career Development and Transition of the Council for Exceptional Children. These outcomes domains also link directly to the seven transition outcomes described by IDEA: (1) postsecondary education, (2) vocational education, (3) integrated employment including supported employment, (4) continuing and adult education, (5) adult services, (6) independent living, and (7) community participation. Adult services, which are more of a process outcome than an impact outcome, align with all three domains.

- A single-case design (SCD)[11]

This report refers to RCTs and QEDs collectively as "group design studies," because they involve comparing outcomes for groups of students. In contrast, SCDs can compare how outcomes change for individual students in response to a program. Table 1 provides brief definitions of these research designs.

Table 1. Eligible research designs

Type of research design	General description
Group designs	**Randomized control trials** are called the "gold standard." They use a lottery-like procedure to randomly assign study participants either to receive new program services (program group or treatment group) or to receive what would normally be available to them (control group). Randomization ensures that the two groups are established by chance so that they are statistically similar on both observable and nonobservable characteristics (motivation, persistence, etc.) before providing program services. This statistical similarity, in turn, makes it most likely that the program is the reason for observed differences on the outcome.
	Quasi-experimental designs use a nonrandom process to select a comparison group that is as similar as possible to a group that has decided to participate in program services. For example, researchers might match participants' observable characteristics (test scores, gender, ethnicity, etc.) to make the program and comparison groups as similar as possible. QEDs are considered less rigorous than randomized control trials because we cannot rule out the possibility that any difference in outcomes reflects some nonobservable difference between the groups that led some students to participate and others to not participate.
Single-case designs	**Single-case designs** involve a single participant or a few participants. Data are collected multiple times before and after a program is introduced, and sometimes again after it is removed and reintroduced, to see how the program affects participant outcomes.

Search strategy

The protocol included the keywords that would be used to search for eligible studies (see Appendix A). A multipronged search strategy helped to ensure that all such studies were identified. First, the research team conducted a keyword search in several electronic databases for published and unpublished studies. Next, the team searched conference proceedings, websites, reports, and reference lists from literature reviews, systematic reviews, and meta-analyses of programs germane to the review. Finally, the team conducted a random search, by year or blocks of years, of 142 issues from 10 disability journals that are recognized as the most rigorous in their acceptance of research studies. The full search identified 10,752 unduplicated references.

Screening

Short study abstracts were obtained for all of the references identified in the literature search. These abstracts were screened to determine whether the studies were eligible for the review. First, the research team read the abstracts and eliminated studies that clearly fell outside the bounds of the protocol (studies in which participants were too young, most

[11] WWC pilot single-case design standards require multiple, independent SCDs of the same intervention to provide a strong basis for evidence. For SCD studies to contribute toward the evidence base on a program, at least five studies must meet pilot design standards, and these studies must have been conducted by at least three distinct research teams at three institutions with at least 20 participants.

participants did not have a disability, no student outcome was measured, program had no secondary school-based component, study had no eligible research design, etc.). Full-text articles were obtained for the remaining 738 references; because these reports provided more information from which to determine eligibility, they could be screened more intensively. The full-text screen removed studies for the same reasons as in the abstract-level screen. Studies at this level of the screening process required full-text screens because the abstract rarely reported all of the information necessary to determine study eligibility. The research team determined that 43 studies[12] were eligible for review using WWC standards. Of these studies, 3 were randomized controlled trials, 19 used quasi-experimental designs, and 21 used single-case designs.

Determining research quality: How did studies meet the evidence standards?

Two WWC-certified reviewers independently reviewed each of the 43 eligible studies.[13] The reviewers then met to reconcile any differences in their study reviews and to prepare a consensus master review. In accordance with the WWC procedures, queries were made to the author if additional study information was needed or if further clarification of questions was necessary to complete the review. A principal investigator or deputy principal investigator worked with reviewers during the reconciliation process to confirm that rating criteria were applied correctly.

The WWC standards that were applied to studies in this review were developed by technical working groups of highly experienced researchers and are documented in the *Procedures and Standards Handbook, Version 2.1* (U.S. Department of Education 2011). This version of the handbook includes evidence standards for group design studies and pilot design standards for SCDs.[14] The SCD standards are different from the standards for group design studies in addition to being at a pilot stage. To underscore these differences, we refer to the "WWC pilot single-case design standards" when referencing SCD standards specifically.

Following WWC procedures, when reviewing studies, we examined the implementation of the study design to identify problems that limited confidence in the estimated effects of

[12] Wagner (1991), Wagner et al. (1993), and Wagner and Blackorby (1996) report on the results of analyses conducted with the same or very similar groups of students. However, Wagner (1991) and Wagner et al. (1993) report several different potential programs and several different outcome domains; all findings of the Wagner and Blackorby (1996) study are reported in one or the other of these two earlier research studies. Because the Wagner et al. (1993) study provides the most detailed technical information, it will be cited for all program and outcome analyses that are common to the other two studies. Wagner (1991) only will be cited for all other analyses. Wagner and Blackorby (1996) will not be cited in text or in the appendices but appears in the reference list because it was reviewed. Notations to this effect appear in the reference list.

[13] WWC reviewer certification requirements are available at http://ies.ed.gov/ncee/wwc/Reviewers.aspx.

[14] To include a type of SCD called a multiple-probe design, we used additional criteria that were being developed by the WWC at the time when we conducted study reviews. Pilot standards that are specific to multiple probe designs were not included in Version 2.1 of the *Handbook*. Official pilot standards for multiple probe designs are planned for the next version of the *Handbook*.

participation in the program. For example, reviewers looked for problems with random assignment and the level of attrition in RCT studies. High attrition—indicated by either a low overall response rate to the study's outcome data collection or big differences in response rates for the program and comparison groups—limits confidence in the study design, because the people who did not respond are often quite different than those who did provide data for the study. For all QEDs and for RCTs with problems in random assignment or attrition, reviewers looked for participant data that affirmatively showed that the program and comparison groups were similar before the program started. For SCDs, reviewers looked for three opportunities to demonstrate a program or treatment effect, indicated by systematic attempts to introduce or withdraw the program, and how many data points are collected during each attempt. Reviewers also looked for how many SCD studies of the same program met the design standards, because SCDs are treated as a collection of small-scale studies for the purpose of contributing evidence to the review. Finally, studies of all design types need to use reliable measures of the outcome variable and be free of other measurement problems.[15]

Studies could fall short of meeting WWC standards for any of these reasons. In this review, the most common reasons for not meeting standards were:

- Group design studies (RCTs and QEDs) that did not provide data demonstrating that program and comparison students were similar before the program was introduced.
- SCD studies that either did not include three attempts to introduce or withdraw the program or did not meet requirements for ensuring that the outcome is a reliable measure.

In some cases, studies provided insufficient information to demonstrate that all standards were met, and the review team was not successful in obtaining the necessary information from the authors. In other cases, study-reported information indicated a standard was not met. Appendix B indicates the reason or "disposition" for each study that did not meet WWC standards.

WWC Evidence Standards Ratings for Group Design Studies. Under the WWC procedures, group design studies are assigned one of the following ratings:

- **Meets evidence standards without reservations:** This rating is for RCT studies with low attrition and no other problems affecting the study design. Studies with this rating provide a high level of assurance that differences in outcomes between the program and comparison groups are attributable solely to the program, because prior to the program the two groups were similar on both measured and unmeasured traits.

[15] The review process did not directly assess implementation fidelity or intervention feasibility. The WWC procedures, which were used to guide this review, do not include explicit standards for assessing implementation.

- **Meets evidence standards with reservations:** QED studies and RCT studies with high attrition and/or a problem with the random assignment process can earn this rating by demonstrating that the program and comparison groups have similar observable characteristics prior to the program. Studies with this rating provide a lower level of assurance that differences in outcomes between the program and comparison groups are attributable solely to the program, because students in the two the groups may differ on unmeasured traits.
- **Does not meet evidence standards:** This rating is for eligible group design studies that do not satisfy one or more of the requirements to meet standards with or without reservations. For example, a QED study in which all students in the program group had an intellectual disability and all students in the comparison group had a learning disability would not meet standards, because the program's effect cannot be separated from other factors associated with the different needs of the two groups of students.

WWC Pilot Single-Case Design Standards Ratings. SCD studies are assigned one of the following ratings:

- **Meets pilot single-case design standards without reservations:** This rating is for studies in which the researcher systematically introduces or withdraws the program at least three times. Outcome data must be collected at least five times before the program is systematically withdrawn or introduced again.[16]
- **Meets pilot single-case design standards with reservations:** This rating is for studies in which the researcher systematically introduces or withdraws the program at least three times. Outcome data must be collected at least three times before the program is systematically withdrawn or introduced again.
- **Does not meet pilot single-case design standards:** This rating is for studies that do not satisfy one or more of the SCD standards with or without reservations. For example, a study with only one or two opportunities to demonstrate a program's effect through systematic attempts to introduce or withdraw the program would not meet pilot SCD standards, because the study cannot sufficiently replicate observations of a program's effect.

Because the pilot SCD standards differ from the standards for group design studies, the rating categories do not imply that a study that meets WWC pilot single-case design standards without reservations is equivalent to a group design study that meets WWC evidence standards without reservations. Following WWC recommendations, SCD studies are treated as a collection of small-scale studies for the purpose of contributing evidence to the review. That is, for SCD studies to contribute toward the evidence base on a program,

[16] For example, a study of a single participant can meet pilot SCD standards without reservations with 20 data points by following an "ABAB" pattern (a minimum of five sessions without any intervention, five sessions with an intervention, five more sessions without any intervention, five more sessions with the intervention).

at least five studies must meet pilot standards, and these studies must have been conducted by at least three distinct research teams at three institutions with at least 20 participants.

Reporting results of the review

In advance of our search and review efforts, we anticipated needing to group studies and summarize findings within these groups to make them most useful to policymakers and practitioners.[17] In part, the more studies that pertain to a program or group of similar programs, the more confidence we could have about the results. In addition, we also wanted to identify which areas of transition programming offered little or no rigorous research. We drew on the transition research literature to identify categories of programs and used WWC summary measures to report on aggregate findings for those categories.

Categories of transition programs

To combine studies in program groups, we focused on categories of programs with a primary goal of facilitating the acquisition of employment, education, or independent living skills upon transition from secondary education. The programs were categorized following recent work of the National Secondary Transition Technical Assistance Center (NSTTAC) at the University of North Carolina at Charlotte (Test et al. 2009). The review team first examined the 16 "predictor categories" identified by Test and colleagues in their review of correlational studies that predict post-high school outcomes. One category, called "Exit Exam Requirements/High School Diploma Status" was eliminated because measured impacts of a state policy rather than a program. The remaining 15 categories were realigned into 8 larger categories of program approaches that are mutually exclusive to better reflect the nature of the programs included in our review (see Table 2). Appendix D presents the full list of NSTTAC predictor categories and shows how we realigned them into 8 categories for this review.

[17] Although WWC reviews individually named programs, our review included named programs and general approaches, such as participation in career and technical education rather than a type of career and technical educational program.

Table 2. Program categories and descriptions

Program category	Description
Career and technical education	Student participation in occupationally-oriented courses that were a part of regular career and technical education delivery. Could range from enrollment in one course appearing on the student's transcript to engagement in a planned program of study focused on an occupational goal. Courses or courses of study may not include a community-based work experience component.
Career awareness or development programs	Programs that provide career-awareness skills development; career-awareness training within a broader transition program.
Community-based work experience programs	Organized work experiences sanctioned by student's school but may be variable in terms of supervision and supports students receive at the job site, connection to students' school curriculum, and whether credits are earned. Programs include cooperative education or other forms of work-based learning, organized work study programs, or supported employment programs.
Employment while in school	Not a program per se but rather a status variable that reflects some level of student experience at either part-time or full-time (summer) work while the student was in school. Not intended to reflect work experience through a school transition program while it is occurring but rather reflective of working while in school.
Functional life-skills development programs	Programs that provide curriculum and instruction on behavioral skills that are critical for productive engagement in employment, postsecondary education, or independent living.
Inclusion in general education	A status variable that reflects student enrollment for most of the school day in a regular or mainstreamed educational environment.
Interagency collaboration programs	Student services provided in an environment in which there is collaboration among stakeholders intended to improve student post-high school success; interagency collaboration can be demonstrated as supporting individual student planning for post-high school activity, or more generally reflecting an agreement between the community agency and a school or school system.
Student support including parent involvement	Programs that involve parents or other community members engaged in transition planning; other natural supports in transition planning; students engaged in a formal or informal support network.

Measures for reporting review findings

We use WWC reporting indicators to summarize the results of studies within program categories. The *rating of effectiveness* indicates whether, on balance, the WWC evidence indicates that the program approach makes a difference for students.[18,19] The *improvement index* gauges how much of a difference the program makes.[20] The *extent of evidence rating* is a rough measure of whether the size of the body of evidence on the program, taken as a whole, is small or medium-to-large, so that readers can judge for themselves how much they can generalize about the findings.

For RCT and QED studies, the rating of effectiveness is based on the number of studies showing statistically significant or large effects relative to studies showing indeterminate effects. Whether studies are RCTs or QEDs also matters, because the highest rating

[18] A limitation of the WWC standards is that all studies with the same rating are treated equally in determining the rating of effectiveness. For example, an RCT involving 20 students is viewed as providing the same level of support as an RCT involving 2,000 students if both studies meet WWC evidence standards without reservations.

[19] The research team needed to go beyond the WWC standards and procedures for rating the effectiveness of program categories when the evidence includes single-case design studies. The WWC has not established a method for combining evidence across SCD studies. The scale used in this report is based on the quality indicators described by Horner et al. (2005) and modeled after the existing WWC rating of effectiveness scale for group design studies (see Appendix C, Tables C1 and C2).

[20] Following the WWC, improvement indices are obtained for group design studies only.

("positive effects") is possible only if the reviewed studies include an RCT. For SCD studies, the rating of effectiveness is based on the number of studies showing a functional relationship between the program and the outcome measure. A rating of effectiveness for SCD studies is calculated only when certain requirements about the number of studies, distinct research teams, and participants are met. The extent of evidence rating is based on all studies reviewed for a program regardless of whether they are RCT, QED, or SCD studies. Ratings consider the number of studies, schools, and participants (see Appendix C for additional information).

Although the WWC uses these indicators to summarize across studies of the same program, we apply them to summarize across similar programs in a particular category. This higher level of aggregation has some risks, but given the small number of studies that met WWC standards in our review, they are not a significant concern here.[21] We briefly describe these terms in Table 3.

Table 3. General description of WWC reporting indicators

Indicator	General description
Rating of effectiveness	An indicator for whether the evidence demonstrates positive effects, potentially positive effects, mixed effects, no discernible effects, potentially negative effects, or negative effects. Ratings are differentiated based on research quality, number of studies, and strength of evidence.
Improvement index	A measure for reporting the size of program effects. It represents how many percentiles an average student is expected to improve or decline due to the program. An improvement index of +10 indicates that a comparison group student at the 50th percentile of the outcome distribution is expected to improve to the 60th percentile after participating in the program.
Extent of evidence	An indicator of the overall size of the evidence base on a program category regardless of study design type. The extent of evidence is considered medium-to-large if evidence comes from more than one study, more than one school, and more than 350 study participants. Otherwise, the extent of evidence is considered small. A medium-to-large extent of evidence rating does not imply that a program generalizes to students with different characteristics or to research settings other than the ones considered in the studies.

Note: Strength of evidence in group design studies is indicated by statistically significant or substantively important findings. Strength of evidence in SCD studies is indicated through visual analysis of the data patterns. Appendix C provides more information on the reporting indicators for group designs and SCDs.

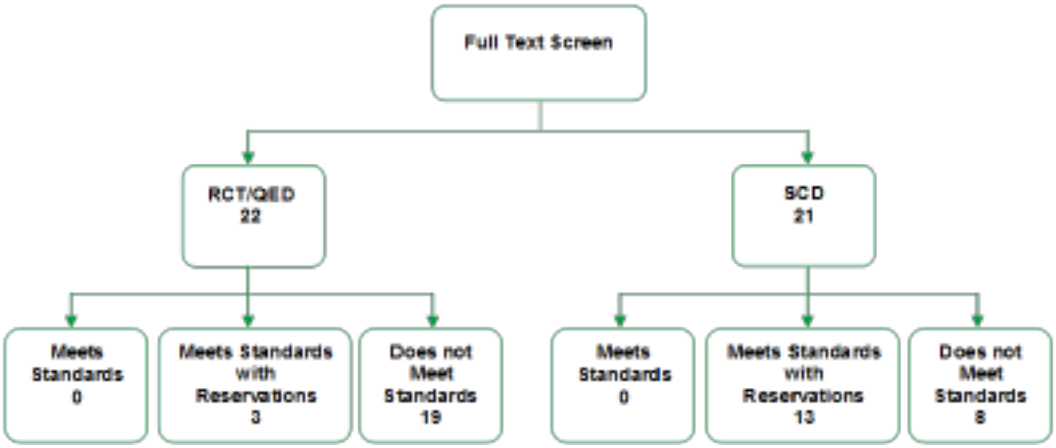
Source: *What Works Clearinghouse Procedures and Standards Handbook, Version 2.1* (U.S. Department of Education 2011), available at http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19.

[21] For example, there is a risk in overgeneralizing strong findings on a highly researched strategy to related strategies that have not been evaluated formally. Also, effectiveness information may be muted in the aggregate if strong findings for one program are combined with findings of no effects for a different program.

Chapter 3. Results of the evidence review

The 43 eligible studies were reviewed and assigned a WWC standards rating. Of the 22 RCT or QED studies, none met WWC evidence standards without reservations, 3 met WWC evidence standards with reservations, and 19 did not meet WWC evidence standards (Figure 2). Of the 21 SCD studies, none met WWC pilot SCD design standards without reservations, 13 met WWC pilot SCD design standards with reservations, and 8 did not meet WWC pilot SCD design standards.

Figure 2. Illustration of the study review process with the number of studies by rating category



Of the 16 studies that met WWC standards with reservations, 3 examined an employment outcome, one examined a postsecondary education outcome, and 13 examined independent living outcomes.

Table 4 summarizes the sample sizes, age ranges, gender, race/ethnicity, and disabilities of student participants across group design and SCD studies that met WWC standards with reservations, by outcome domain. Across the studies, employment outcomes were measured for a total of 666 students, postsecondary education outcomes were measured for 215 students, and independent living outcomes were measured for 42 students. Nearly all of the students (98 percent) that participated in the studies measuring employment outcomes had either an intellectual disability (66 percent) or an emotional or behavioral disturbance (32 percent), although these studies also look at other disabilities. The students that participated in the study measuring postsecondary education had emotional or behavioral disturbances. Of 42 participants in the studies measuring independent living outcomes (the SCD studies), 40 had an intellectual disability, although 15 had other types of disabilities, as well. The majority of students across studies were male. The race/ethnicity breakdown varied across studies and was not consistently provided.

See Appendix E for study-level information on student characteristics, sample sizes, program descriptions, program settings, and outcomes.

Table 4. Sample sizes and participant characteristics for studies that met WWC standards with reservations, by outcome domain and design type

Outcome domain (design type)	Number of studies	Number of participants	Ages	% Male	Race/ ethnicity	Disabilities represented
Employment (group design)	3	666	14–22	64	66% Caucasian	Autism, ADHD, EBD, ID, LD, PD, SI, TBI
Employment (SCD)	0	n.a.	n.a.	n.a.	n.a.	n.a.
Postsecondary education (group design)	1	215	18–22	76	21% Caucasian	EBD
Postsecondary education (SCD)	0	n.a.	n.a.	n.a.	n.a.	n.a.
Independent living (group design)	0	n.a.	n.a.	n.a.	n.a.	n.a.
Independent living (SCD)	13	42	14–20	55	NR[a]	Autism, ADHD, AT, CP, D, D-B, DS, EP, ID, LD, SD, TS, VI

Note: n.a.=not applicable; ADHD=attention deficit hyperactivity disorder; AT=ataxia; CP=cerebral palsy; D=deaf; D-B=deaf-blind; DS=Down syndrome; EBD=emotional or behavioral disability; EP=epilepsy; ID=intellectual disability; LD=learning disability; PD=physical disability; SD=seizure disorder; SI=sensory impairment; TS=Tourette syndrome; TBI=traumatic brain injury; VI=visual impairment.

[a] NR means not reported, because fewer than half of the studies reported information for the table cell.

Source: Appendix E.

Studies with employment outcomes

Three eligible studies, all of which were group designs with outcomes in the employment domain, met WWC evidence standards with reservations. All three studies provided evidence of the effectiveness of community-based work experience programs (Baer et al. 2011; Cimera 2010; Karpur et al. 2005). Table 5 summarizes the WWC effectiveness ratings, improvement indices, and extent of evidence ratings for these three studies, and highlights voids in the research literature where no program studies met WWC evidence standards in the seven other categories of programs.

Table 5. Summary of evidence for the employment domain, by program category

Program category	Type of design	Rating of effectiveness	Improvement index (percentile points)	Extent of evidence			
			Average	Range	No. of studies	No. of students	Rating
Career and technical education	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Career awareness or development programs	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Community-based work experience programs	Group	Mixed effects	+2	-2 to +17	3	666	Medium-to-Large
Employment while in school	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Functional life-skills development programs	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Inclusion in general education	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Interagency collaboration programs	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Student support including parent involvement	n.a.	n.a.	n.a.	n.a.	0	0	n.a.

Notes: n.a.=not applicable. See Appendix E for study-level information on participant characteristics. See Appendix C for criteria for the rating of effectiveness and the extent of evidence rating.

Sources: Appendix E and F.

The summary information in Table 5 above masks distinct differences in the types of programs being evaluated in these three studies and in the nature of the disabilities for the students in the study sample. Below, we highlight design features of each of the three studies as well as program characteristics under study, and then summarize important differences that merit consideration by education decision makers.

Baer et al. (2011) was the largest QED study in terms of sample size and followed the study sample for the longest period of time. However, the study was limited to students with intellectual disabilities. Baer and his colleagues followed 409 former students who graduated or aged out of special education in 2005, 2006, 2007, and 2008 in a single Midwestern state and assessed whether they had by one year after graduation ever worked 35 or more hours per week. Although the study examined several school-based programs, the only program approach for which the study met WWC evidence standards was (in the words of the study authors) "whether… the students had participated in any work study programs while in high school." The study authors did not give details of the characteristics of the work study program; the school records indicated that the students had participated in a work study program. The study authors found that participating in a work study program had no statistically significant effect on full-time employment rates following school for students with intellectual disabilities.

Cimera (2010) reported the effects of a small evaluation that used a QED to examine the employment-related outcomes for 21 students who received a community-based transition program (referred to by the author as "supported employment") versus 21 students who received "no transition services." The students were selected so that pairs of students—one from the program group and one from the comparison group—had the same gender, age, primary disability, presence of a secondary disability, and adult service provider. Those in the program group received a menu of services including job shadowing and sampling, career and technical assessments, work adjustment supports, payment for work, and other IEP-specified community-based services. Students in the "no transition services" comparison group had been diagnosed with their disability after leaving high school, were homeschooled or were in school prior to the 1990 authorization of the IDEA, and had IEPs that did not reflect receipt of any transition services while they were in school. The outcome measure from this study that conformed with WWC evidence standards was "average months employed per year after graduation." This evaluation demonstrated statistically significant positive effects with an improvement index for the program group of 17 percentage points over the "no transition services" comparison group.

The third study that examined the effects of a community-based work experience program on employment-related outcomes was conducted by Karpur et al. (2005). Karpur and her colleagues assessed the efficacy of the Steps-to-Success program in Miami-Dade County on 43 students with emotional or behavioral disorders using a matched pairs QED. They selected 172 comparison students who were matched on type and severity of disability and who did not participate in this program. The Steps-to-Success program contained six components: (1) person-centered planning, (2) community/career and technically oriented curriculum and employability training, (3) progressive inclusion in career and technical/technical courses, (4) paid or unpaid work-based practicum, (5) wraparound supports for school and work experience endeavors, and (6) group therapy and counseling. When assessed at follow-up on whether these former students from both groups were employed, there was no statistically significant difference—that is, the study found no discernible employment-related effects of the Steps-to-Success program at follow-up.

Although one study showed a statistically significant effect, two studies showed an effect that was not statistically significant. Taking these three studies as a group, the effectiveness of community-based work experience programs is considered to demonstrate "mixed effects" (see Appendix C, Table C.1). The extent of evidence is rated "medium-to-large" for employment outcomes, because it comes from more than one study, more than one school, and at least 250 students (see Appendix C, Table C.3).

Studies with postsecondary education outcomes

Only one study—the same Karpur et al. (2005) study reviewed earlier in the employment outcomes section—satisfied WWC evidence standards with reservations with regard to postsecondary outcomes. Karpur et al. assessed the effects on postsecondary enrollment of the Steps-to-Success program, a community-based work experience program with additional

supports. The effects of this program were dramatically different for postsecondary education than they were for employment outcomes. Steps-to-Success participants were three times as likely to be enrolled in postsecondary academic or career and technical education programs as were nonprogram comparison group students.

With no additional studies from which to make comparisons, we have no additional contextual information to offer local program developers. Karpur et al. repeatedly state that their program was based on an earlier Transition to Independence Process (TIP) model, and list seven guiding principles of this model in their description of their Steps-to-Success program model. No evaluations of the TIP model met the WWC evidence standards used for our review, however.

Table 6 summarizes the WWC effectiveness rating, improvement index, and extent of evidence rating for the single study that met WWC evidence standards, and again highlights the lack of WWC evidence-rated studies for other categories of programs that focused their efforts on improving postsecondary outcomes. Based on the one study, community-based work experience programs are rated as providing "potentially positive effects" for postsecondary education outcomes and the extent of evidence is rated as "small."

Table 6. Summary of evidence for the postsecondary education domain

			Improvement index (percentile points)	Extent of evidence			
Program category	Type of design	Rating of effectiveness	Average	Range	No. of studies	No. of students	Rating
Career and technical education	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Career awareness or development programs	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Community-based work experience programs	Group	Potentially positive effects	+16	n.a.	1	215	Small
Employment while in school	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Functional life-skills development programs	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Inclusion in general education	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Interagency collaboration programs	n.a.	n.a.	n.a.	n.a.	0	0	n.a.
Student support including parent involvement	n.a.	n.a.	n.a.	n.a.	0	0	n.a.

Notes: n.a.=not applicable. See Appendix E for study-level information on participant characteristics. See Appendix C for criteria for the rating of effectiveness and the extent of evidence rating.

Sources: Appendix E and F.

Studies with independent living outcomes

There were 13 single-case design studies that examined one or more independent living outcomes and met WWC pilot single-case design standards with reservations. Each study used a sample of fewer than 6 students and investigated the effects of developing a specific functional life skill. Because the sample sizes are small—only 42 total students across the 13 studies—and a range of programs and outcomes are examined, we do not generalize study findings beyond the student participants themselves in the settings in which they participated. Notably, participants varied considerably across studies in their functional abilities. In any given study, the participants struggled initially with the particular skill that the study examined and were selected because they were expected to be able to participate meaningfully in a program designed to improve on that skill. All but two of the study participants had an intellectual disability. The functional abilities of these students may or may not resemble the functional abilities of other students with intellectual disabilities and are not representative of functional abilities for students with disabilities overall.

Table 7. Summary of evidence for the independent living domain

Program category	Type of design	Rating of effectiveness	Extent of evidence		
			No. of studies	No. of students	Rating
Career and technical education	n.a.	n.a.	0	0	n.a.
Career awareness or development programs	n.a.	n.a.	0	0	n.a.
Community-based work experience programs	n.a.	n.a.	0	0	n.a.
Employment while in school	n.a.	n.a.	0	0	n.a.
Functional life-skills development programs	SCD	Potentially positive effects	13	42	Small
Inclusion in general education	n.a.	n.a.	0	0	n.a.
Interagency collaboration	n.a.	n.a.	0	0	n.a.
Student support including parent involvement	n.a.	n.a.	0	0	n.a.

Notes: n.a.=not applicable. Improvement index values are not calculated for SCD studies.

Sources: Appendix E and F.

To combine SCDs into a domain-level rating, we examined the collection of studies to see how program effects were replicated across studies. The 13 studies of functional life-skills programs that met WWC pilot single-case design standards with reservations were conducted by nine research teams and involved a total of 42 students. We found "potentially positive effects"[22] of functional life-skills development programs on direct measures of students' independent living potential.[23] However, there is no standard way of aggregating findings from single-case design studies. In this section, we describe strategies that were used in multiple studies and reiterate that the findings from these studies are not intended to generalize beyond the students—most of whom have an intellectual disability and highly specialized support needs—who were studied in the settings in which they participated.

It was common for programs across the SCD studies to include multiple components that are tailored to the individual needs of the student participants. In such instances, the separate effects of each component typically cannot be distinguished. Although this factor limits our ability to determine whether certain practices are effective on their own, we believe it is still useful to highlight descriptively the common features of programs that researchers have used that may be promising for assisting future students with intellectual disabilities with their transitions to post-high school life.

Ten of the 13 studies examined functional life-skills development programs that used (a) multimedia computer-based instruction, (b) a system of least prompts to reinforce using the least support needed for the student to complete a task, and/or (c) a constant time-delay procedure extending the delay in cueing the student to demonstrate the skills as part of delivering programs for students. Table 8 and the text below describe these three strategies and the study findings involving them. The remaining three studies (Ganz and Sigafoos 2005; Taras 1993; Test et al. 1993) examined unrelated program components that involved self-monitoring, a multicomponent "independence training" strategy, and the one-more-than money-counting technique, respectively.[24] Appendix E and Appendix F summarize these studies and findings.

---

[22] WWC has not established procedures for rating the effectiveness of interventions based on SCD research. We developed SCD ratings of effectiveness for use in this review, as an exploratory way to rate the effects of a program based on single-case design studies. The ratings were modeled after WWC ratings of effectiveness based on RCTs and QEDs and developed to be consistent with the WWC pilot standards for SCDs. Appendix C, Table C.2 describes the rating criteria. Functional life-skills development programs are rated as "potentially positive," because there are at least three studies that document a positive functional relationship and no studies that document a negative functional relationship (a relationship opposite to the one that is theorized).

[23] Examples of direct measures of independent living potential include using an ATM, fax machine, or copy machine; making a debit card purchase; selecting grocery items; cooking; and doing laundry.

[24] The self-monitoring training in Ganz and Sigafoos (2005) used a reward system to encourage students to complete tasks. Taras (1993) examined a multicomponent independence training program that targeted blind students and combined group training, self-evaluation, peer evaluation, reinforcement, prompting, and modeling. The one-more-than money-counting technique used by Test et al. (1993) taught students to put one dollar aside for the change or "cents pile" and count the number of whole dollars necessary for a purchase.

Table 8. Strategies used in multiple studies of functional life-skills development programs

Strategy	General description
Multimedia computer-based instruction	Multimedia computer-based instruction uses computer-based instruction and some combination of video modeling, animated prompting, images, or interactive practice to teach students how to complete tasks such as shopping or using an ATM. Students are then observed undertaking the tasks in real environments (for example, at the store or ATM).
System of least prompts	In a system of least prompts, students demonstrate a targeted behavior or skill by completing a task with the least support necessary. Students are initially given a prespecified amount of time to complete the task independently. If a correct response is not observed, prompts are introduced that provide increasing levels of support. Prompts can be verbal or include gestures, modeling, or physical assistance.
Constant time delay	Constant time delay typically is used as a prompt-fading approach. Students are shown how to complete a skill such as doing laundry or making copies, and then they are asked to demonstrate the skill. A natural cue is selected at each task step that is meant to elicit the correct response when introduced. Initially, the students are prompted immediately after the natural cue (called a zero-second delay). The delay is then extended, typically to two to five seconds, to allow students an opportunity to self-initiate and demonstrate the skill without the prompts. Students are then given feedback on their performance.

Multimedia computer-based instruction

Four studies (Hutcherson et al. 2004; Mechling and Gast 2003; Mechling et al. 2002; Mechling et al. 2003) with 10 total students with intellectual disabilities examined the effects of multimedia computer-based instruction on grocery-store item selection and debit card purchases.[25] The interventions in the three Mechling et al. studies, which also integrated components like a system of least prompts or a constant time-delay procedure, led to improved community participation outcomes for nearly all of the student participants.[26] However, we concluded that students in Hutcherson et al. (2004) did not show a clear collective response to the intervention. Consistent with our findings, Hutcherson et al. reported that some improvements were either marginal or not immediate, and some students were already performing at high levels prior to the intervention. Thus, all of the studies that present evidence of a positive response to multimedia computer-based instruction were conducted by Mechling and coauthors. Although the findings provide some preliminary evidence, their replication by other researchers would greater support a conclusion that multimedia computer-based instruction is effective for students with intellectual disabilities who need help improving community participation outcomes.

[25] One student participated in two of the studies (Mechling and Gast 2003 and Mechling et al. 2002) and is counted once in the total number of students.

[26] We examined the data patterns in each study using visual analysis, which is a process whereby researchers can infer whether the intervention had an effect on student outcomes. Visual analysis involves examining several characteristics about students' data patterns before and after an intervention, such as changes in the level, trend, and variability of the data. Following the WWC pilot standards for SCDs, intervention effects must be observed at three different points in time to conclude that it had an effect on student outcomes. See Appendix A for further details.

System of least prompts

Six studies with a total of 22 students with either a mild or moderate intellectual disability used a system of least prompts as part of the intervention procedure (Cihak et al. 2004; Mechling et al. 2002; Mechling et al. 2008; Mitchell et al. 2000; Taber et al. 2003; Taylor et al. 2002). All but one of the students had better independent living outcomes following intervention programs that included a system of least prompts. The outcomes—doing laundry, using an ATM, making copies, sending a fax, making a purchase with a debit card, and/or using a cell phone to get help when lost—were selected by the authors based on students' individual experiences, abilities, or needs. Across the studies, the system of least prompts was used in functional skills-development programs delivered by DVD, an auditory prompting system, or photographs. Although it is promising that a system of least prompts is associated with successful intervention programs that are delivered in different ways and that target different functional skills, we cannot ascertain the degree to which it contributes to these positive findings for students.

Constant time delay

Three studies with eight total students used constant time delay as part of the delivery of functional life-skills development programs (DiPipi-Hoy and Jitendra 2004; Mechling and Gast 2003; Mechling et al. 2003). Although the outcomes in all three studies involved purchasing skills, the intervention in DiPipi-Hoy and Jitendra (2004) was parent-delivered rather than multimedia-based as in the Mechling et al. studies. That is, the parents of three students were trained to implement for their children a constant time-delay procedure. DiPipi-Hoy and Jitendra found that the parents quickly learned the procedure and implemented it with a high degree of fidelity. The researchers also found that the parent-delivered instruction improved the three students' purchasing skills at local grocery and retail stores.

Chapter 4. Hypotheses for future program development and research on transitions

Our review of transition research studies from the past two decades indicates that relatively few studies meet the WWC standards for credible evidence of effectiveness. Many studies do not include a control or comparison group and, therefore, were screened out early in our review process. Studies with a comparison group often had design or implementation weaknesses of sufficient severity that their findings are not judged to be credible estimates of the intervention's impacts. Finally, perhaps because of pressures to keep costs low and turnaround of results quick, many researchers do not follow students beyond high school to ascertain whether secondary school interventions have the desired ultimate effects after the transition from high school. In our review, no group design studies met the WWC evidence standards without reservations, and only three group design studies (randomized control trials or quasi-experimental studies) met evidence standards with reservations. A set of 13 SCD studies of functional life-skills development programs met the WWC pilot single-case design standards with reservations (see Figure 2). These studies were based on student samples that ranged from about 40 to 400 for the group designs and one to 4 for the SCDs. In total, the existence of evidence across eight program categories allowed us to review only two of the categories, providing little support from high-quality intervention research for identifying a wide range of programs to help students with disabilities to transition successfully to employment, postsecondary education and training, or independent living.

The lack of studies that met the WWC's standards led us to also consider empirical studies that fell somewhat short of meeting the standards as a source of hypothesis-generating information on approaches to students' secondary transitions. The research team felt that these studies might offer some additional insights to inform policy directions and practices concerning students' secondary transitions, provided that the research designs in these studies had many of the features that the WWC standards require. The team defined a set of rating criteria resulting in the inclusion of eight additional group design intervention studies as providing "exploratory results" (see Appendix A); these criteria allowed group design studies to be included if they did not demonstrate baseline equivalence but met all other requirements for a rating of meeting WWC evidence standards with reservations. Only group design studies were considered for this lower tier; SCD studies were not considered, because the WWC standards for SCDs are in a pilot stage. Table 9 lists the program categories in which these exploratory studies fell.

Table 9. Study counts that include exploratory studies by program category

Program category	Reviewed studies (n=43)	Met WWC standards (n=16)	Exploratory studies (n=10)
Career and technical education	7	0	4
Career awareness or development programs	4	0	1
Community-based work experience programs	10	3	3
Employment while in school	7	0	2
Functional life-skills development programs	20	13	0
Inclusion in general education	5	1	2
Interagency collaboration	1	0	1
Student support including parent involvement	0	0	0

Note: The number of studies in each column may sum to more than the total number of studies because several studies examine programs in more than one category. Appendix G has broken out all the group design studies by program categories and the domains where they appear.

Because these additional studies did not meet WWC standards, we do not have the same level of confidence in the causal connection between the implementation of the interventions and the reported effects that could be expected to accrue from the interventions. For that reason, we consider any hypotheses we glean from these studies to be "exploratory" and useful mostly for thinking about the ongoing development of program strategies rather than yielding a set of confident recommendations for "what works." Our aim is to help policymakers and practitioners as they consider different programmatic options for transitioning high school students with disabilities.

Hypotheses about characteristics of programs

First, it appears that a growing number and greater variety of transition programs have been developed and described in the research literature since Alwell and Cobb produced their 2006 review of the intervention research literature in special education transition. Local policymakers now appear to have greater numbers of, and more widely varying, "named" or "branded" curricula and transition support models than they had just 10 years ago as they consider their own programmatic development needs.

Second, many of these model programs "bundle" combinations of components that can target different types of students and outcomes. We see this pattern especially for functional life-skills development programs, where program developers and researchers often focus at least as much on identifying promising strategies for particular students (for example, students with autism or intellectual disabilities) as on identifying promising strategies for general populations of students. This bundling of components is clearly evident, for example, in the several studies by Mechling et al. that combine multimedia computer-based instruction with other program components, such as a system of least prompts or a constant time-delay procedure. Bundling of components is also evident in the suite of wraparound services Karpur et al. (2005) put together to address the unique needs of students with emotional and behavioral disorders who were placed in community-based work experience programs. Although Karpur's intervention was not found to be effective in increasing employment rates at follow-up, it illustrates the willingness of local program developers to experiment with different kinds of community-based work supports that

target students with varying types and severities of disability. Another facet of this "targeting" process is also evident in the Karpur et al. study; their program was not found to be effective in the employment outcome domain, but it was highly successful in the postsecondary education outcome domain. It may illustrate that "bundled" services should target outcome domains as well as student needs.

We hypothesize that this increased variability of options is the result of school personnel having more information about a greater variety of program options as well as having more "close to the ground" experience about how to combine programmatic components to best serve students with differing challenges and transition outcomes in mind. Secondary education with a focus on transition outcomes for students with disabilities is still a relatively recent policy focus when compared with, for example, academic outcomes. As local policymakers and program developers consider their own schools, communities, and adult services contexts, it seems they have a lot more published input from which to choose. Expanded program options notwithstanding, it is evident from our review that high-quality evaluations of program effectiveness have not kept pace with program development. The next section offers a series of methodological recommendations to help alleviate this issue.

Hypotheses about potential avenues for program development

Combined, the studies that met WWC standards and those that we considered as exploratory[27] generated several hypotheses for which there is suggestive evidence that federal, state, and local educators might consider. These possibilities include:

- **How program components *are linked* may be critical.** One area for exploration is how program components are constructed and linked to improve student outcomes, as an alternative to offering a menu of potentially disconnected program components. For example, Baer et al. (2011)[a] found no evidence of positive benefits for community-based work experience programs for students with intellectual disabilities. Similarly, although her study is drawn from correlational survey evidence, Wagner et al. (1993)* found similar null results for all of the students with disabilities in their national sample. Yet Shandra and Hogan (2008)*, drawing exclusively from correlational analysis of a different national survey sample, report just the reverse (as did Wagner for the small subset of students with sensory impairments in her survey sample). We point out these contradictions with caution, as, unlike with the Baer study, we do not know the specific school contexts supporting the samples in the Wagner study or the Shandra and Hogan study. This caution notwithstanding, there is some suggestion that work experience programs must be integrated into (or complemented with) other program components such as regular career and technical education or

[27] For this section: [a]studies that met WWC evidence standards; *exploratory studies.

workplace supports for them to produce expected positive benefits. The importance of this connection to other programming, whether in school or in the form of workplace supports, may explain why Cimera (2010) found recipients of supported employment services stayed employed more than twice as long as "no treatment" controls, whereas Baer et al. stated, "…it (null effects in their study) may also have resulted from the failure of many work study programs to integrate work study and the general curriculum" (pp. 8).

- **Participation in career and technical education may be important for promoting employment outcomes.** Three exploratory studies examined the links between career and technical education and a variety of employment-related outcomes, including employment levels (part-time, full-time, or employed at all), wages, fringe benefits, and employment stability (consistency). Although all of these studies used relatively sophisticated statistical controls to isolate the relationship between participation in career and technical education and employment outcomes, none of them did so adequately to meet WWC evidence standards, and none involved actual observations of the career and technical education program environments (they were all analyses of national survey data). However, the results from these three large, nationally representative survey studies suggest that participation in mainstream career and technical education may be positively associated with student employment outcomes (see Harvey 2002*; Shandra and Hogan 2008*; Wagner et al. 1993*). Although two other more geographically focused surveys contradicted these findings, both offered caveats with their results. Baer et al. (2011)* suggested that the career and technical education received by the students with intellectual disabilities in their study was likely not offered in inclusive settings (that is, it was not *regular* career and technical education). Fourqurean et al. (1991)* indicated that few students in their study sample *did not* receive regular career and technical education, and this lack of variability resulted in the "no positive effect" finding.

  An important caution about drawing even tentative lessons from these studies is that the students with disabilities who were interested and able to participate in regular career and technical education might have been more motivated to become employed or to work in higher earning jobs than the students with disabilities who did not choose to participate in regular career and technical education.

  If the post-high school goal for students is to attend postsecondary education or to live independently in the community rather than to go to work, one exploratory study (Wagner 1993*) suggests participation in regular career and technical education neither helps nor hinders students in achieving this objective. Similarly, Harvey (2002)* reported negative correlations between enrollment in regular career and technical education and postsecondary education participation rates, again drawing attention to the likely importance of aligning high school programming with post-high school goals.
- **Employment in at least one job before students with disabilities leave high school may be an integral part of transition support.** There have been

correlational studies since the mid-1980s supporting a positive relationship between working in some capacity while in school and greater employment outcomes after leaving school for students with disabilities (Hasazi et al. 1985; Mithaug et al. 1985; Wehman et al. 1985). No studies could improve upon the *quality* of the evidence supporting these correlational findings, but several exploratory studies offered findings that were consistent with those earlier results (Harvey 2002*; Shandra and Hogan 2008*; Wagner et al. 1993*).

- **Inclusive education settings may be a key dimension of transitioning to postsecondary education.** Although only one study meets WWC evidence standards with reservations, and one additional exploratory study in our review examined the relationship between inclusive education and postsecondary enrollment, these studies do suggest a positive direction for inclusive education. Baer et al. (2011)* found positive correlations for students with intellectual disabilities. Wagner et al. (1993)* reported positive correlations for students with mild disabilities. These findings require the same cautions as those noted above for studies that examined participation in career and technical education. We cannot know for sure that students in inclusive settings do better because of the setting or because participants in inclusive education are different in ways that researchers could not control for in the studies.
- **Several strategies, such as computer-based instruction and prompting, may help students with intellectual disabilities to live more independently by increasing their functional skills.** Thirteen studies that met WWC pilot SCD standards with reservations in our review examined the relationship between functional life-skills development programs and direct measures of independent living outcomes, and most of them found positive functional relationships. The study by DiPipi-Hoy and Jitendra (2004), for example, found a positive relationship between a constant time-delay procedure and purchasing in the community. The researchers had focused on making the intervention easy to deliver for parents who had a large role in implementing it. Unlike many constant time-delay procedures, the intervention used in this study did not involve zero-second delay trials so that the procedure could be more naturalistic. We hypothesize that program developers and schools may find that the degree to which programs can be implemented easily is an important consideration, especially when the youth's family or peers are helping to provide the intervention.

  Although the findings from the 13 studies are promising about the effectiveness of functional life-skills development programs, the studies targeted different skills (for example, purchasing, selecting food items, cooking, or doing laundry) and selected youth who could benefit from receiving the program services. Studies also tended to bundle program components, making it impossible to identify the separate effect of each component. Thus with the available information we cannot say for sure that any single program examined by studies included in this review—including multimedia computer-based instruction, system of least prompts, and constant time delay, which were examined as part of multiple studies—is either

effective on its own or would be effective for students with characteristics that differ greatly from study participants with intellectual disabilities.

Research design considerations

Finally, the results from this review suggest several recommendations for researchers who are dedicated to contributing information about what does and does not work for student transition.

- **Greater attention should be paid to a few key design elements for group design research.** Randomly assigning eligible participants into a program and comparison group[28], using comparison groups that are shown to be equivalent to the program group before the program begins, and including preintervention measures of all outcomes in the analysis is time-consuming and sometimes expensive, but it is well worth the effort, both for the interpretability of the individual study findings and their utility in larger evidence reviews. If random assignment is not feasible, researchers are encouraged to include in their studies a comparison group as similar as possible to the program group. When comparing pre and post data, for example, a comparison group will help to strengthen the usefulness of study findings in evidence reviews. This request to pay attention to design elements also holds true for single-case design studies. Researchers should become familiar with the WWC's pilot standards for single-case designs and consider those criteria in their studies. One of the most important takeaway points from the pilot standards is that study designs should always allow for at least three opportunities to demonstrate an intervention effect at three different points in time.
- **Group design studies that focus on post-high school employment outcomes should include as a control variable data on "employment while in school."** Had study authors whose interventions focused on employment outcomes provided data on whether students had ever had a job before they received the intervention (and used this information as a control variable in their analyses), this review would likely have included many more studies. The correlational evidence showing the positive relationship between having a job while in school and having one after leaving school has been consistent since the follow-up studies of the 1980s. As a result, studies that fail to control for or take into account students' prior job experience cannot isolate the effects of a different work-related intervention on post-high school employment outcomes.
- **Unless it is a randomized control trial with low attrition, a group design study should always control for type and severity of disability in analyses**. As you might expect, no student attribute variables were more consistently important in both studies that met WWC evidence standards and exploratory studies in differentiating post-high school results from in-school experiences than disability

[28] Using a wait-list control group is an option if random assignment is not feasible.

type and severity. Although many of the studies we looked at included these types of data in the sample information, the analyses often did not. Drawing the causal connections between intervention and outcomes is much more difficult if equivalence of type and severity of disability in the program and comparison group samples is not explicitly made. Our understanding about the effectiveness of different programs will be much clearer and more policy-relevant if researchers on the secondary transitions of students with disabilities break out information around type and severity of disability, gender, and race/ethnicity, when describing the characteristics of the sample and in the analyses that they conduct.

- **Studies should measure post-high school outcomes directly and connect those measures analytically to indirect or interim measures (locus of control) that are taken while the students are still in school.** Many studies were initially identified in a number of the Test et al. (2009) review areas. We did not include them in this review because they did not examine direct measures of post-high school outcomes of interest. Perhaps most striking for their absence from this review were studies that focused on the effects of self-determination curriculum and instruction on transition outcomes. We were interested in reporting on the effects of these self-determination interventions but found that none of the reported research fit the central requirements of our review: that outcome measurement of the effects of these interventions must have focused on direct measurement of employment, postsecondary participation, or independence of living environments. We encourage the special-education research community to extend its outcome measurement for such curricular interventions to include actual measures of post-high school employment, postsecondary education, or independent living. This approach would authenticate the effects of these types of in-school curriculum and instruction on post-high school outcomes. Fulfilling this goal is increasingly possible; post-high school outcomes are becoming easier and less costly to obtain. State longitudinal data systems supported by ED grants can be accessed by researchers who convince state and local education officials of the importance of their analyses, and other administrative databases (for example, the National Student Clearinghouse for postsecondary enrollment information) can also be tapped. IES' National Center for Special Education Research runs various competitions to provide grant funds for rigorous research including in the area of post-high school transitions.
- **We need more work on scale development for direct measures of independent living, particularly for students with more moderate and severe intellectual disabilities.** From the studies we reviewed, research on the direct measurement of living independently seems to have lagged well behind that of employment or postsecondary education participation. Heal and Rusch (1994)* reported a ranked scale based on Guilford's (1954) work (as cited in Heal and Rusch) but we found no studies subsequently used this scaling technique. More frequently, estimates of independent living were simply reported as "yes/no" observations on a number of status issues, such as living alone; living with a spouse or roommate; living in a

college dormitory; living in military housing (Wagner 1993)*; being married; being registered to vote; participating in social groups; or holding one of several items such as a driver's license, credit card, or checking account (Izzo et al. 2000)*. Future group design research on independent living would benefit from the development of a composite scale from these disparate items that could be used consistently in the field.

- **To the extent possible, researchers studying the effects of transition programs should include any information they may be able to access on "treatment integrity" or "treatment fidelity."** We are fully aware that much of the research on post-high school outcomes associated with school-based transition programs has used *ex post facto* designs that leave follow-up researchers with little or no capacity to assess treatment integrity directly. When designing future studies, it would be useful for transition researchers to build into group design studies measures of treatment integrity and collection and reporting of more details about transition programs including treatment components, their implementation, and their consistency and integrity. Such information adds precision to our understanding of variations in outcome effects across programs that appear to be similar but have substantive differences in features or implementation processes.

Fortunately, several important evaluations on the horizon may address many of the aforementioned research design considerations and could provide useful information for policymakers and researchers who have an interest in this transition area. These studies of several demonstration programs funded by the U.S. Social Security Administration were too early into data collection and analysis to provide information on the effects of the programs on post-high school outcomes.[29] We mention them here, however, for several reasons: (a) these study designs provide examples of how to do the difficult work of conducting random assignment of students with disabilities and tracking their outcomes in secondary school and beyond; (b) early reports from these projects have also focused on issues of implementation fidelity; and (c) follow-up reports from these large-scale evaluations hold great promise for understanding the effects of the program implemented as well as the differential effects of varying components on students with different types of disabilities and in a number of implementation contexts. These types of next-generation research studies in special education transition may inform local policymakers about sequencing and intensity of program components and how these components vary in their effect on different transition outcome domains. These studies, coupled with the existing empirical research base, may also provide transition researchers and program designers with the tools to put together logic models that can clarify the strength and directionalities of program components and other inputs to different transition outcomes.

[29] See the three Fraker et al studies from 2011. While these interim reports had 12-month impact analyses, they had to be excluded because significant numbers of each of the randomly selected treatment members were still in school and still eligible for and receiving services at the time of the 12-month follow-up measurement.

Appendix A. Evidence review protocol

Overview

Evidence review focus

This evidence review focuses on school-based programs (strategies, interventions, or sets of services) designed for use with adolescents of secondary age with differing types of disabilities. This review focuses only on studies that examine programs that are implemented (a) in a school; (b) in other locations (such as community, home, or job sites) if implemented under the direction of, or in collaboration with, a school program funded through the Individuals with Disabilities Education and Improvement Act (IDEA); or (c) by a researcher in a clinical or other nonschool setting if the program could clearly be used in a typical school-age program, and the outcome measurement is aligned with one or more of the outcomes of interest in this review, as determined by the principal investigator (PI). Outcome domains include employment, postsecondary education, and independent living.

The review of evidence in this topic area seeks to address the following central questions:

- Which programs improve the employment, postsecondary education, or independent living outcomes for adolescents with disabilities?
- How does the effectiveness of differing secondary programs for adolescents with all types of disabilities vary by outcome domain?
- How does the effectiveness of differing secondary programs for adolescents with all types of disabilities vary by type and severity of disability?

Based on the availability of evidence, the review report may describe additional subgroup analyses, such as by secondary school level (for example, middle versus high schools), gender, or the setting in which program was delivered (for example, in school versus in the community). In addition, if a sufficient number of studies are available, the evidence review report may examine differences in magnitude of effect or improvement index for outcomes measured immediately after treatment (proximal effects) versus maintenance (distal) effects.

Key definitions

The evidence review uses the following disability definitions and classifications, which come from the Individuals with Disabilities Education Improvement Act of 2004 (IDEA) as they pertain to students of ages 13 to 21. IDEA is the federal law ensuring public education and related services to children and youth with a disability in the United States. Under IDEA, states and public agencies provide early intervention, special education, and related services to eligible infants, toddlers, children, and youth with disabilities. Children and youth (ages 3 to 21) receive special education and related services under IDEA, Part B. Our review includes 12 disability categories under which a student may be found eligible for special education and related services: (1) autism, (2) deaf-blindness, (3) emotional

disturbance, (4) hearing impairments, including deafness, (5) intellectual disabilities, (6) multiple disabilities, (7) orthopedic impairment, (8) other health impairment, (9) specific learning disability, (10) speech or language impairment, (11) traumatic brain injury, and (12) visual impairment, including blindness.

Consistent with the IDEA eligibility categories, this review examines the evidence for programs developed for and implemented with students served under *any* of the preceding categorical areas. Coupled with the IDEA definitions that follow are additional commonly used terms for specific disabilities. For example, *hard of hearing* is widely used in the deaf community and is therefore included in our search terms.

Table A1. Definitions of disability classifications based on IDEA

Classification	Definition
Autism, or autistic disorder	A developmental disability significantly affecting verbal and non-verbal communication and social interaction, generally evident before age 3, that adversely affects a child's educational performance. Autism is also associated with repetitive activities and stereotyped movements, resistance to environmental change or changes in daily schedules/routines, and atypical responses to sensory input. Related terms include *Autism Spectrum Disorders, Asperger's syndrome or disorder, Rett's disorder or syndrome, Pervasive Developmental Disorders – Not Otherwise Specified* (PDD-NOS), and *childhood disintegrative disorder.*
Deaf-blindness	Concomitant hearing and visual impairments, the combination of which causes such severe communication and other developmental and educational needs that they cannot be accommodated in special education programs solely for children with either deafness or blindness.
Emotional disturbance	A condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree that adversely affects a child's educational performance: (a) an inability to learn that cannot be explained by intellectual, sensory, or health factors; (b) an inability to build or maintain satisfactory interpersonal relationships with peers and teachers; (c) inappropriate types of behavior or feelings under normal circumstances; (d) a general pervasive mood of unhappiness or depression; and/or (e) a tendency to develop physical symptoms or fears associated with personal or school problems. Related terms in the literature are *emotional disorder* and *behavior disorder*.
Hearing impairments, including deafness	An impairment in hearing, whether permanent or fluctuating, that adversely affects a child's educational performance but does not meet the criteria for deafness. Related terms include *hard of hearing*. The term *deafness* refers to a hearing impairment that is so severe that the child is impaired in processing linguistic information through hearing, with or without amplification, and this impairment adversely affects the child's educational performance.
Intellectual disabilities	Significantly sub-average general intellectual functioning, existing concurrently with deficits in adaptive behavior and manifested during the developmental period that adversely affects the child's educational performance. These terms can also be found in the research literature: *mental retardation*, *intellectual disability*, *severe disability, developmental disability*, or *cognitive disability*.
Multiple disabilities	Concomitant impairments (intellectual disabilities-blindness, intellectual disabilities-orthopedic impairment, etc.), whose combination causes such severe educational needs that they cannot be accommodated in special education programs solely for one of the impairments. This does not include deaf-blindness.
Orthopedic impairment	A severe orthopedic impairment that adversely affects a child's educational performance. The term includes impairments caused by congenital anomaly (e.g., club foot, absence of limb), impairments caused by disease (e.g., poliomyelitis, bone tuberculosis), and impairments from other causes (e.g., *cerebral palsy,* amputations, and fractures or burns that cause contractures). The literature cross-references the term *physical disabilities*.
Other health impairment	Limited strength, vitality, or alertness, including a heightened alertness to environmental stimuli, that results in limited alertness with respect to the educational environment, that (a) is due to chronic or acute health problems such as *asthma*, *attention deficit disorder* or *attention deficit hyperactive disorder*, *diabetes*, *epilepsy*, a heart condition, hemophilia, lead poisoning, *leukemia*, nephritis, rheumatic fever, sickle cell anemia, and *Tourette syndrome;* and (b) adversely affects a child's educational performance.
Specific learning disability	A disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as *perceptual disabilities*, brain injury, *minimal brain dysfunction*, *dyslexia*, and *developmental aphasia*. The term does not include learning problems that are primarily the result of visual, hearing, or motor disabilities, of intellectual disabilities, of emotional disturbance, or of environmental, cultural, or economic disadvantage.
Speech or language impairment	A communication disorder, such as *stuttering*, impaired articulation, language impairment, or a voice impairment, that adversely affects a child's educational performance.
Traumatic brain injury	An acquired injury to the brain caused by an external physical force, resulting in total or partial functional disability or psychosocial impairment or both, that adversely affects a child's educational performance. The term applies to open or closed head injuries resulting in impairments in one or more areas, such as cognition; language; memory; attention; reasoning; abstract thinking; judgment; problem-solving; sensory, perceptual, and motor abilities; psychosocial behavior; physical functions; information processing; and speech. The term does not apply to brain injuries that are congenital or degenerative, or to brain injuries induced by birth trauma.
Visual impairments, including blindness	Impairment in vision that, even with correction, adversely a child's educational performance. The term includes both partial sight and blindness; a related term in the literature is *low vision*.

Source: Individuals with Disabilities Education Improvement Act, 34 C.F.R. Part 300 § 300.8 (C). September 17, 2011.

Eligibility criteria

Sample

This review includes students of ages 13 to 21 with disabilities, who are in secondary school settings, and who are eligible for special education services under IDEA. These settings must be in the United States or its territories or in U.S. schools abroad (for example, military or diplomatic schools). If a term other than the disability designations under IDEA (autism, deaf-blindness, deafness, developmental delay, emotional disturbance, hearing impairment, intellectual disability, multiple disabilities, orthopedic impairment, other health impairment, specific learning disability, speech or language impairment, traumatic brain injury, visual impairment/blindness) is used to describe students in the sample, sufficient information must be provided to confirm alignment with IDEA definitions. If a program appears to be designed for students with disabilities but the study does not identify the population as such, per se, information about the study population will be requested from the study authors.

In studies including adolescents without disabilities, the review focuses on findings for the subgroup of students with disabilities. If only aggregate findings are available, at least 50 percent of the study sample must be composed of students with disabilities. In single-case design (SCD) research, the review focuses only on data for the individual students with disabilities. If the sample for the SCD is a group of students, then at least half the group must be students with disabilities.

Results that are available for the subgroups of students defined by disability type, age, gender, socioeconomic status, race/ethnicity, English language learners, co-morbidity, or severity of disability are documented in the evidence review report. When results are available for subgroups of settings based on the following characteristics, they are documented in the topic report: level of secondary school (middle/junior high, high school), location (urban, suburban, rural), setting (special education class, general education class, community, other), and staff education/qualifications or training (for example, certification or years of experience).

Publication date and language

The study must have been publicly released between 1991 and June 30, 2011, and obtained by the review team prior to drafting the evidence review report. This time frame was established to define a realistic scope of work for the review and to complete the review and writing work by the contract deadline. Studies must be published in English.

Outcome domains

The review includes three outcome domains. These outcome domains focus on the measurement of students' post-high school employment, pursuit of postsecondary education, and capacity for independent living as adults. For an outcome to be eligible for

the review, it must be a direct measure of skill acquisition within the three outcome domains listed above that is very likely to be transferable to adult life (for example, money counting, housekeeping). Studies are required to focus on student outcomes, not teacher outcomes or other nonstudent outcomes, such as parent or school system outcomes. Studies must include at least one relevant outcome for which adequate content validity and reliability (as defined below) have been demonstrated. Studies that focus on outcomes measured in languages other than English are excluded (for example, Spanish language skills). The following table provides examples of outcomes in each domain.

Table A2. Outcome domains considered in the evidence review and example outcomes

Domain	Outcome examples
Employment	Involvement in competitive employment, supported employment, community-based career and technical instruction or training (if connected directly to a planned, subsequent employment placement), sheltered employment or workshops, independent or self-employment, as well as specific features associated with these placements, such as earnings, hourly wages, or levels of benefits. All specific outcome placements must result in pay, with the exception of work on a family farm or other family-owned business.
Postsecondary education	Participation or matriculation in discrete classes or degree programs in community colleges, tribal colleges, or four-year colleges or universities; participation or enrollment in adult-education classes, career and technical training programs, or private certification postsecondary institutions.
Independent living	Self-care or daily living skills (for example, cooking, grooming, housekeeping, following a budget); community participation outcomes (for example, voting, using restaurants, stores, libraries, recreation centers, and community services); and finally, other types of outcomes associated with independent living, such as home ownership or home maintenance.

For studies that are clearly redundant (samples and treatments the same), the review privileges the study that reports on an outcome most proximal in time to the program. Should this study not meet standards, then the review uses the study whose outcome is the next most proximal, and continues in this way until one study meets standards or none of the studies is found to meet standards.

Programs

This review encompasses programs that have a secondary school-based component, delivered in the secondary school (by research or school staff) or coordinated and/or delivered by secondary school staff in other settings. Studies in which researchers or community personnel deliver programs outside the secondary school setting (for example, career training provided by a community organization) are not included unless the study was conducted with secondary students of ages 13 to 21 and the program clearly could be used in a typical school program, as determined by the review PI. The review does not include studies of programs conducted after the student completes grade 12, receives a GED, or graduates from high school.

Programs must have as their primary goal the acquisition of employment, education, or independent living skills upon transition from secondary education. Skill acquisition is distinguished from therapeutic outcomes (for example, range of motion) and from reductions in behavior problems without skill acquisition (for example, reduction in self-injury that does not include instruction in replacement behavior to achieve the same

function). Programs may include branded curricula and specific treatments such as time delay, peer tutoring, or computer-assisted instruction. The following table describes examples of the types of programs that are eligible for review.

Table A3. Examples of programs eligible for review by outcome domain

Domain	Program examples
Employment	Career or employment transition programs target skills and create opportunities for future employment. These programs may be aimed at improving job-specific skills (for example, technology training) or facilitating career planning and job placement.
Postsecondary education	Postsecondary education transition programs target skills for future education. These programs may be aimed at improving specific academic skills (targeting language arts, mathematics, science, or social studies) or preparing students for career and technical or postsecondary education.
Independent living	Independent living transition programs target skills for overall functioning in home, job, leisure activities, and the community. These programs may target mastering a specific activity such as crossing the street or broader adaptive behaviors such as self-determination or making choices. They must involve direct teaching and outcome measurement of the targeted skills, as opposed to teaching interventions and measuring constructs (for example, self-esteem) that are hypothesized to support independent living skills.

To be reviewed, programs for students with disabilities must be *replicable*. If the program is branded, information about how to obtain the program must be available. Studies of programs that are not branded must describe the program, including the skills being targeted, the approach to enhancing the skills, the target population, components or features of the program that were implemented, characteristics of the settings in which it was implemented, the duration and intensity of the program, and the characteristics and training of the people administering the program.

Research design

Only empirical studies using quantitative methods and inferential statistical analysis and that take the form of a randomized controlled trial (RCT), a quasi-experimental matched comparison group design (QED), or an SCD are eligible for this review. The review includes only research designs that compare the effects of a program with a baseline or business-as-usual scenario. Alternating treatment SCDs are excluded unless the design is intended to evaluate the effectiveness of a program relative to a baseline condition.

Longitudinal surveys (for example, NLTS and NLTS 2) and secondary analyses of extant data are eligible for inclusion only to the extent that the studies (a) include "program" data (for example, participated in career and technical education) that are coded dichotomously, and (b) satisfy all other conditions for methodological rigor.

Literature search strategy

The literature search strategy for this review has three components to identify all published and unpublished intervention studies that may be eligible for review. The first component is a keyword search based on parameters developed by the review team. The second component is a comprehensive grey literature search. The third component is a hand search of random issues from top field journals. Each type of search is described below.

Keyword search

The primary objective of the keyword search is to identify potentially eligible studies on interventions for secondary students with disabilities and assess the likely extent of studies on each intervention, so that program categories can be developed. The focus is on breadth rather than depth. After the keyword search, subsequent searches will focus on the selected interventions and will be designed to capture all potentially eligible studies, including any that the keyword search did not identify.

The following keywords are meant to capture literature that falls within the scope of the protocol. Given the objective stated above, targeted outcomes and study design terms are included to focus the search on identifying literature that will support the review report. The keyword list is followed by a list of databases to be searched. These keywords were developed by the review team and may be modified or supplemented by library staff if it is determined that other keywords would better target relevant literature. An asterisk means that all words starting with the letters preceding the asterisk are included as terms.

Disability: "disab*" OR **"**autis*" OR "Asperger*" OR "Rett*" OR "pervasive developmental disorder" OR "childhood disintegrative disorder" OR "asthma" OR "attention deficit disorder" OR "attention deficit hyperactive disorder" OR "deaf-blindness" OR "emotional disturb*" OR "behavior disorder" OR "hearing impair*" OR "deaf" OR "hard of hearing" OR "mental retardation" OR "intellectual disability" OR cognitive disability OR developmental disability OR physical disability OR "Down* syndrome" OR "multiple disabilities" OR orthopedic impair*" OR "cerebral palsy" or "spina bifida" OR "other health impair*" OR "cancer" OR "epilepsy" OR "specific learning disability" OR "dyslexia" OR "aphasia" OR "minimal brain dysfunction" OR "auditory processing disorder" OR "perceptual disorder" OR "stutter*" OR "speech or language impair*" OR "traumatic brain injury" OR "visual impair*" OR "blind" OR "low vision" OR "special education."

AND

Population: adolescent* OR child* OR "young adult*" OR youth* OR student* OR teen* OR "transition age"

AND

Outcome: "post-school" OR "post-secondary" OR "postsecondary" OR "higher education" OR "bachelor's degree" OR college OR university OR "associate's degree" OR vocational OR "certificate program" OR employment OR career* OR "independent living" OR "quality of life" OR "school to work" OR "school-to-work" OR "transition" OR "job*" OR "work*" OR ("relationship?" OR "friend")

AND

Program: (career n3 awareness) OR "community services" OR "exit exam" OR inclusion OR "general education" OR "interagency collaboration" OR "occupational course*" OR "unpaid employment" OR "paid employment" OR "work experience" OR "parental involvement" OR "program* of study" OR "self-advocacy" OR "self-determination" OR "self-care" OR "social skills" OR "independent living" OR "student support*" OR "transition* program*" OR "vocational education" OR "special education" OR "best practice*" OR "school counsel*" OR "dropout prevention" OR "job coach*" OR "supported employment" OR (("community based" OR "community-based") AND (instruction OR training)) OR "behavior management" OR "behavior support" OR "inclusive education" OR "assistive technology" OR "speech therapy" OR "vocational rehabilitation" OR "self-directed IEP" OR "IEP*" OR "individual* education program*" OR "transition planning" OR "family involvement" OR "school experience*" OR "work-awareness" OR "work awareness" OR "student-focused planning"

AND

Research design: (control OR comparison OR matched) AND group) or treatment OR random* OR assignment OR baseline OR experiment* OR evaluation* OR impact* OR effectiveness OR causal OR posttest OR "post-test" OR pretest OR "pre-test" OR "random* control trial*" OR "RCT" OR "quasi-experimental" OR "QED" OR "regression discontinuity" OR "RDD" OR "multiple baseline" OR "multiple probe" OR "single case" OR "single subject" OR "ABAB" OR "alternating treatment" OR "meta-analysis" OR "meta analysis" OR "reversal design" OR "withdrawal design" OR descriptive OR correlation* OR "pre-post" OR predictor OR predictive OR quantitative.

The keyword search includes the following electronic databases: ERIC, PsycINFO, Campbell Collaboration, Academic Search Premier, EconLit, Business Source Corporate, SocINDEX with Full Text, EJS E-Journals, Education Research Complete, WorldCat, Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, Cochrane Methodology Register, What Works Clearinghouse, EPPI-Centre, and Pro-Quest.

Fugitive or grey literature search

In addition to the keyword search, the review team seeks to identify relevant unpublished studies through the following approaches:

- Searching institutional repositories such as OpenDOAR: The Directory of Open Access Repositories. OpenDOAR is a comprehensive international directory of academic open-access sources, of which 23 percent are North American. Publications are typically a mixture of published and grey literature.

- Searching conference abstracts and proceedings. According to the Campbell Collaboration, more than one-half of studies presented at academic conferences remain unpublished. For this review, the abstracts and/or proceedings of four major education/disability conferences are searched for program studies pertinent to this review. In some cases (for example, AERA 2010 and 2011), full-text papers are available to conference attendees on the conference website. When full-text papers are unavailable, authors are contacted to provide full-text versions of papers that pass an initial abstract screen. The review team conducts a search of the conference abstracts and proceedings for the following associations:

    - American Educational Research Association (AERA)
    - Council for Exceptional Children (CEC)
    - Division on Career Development and Transition (CEC-DCDT)
    - TASH (formerly, the Association for Persons with Severe Handicaps)

- Searching unpublished literature databases.
- PAIS International: This ProQuest database includes published and unpublished literature from more than 120 countries. It is recommended by the Campbell Collaboration as a reliable search source for unpublished literature.
- Other sources:

    - Soliciting unpublished work from key researchers identified by the review team
    - Checking references of prior reviews, research syntheses, and studies included in final review (using the reference lists of prior reviews, research syntheses, and included studies to ensure that key studies have not been omitted)
    - Searching websites of all the developers of relevant programs or practices for any research or implementation reports.
    - Checking websites summarizing research on programs for children and students (see Table A4)

Table A4. Websites included in the fugitive or grey literature search

Website	URL
Alliance for Excellent Education	http://www.all4ed.org/
American Speech-Language-Hearing Association	http://www.asha.org/
American Youth Policy Forum	http://www.aypf.org/
Appalachian Education Laboratory	http://www.cna.org/centers/education/rel/
ASCD	http://www.ascd.org
American Council on Education	http://www.acenet.edu
American Institutes for Research	http://www.air.org
Brown Center on Education Policy at Brookings	http://www.brookings.edu/brown.aspx
Carnegie Corporation of New York	http://carnegie.org/
Center for Public Education	http://www.centerforpubliceducation.org/
Center for Research on Learning and Training	http://www.crlt.umich.edu/index.php
Center on Education Policy	http://www.cep-dc.org/
Center on Instruction	http://www.centeroninstruction.org/
Chaplin Hall Center for Children	http://www.chapinhall.org/
Congressional Research Service	http://www.loc.gov/crsinfo/
Council for Learning Disabilities	http://www.cldinternational.org/
Consortium for Policy Research in Education	http://cpre.wceruw.org/
Educause	http://www.educause.edu
IES	http://ies.ed.gov/
Learning Disabilities Association of America	http://www.ldanatl.org/
Idonline	http://www.ldonline.org
MDRC	http://www.mdrc.org
Mid-continent Research for Education and Learning	http://www.mcrel.org/
National Center for Learning Disabilities	http://www.ncld.org/
National Center on Secondary Education and Transition	http://www.ncset.org/
Pacific Resources for Education and Learning	http://www.prel.org/
Pathways to College Network	http://www.pathwaystocollege.net/
RAND	http://www.rand.org
Regional Education Laboratories	http://ies.ed.gov/ncee/edlabs/
SSRN	http://www.ssrn.com
Urban Institute	http://www.urban.org/
United States Department of Education	http://www.ed.gov

Hand search of issues from disability journals

The review team will conduct a random search, stratified by year or blocks of years, of issues from 10 top disability journals.

Screening strategy

References captured through the literature search were de-duplicated and then screened for eligibility in two stages. First, an abstract-level screen will assess whether references are for quantitative studies, studies that examine a program or outcome germane to the review, and studies that include students with disabilities within the relevant age range. Full-text articles will be obtained for the remaining studies following the abstract-level screen. The second stage of the screening process will be based on the full text. Because abstracts do not always provide enough information to determine whether a study is eligible for review, the full-text screen will apply the same screening criteria. The full-text screen will also assess whether studies use an eligible quantitative research design (that is, an RCT, QED, or SCD). Remaining studies following the full-text screen will be reviewed against WWC evidence standards.

WWC evidence standards for group designs and pilot standards for SCDs

This review applies the WWC evidence standards and WWC pilot standards for SCDs as documented in the *Procedures and Standards Handbook, Version 2.1* (U.S. Department of Education, 2011). The following sections describe the application of WWC standards in this review and any additional requirements not specifically addressed in the *Handbook* that are used for rating evaluating studies (for example, measures used for demonstrating baseline equivalence, additional requirements for multiple probe SCDs, and criteria for exploratory evidence). All standards apply to overall findings as well as to analyses of subsamples. Table A.5 briefly describes the categories of evidence according to the WWC. RCT and QED studies that do not meet WWC evidence standards in this review may be included as providing exploratory evidence by satisfying weaker criteria developed by the research team.

Table A5. WWC standards ratings and criteria

Study rating	Designs eligible for rating	Criteria
Meets WWC evidence standards without reservations	RCT	A well-implemented RCT with low attrition.
Meets WWC evidence standards with reservations	QED, RCT	A QED or a RCT with high attrition that has established equivalence of the analytic samples.
Meets WWC pilot single-case design standards without reservations	SCD	A well-implemented SCD with at least five data points per phase.
Meets WWC pilot single-case design standards with reservations	SCD	A well-implemented SCD with at least three data points per phase.

Note: The term well-implemented means that a study design satisfies other requirements that protect against biased inferences, such as the presence of confounding factors, for example. The review process did not assess implementation fidelity or intervention feasibility directly because the WWC does not have explicit standards for these factors.

Attrition in RCTs

Systematic reviews such as this one must be concerned about overall and differential attrition from the intervention and comparison groups for RCTs, as both contribute to the potential bias of the estimated effect of an intervention. Only RCTs with low attrition can meet WWC evidence standards without reservations. The attrition bias model developed by the WWC will be used in determining whether a study meets evidence standards for this review (see Appendix A of the *Handbook*).

When the combination of overall and differential attrition rates causes an RCT study to meet the liberal attrition standard (illustrated heuristically by the green and white areas on the diagram shown below), the attrition will be considered "low" and the level of bias acceptable. However, for RCTs with combinations of overall and differential attrition rates in the red area, the attrition will be considered "high" with potentially high levels of bias and, therefore, must demonstrate equivalence. This reflects the reasonably high probability that attrition is due directly to student choices to withdraw from treatment conditions or outcome measurement.

Many studies reviewed by the WWC are based on designs with multiple levels. Bias can be generated not only from the loss of clusters (such as schools), but also from sample members within the clusters (such as students), if those sample members withdraw because of their treatment status. The attrition standard applies to both levels. To meet the standard, a study must first pass at the cluster level, using the designated attrition boundary. Second, the study must pass at the subcluster level, using the same attrition boundary, with attrition based only on the clusters still in the sample. That is, the denominator for the subcluster attrition calculation includes only sample members at schools or classrooms that remain in the study after cluster attrition.

Figure A1. Acceptable and unacceptable combinations of overall and differential attrition

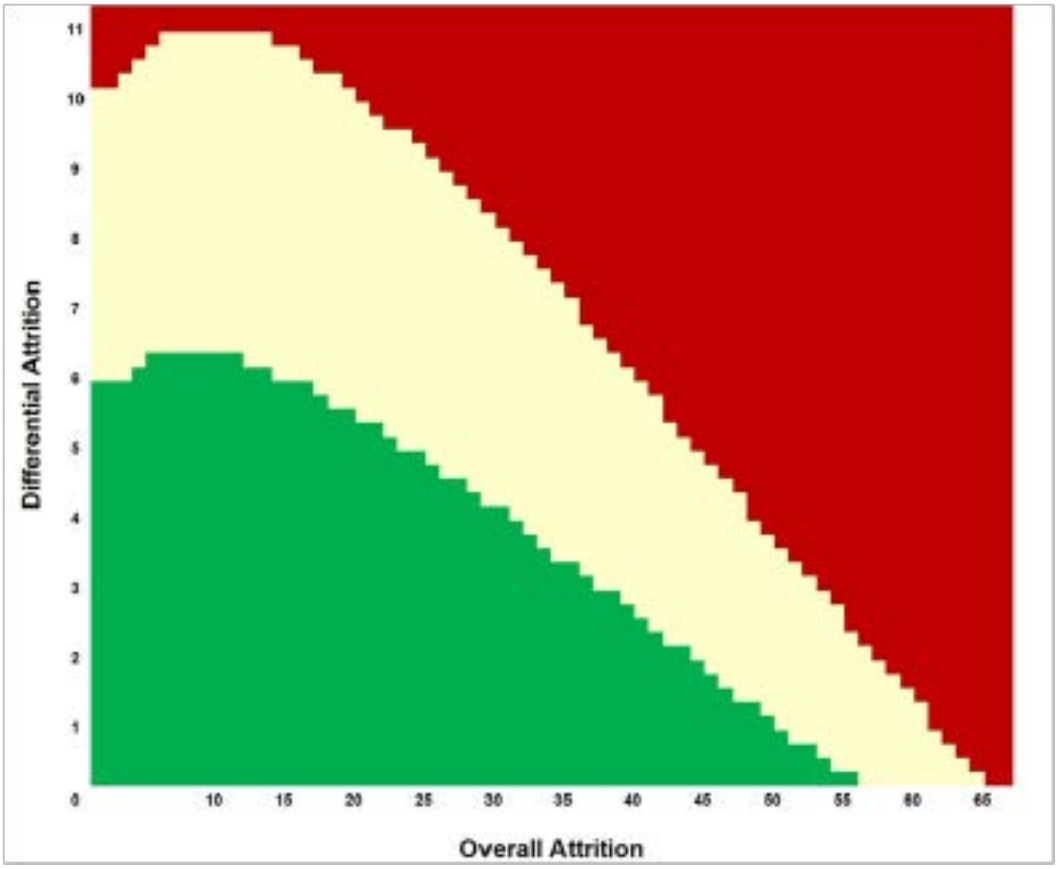


Table A6. Highest level of differential attrition allowable to meet the attrition standard based on the liberal attrition standard

Overall attrition	Allowable differential attrition	Overall attrition	Allowable differential attrition	Overall attrition	Allowable differential attrition	Overall attrition	Allowable differential attrition
0	10.0	17	10.5	34	7.4	51	3.5
1	10.1	18	10.3	35	7.2	52	3.2
2	10.2	19	10.2	36	7.0	53	3.0
3	10.3	20	10.0	37	6.7	54	2.8
4	10.4	21	9.9	38	6.5	55	2.6
5	10.5	22	9.7	39	6.3	56	2.3
6	10.7	23	9.5	40	6.0	57	2.1
7	10.8	24	9.4	41	5.8	58	1.9
8	10.9	25	9.2	42	5.6	59	1.6
9	10.9	26	9.0	43	5.3	60	1.4
10	10.9	27	8.8	44	5.1	61	1.1
11	10.9	28	8.6	45	4.9	62	0.9
12	10.9	29	8.4	46	4.6	63	0.7
13	10.8	30	8.2	47	4.4	64	0.5
14	10.8	31	8.0	48	4.2	65	0.3
15	10.7	32	7.8	49	3.9	66	0.0
16	10.6	33	7.6	50	3.7	67	-

Group equivalence in RCTs with high attrition and QEDs

To meet WWC evidence standards with reservations

If the study design is a RCT with high levels of attrition, or a QED, the study must demonstrate baseline equivalence of the intervention and comparison groups on relevant or correlated factors for the analytic sample to meet WWC evidence standards with reservations. The onus for demonstrating equivalence on important/relevant factors in these studies rests with the authors of the study. Sufficient reporting of preintervention data should be included in the study report (or obtained from the study authors) to allow the review team to draw conclusions about the equivalence of the intervention and comparison groups. In comparison group studies that include both students with and without disabilities, the intervention and comparison groups must include similar percentages of students with disabilities, as determined by the PI. For this review, it is possible for a study to meet WWC evidence standards with reservations in one outcome domain and not in another. Thus, rules for establishing baseline equivalence should be applied within each domain.

Eligible studies for which the equivalence standard applies must demonstrate equivalence on the type or severity of disability and on the factors listed in one of the following two options:

- Option 1: a pre-intervention measure of the outcome variable
- Option 2: gender and either race/ethnicity or a measure of low socioeconomic status

Preintervention measures of the outcome are the preferred variables for demonstrating baseline equivalence for the WWC. This review provides for a second option because preintervention measures cannot be obtained for many outcomes in the postsecondary

education and independent living domains. In the employment domain, studies rarely report preintervention measures such as high school employment. Studies that demonstrate equivalence through one set of variables (for example, option 1) but also present evidence of significant differences across treatment and comparison groups with respect to the other set of variables (for example, option 2) do not meet WWC evidence standards. Studies that include only students of one disability category or level of severity meet the requirement to demonstrate equivalence on type or severity of disability.

Groups are considered equivalent if the reported differences in their preintervention characteristics are less than or equal to 5 percent of the pooled standard deviation in the sample, regardless of statistical significance. If differences are greater than 5 percent and less than or equal to 25 percent of the pooled standard deviation in the sample, the analysis must control analytically for the individual-level preintervention characteristics on which the groups differ. If preintervention differences are greater than 0.25 for *any* of the relevant pre-intervention characteristics, the study does not meet standards for that domain.

To meet review criteria for exploratory evidence

Studies that do not meet the standard for demonstrating baseline equivalence but nevertheless have many desirable research design qualities are included in the report as exploratory evidence. To be included at this level, a study must meet the following criteria:

- Study is an RCT with high attrition or a QED that meets all requirements for a rating of "meets WWC evidence standards with reservations" except for the requirement to demonstrate baseline equivalence.
- Study meets one of three following conditions:

    1. It provides information about baseline equivalence on severity of disability (reported differences between groups are less than 25 percent of the pooled standard deviation in the sample, the sample contains only one disability severity and/or type, or case-by-case matching is used to create comparison groups), and it includes statistical controls in the analysis if reported differences between groups are between 5 percent and 25 percent of the sample pooled standard deviation (criteria is ignored if only one disability type/severity is used or matching is used).
    2. Lacking both of the above, it contains a sample drawn from a large database in which it is likely that groups are similar with respect to the characteristics indicated above.
    3. Study reports bivariate correlations between the treatment variable and both severity/type of disability and one additional demographic variable relevant for the review that are close to zero (-0.05 to 0.05), AND those demographics are controlled for in the analyses.

Psychometric properties for outcome measures

Eligible outcomes must be in a relevant domain, cannot be overaligned with the intervention, and must meet standards of reliability and validity.

Alignment

Outcome measures that are closely aligned or actually tailored to the intervention are likely to demonstrate larger effect sizes than those that are less closely aligned with the intervention when a group comparison is used. When the outcome measure (a) includes some of the same materials (such as books or passages) that are used in the intervention, or (b) is administered to the intervention group as part of the intervention, it is generally considered to be overaligned. Outcome measures that are overaligned with the intervention in group comparison studies will not be included in determining intervention ratings for this review. Although this phenomenon occurs frequently in studies of the effects of interventions on academic achievement in schools, particularly when the outcome measure is administered immediately at the end of the intervention period, it is likely to occur less frequently in this review, a result of the chronological separation between intervention delivery and post-high school outcome measurement.

When the intervention involves placement and training, with support, in community-based competitive or sheltered employment settings (for example, supported employment), actual placement in competitive employment immediately upon graduation from high school will be considered overaligned. However, a more distal measure of placement in either competitive or sheltered employment (for example, a year after graduation from high school) will be considered a measure of maintenance of effect and will be included as an appropriate outcome measure, as would less aligned measures, such as proportion of time employed in competitive employment in the year after graduation, or annual earnings in the year after graduation. Similarly, with postsecondary education, if the intervention involves adding a member from a postsecondary institution on a student's transition-planning Individualized Education Program (IEP), an outcome measure of application to, or acceptance in, that IEP member's postsecondary institution will be considered overaligned. As with employment measures, however, more distal measurement timing will result in including the studies in the review.

Reliability and validity

The study must include at least one former student outcome measure with evidence of face validity and, to the extent the outcome is measured on a scale basis, some estimate of the reliability of the outcome. Standardized assessment measures are assumed to have adequate reliability and validity, but most outcomes measures in this review will not be tests but rather simple percentages (for example, placement in competitive employment, or maintenance in employment setting), wage- or benefit-level scale data, binomial outcomes (for example, enrollment in postsecondary institution), or observational data. In those

cases, the review team will determine the reliability of outcomes based on the following table.

Table A7. Reliability requirements for nonstandardized outcome measures by design type

Design	Reliability requirements
Group design (RCT, QED)	Inter-rater reliability is at least 0.50 using percentage of agreement, correlational, or Kappa statistics, OR Internal consistency score reliability is at least 0.60, OR Temporal stability/test-retest score reliability is at least 0.40.
Single-case design	Inter-rater reliability is at least 0.80 using percentage of agreement or 0.60 using Kappa statistics, AND The outcome is measured systematically over time by more than one assessor, with inter-assessor agreement collected in all phases and in at least 20 percent of all sessions (total across phases) for a condition (for example, baseline, intervention). The review will include and indicate with a footnote studies where inter-assessor agreement was collected on at least 20 percent of sessions but where it is not clear from the study text whether 20 percent of sessions in each condition included assessment, provided that all other design requirements are satisfied.

If the research article does not specify the score reliability of each outcome measure, data from the publisher of the test or scale or other sources may be used to establish the score reliability of an outcome measure for the study population. If studies did not analyze the score reliability of outcome measures using study data, and analyses by test publishers or other researchers did not include adolescents with disabilities, any other available evidence of score reliability and validity of the measure for the study population will be considered, and a decision about the adequacy of the outcome measure will be made on a case-by-case basis in consultation with experts.

If an outcome measure is composed of different tests for different adolescents in the sample, it will be considered a valid outcome if it meets the following criteria:

- The measure purports to estimate a similar construct and was standardized on a similar population, as reflected in the test manual or empirical studies focused on the test.
- The test must meet the thresholds for reliability described above.
- There must be clear rules for which test is administered to which individual, and the rules must be applied in the same way to the treatment and control groups.
- The distribution of tests administered at baseline to the treatment and control groups must be similar.

If information necessary to apply these criteria is not available in the article, an author query will be initiated to obtain the information.

Statistical and analytic issues

RCTs with low attrition do not need to use statistical controls in their analyses, though statistical adjustment for well-implemented RCTs is permissible and can help generate more precise estimates of effect size. For RCTs, such estimates will be adjusted for differences in preintervention characteristics at baseline (if available) using a difference-in-differences method if the authors did not adjust for pretest (see Appendix B in version 2.1 of the *Handbook*). Beyond the preintervention characteristics required by the equivalence standard, statistical adjustment can be made for other measures in the analysis, though they are not required.

For this systematic review, the preference will be to report on and calculate effect sizes for postintervention means adjusted for the preintervention measure. If a study reports both unadjusted and adjusted postintervention means, we will report the adjusted means and unadjusted standard deviations. If adjusted postintervention means are not reported, they will be requested from the authors.

The p-value of group differences will be recalculated if (a) the study authors did not calculate statistical significance, (b) the study authors did not account for clustering when there is a mismatch between the unit of assignment and unit of analysis, or (c) the review team implements a difference-in-differences adjustment with an RCT. Otherwise, the review team will accept the calculations provided in the study. For assessing statistical significance, the p-value will be adjusted for multiple comparisons when appropriate.

When a misaligned analysis is reported (that is, the unit of analysis is not the same as the unit of assignment), and the author is unable to provide a corrected analysis, the effect sizes we compute will incorporate a statistical adjustment for clustering. The default intraclass correlations used for this review are 0.20 for cognitive outcomes (in the rare instances where we encounter such an outcome), and 0.10 for outcomes in social-emotional development, behavior, and functional ability. For an explanation of the clustering correction, see Appendix C in version 2.1 of the *Handbook*.

When multiple comparisons are made (multiple outcome measures are assessed within an outcome domain in one study) and not accounted for by the authors, we will account for the multiplicity by adjusting the reported statistical significance of the effect using the Benjamini-Hochberg correction. See Appendix D in version 2.1 of the *Handbook* for the formulas the WWC uses to adjust for multiple comparisons.

Single-case designs

The review will include the following types of SCDs: reversal/withdrawal designs, changing-criterion designs, multiple-baseline designs (across participants or conditions), and multiple-probe designs (across participants or conditions). The following criteria apply for SCD research studies:

- The independent variable (the intervention) must be systematically manipulated, with the researcher determining when and how the independent variable conditions change.
- The study must include at least three attempts to demonstrate an intervention effect or shifts in the criterion at three different points in time.
- Phases used as part of the three attempts to demonstrate an intervention effect or shift the criterion at three different points in time must have:

    - At least three data points each for the design to meet WWC pilot standards for single-case designs with reservations.
    - At least five data points each for the design to meet WWC pilot standards for single-case designs without reservations.

- Multiple-probe designs need to meet additional criteria[30] in this review because some data points are intentionally missing, and as a result the timing of when data are collected is an important determinant of whether the study designs meet pilot SCD standards:

    - The design must include three consecutive probe points for each case at the start of the study to meet WWC pilot SCD standards without reservations. To meet WWC pilot SCD standards with reservations, the design must include at least one probe point for each case within the first three sessions.
    - The data pattern for each case must include three consecutive probe points just prior to introducing the independent variable to meet pilot SCD standards without reservations. The data pattern for each case must include one probe point just prior to introducing the independent variable to meet pilot SCD standards with reservations.
    - Whenever the intervention is introduced, the data pattern for subsequent participants (or conditions) must include a probe point prior to the first intervention session or once the pre-specified intervention criterion is reached.
    - Probe points must be available at least every eight preintervention sessions in multiple-probe (days) designs.

- For this review, there may be occasions when fewer than three data points in a phase will not require the study to be rated as not meeting WWC pilot standards for single-case designs. The following are exceptions:

[30] To include multiple probe designs, we used additional criteria that were being developed by the WWC at the time when we conducted study reviews. Pilot standards that are specific to multiple probe designs were not included in Version 2.1 of the *Handbook*. Official pilot standards for multiple probe designs are planned for the next version of the *Handbook*.

- Interventions for severe problem behavior, such as aggression and self-injury, for which extended baselines or reversal conditions pose serious ethical and procedural concerns.
- Interventions on "zero baseline" behaviors, when there is no logical reason to believe that further assessment would yield other than zero baseline performance. In such cases, a multiple-probe design may be used to alleviate potential "punishing" effects of repeated failure experiences.

- The data patterns for SCDs that meet pilot standards with or without reservations are examined by visual analysis and rated as providing strong evidence, moderate evidence, or no evidence of a causal relationship. Visual analysis accounts for the level, trend, and variability of data, as well as the percentage of overlap, the immediacy of effect, and the degree of consistency across similar phases. Strong evidence is defined as the demonstration of at least three intervention effects and no non-effects. Moderate evidence is defined as the demonstration of at least three intervention effects but also the existence of at least one non-effect. No evidence is defined as the lack of demonstration of at least three intervention effects.
- Single-case design studies can contribute toward a rating of effectiveness for a program in this review, but the effects of a program based on SCDs will be evaluated separately from the effects of a program based on RCTs and QEDs. We developed SCD rating categories for use in this review because the WWC has not established procedures for rating the effectiveness of interventions based on a large body of SCD research. Consistent with WWC procedures, SCD research studies meeting standards must collectively satisfy the WWC reporting requirements to be eligible for an effectiveness rating (see Table C2).

Appendix B. Dispositions for studies that do not meet WWC standards

Table B1. Dispositions for group design studies that do not meet WWC standards

Short citation	Baseline equivalence disposition	Other disposition (if any)
Baer et al. (2003)	Not established for severity/type of disability	
Benz et al. (1997)	Not established for severity/type of disability, gender or ethnicity	
Betz et al. (2010) [1]	Attrition exceeded WWC standards and baseline equivalence was not established for ethnicity	
Doren and Benz (1998)	Not established for severity/type of disability	
Doren et al. (2011) [1]	Not established for ethnicity	
Dunn and Schumaker (1997)	Not established for gender or ethnicity	
Farley and Johnson (1999)	Not established for severity/type of disability, gender or ethnicity	N=1 confound
Flexer et al. (2011)	Not established for gender	
Fourqurean (1991) [1]	Not established for ethnicity	
Harvey (2002) [2]	Not established for type of disability, but proxy measures deemed adequate	
Izzo (2000) [1]	Attrition exceeds WWC standards and not established for severity/type of disability, gender or ethnicity	
McDonnall (2009)	Not established for severity of disability, gender or ethnicity	
Newman and Cemeto (1993)	Not established for severity/type of disability, gender or ethnicity	
Rabren et al. (2002)	Not established for severity/type of disability, gender, or ethnicity	
Shandra and Hogan (2008) [2]	Not established for severity/type of disability, gender, or ethnicity	
Siegel et al. (1992)	Not established for gender or ethnicity	Intervention is combined with another intervention
Wagner and Blackorby (1996) [2]	Not established for severity/type of disability, gender, or ethnicity	
Wagner (1993) [2]	Not established for severity/type of disability, gender, or ethnicity	
Wagner (1991) [2]	Not established for severity/type of disability, gender, or ethnicity	
Winsor et al. (2011) [1]	Not established for gender or ethnicity	
Wolffe and Kelly (2011)	Not established for severity/type of disability, gender, or ethnicity	Intervention is combined with another intervention

Note: [1]Study met exploratory standards by demonstrating baseline equivalence on type and/or severity of disability.

[2]Study met exploratory standards by containing a sample drawn from a large database in which it is likely that groups were similar with respect to severity/type of disability, gender, and ethnicity.

Table B2. Dispositions for single-case design studies that do not meet WWC pilot standards

Short citation	Inter-assessor agreement disposition	Other disposition (if any)
Ayres et al. (2006)	n.a.	Insufficient number or placement of data points
Bailey and Angell (2005)	Not established on the frequency of sessions with agreement data	Fewer than three attempts to demonstrate an intervention effect at three different points in time
Burke et al. (2010)	Not established on the frequency of sessions with agreement data	Fewer than three attempts to demonstrate an intervention effect at three different points in time; insufficient number or placement of data points
Ferguson et al. (2005)	Not established on the frequency of sessions with agreement data	n.a.
Graves (2005)	Not established on the frequency of sessions with agreement data	Fewer than three attempts to demonstrate an intervention effect at three different points in time
Haring et al. (1995)	Not established on the frequency of sessions with agreement data	Fewer than three attempts to demonstrate an intervention effect at three different points in time; insufficient number or placement of data points
Schloss and Alper (1997)	Not established on the frequency of sessions with agreement data	Insufficient number or placement of data points
Van Laarhoven et al. (2009)	Not established on the frequency of sessions with agreement data	Fewer than three attempts to demonstrate an intervention effect at three different points in time; insufficient number or placement of data points

Note: n.a. (not applicable) indicates there is no disposition for a study in a particular column.

Appendix C. Rating criteria for studies and programs

Table C1. Rating of effectiveness criteria for program categories based on QEDs and RCTs

Rating of effectiveness	Criteria
Positive effects based on group designs	Two or more studies show statistically significant positive effects, at least one of which met WWC evidence standards for a strong design, AND
	No studies show statistically significant or substantively important negative effects.
Potentially positive effects based on group designs	At least one study shows a statistically significant or substantively important positive effect, AND
	No studies show a statistically significant or substantively important negative effect AND fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects.
Mixed effects based on group designs	(1) At least one study shows a statistically significant or substantively important positive effect AND at least one study shows a statistically significant or substantively important negative effect, but no more such studies than the number showing a statistically significant or substantively important positive effect, OR
	(2) At least one study shows a statistically significant or substantively important effect AND more studies show an indeterminate effect than show a statistically significant or substantively important effect.
Potentially negative effects based on group designs	(1) One study shows a statistically significant or substantively important negative effect and no studies show a statistically significant or substantively important positive effect, OR
	(2) Two or more studies show statistically significant or substantively important negative effects, at least one study shows a statistically significant or substantively important positive effect, and more studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects.
Negative effects based on group designs	Two or more studies show statistically significant negative effects, at least one of which met WWC evidence standards for a strong design, AND
	No studies show statistically significant or substantively important positive effects.
No discernible effects based on group designs	None of the studies shows a statistically significant or substantively important effect, either positive or negative.

Source: U.S. Department of Education (2011).

For the effectiveness of a program category in a domain to be rated based on SCD research studies, the studies must collectively satisfy WWC reporting requirements. First, there must be at least five SCD studies examining the program that met WWC SCD standards with or without reservations. Second, these studies must have been conducted by at least three research teams with no overlapping authorship at three different institutions. Third, the combined number of participants must total at least 20. If any of these requirements are not met, then the program category is not rated based on SCD research studies.

The WWC has not established procedures for rating the effectiveness of interventions when the group of SCD research studies that meets standards satisfies the WWC reporting requirements. We modeled rating categories—described in Table C.2—after the WWC as an exploratory way to rate the effects of a program based on single-case design studies. We also sought to design ratings categories that were consistent with the WWC pilot SCD standards (U.S. Department of Education 2011) and the quality indicators proposed by Horner et al. (2005). The rating scale was reviewed by members of the expert panel that developed the WWC pilot SCD standards and by members of the project team that operates the WWC. Nevertheless, the rating scale that we use in this review is not necessarily the one that the WWC may adopt in the future.

Table C2. Rating of effectiveness criteria for program categories based on SCDs that satisfy WWC reporting requirements

Rating of effectiveness	Criteria
Positive effects based on single-case designs	Five or more studies met WWC SCD standards without reservations, AND each of these studies documents a positive functional relationship between the intervention and outcome measures for all primary designs in the studies, AND no studies document a negative functional relationship.
Potentially positive effects based on single-case designs	Three or more studies document a positive functional relationship between the intervention and outcome measures, AND no studies document a negative functional relationship.
Mixed effects based on single-case designs	(1) At least one study documents a positive functional relationship between the intervention and outcome measures AND at least one study documents a negative functional relationships, AND at least as many studies document a positive functional relationship as document a negative functional relationship, OR (2) At least one study documents a positive functional relationship AND one or more do not document a functional relationship, either positive or negative, AND no studies document a negative functional relationship.
Potentially negative effects based on single-case designs	(1) One or more studies document a negative functional relationship between the intervention and outcome measures, AND no studies document a positive functional relationship, OR (2) Two or more studies document a negative functional relationship between the intervention and outcome measures, AND one or more studies document a positive functional relationship, AND more studies document a negative than a positive functional relationship.
Negative effects based on single-case designs	Five or more studies document a negative functional relationship between the intervention and outcome measures, AND no studies document a positive functional relationship.
No discernible effects based on single-case designs	None of the studies documents a functional relationship between the intervention and outcome measures, either positive or negative.

Table C3. Extent of evidence criteria for program categories

Extent of evidence	Criteria
Medium-to-large	The domain includes more than one study, AND The domain includes more than one school, AND The domain findings are based on a total sample size of at least 350 students, OR, assuming 25 students in a class, a total of at least 14 classrooms across studies.
Small	The domain includes only one study, OR The domain includes only one school, OR The domain findings are based on a total sample size of fewer than 350 students, AND, assuming 25 students in a class, a total of fewer than 14 classrooms across studies.

Note: The extent of evidence rating applies to the entire body of research reviewed for a particular program category and does not distinguish between group design and single-case design studies.

Source: U.S. Department of Education (2011).

Appendix D. Alignment of Test, Fowler, Kohler, and Kortering (2010) taxonomy to program categories in this review

Table D1. Alignment of NSTTAC taxonomy to program categories in this review

Current review category	Example definition	NSTTAC predictor category
Career and technical education	Secondary program includes occupational/ career and technical courses	Vocational Education OR Occupational Courses
Career awareness or development programs	Career awareness skills-development or program; career awareness training within a broader transition program; transition program includes transition services and collaboration among stakeholders to help students achieve post-high school goals	Transition Program
Community-based work experience program	Organized work experiences sanctioned by student's school that may vary in terms of supervision and supports students receive at the job site, the connection to students' school curriculum, and whether credits are earned. These programs include cooperative education or other forms of work-based learning, organized work study programs, or supported employment programs.	Community Experiences OR Work Study
Employment while in school	Students have paid work while in high school	Paid Employment
Functional life-skills development	Self-care and independent living skills demonstrated in high school; responsible behaviors and successful adaptive social skills; self-advocacy and other self-determination skills (for example, decision making, goal-setting, problem solving)	Self-care/Independent Living Skills Development OR Self-advocacy/Self-determination Skills Development OR Social Skills Development
Inclusion in general education	Enrollment in regular academic courses; integration of students with disabilities; regular education courses taken in high school	Inclusion in General Education
Interagency collaboration	Collaboration among stakeholders intended to improve student post-high school success; interagency collaboration demonstrated at the student and district level	Interagency Collaboration
Student support including parent involvement	Parents engaged in IEP or transition planning; other community members engaged in transition planning; other natural supports in transition planning; students engaged in a formal or informal support network	Student Support OR Parental Involvement
Category Not Germane to Current Review		
Not applicable	Demonstrated academic achievement; secondary courses passed; graduation with diploma	Exit Exam Requirements/ High School Diploma Status

Notes: NSTTAC=National Secondary Transition Technical Assistance Center. NSTTAC categories are reported in Test et al. (2010).

Appendix E. Samples, interventions, settings, and outcomes for individual studies

Table E1. Samples, interventions, settings, and outcomes for group design studies examining outcomes in the employment domain

Sample size	Age range	Disability	Male	Race/ ethnicity	Program	Settings	Outcomes
Baer et al. (2011)—Meets WWC Evidence Standards with Reservations							
409	NR	ID (mild, moderate, and severe, and multiple disabilities)	52%	60% Caucasian	**Community-based work experience program.** Examined EMIS and school records for whether student ever participated in any work study programs	Schools	Competitive employment—working for pay 35 hours per week within one year of exiting high school
Cimera (2010)–Meets WWC Evidence Standards with Reservations							
42	NR	ADHD, autism, EBD, ID (mild to severe), PD, SI, TBI	66.7%	NR	**Community-based work experience program.** Examined IEPs to identify students who received job shadowing, job sampling, career and technical assessments, work adjustment, paid jobs, or other services in the community while in high school.	School, work site, and community	Months employed
Karpur et al. (2005)—Meets WWC Evidence Standards with Reservations							
215	18–22	EBD	76%	21% Caucasian	**Community-based work experience program.** Steps-to-Success, individually tailored program offering job training, placement, and support; career and technical education; paid and unpaid work; and counseling.	School, work site, and community	Employment status

Note: ADHD=attention deficit hyperactivity disorder; EBD=emotional/behavioral disturbance; ID=intellectual disability (some authors reported the severity of students' intellectual disability—mild, moderate, or severe); PD=physical disability; SI=sensory impairment; TBI=traumatic brain injury.

NR indicates that the age range was not reported.

[a] The study reported the frequency of inter-assessor agreement across all sessions and, not by condition. Data was collected during 25 percent of the sessions for each student, so the inter-assessor agreement standards are likely to be met, but this determination cannot be made completely.

[b] Students were described as having mild to severe retardation, but the study did not indicate the level of retardation for each student individually.

[c] The study reported data on four target behaviors: (1) social greetings and departures, (2) tasks completed or interrupted, (3) help or bathroom needed, and (4) materials needed. Data on social greetings and departures was eligible for this evidence review.

Table E2. Samples, interventions, settings, and outcomes for group design studies examining outcomes in the postsecondary education domain

Sample size	Age range	Disability	Male	Race/ ethnicity	Program	Settings	Outcomes
Karpur et al. (2005)—Meets WWC Evidence Standards with Reservations							
215	18–22	EBD	76%	21% Caucasian	**Community-based work experience program.** Steps-to-Success, individually tailored program offering job training, placement, and support; career and technical education; paid and unpaid work; and counseling	School, work site, and community	Postsecondary enrollment

Table E3. Samples, interventions, settings, and outcomes for studies examining outcomes in the independent living domain

Sample size	Age range	Disability	Male	Race/ ethnicity	Program	Settings	Outcomes
Cihak et al. (2004)—Meets WWC Pilot Single-Case Design Standards with Reservations							
5	17–19	5 (moderate ID)	5	NR	**Functional life skills**: Students were repeatedly assessed using a system of least prompts before and after engaging in different combinations of simulated-only instruction (SOI), in which photographs were used to teach the steps necessary to complete a task, and community-based instruction (CBI), training that demonstrates the steps necessary to complete a task. Students were trained using SOI, CBI, SOI and CBI on consecutive days (CCD), and SOI and CBI on the same day (CSD).	School resource classroom	Completion of steps necessary to send a fax, withdraw $20 from an ATM, make a debit card purchase, use a copy machine (percentage of steps performed independently)
DiPipi-Hoy and Jitendra (2004)—Meets WWC Pilot Single-Case Design Standards with Reservations							
3	16–20	1 (LD, cerebral palsy, seizure disorder); 1 (mild ID); 1 (moderate ID, down syndrome)	0	NR	**Functional life skills**: Students were repeatedly assessed before and after constant time-delay training on purchasing skills.	Local grocery and retail stores	Unprompted correct responses during purchases in community (percentage of responses)
Ganz and Sigafoos (2005)—Meets WWC Pilot Single-Case Design Standards with Reservations							
1[a]	20	1 (severe ID, autism, S/L, VI)	1	1 (C)	**Functional life skills**: Student was repeatedly assessed before and after training on self-monitoring, a cognitive-behavioral practice designed to modify a person's behavior.	School classroom	Number of tasks completed (for example, sorting objects into containers or by color)
Hutcherson et al. (2004)—Meets WWC Pilot Single-Case Design Standards with Reservations [b]							
4	14–16	4 (ID)	1	NR	**Functional life skills**: Students were repeatedly assessed before and after Project SHOP, computer-based instruction used to teach students how to select grocery store items.	School classroom	Grocery-store item selection (percentage of items correctly selected)

Sample size	Age range	Disability	Male	Race/ ethnicity	Program	Settings	Outcomes
Mechling and Gast (2003)—Meets WWC Pilot Single-Case Design Standards with Reservations							
2 [c]	16–18	1 (mild ID, cerebral palsy); 1 (moderate ID)	1	NR	**Functional life skills**: Students were repeatedly assessed before and after multimedia computer-based instruction with video modeling, video prompting, still photographs and constant time delay used to teach individuals to locate items in a grocery store.	Private office of principal investigator; grocery	Grocery-item selection (percentage of items located correctly)
Mechling et al. (2002) – Meets WWC Pilot Single-Case Design Standards with Reservations							
2[d]	14–17	1 (moderate ID, cerebral palsy); 1 (moderate ID, down syndrome	0	NR	**Functional life skills**: Students were repeatedly assessed before and after multimedia computer-based instruction with a system of least prompts to teach individuals to locate items in a grocery store.	School or private center; grocery	Grocery-item selection (number of items located correctly with a photo or word list)
Mechling et al. (2003)—Meets WWC Pilot Single-Case Design Standards with Reservations							
3	16–18	1 (moderate ID, Tourette syndrome); 1 (moderate ID, Ataxia); 1 (moderate ID, ADHD, microcephaly, S/L)	2	NR	**Functional life skills**: Students were repeatedly assessed before and after multimedia computer-based instruction with video modeling, video prompting, still photographs and constant time delay to teach subjects to use a debit card to make purchases at an automatic payment machine.	School	Making a debit card purchase (percentage of steps performed correctly)
Mechling et al. (2008)—Meets WWC Pilot Single-Case Design Standards with Reservations							
2[e]	19–20	1 (moderate ID, down syndrome); 1 (moderate ID)	0	NR	**Functional life skills**: Students were repeatedly assessed before and after training on using a DVD player as a self-operated prompting device to complete multistep tasks with a system of least prompts.	School-rented apartment kitchen	Cooking skills (percentage of steps completed correctly)

Sample size	Age range	Disability	Male	Race/ ethnicity	Program	Settings	Outcomes
Mitchell et al. (2000)—Meets WWC Pilot Single-Case Design Standards with Reservations							
3	14–16	2 (mild ID, epilepsy); 1 (mild ID, cerebral palsy)	1	NR	**Functional life skills**: Students were repeatedly assessed before and after participating in training on the use of an auditory prompting system (cassette player and headphones) and a system of least prompts to complete task steps	School classroom bathroom	Bathroom cleaning skills (percentage of correct responses)
Taber et al. (2003)—Meets WWC Pilot Standards for Single-Case Designs with Reservations							
6	14–18	6 (moderate ID)	3	NR	**Functional life skills**: Students were repeatedly assessed before and after training with a system of least prompts on task analysis steps for answering a cell phone and providing location information when lost or identifying when lost and using the speed-dial function of the phone and providing location information.	School	Task analysis steps completed independently (percentage of steps)
Taras et al. (1993)—Meets WWC Pilot Single-Case Design Standards with Reservations							
4[f]	19–20	1(VI); 2 (borderline ID, VI); 1 (mild ID, VI)	4	3 (C); 1 (B)	**Functional life skills**: Students were repeatedly assessed before and after multicomponent independence training, a group training technique that combines social learning components such as self-evaluation, peer evaluation and reinforcement with traditional operant procedures such as prompting and modeling.	School classroom	Task-analyzed leatherwork steps completed successfully (number of steps)
Taylor et al. (2000)—Meets WWC Pilot Single-Case Design Standards with Reservations							
4	16–20	4 (moderate ID)	3	NR	**Functional life skills**: Students were repeatedly assessed before and after participating in training that used a system of least prompts to teach individuals the steps necessary to clean laundry.	School family living classroom	Laundry skills (percentage of steps completed correctly)
Test (1993)—Meets WWC Pilot Single-Case Design Standards with Reservations [g]							
3[h]	16	3 (moderate ID)	2	1 (C); 2 (B)	**Functional life skills**: Students were repeatedly assessed before and after training on counting money by adding one additional dollar for the amount owed in cents. Students were trained to put one dollar aside for the change or "cents pile" and then count the number of whole dollars necessary for the purchase.	School library	Number of correct responses on a 15-item money probe.

Note: SCD=single-case design. ID=Intellectual Disability. Some authors reported the severity a student's retardation (borderline, mild, moderate, severe, or profound). LD=learning disability.VI=visual impairment. S/L=speech or language impairment.

[a] The study design for three additional students did not meet the WWC pilot standards for single-case designs.

[b] The study reported the mean rate (80 percent) and range (64 to 96 percent) of inter-assessor agreement across all participants during all grocery store probes. The author was unable to provide additional detail on inter-assessor agreement. It appears this is the mean rate and range overall sessions and not individual cases, so the inter-assessor agreement standards are likely to be met, but this determination cannot be made completely.

[c] One additional 12-year old student was not eligible for this evidence review.

[d] Two additional individuals, one 12 years old and one 9 years old, were not eligible for this evidence review.

[e] One additional 22-year-old individual was not eligible for this evidence review.

[f] Three additional 1st grade students were not eligible for this evidence review.

[g] This rating is based on the presentation of data across students. The alternative presentation of the data, which included a figure for each student across money groups, was not the primary analysis and was excluded from this evidence review.

[h] Two additional individuals, both adults not in secondary school, were not eligible for this evidence review.

APPENDIX F. Description of group design study findings

Table F1. Group design study findings for employment outcomes

Outcome	Intervention	Sample size	Statistics means (standard deviations), *t*-values, or odds ratio		Evidence review calculations			*p-value*
			Treatment group	Comparison group	Mean or % difference	Effect size (confidence interval)	Improvement index	
Baer et al. 2011								
Employment status	Community-based work experience program	409	Odds ratio=0.857		n.a.	–0.09 (–0.16/–0.02)	-9.1%	p=0.09
Cimera 2010								
Months Employed	Community-based work experience program	42	*t*=2.08		4.08	+0.45 (+0.02/+0.88)	17.20%	*p*=0.02
Karpur et al. 2005								
Employment Status	Community-based work experience program	215	41.9%	50.6%	–8.7%	–0.04 (–0.13/+0.05)	–1.60%	*p* = .18

Table F2. Group design study findings for postsecondary education outcomes

Outcome	Intervention	Sample size	Statistics means (standard deviations), *t*-values, or odds ratio		Evidence review calculations			*p-value*
			Treatment group	Comparison group	Mean or % difference	Effect size (confidence interval)	Improvement index	
Karpur et al. 2005								
Postsecondary education enrollment	Community-based work experience program	215	27.9%	8.7%	19.2%	+0.40 (+0.31/+0.49)	15.54%	$p<0.000$

Table F3. Single-case design study findings for independent living outcomes

Outcome	Intervention	SCD type	Sample size	Visual analysis rating from evidence review
Cihak et al. (2004)				
Sending a fax (percentage of steps performed independently)	Simulated instruction (SI) using a system of least prompts	Multiple probe across students	5	Strong (+)
Making an ATM withdrawal (percentage of steps performed independently)	Community-based instruction (CBI) using a system of least prompts	Multiple probe across students	5	Strong (+)
Making a debit card purchase (percentage of steps performed independently)	SI and CBI on consecutive school days using a system of least prompts	Multiple probe across students	5	Strong (+)
Using a copy machine (percentage of steps performed independently)	SI and CBI on the same school day using a system of least prompts	Multiple probe across students	5	Strong (+)
DiPipi-Hoy and Jitendra (2004)				
Unprompted correct responses during purchases in community (percentage of responses)	Constant time delay	Multiple probe across students	3	Strong (+)
Ganz and Sigafoos (2005)				
Number of tasks completed (for example, sorting objects into containers or by color)	Self-monitoring	Changing criterion	1	Strong (+)
Hutcherson et al. (2004)				
Grocery store item selection for Abby (percentage of items correctly selected)	Project SHOP computer-based instruction	Multiple probe across three foods	1	No evidence[a]
Grocery store item selection for Kate (percentage of items correctly selected)	Project SHOP computer-based instruction	Multiple probe across three foods	1	No evidence[a]
Grocery store item selection for Sue (percentage of items correctly selected)	Project SHOP computer-based instruction	Multiple probe across three foods	1	No evidence[a]
Grocery store item selection for Brad (percentage of items correctly selected)	Project SHOP computer-based instruction	Multiple probe across three foods	1	No evidence[a]
Mechling and Gast (2003)				
Grocery store item selection for Carol (percentage of items correctly selected in the store)	Multimedia computer-based instruction with constant time delay	Multiple probe across three food sets	1	Strong (+)
Grocery store item selection for Daryl (percentage of items correctly selected in the store)	Multimedia computer-based instruction with constant time delay	Multiple probe across three food sets	1	No evidence
Mechling et al. (2002)				
Grocery item selection for Carol (number of items located correctly with a photo list)	Multimedia computer-based instruction with a system of least prompts	Multiple probe across three sets of words	1	Moderate (+)

Outcome	Intervention	SCD type	Sample size	Visual analysis rating from evidence review
Grocery item selection for Stella (number of items located correctly with a photo list)	Multimedia computer-based instruction with a system of least prompts	Multiple probe across three sets of words	1	No evidence
Grocery item selection for Carol (number of items located correctly with a word list)	Multimedia computer-based instruction with a system of least prompts	Multiple probe across three sets of words	1	Moderate (+)
Grocery item selection for Stella (number of items located correctly with a word list)	Multimedia computer-based instruction with a system of least prompts	Multiple probe across three sets of words	1	Moderate (+)
Mechling et al. (2003)				
Making a debit card purchase (percentage of steps performed correctly)	Multimedia computer-based instruction with constant time delay	Multiple probe across students	3	Strong (+)
Mechling et al. (2008)				
Cooking skills for Sue (percentage of steps completed correctly)	Multimedia DVD-based instruction with a system of least prompts	Multiple probe across three cooking skills	1	Strong (+)
Cooking skills for Molly (percentage of steps completed correctly)	Multimedia DVD-based instruction with a system of least prompts	Multiple probe across three cooking skills	1	Strong (+)
Mitchell et al. (2000)				
Bathroom cleaning skills for Lynn (percentage of correct responses)	Auditory prompting system with a system of least prompts	Multiple probe across three cleaning skills	1	Strong (+)
Bathroom cleaning skills for Yvonne (percentage of correct responses)	Auditory prompting system with a system of least prompts	Multiple probe across three cleaning skills	1	Strong (+)
Bathroom cleaning skills for Doug (percentage of correct responses)	Auditory prompting system with a system of least prompts	Multiple probe across three cleaning skills	1	Strong (+)
Taber et al. (2003)				
Task analysis steps completed independently (percentage of steps)	Answer cell phone and provide location information with a system of least prompts	Multiple probe across students	3	Strong (+)
Task-analysis steps completed independently (percentage of steps)	Use cell phone speed dial to call for assistance with a system of least prompts	Multiple probe across students	3	Strong (+)
Taras et al. (1993)				
Task-analyzed leatherwork steps completed successfully for John (number of steps)	Multicomponent independence training	Multiple baseline across three leather items	1	Strong (+)

Outcome	Intervention	SCD type	Sample size	Visual analysis rating from evidence review
Task-analyzed leatherwork steps completed successfully for Mark (number of steps)	Multicomponent independence training	Multiple baseline across three leather items	1	No evidence
Task-analyzed leatherwork steps completed successfully for Lee (number of steps)	Multicomponent independence training	Multiple baseline across three leather items	1	No evidence
Task-analyzed leatherwork steps completed successfully for Len (number of steps)	Multicomponent independence training	Multiple baseline across three leather items	1	No evidence
Taylor et al. (2002)				
Laundry skills (percentage of steps completed correctly)	System of least prompts	Multiple probe across students	4	Strong (+)
Test et al. (1993)				
Correct responses on 15-item money-counting probes (number of responses)	One-more-than technique	Multiple probe across students	3	Strong (+)

Notes: SCD=single-case design. This table presents findings considered for the SCD rating of effectiveness. We consider strong and moderate visual analysis ratings to be indicative of functional relationships demonstrated in the study. No evidence visual analysis ratings indicate that the evidence review concluded that the data did not document three demonstrations of an intervention effect in the same direction.

[a] Evidence review rating differs from study-reported findings, which indicate that the percentage of correct responses in the community increased following intervention.

Appendix G. Narrative description of group design study findings for studies that meet WWC standards with reservations and exploratory studies

Table G1. Group design study findings for employment outcomes, by program category

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
Studies that Examine the Effects of ***CAREER AND TECHNICAL EDUCATION***					
Harvey (2002) Career and technical education—students who participated in any career and technical coursework while in high school	QED—NELS: 88 national stratified probability sample from 1,052 schools	All	Exploratory	Baseline equivalence not established for type of disability, but proxy measures deemed adequate	Positive effect on employment rates (55 percent for participants in career and technical education versus 46 percent for nonparticipants) and on annual earnings ($8,603 for participants in career and technical education versus $6,873 for nonparticipants)
Shandra and Hogan (2008) Students took sequence of courses based on an occupational goal, cooperative education, a school-sponsored enterprise, and a "tech-prep" program	QED—NLSY97 (U. S. Department of Labor national survey of 8,984 children ages 12–16 in 1997	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who participated in school-based school-to-work programs had 1.3 times the odds of having stable employment, 1.4 times the odds of full-time employment, and 1.2 times the odds of receiving insurance ($p<0.05$ for all comparisons). School-based learning had no discernible relationship to hourly compensation or sick days (p=ns for all comparisons). Participation in cooperative education is positively and significantly associated with annual income ($p<0.01$), full-time work ($p=0.05$), holding a job with employer-offered health insurance ($p=<0.05$), and the receipt of paid sick days ($p<0.05$). Participation in career major and school-sponsored enterprise is positively associated with stable employment ($p<0.01$). Participation in technical preparation is positively associated with the likelihood of full-time employment ($p<0.05$).

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
					Students who had taken career and technical education in their last year in secondary school were 9.3 percentage points more likely to be competitively employed than youth who had not taken career and technical education (p=0.05). Students whose secondary career and technical education involved work experiences were 13.9 percentage points more likely be employed (p<0.01). Students who took concentrated career and technical education classes had an employment rate that was 19.0 percentage points higher than students who did not take concentrated classes (p<0.05). Students who took survey career and technical courses had an employment rate that was 19.8 percentage points higher than students who did not take survey career and technical classes (p<0.05). Students who took a concentrated career and technical education classes had total compensation that was $1,851 higher than students who did not take concentrated career and technical education classes (p<0.10). Students who took survey career and technical courses had a total compensation that was $1,097 higher than students who did not take survey career and technical classes (p=ns).
Wagner (1991) Wagner et al. (1993) Participation in regular career and technical education	QED—NLTS 1 survey of nationally representative sample of former special education students	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	
Baer et al. (2011) Career and technical education (Y or N) as determined by the student's school district MIS data record	QED—Follow-up of 2005, 2006, 2007, and 2008 special education grads in large state	Students with Intellectual Disabilities	Exploratory	Baseline equivalence exceeded WWC standards for gender	Career and technical education did not predict employment for students with ID (p=ns).
Studies that Examine the Effects of ***CAREER AWARENESS or DEVELOPMENT PROGRAMS***					
No group design studies were identified					
Studies that Examine the Effects of ***COMMUNITY-BASED WORK EXPERIENCE PROGRAMS***					
Baer et al. (2011) Work study participation (Y or N) as determined by the student's school district MIS data record	QED—Follow-up of 2005, 2006, 2007, and 2008 special education grads in large state	Students with Intellectual Disabilities	Meets with reservations	n.a.	Work study was not a significant predictor of employment at follow-up (p=ns).

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
Cimera (2010) Community-based transition services delivered in high school that involved career and technical assessment and other supports but primarily helped students obtain and sustain paid employment	QED—Follow-up in 2008 of all recipients of adult services from nine sites across mid-Atlantic states	All served by Adult Services (primarily students with mild to severe MR, autism, or mental illness	Meets with reservations	n.a.	Students who were supported by community-based transition services (supported employment) during high school were employed twice as many months as students who received transition services in high school only (p=0.001) and students who received no transition services in high school (p=0.0006).
Karpur et al. (2005) Steps-to-Success Program, a transition program including planning, career and technical education and training, paid and unpaid work experience, school and work supports, therapeutic counseling and social skill development	QED—follow-up of students who either graduated high school or dropped out with one year of program exposure	EBD	Meets with reservations	n.a.	Students who participated in the Steps-to-Success program were no more likely (p=ns) to be employed than a matched sample of students who did not participate in Steps-to-Success.
Shandra and Hogan (2008) Work-based school-to-work programs with emphasis on mentoring, job shadowing, and internship placement	QED—NLSY97 (U.S. DOL national survey of 8,984 children age 12–16 in 1997)	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who participated in work-based school-to-work programs had 1.2 times the odds of receiving fringe benefits than students who did not participate in work-based programs (p<0.05). Work-based learning had no discernible relationship to annual income, hours worked per week, or work status (p=ns for all comparisons).
Doren et al. (2011) Oregon "Youth Transition Program" All students in the program receive a comprehensive package of services designed to address a broad array of transition needs including: (a) individualized planning, (b) instruction in academic, career and technical, independent living, and personal social skills, (c) career development services, (d) connections with local employers, development of school-based businesses, on-the-job assessment and training, (e) support services such as individualized mentoring or referrals for additional specific interventions, and (f) follow-up support for one year after leaving the program	QED—Follow-up of sample of Oregon YTP graduates served from 1990–1997, and graduated by 1998	All, but females only	Exploratory	Baseline equivalence not established for ethnicity	Females with disabilities earned significantly lower starting wages ($4360 lower) at exit from the YTP program than males with disabilities (p<0.001).

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
Wagner et al. (1993) Participation in high school work study program	QED—NLTS 1 survey of nationally representative sample of former special education students	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who participated in work study programs had an employment rate that was 2.0 percentage points lower than students who did not participate in work study programs (*p*=ns) Students who participated in work study programs had a total compensation that was $545 higher than students who did not participate in work study programs (*p*=ns).
Studies that Examine the Effects of ***EMPLOYMENT while in SCHOOL***					
Doren et al. (2011) Oregon "Youth Transition Program"	QED—Follow-up of 521 participants in the Oregon Youth Transition Program from 1990 to 1997	All	Exploratory	Baseline equivalence not established for ethnicity	Higher wages at the time students exited from the Oregon YTP program for males predicted higher wages for males at follow-up
Fourqurean et al. (1991) Worked while in school	QED—Follow-up of 123 graduates from four high schools in Houston, TX, from 1986 to 1989	Students with learning disabilities	Exploratory	Baseline equivalence not established for ethnicity	Students who had been employed while in high school had greater employment stability than students who had not been employed while in high school.
Studies that Examine the Effects of ***FUNCTIONAL LIFE SKILLS DEVELOPMENT***					
No group design studies were identified					
Studies that Examine the Effects of ***INCLUSION in GENERAL EDUCATION***					
Wagner et al. (1993) Participation in regular education	QED—NLTS 1 survey of nationally representative sample of former special education students	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who spent all their time in regular education had a competitive employment rate that was 11.2 percentage points higher than students who spend half their time in regular education (p<0.01). Students who spent all their time in regular education had a total compensation that was $2095 higher than students who spend half their time in regular education (p<0.001).
Studies that Examine the Effects of ***INTERAGENCY COLLABORATION***					
Winsor et al. (2011) *Jobs by 21 Partnership Project* involving collaboration between differing county agencies	390 graduates from project versus 656 comparison graduates from the same county.	Students with develop-mental disabilities	Exploratory	Baseline equivalence not established for gender or ethnicity	*Jobs by 21 Partnership Project* participants earned greater wages and were employed at a greater rate than were nonparticipants from the same counties where the PP intervention was implemented.
Studies that Examine the Effects of ***STUDENT SUPPORT INCLUDING PARENT INVOLVEMENT***					
No group design studies were identified					

Table G2. Group design study findings for postsecondary education outcomes, by program category

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
Studies that Examine the Effects of ***CAREER AND TECHNICAL EDUCATION***					
Harvey (2002) Career and technical education—students who participated in any career and technical coursework while in high school	QED—NELS: 88 national stratified probability sample from 1,052 schools	All	Exploratory	Baseline equivalence not established for type of disability, but proxy measures deemed adequate	Students with disabilities who participated in career and technical education had a 56.9 percent participation rate in postsecondary education as compared with a 68.2 percent participation rate among students with disabilities who did not participate in career and technical education. Students who had taken career and technical education in their last year in secondary school had a postsecondary career and technical school attendance rate that was 7.7 percentage points higher than students who had not taken career and technical education in their last year in high school (p<0.01). Students who took survey career and technical courses had a postsecondary academic enrollment that was 10.3 percentage points higher than students who did not take survey career and technical courses (p=ns). Students who took concentrated career and technical courses had a postsecondary academic enrollment that was 2.0 percentage points higher than students who did not take concentrated career and technical courses (p=ns). Students who took survey career and technical courses had a postsecondary career and technical enrollment that was 10.4 percentage points higher than students who did not take survey career and technical courses (p=ns).
Wagner et al. (1993) Career and technical education—both survey (taking unrelated courses) and concentrated	QED—NLTS I, nationally representative sample	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who took concentrated career and technical courses had a postsecondary career and technical enrollment that was 10.3 percentage points higher than students who did not take concentrated career and technical courses (p=.ns).
Studies that Examine the Effects of ***CAREER AWARENESS or DEVELOPMENT PROGRAMS***					
No group design studies were identified					

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
Studies that Examine the Effects of ***COMMUNITY-BASED WORK EXPERIENCE PROGRAMS***					
Karpur et al. (2005) Steps-to-Success Program, a comprehensive transition program including planning, career and technical education and training, paid and unpaid work experience, school and work supports, therapeutic counseling and social skill development	QED—follow-up of students who either graduated high school or dropped out with one year of program exposure	EBD	Meets with reservations	n.a.	Students who participated in the Steps-to-Success program were more likely to be enrolled in postsecondary education than a matched sample of students who did not participate in Steps-to-Success (p=0.001).
Wagner et al. (1993) Participation in a high school work experience program	QED—NLTS I, nationally representative sample	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who took a high school work experience program were less likely than students who did not take a high school work experience program to enroll in postsecondary education (p<0.05). Students who participated in a work experience program while in high school had postsecondary academic enrollment that was 9.3 percentage points lower than students who did not participate in a work experience program (p<0.05). Students who participated in a work experience program while in high school had postsecondary career and technical enrollment that was 5.6 percentage points higher than students who did not participate in a work experience program (p=ns).
Studies that Examine the Effects of ***EMPLOYMENT while in SCHOOL***					
No group design studies were identified					
Studies that Examine the Effects of ***FUNCTIONAL LIFE SKILLS DEVELOPMENT***					
No group design studies were identified					
Studies that Examine the Effects of ***INCLUSION in GENERAL EDUCATION***					
Baer et al. (2011) Inclusion defined as being in regular classes at least 80 percent of the time according to the student's school district MIS data record	QED—Follow-up of 2005, 2006, 2007, and 2008 special education grads in large state	Students with Intellectual Disabilities	Exploratory	Baseline equivalence exceeded WWC standards for ethnicity	Inclusion for students with ID had an a risk-odds ratio of 1.94, meaning inclusion almost doubled the chances for students with ID to enroll in postsecondary education when controlling for gender and ethnicity (p<0.05).
Wagner et al. (1993) Participation in regular education classes	QED—NLTS I, nationally representative sample	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who spent all their time in general education had a postsecondary career and technical enrollment that was 3.2 percentage points higher than students who spent half of their time in general education (p=ns). Students who spent all their time in general education had a postsecondary academic enrollment that was 4.2 percentage points higher than students who spent half of their time in general education (p=ns).

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
Studies that Examine the Effects of ***INTERAGENCY COLLABORATION***					
No group design studies were identified					
Studies that Examine the Effects of ***STUDENT SUPPORT INCLUDING PARENT INVOLVEMENT***					
No group design studies were identified					

Table G3. Group design study findings for independent living outcomes, by program category

Intervention description and brief citation	Method	Disability groups	Evidence rating	Reason study did not meet standards	Reported findings
Studies that Examine the Effects of ***CAREER AND TECHNICAL EDUCATION***					
Wagner et al. (1993) Participation in regular career and technical education	QED—NLTS 1 survey of nationally representative sample of former special education students	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who took a concentrated career and technical education classes had an independent living outcome that was 5.2 percentage points higher than students who did not take concentrated classes (*p*=ns). Students who took survey career and technical courses had an independent living outcome that was 5.6 percentage points higher than students who did not take survey career and technical classes (*p*=ns).
Studies that Examine the Effects of ***CAREER AWARENESS and DEVELOPMENT PROGRAMS***					
Betz et al. (2010) Three-session, eight-module, Transition Preparation Training (TPT) program including spina bifida management that included transition planning, career and future goal assessment, and practice of strategies for obtaining services to meet identified goals	RCT	Spina Bifida	Exploratory	Attrition exceeded WWC standards and baseline equivalence not established for ethnicity	Treatment had no impact on any of the independent living outcomes measured (*p*=ns).
Studies that Examine the Effects of ***COMMUNITY-BASED WORK EXPERIENCE PROGRAMS***					
Wagner et al. (1993) Participation in high school work study program	QED—NLTS 1 survey of nationally representative sample of former special education students	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who participated in a high school work study program had an independent living outcome that was 6.2 percentage points higher than students who did not participate in work study programs (*p*=ns).
Studies that Examine the Effects of ***EMPLOYMENT while in SCHOOL***					
No group design studies were identified					
Studies that Examine the Effects of ***FUNCTIONAL LIFE SKILLS DEVELOPMENT***					
No group design studies were identified					
Studies that Examine the Effects of ***INCLUSION in GENERAL EDUCATION***					
Wagner et al. (1993) Participation in regular education; taking advanced coursework	QED—NLTS 1 survey of nationally representative sample of former special education students	All	Exploratory	Baseline equivalence not established for severity/type of disability, gender, or ethnicity	Students who spent all their time in regular education had an independent living outcome that was 5.4 percentage points higher than students who spend half their time in regular education (*p*=ns).
Studies that Examine the Effects of ***INTERAGENCY COLLABORATION***					
No group design studies were identified					
Studies that Examine the Effects of ***STUDENT SUPPORT INCLUDING PARENT INVOLVEMENT***					
No group design studies were identified					

Appendix H. Glossary

Table H1. Glossary

Terminology	Definition
Attrition	Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. We consider the total attrition rate and the difference in attrition rates across groups within a study.
Case	The unit of analysis in a SCD study, which can be a single participant or a group of participants.
Changing criterion design	An SCD that features repeated measurement of an outcome under intervention criteria that are systematically and progressively changed.
Confounding factor	A confounding factor is a component of a study that aligns completely with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.
Design	The design of a study is the method by which intervention and comparison groups (QED/RCT), or intervention and comparison conditions (SCD), were assigned.
Domain	A domain is a group of closely related outcomes.
Eligibility	A study is eligible for review and inclusion in this report if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design or a single-case design.
Equivalence	A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.
Evidence rating	The evidence rating for a SCD is a summary of the visual analysis examination of the findings. Evidence is rated as providing strong, moderate, or no evidence.
Extent of evidence	An indication of how much evidence supports the findings.
Functional relationship	An SCD that demonstrates strong or moderate evidence of a causal relationship.
Improvement index	Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.
Moderate evidence	An SCD in which the data demonstrate at least three intervention effects but also at least one non-effect. Moderate is also the highest rating for multiple baseline SCDs in which an intervention effect for one case is associated with a change in the baseline data pattern for another case.
Multiple baseline	An SCD that features staggered introduction of the intervention across different points in time.
Multiple comparison adjustment	When a study includes multiple outcomes or comparison groups, we will adjust the statistical significance to account for the multiple comparisons, if necessary.
Multiple probe	A multiple baseline SCD variant in which baseline sessions occur only periodically and at each time the intervention is introduced.
No evidence	An SCD in which the data do not demonstrate three intervention effects at three points in time.
Quasi-experimental design (QED)	A research design in which subjects are assigned to treatment and comparison groups through a process that is not random.
Randomized controlled trial (RCT)	An experiment in which investigators randomly assign eligible participants into treatment and comparison groups.
Rating of effectiveness	We follow the WWC in rating the effects of an intervention in each domain based on the quality of the research design and the magnitude, statistical significance or visual analysis, and consistency in findings.
Research team	The list of authors on an individual publication.
Reversal/withdrawal	A reversal/withdrawal SCD is based on repeated introduction and withdrawal of the intervention for a particular case.
Single-case design (SCD)	A research approach in which an outcome variable is measured repeatedly within and across conditions that are defined by the presence or absence of an intervention.
Standard deviation	The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample tend to be spread out over a large range of values.
Statistical significance	The probability that the difference between groups is a result of chance rather than a real difference between the groups. The review labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5 percent ($p<0.05$).
Strong evidence	An SCD in which the data demonstrate at least three intervention effects and no non-effects.
Study	The analysis of a set of research questions. Typically refers to an individual publication.
Substantively important	A finding that has an effect size of 0.25 or greater, regardless of statistical significance.
Visual analysis	The process used by the WWC for rating evidence in a SCD, which accounts for the level, trend, and variability of data, as well as the percentage of overlap, the immediacy of effect, and the degree of consistency across similar phases.

References

References are listed in several sections. The first section includes references for studies that are cited in the narrative but are not part of the evidence review. References in the next section include studies that are included in the evidence review and meet WWC standards. The final section includes references for studies that do not meet WWC standards. References in the final section that are marked with an asterisk indicate studies that are included in the review as exploratory.

Studies cited in report but not included in the evidence review

Alwell, M., and Cobb, B. (2006). A map of the intervention literature in secondary special education transition. *Career Development for Exceptional Individuals, 29*(1), 13–27.

Cobb, B., and Alwell, M. (2009). Transition planning/coordinating interventions for youth with disabilities: a systematic review. *Career Development for Exceptional Individuals*, *32*(2), 70–81.

Fraker, T., Baird, P., Black, A., Mamun, A., Manno, M., Martinez, J., Rangarajan, A., and Reed, D. (2011, April). *The Social Security Administration's Youth Transition Demonstration Projects: Interim report on Colorado Youth WINS.* Retrieved from http://ssrn.com/abstract=1849507.

Fraker, T., Black, A., Broadus, J., Mamun, A., Manno, M., Martinez, J., McRoberts, R., Rangarajan, A., and Reed, D. (2011, April). *The Social Security Administration's Youth Transition Demonstration Projects: Interim report on the City University of New York's Project.* Retrieved from http://ssrn.com/abstract=1849506.

Fraker, T., Black, A., Mamun, A., Manno, M., Martinez, J., O'Day, B., O'Toole, M., Rangarajan, A., and Reed, D. (2011, February). *The Social Security Administration's Youth Transition Demonstration Projects: Interim report on Transition WORKS.* Retrieved from http://ssrn.com/abstract=1849503.

Halpern, A.S. (1994). The transition of youth with disabilities to adult life: A position statement of the Division on Career Development and Transition, the Council for Exceptional Children. *Career Development for Exceptional Individuals. 17*, 115–124.

Hasazi, S., Gordon, L., and Roe, C. (1985). Factors associated with the employment status of handicapped youth exiting high school from 1979 to 1983. *Exceptional Children, 51*, 455–469.

Heal, L.W., and Rusch, F.R. (1994). Prediction of residential independence of special education high school students. *Research in Developmental Disabilities, 15*(3), 223–243.

Horner, R.H., Carr, E.G., Halle, J., McGee, G., Odom, S., and Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179.

Mithaug, D.E., Horiuchi, C., and Fanning, P.N. (1985). A report on the Colorado statewide follow-up survey of special education students. *Exceptional Children, 51*, 397–404.

Newman, L., Wagner, M., Knokey, A.-M., Marder, C., Nagle, K., Shaver, D., Wei, X., with Cameto, R., Contreras, E., Ferguson, K., Greene, S., and Schwarting, M. (2011). *The post-high school outcomes of young adults with disabilities up to 8 years after high school. A report from the National Longitudinal Transition Study-2 (NLTS2)* (NCSER 2011-3005). Menlo Park, CA: SRI International.

Test, D.W., Mazzotti, V.L., Mustian, A.L., Fowler, C.H., Kortering, L., and Kohler, P. (2009). Evidence-based secondary transition predictors for improving postschool outcomes for students with disabilities. *Career Development for Exceptional Individuals, 32*(3), 160–181.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2011, September). *Procedures and standards handbook (Version 2.1)*. Retrieved from http://whatworks.ed.gov.

Wehman, P., Kregal, J., and Seyfarth, J. (1985). Transition from school to work for individuals with severe handicaps: a follow-up study. *Journal of the Association for Persons with Severe handicaps, 10*, 132–136.

Meets WWC evidence standards without reservations (group design studies)

None.

Meets WWC evidence standards with reservations (group design studies)

Baer, R.W., Daviso, A.W., Flexer, R.W., Queen, R.M., and Meindl, R.S. (2011). Students with intellectual disabilities: predictors of transition outcomes. *Career Development for Exceptional Individuals, 34*(3), 132–141.

Cimera, R.E. (2010). Can community-based high school transition programs improve the cost-efficiency of supported employment? *Career Development for Exceptional Individuals, 33*(1), 4–12.

Karpur, A., Clark, H.B., Caproni, P., and Sterner, H. (2005). Transition to adult roles for students with emotional/behavioral disturbances: a follow-up study of student exiters from Steps-to-Success. *Career Development for Exceptional Individuals, 28*(1), 36–46.

Meets WWC pilot single-case design standards without reservations

None.

Meets WWC pilot single-case design standards with reservations

Cihak, D.F., Alberto, P.A., Kessler, K.B., and Taber, T.A. (2004). An investigation of instructional scheduling arrangements for community-based instruction. *Research in Developmental Disabilities, 25*(1), 67–88.

DiPipi-Hoy, C., and Jitendra, A. (2004). A parent-delivered intervention to teach purchasing skills to young adults with disabilities. *Journal of Special Education, 38*(3), 144–157.

Ganz, J.B., and Sigafoos, J. (2005). Self-monitoring: Are young adults with MR and autism able to utilize cognitive strategies independently? *Education and Training in Developmental Disabilities, 40*(1), 24–33.

Hutcherson, K., Langone, J., Ayres, K., and Clees, T. (2004). Computer assisted instruction to teach item selection in grocery stores: an assessment of acquisition and generalization. *Journal of Special Education Technology, 19*(4), 33–42.

Mechling, L.C., and Gast, D.L. (2003). Multimedia instruction to teach grocery word associations and store location: a study of generalization. *Education and Training in Developmental Disabilities, 38*(1), 62–76.

Mechling, L.C., Gast, D.L., and Barthold, S. (2003). Multimedia computer-based instruction to teach students with moderate intellectual disabilities to use a debit card to make purchases. *Exceptionality, 11*(4), 239–254.

Mechling, L.C., Gast, D.L., and Fields, E. A. (2008). Evaluation of a portable DVD player and system of least prompts to self-prompt cooking task completion by young adults with moderate intellectual disabilities. *Journal of Special Education, 42*(3), 179–190.

Mechling, L.C., Gast, D.L., and Langone, J. (2002). Computer-based video instruction to teach persons with moderate intellectual disabilities to read grocery aisle signs and locate items. *Journal of Special Education, 35*(4), 224–240.

Mitchell, R.J., Schuster, J.W., Collins, B.C., and Gassaway, L.J. (2000). Teaching vocational skills with a faded auditory prompting system. *Education and Training in Mental Retardation and Developmental Disabilities, 35*(4), 415–427.

Taber, T.A., Alberto, P.A., Seltzer, A., and Hughes, M. (2003). Obtaining assistance when lost in the community using cell phones. *Research and Practice for Persons with Severe Disabilities, 28*(3), 105–116.

Taras, M.E., Matson, J.L., and Felps, J.N. (1993). Using independence training to teach independent living skills to children and young men with visual impairments. *Behavior Modification, 17*(2), 189–208.

Taylor, P., Collins, B.C., Schuster, J.W., and Kleinert, H. (2002). Teaching laundry skills to high school students with disabilities: generalization of targeted skills and nontargeted information. *Education and Training in Mental Retardation and Developmental Disabilities, 37*(2), 172–183.

Test, D.W., Howell, A., Burkhard, K., and Beroth, T. (1993). The one-more-than technique as a strategy for counting money for individuals with moderate mental retardation. *Education and Training in Mental Retardation, 28*, 232–241.

Does not meet WWC evidence standards (group design studies) (studies denoted with an asterisk were exploratory studies)

Baer, R.M., Flexer, R.W., Beck, S., Amstutz, N., Hoffman, L., Brothers, J., Stelzer, D., and Zechman, C. (2003). A collaborative follow-up study on transition service utilization and post-school outcomes. *Career Development for Exceptional Individuals, 26*(1), 7–25.

Benz, M.R., Yovanoff, P, and Doren, B. (1997). School-to-work components that predict postschool success for students with and without disabilities. *Exceptional Children, 63*(2), 151–165.

*Betz, C.L., Smith, K., and Macias, K. (2010). Testing the transition preparation training program: a randomized controlled trial. *International Journal of Child and Adolescent Health, 3*(4), 595–607.

Doren, B., and Benz, M. (1998). Employment inequality revisited: predictors of better employment outcomes for young women with disabilities in transition. *Journal of Special Education, 31*(4), 425–442.

*Doren, B., Gau, J.M., and Lindstrom, L. (2011). The role of gender in the long-term employment outcomes of young adults with disabilities. *Journal of Vocational Rehabilitation, 34*(1), 35–42.

Dunn, C., and Shumaker, L. (1997). A follow-up study of former special education students from a rural and urban county school system. *Career Development for Exceptional Individuals, 20*(1), 43–54.

Farley, R.C., and Johnson, V.A. (1999). Enhancing the career exploration and job-seeking skills of secondary students with disabilities. *Career Development for Exceptional Individuals, 22*(1), 43–54.

Flexer, R.W., Daviso, A.W., Baer, R.M., Queen, R.M., and Meindl, R.S. (2011). An epidemiological model of transition and postschool outcomes. *Career Development for Exceptional Individuals, 34*(2), 83–94.

*Fourqurean, J.M., Meisgeier, C., Swank, R.R., and Williams, R.E. (1991). Correlates of postsecondary employment outcomes for young adults with learning disabilities. *Journal of Learning Disabilities, 24*, 400–405.

*Harvey, M.W. (2002). Comparison of postsecondary transitional outcomes between students with and without disabilities by secondary vocational education participation: findings from the national education longitudinal study. *Career Development for Exceptional Individuals, 25*(2), 99–122.

*Izzo, M., Cartledge, G., Miller, L., Growick, B., and Rutkowski, S. (2000). Increasing employment earnings: extended transition services that make a difference. *Career Development for Exceptional Individuals, 23*(2), 139–156.

McDonnall, M.C., and Crudden, A. (2009). Factors affecting the successful employment of transition-age youths with visual impairments. *Journal of Visual Impairment and Blindness, 103*(6), 329–341.

Newman, L., and Cameto, R. (1993). What makes a difference? Factors related to postsecondary school attendance for young people with disabilities. SRI International: Menlo Park, CA.

Rabren, K., Dunn, C., and Chambers, D. (2002). Predictors of post-high school employment among young adults with disabilities. *Career Development for Exceptional Individuals, 25*(1), 25–40.

*Shandra, C.L., and Hogan, D.P. (2008). School-to-work program participation and the post-high school employment of young adults with disabilities. *Journal of Vocational Rehabilitation, 29*(2), 117–130.

Siegel, S., Robert, M., Waxman, M., and Gaylord-Ross, R. (1992). A follow-along study of participants in a longitudinal transition program for youths with mild disabilities. *Exceptional Children, 58*(4), 346–356.

*Wagner, M., Blackorby, J., Cameto, R, and Newman, L. (1993). *What makes a difference? Influences on postschool outcomes of youth with disabilities. The third comprehensive report from the National Longitudinal Transition Study of Special Education Students*. SRI International: Menlo Park, CA.

Additional citation

*Wagner, M.M., and Blackorby, J. (1996). Transition from high school to work or college: how special education students fare. *Future of Children, 6*(1), 103–120.

*Wagner, M. (1991). The benefits of secondary vocational education for young people with disabilities. Findings from the National Longitudinal Transition Study of Special Education Students. SRI International: Menlo Park, CA.

Additional citation

*Wagner, M.M., and Blackorby, J. (1996). Transition from high school to work or college: how special education students fare. *Future of Children, 6*(1), 103–120.

*Winsor, J.E., Butterworth, J., and Boone, J. (2011). Jobs by 21 partnership project: impact of cross-system collaboration on employment outcomes of young adults with developmental disabilities. *Intellectual and Developmental Disabilities, 49*(4), 274–284.

Wolffe, K., and Kelly, S.M. (2011). Instruction in areas of the expanded core curriculum linked to transition outcomes for students with visual impairments. *Journal of Visual Impairment and Blindness, 105*(6), 340–349.

Does not meet WWC pilot single-case design standards

Ayres, K.M., Langone, J., Boon, R.T., and Norman, A. (2006). Computer-based instruction for purchasing skills. *Education and Training in Developmental Disabilities, 41*(3), 253–263.

Bailey, R.L., and Angell, M.E. (2005). Improving feeding skills and mealtime behaviors in children and youth with disabilities. *Education and Training in Developmental Disabilities, 40*(1), 80–96.

Burke, R.V., Andersen, M.N., Bowen, S.L., Howard, M.R., and Allen, K.D. (2010). Evaluation of two instruction methods to increase employment options for young adults with autism spectrum disorders. Research in *Developmental Disabilities, 31*(6), 1223–1233.

Ferguson, H., Myles, B.S., and Hagiwara, T. (2005). Using a personal digital assistant to enhance the independence of an adolescent with Asperger syndrome. *Education and Training in Developmental Disabilities, 40*(1), 60–67.

Graves, T.B., Collins, B.C., Schuster, J.W., and Kleinert, H. (2005). Using video prompting to teach cooking skills to secondary students with moderate disabilities. *Education and Training in Developmental Disabilities, 40*(1), 34–46.

Haring, T.G., Breen, C.G., Weiner, J., Kennedy, C.H. and Bednersh, F. (1995). Using videotape modeling to facilitate generalized purchasing skills. *Journal of Behavioral Education, 5*(1), 29–53.

Schloss, P.J., and Alper, S. (1997). Teaching meal planning to youth with mental retardation in natural settings. *Remedial and Special Education, 18*(3), 166–173.

Van Laarhoven, T., Johnson, J.W., Van Laarhoven-Myers, T., Grider, K.L., and Grider, K.M. (2009). The effectiveness of using a video iPod as a prompting device in employment settings. *Journal of Behavioral Education, 18*(2), 119–141.