

Afterschool Programs: A Review of Evidence Under the Every Student Succeeds Act

Ruth Curran Neild



Sandra Jo Wilson



Wendy McClanahan



Acknowledgments

We are grateful to Elizabeth Ty Wilde, program officer at The Wallace Foundation, who provided valuable guidance throughout the project and thoughtful feedback on protocol development and report drafts. Ed Pauly and Ann Stone also provided helpful input at critical junctures. Kathryn Young of Education Counsel researched how various programs under ESSA can support afterschool programming and provided precise language for each program's purpose and evidence requirements.

We also thank the many individuals who conducted study reviews and provided feedback on drafts of the protocol and report; special thanks go to Sarah Sahni, Tracey Hartmann, and Monica Mielke. Kathryn Carter and Lindsey Liu provided helpful feedback on writing and graphic presentation.

We are deeply grateful for the wisdom of afterschool field leaders who provided feedback on our ideas for evidence review content and displays: Jessica Donner, Every Hour Counts; Julie Dopler, Cincinnati Public Schools; James Doyle, Higher Achievement; Ellen Gannett and Georgia Hall, National Institute on Out-of-School Time at Wellesley College; Rachel Gwaltney, National Summer Learning Organization; Anna Harutyunyan, Nashville After Zone Alliance; Alli Lidie, New York State Network for Youth Success; Greg MacPherson, Big Thought; Allison Overseth, Partnership for After School Education; Jen Rinehart, Afterschool Alliance; Hillary Salmons, Providence After School Alliance; and Gina Warner, National AfterSchool Association.

Funding and conflict of interest statement

The Wallace Foundation provided funding for this evidence review. However, the content of this review does not necessarily reflect the views of the Foundation.

The Wallace Foundation supported implementation of some afterschool programs described in this review as well as studies of these programs. Decisions about study inclusion criteria and ratings were made solely by the authors.

Studies by Research for Action and Abt Associates, Inc., are included in this review. To manage potential conflicts of interest, study ratings were determined by reviewers who were not involved in these studies.

Electronic access

This review is available for web download at <https://www.researchforaction.org/projects/afterschoolessa/>.

A companion guide with detailed evidence summaries for individual afterschool programs is available for web download at <https://www.researchforaction.org/projects/afterschoolessa/>.

Suggested citation

Neild, R.C., Wilson, S.J., & McClanahan, W. (2019). *Afterschool programs: A review of evidence under the Every Student Succeeds Act*. Philadelphia: Research for Action.

Afterschool Programs: A Review of Evidence Under the Every Student Succeeds Act

Ruth Curran Neild



Sandra Jo Wilson



Wendy McClanahan



Summary	i
Chapter 1: Introduction	1
The purpose and content of this review	2
Why afterschool programming matters	3
Evidence requirements under ESSA	3
How this review is organized	4
Chapter 2: How We Conducted the Review	7
How we conducted a systematic and comprehensive literature search	7
How we rated studies against standards for evidence of cause-and-effect relationships	11
How we report study findings and program information	16
Chapter 3: The Landscape of Afterschool Studies.....	19
There are more studies of the effects of Multicomponent and Academic afterschool programs than of other types of programs	20
There are at least 25 studies at each grade level, though there are twice as many studies of elementary programs as high school programs	22
About 30 percent of studies have samples of fewer than 100 students, and over half include fewer than 350 students	23
Mathematics Achievement and Reading/English Language Arts (ELA) achievement are by far the most common outcome domains studied	24
State standardized tests were by far the most common outcome measures	26
Outcome domains vary in the percentage of statistically significant improved outcomes	27
Average impacts of afterschool programs are positive across most outcome domains	27
Most programs were implemented in 2006 or later, and most studies were published during that time frame	30
Chapter 4: Evidence-Based Afterschool Programs under the Every Student Succeeds Act	33
Of the 124 afterschool programs with evidence meeting Cause-and-Effect requirements for ESSA Tiers I-III, half have at least one statistically significant improved outcome	35
Almost all program types have one or more programs with evidence meeting Cause-and-Effect requirements for ESSA Tiers I-III and at least one statistically significant improved outcome	36
Few afterschool programs meet Tiers I or II when Broad Application criteria (number of study sites and sample size) are applied	38
At-a-Glance: Afterschool programs with evidence of improved outcomes at Tiers I-III	41
Chapter 5: Key Findings and Recommendations.....	47
Key findings	47
Recommendations for afterschool program providers	47
Recommendations for states	50
Recommendations for the federal government	51
Recommendations for evaluators	51
References	52
Appendix: Review Methods	66

Summary

When the dismissal bell rings and school ends for the day, millions of American students make their way to afterschool programs. These programs, offered in schools or at other community locations, provide a safe and supervised place for children and youth during the afterschool hours when many parents or other caregivers are still at work. But afterschool programming can offer more than safety and supervision. It can also complement the regular school day by providing academic support as well as enrichment activities consistent with a well-rounded education. In afterschool programs, students can explore interests and careers, develop skills, enhance social and emotional competencies, get physical exercise, and learn about healthy behaviors.

Afterschool program models differ in their focus, staffing, components, resources, intended participants, and program length, among other features. When deciding which type of afterschool program to implement, communities should consider their own unique needs, resources, and goals—as well as whether any programs of interest are supported by evidence of their effectiveness. This review provides one type of information needed for these decisions: research evidence about whether specific programs have improved student outcomes.

The purpose and content of this review

This review summarizes virtually all available evidence on the effectiveness of specific afterschool programs, based on a comprehensive literature search and review of studies published in 2000 or later. The review is motivated, in part, by a growing emphasis on using rigorous assessment of program impacts to inform decision-making in education and youth development programs. Consistent with this emphasis, the Every Student Succeeds Act (ESSA), the 2015 reauthorization of the Elementary and Secondary Education Act, encourages—and in some cases requires—states and districts to use approaches backed by rigorous research showing improved outcomes. Under the law, individual states may raise the bar and require even stronger evidence than ESSA does.

The major federal funding source for afterschool programs, the 21st Century Community Learning Centers program, is authorized under ESSA. Other titles and programs in the law, including Title I and federal discretionary grant programs, also can be used to support afterschool activities. Because of the variety of ways for afterschool programs to be funded under ESSA, this review also explains ESSA's evidence framework and how it applies to key sources of afterschool program funding authorized under the law.

The primary audience for this review is afterschool program providers, policymakers at state and local education agencies, and others who make decisions about whether to provide funding for afterschool programs and which programs to offer. Secondary audiences include evaluators of afterschool programs and others who seek to build rigorous evidence about the effects of afterschool program models.

ESSA's framework for evidence of program effectiveness

This review uses the ESSA evidence framework to assess the evidence of afterschool program effectiveness. ESSA describes a tiered evidence framework of four levels, or tiers: *Strong* (Tier I), *Moderate* (Tier II), *Promising* (Tier III), and a fourth category that has been titled *Demonstrates a Rationale* (Tier IV) in September 2016 guidance from the U.S. Department of Education. The law provides the basic definitions for the tiers, specifying, for example, that Tier I evidence must come from at least one experimental study showing an improved outcome and that Tier II evidence requires a quasi-experiment.

In its guidance, the Department made additional recommendations for applying the evidence tiers in practice. Notably, the Department sought to frame evidence use as part of a sensible cycle of continuous improvement in education. The guidance also articulated specific criteria for the rigor and relevance of studies, especially for Tiers I and II. For example, the Department recommends that Tier I evidence come from a study (or studies) with at least 350 students, conducted in more than one school district. In this review, we use ESSA's evidence requirements and the Department's guidance to identify afterschool programs that we judge to be supported by evidence at each tier. We also provide further definition for Tiers III and IV, which are described only generally in both ESSA and the guidance.

What this review does not do

Because the key purpose of this review is to help decision-makers identify effective afterschool programs to implement, we do not address the larger question of whether afterschool programs, as a class, are effective at improving student outcomes. Further, we do not rigorously examine which parts of individual programs were effective or seek to generalize across programs about the effectiveness of specific components.

Key findings

There is a substantial set of afterschool programs with evidence of effectiveness meeting ESSA Tiers I-III, including branded and unbranded programs.

We identified more than 60 programs with research evidence at Tiers I-III showing one or more improved outcomes for students. These programs include a mix of branded programs or well-developed models (for example, Building Educated Leaders for Life [BELL], AfterZone, and Chicago After School Matters) and unbranded, "homegrown" programs (for example, 21st Century Community Learning Centers programs in Philadelphia). Taken together, the programs improved a variety of outcomes, ranging from mathematics and reading/ELA achievement to physical activity/health, school attendance, promotion and graduation, and social and emotional competencies. At the same time, we found relatively few instances of statistically significant negative outcomes for students.

Detailed information about the goals, components, and studies of these programs, as well as of other programs with similar-quality research that did not show positive outcomes, is presented in a companion evidence guide.

Effective afterschool programs can be found at each grade level and within almost every program type.

There are more programs with evidence of effectiveness for students in the elementary and middle grades, but afterschool programs meeting ESSA Tier I-III criteria can be found across the grade levels and for almost all program types. Academic programs, which focus on improvement of academic skills, and multicomponent programs, which offer a wide range of activities, have the largest number of effective program options. There are more studies of these programs than others, which may be a function of their being more commonly offered.

There are more afterschool program options for improving mathematics and reading/ELA achievement than for improving other outcomes.

Mathematics and reading/ELA achievement, as measured by standardized test scores, are the most commonly studied outcomes, by far. For this reason, even though the percentage of positive findings was lower for achievement than for outcomes like school attendance and physical activity/health, there are more programs with evidence of having improved mathematics and reading/ELA achievement. Because improving academic outcomes is a key goal of the 21st Century Community Learning Centers program, it is important to note that numerous programs have demonstrated effects on mathematics and reading/ELA achievement, including improvements in standardized test scores.

Recommendations for afterschool program providers

This set of recommendations is geared to readers who make decisions about which afterschool programming to offer and how to evaluate program impact. This includes district and school personnel as well as community groups, nonprofits, and others who provide afterschool programs.

Think of research evidence as a tool to help you make the best bets about what will work for your students.

Research evidence should complement, not replace, program providers' craft knowledge and good judgment. Research studies, even the most rigorous ones, provide information only about whether a program has worked in a place (or places) where it has been tried in the past. For this reason, the best way to think about evidence is as a tool that helps program providers increase the likelihood of offering effective programs. The more studies of a program, the larger the sample sizes, and the more consistent

the results, the more confident we can be about the amount of evidence for a program. Larger amounts of evidence provide more assurance that an approach reliably produces certain results, and such evidence should be taken especially seriously, even if it challenges one's beliefs and assumptions.

Most important, even an afterschool program with strong evidence of effectiveness is unlikely to generate benefits if its implementation is not of high quality. An evidence-based program must be replicated with high fidelity to the model to have a reasonable chance of achieving its intended impacts.

Check the rigor of research provided by vendors against the program summaries provided in this review and accompanying volume.

Program providers can check the rigor of any vendor-supplied evidence against program summaries presented in this review and the companion evidence guide. Because this review is based on a thorough, comprehensive literature search, we believe that we have identified and assigned an ESSA tier rating to virtually every study from 2000 through 2017 examining the effects of afterschool programs on the outcomes we specified. Vendor-supplied evidence meeting these inclusion criteria that is not reviewed in this report is likely not to use a comparison group design, an essential element for Tiers I-III under ESSA.

It is common for afterschool programs to be effective in some domains but not others, so consider overall evidence of effectiveness. Pay special attention to specific outcomes that relate to your improvement priorities.

Relatively few programs in this review resulted in negative outcomes, meaning that they harmed students. Nevertheless, it is important to use good judgment by examining the overall effectiveness of a program across all outcomes that were measured, rather than focusing selectively on positive findings. For example, a program may have produced health benefits but no improvements in mathematics or reading/ELA achievement. If improving academic achievement and health are both high priorities, you will need to find a program that is likely to improve both types of outcomes or else implement more than one afterschool program model.

In addition, for any program that is of initial interest, we recommend examining the specific outcomes and measures used in a study. Even within a given outcome area, some outcomes may be more global than others. For example, a program may have evidence that student participation has a large effect on vocabulary development but no effect on a state standardized reading/ELA assessment. It is important to think about the specifics of the outcome you are seeking to achieve as you consider the evidence presented in this review.

If existing afterschool programs with evidence at Tiers I-III do not fit with needs and resources, use the flexibility built in to ESSA Tier IV.

Tier IV of ESSA's evidence framework provides a door through which new afterschool program models can be introduced and evaluated for evidence of effectiveness. If no evidence exists at Tiers I-III, a program can meet Tier IV criteria if there is a well-specified logic model, informed by research, that provides reason to believe that the program will improve outcomes. Neither ESSA nor the Department's guidance specify the type or qualities of research that should inform the logic model. However, for program providers wishing to use the Tier IV option, a starting place is studies that meet Tier IV requirements and are presented in Appendix EG-2 of a companion evidence guide. As comparison group studies, they are superior to studies that use a pretest-posttest design without a comparison group.

Recommendations for states

States are justified in using a more liberal tier standard (Tiers I-IV) for afterschool programs until there are more programs with evidence meeting Tiers I and II.

With some exceptions, ESSA allows states full discretion about which level of evidence (Tiers I-IV) to require for activities funded under the law. States can select the level of evidence for the major source of afterschool funding under ESSA, the 21st Century Community Learning Centers program. Based on our review of the afterschool literature, we believe that states are justified in allowing programs to cite evidence at ESSA Tiers I-IV until there is more evidence at Tiers I-II. For example, although 20 afterschool programs meet Tiers I-II research design requirements and improved at least one student outcome, only five programs meet these tiers when all of the Department's recommendations, such as minimum study sample size, are applied.

States should strongly encourage districts and providers to conduct evaluations designed to meet Tier I and Tier II criteria for research quality and should provide adequate evaluation funding.

This review identified 57 programs with research judged to be consistent with What Works Clearinghouse (WWC) quality standards for experiments or quasi-experiments. These findings tell us that rigorous studies of effectiveness can be done, and are being done, in the afterschool field. The WWC, an initiative of the Department of Education's Institute of Education Sciences, reviews studies of the effectiveness of education interventions, and its standards are a benchmark in the field. Designing and implementing studies to meet WWC standards is a reasonable and achievable goal.

Additional studies, particularly those designed to meet Tiers I-II, are needed to identify more afterschool programs that improve outcomes for students. States distribute 21st Century Community Learning Centers funding and are in a unique position to encourage systematic, rigorous research about afterschool programs. By incentivizing program providers to clarify their logic models, implement their programs with fidelity, and conduct rigorous studies of the effectiveness and implementation of their afterschool programs, states could increase the field's understanding of what works in afterschool programming and what it takes to achieve desired results.

Recommendations for the federal government

The U.S. Department of Education’s 21st Century Community Learning Centers program should encourage evaluations that investigate a range of student outcomes, including academic achievement measured other than by state standardized tests.

The U.S. Department of Education describes a broad vision for the 21st Century Community Learning Centers program, including providing opportunities for academic enrichment and a wide array of additional services, programs, and activities for students and families. Given this broad mandate and focus on enrichment, the indicators used by the federal government to judge the program’s performance may capture only a subset of potentially important outcomes. **We particularly note that using state standardized test scores as key performance measures may set unrealistic expectations for what afterschool programs can and should accomplish.** We recommend that the Department encourage states and evaluators to use measures that are clearly aligned to the purposes of the programs and sensitive to changes that they realistically could produce.

Recommendations for evaluators

Thoroughly report afterschool program components, implementation challenges, and cost.

To use research in practice, providers need to know what the program offered, how it offered these components, and what they cost. Providers can use information about a program’s implementation to assess whether the program is suitable for their setting, whether their staff have the training or capacity to implement the activities, what implementation challenges to expect, and the like. To respond to this need, we collected and summarized information available from the studies included in the review, but we found that too often studies provide minimal descriptions of the program. If research is to support decision-making, evaluators need to provide much more detail about the features of programs and settings they have studied.

Provide all information needed to conduct a study review for ESSA.

To determine the ESSA tier for each study, we used only the information that study authors provided about their findings. We did not contact authors for additional information if an important piece of technical detail, such as a standard deviation or sample attrition, was missing. In a few cases, the lack of this information meant that we could not provide an ESSA rating for the study, and in other cases, studies might have been rated at a lower tier than they would have been if we had all of the information needed.

The information needed to review studies against What Works Clearinghouse (WWC) standards is described in the WWC Reporting Guide for Study Authors. We recommend that study authors check their reporting against this document (https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_gsa_v1.pdf).

About this review

This is one of five evidence reviews commissioned by The Wallace Foundation summarizing programs that meet ESSA evidence requirements across various priority topics in education. Readers of this review may also be interested in a similar review of the effectiveness of summer learning programs (forthcoming). Other Wallace-funded reviews have addressed education leadership development, arts integration, and social and emotional learning interventions (Grant, et al., 2017; Herman et al., 2017; Ludwig, Boyle, & Lindsay, 2017).

The Wallace Foundation supported some afterschool programs described in this review and funded some studies of these programs. Decisions about study inclusion criteria and ratings were made solely by the authors of this review. Studies by Research for Action and Abt Associates, Inc., are included in this review; to manage potential conflicts of interest, ESSA ratings were determined by reviewers who were not involved in any of these studies.

Introduction

When the dismissal bell rings and school ends for the day, millions of American students make their way to afterschool programs. These programs, offered in schools or at other community locations, provide a safe and supervised place for children and youth during the afterschool hours when many parents or other caregivers are still at work. But afterschool programming can offer more than safety and supervision. It can also complement the regular school day by providing academic support as well as enrichment activities consistent with a well-rounded education. In afterschool programs, students can explore interests and careers, develop skills, enhance social and emotional competencies, get physical exercise, and learn about healthy behaviors.

There are numerous approaches to afterschool programming, just as there are many ways to organize the regular school day. Afterschool program models differ in their focus, staffing, components, resources, intended participants, and program length, among other features. When deciding which type of afterschool program to implement, communities should consider their own unique needs, resources, and goals—as well as whether any programs of interest are supported by evidence of their effectiveness. This review provides one type of information needed for these decisions: research evidence about whether specific programs have improved student outcomes.

This review summarizes virtually all available evidence on the effectiveness of specific afterschool programs, based on a comprehensive literature search and review of studies published in 2000 or later. The review is motivated, in part, by a growing emphasis on using rigorous assessment of program impacts to inform decision-making in education and youth development programs. Consistent with this emphasis, the Every Student Succeeds Act, or ESSA (Public Law No. 114-95, 2015), encourages—and in some cases requires—states and districts to use approaches backed by at least one rigorous study showing improved outcomes. Under the law, individual states may raise the bar and require even stronger evidence than ESSA does.

The largest source of federal funding for afterschool programs, the 21st Century Community Learning Centers program, is authorized under ESSA. Other programs authorized under this legislation also can support afterschool activities (see Box 1.1). Because the authorizing legislation emphasizes evidence use, it is important for decision-makers to have a clear understanding of which afterschool programs have a track record of improving student outcomes and the number of evidence-based afterschool options for different types of programs and student populations.

The purpose and content of this review

This review summarizes evidence of the effectiveness of a wide array of afterschool programs, offered across many different contexts and grades and involving students with a range of academic achievement levels, strengths, and challenges. To assess the strength of the evidence, we apply ESSA's tiered evidence framework and the U.S. Department of Education's recommendations for using the evidence tiers in practice. Although ESSA and its evidence framework figure prominently in this review, we believe that the information is also valuable for decision-making outside of the ESSA context.

The primary audience for the report is afterschool program providers, state and school district policymakers, and others who make decisions about whether to provide funding for afterschool programs and which programs to offer. Secondary audiences include evaluators of afterschool programs and others who seek to build rigorous evidence about the effects of afterschool program models.

The purpose of this review is to summarize evidence of effectiveness for distinct, individual afterschool programs and models. We present a high-level version of these findings in this document and provide more detail about each program in a companion evidence guide. Although we make some general observations about research in the afterschool field, it is not our purpose to address the larger question of whether afterschool programs, as a class, are effective at improving student outcomes. Further, we do not rigorously examine which parts of individual programs were effective or seek to generalize across programs about the effectiveness of specific components.

This is one of five evidence reviews commissioned by The Wallace Foundation. These reviews assess the likelihood that programs could meet ESSA evidence requirements across various priority topics in education. Readers of this report may also be interested in a similar review of the effectiveness of summer learning programs (forthcoming). Other Wallace-funded reviews have addressed education leadership development, arts integration, and social and emotional learning interventions (Grant, et al., 2017; Herman et al., 2017; Ludwig, Boyle, and Lindsay, 2017).

The Wallace Foundation supported some afterschool programs described in this review and funded some studies of these programs. Decisions about study inclusion criteria and ratings were made solely by the authors of this review. Studies by Research for Action and Abt Associates, Inc., are included in this review; to manage potential conflicts of interest, ESSA ratings were determined by reviewers who were not involved in any of these studies.

Why afterschool programming matters

Many families, especially those with children in elementary school, rely on afterschool programs for safe and reliable childcare during non-school hours. Federally funded programs are particularly important for enabling low-income families to access afterschool care.

A key function of afterschool programs is to bridge the time gap between the end of the school day and when a parent or other caregiver arrives home from work. At the most basic level, afterschool programs provide a safe and supervised environment, especially for younger children. Federally funded afterschool programs are especially important for low-income parents, who might otherwise have difficulty accessing reliable care for their children in the non-school hours (Earle, 2009).

In 2014, 10.2 million children in the U.S. (18 percent) and one-quarter of families relied on afterschool programs (Afterschool Alliance, 2014). A survey of parents suggests that an additional 19.4 million children would be enrolled in an afterschool program if one were available to them (Afterschool Alliance, 2014).

Afterschool programs offer opportunities for students to develop academic skills.

Afterschool programs have the potential to close academic skill gaps and help students keep pace with the instruction offered during the school day. A common activity in afterschool programs is supervision of and help with homework completion. Some programs offer more intensive academic support, including tutoring offered one-on-one and in small groups. In this way, afterschool programs can serve as an extension of the school day and an opportunity to tailor instruction to areas of academic need.

Among the many activities that ESSA allows under the 21st Century Community Learning Centers program, academic enrichment heads the list, including tutoring to help students meet challenging academic standards. Low-income students, who are more likely than other students to participate in federally funded afterschool programs, have persistent academic skill gaps relative to their higher-income peers (Reardon, 2011).

Afterschool programs can contribute to a well-rounded education.

Throughout ESSA, states are encouraged to offer students a “well-rounded education,” including opportunities to develop knowledge and skills related to health, music and the arts, engineering, technology, computer science, physical education, and career and technical education. ESSA permits states to include additional subjects on this list “with the purpose of providing all students access to an enriched curriculum and educational experience” (Jones & Workman, 2016).

ESSA encourages states to integrate these subjects into the school day to enable richer intellectual connections and use school time more efficiently. Nevertheless, schools are pressed to offer a range of subjects and enrichment experiences within regular school hours. Afterschool programs, including school-sponsored extracurricular activities, provide additional time for enrichment that extends and complements what schools can offer in their regular curriculum.

For low-income students, afterschool programs can provide an opportunity to experience arts, sports, and technology activities that their families would not be able to purchase privately. Afterschool programs have a potentially important role in mitigating the gap between high-income and low-income families in expenditures on education enrichment activities, which has widened consistently and substantially since the 1970s (Duncan & Murnane, 2011).

Afterschool programs can provide opportunities to develop social and emotional skills.

Although afterschool programs are, by definition, somewhat structured, they have the potential to be less structured than school, allowing for greater student choice, interest exploration, and leadership. Through clubs, sports, and pursuit of individual interests, students can develop what are sometimes called “soft skills” or “21st century skills,” including the ability to interact positively with peers and adults, plan and execute projects, lead groups, solve problems, and persist in the face of difficulty.

Evidence requirements under ESSA

This review uses the ESSA evidence framework to assess the evidence of afterschool program effectiveness. ESSA describes a tiered evidence framework of four levels, or tiers: *Strong* (Tier I), *Moderate* (Tier II), *Promising* (Tier III), and a fourth category that has been titled *Demonstrates a Rationale* (Tier IV) in guidance from the U.S. Department of Education (2016). Each of these tiers is described in more detail in Chapter Two.

Box 1.1 summarizes ESSA titles and programs that are most relevant to afterschool programming, along with their evidence requirements. Under some sections of ESSA, one or more activities must be supported by evidence of effectiveness at Tiers I-III. For other programs under ESSA, fund recipients must use evidence-based approaches (Tiers I-IV) when evidence “is reasonably available,” as determined by states. Other ESSA programs provide competitive preference for implementing evidence-based interventions that meet Tiers I-III. Because of these different requirements and incentives, it is important for decision-makers and afterschool providers to understand which programs authorized under ESSA can support afterschool activities and whether their states have set a minimum evidence level for these programs.

How this review is organized

In Chapter Two, we describe how the review was conducted, explain ESSA’s evidence framework, and introduce vocabulary used in this review to describe the evidence levels. In Chapter Three, we explain what we learned about the characteristics of studies we identified, and in Chapter Four, we discuss the evidence for specific afterschool programs under ESSA. Chapter Five summarizes key findings and presents recommendations for program providers, state and local education agencies, and evaluators of afterschool programs. An abbreviated version of the protocol that guided this review can be found in the Appendix.

A companion evidence guide (available at <https://www.researchforaction.org/projects/afterschoolessa/>) provides detailed summaries of afterschool programs with studies meeting Tier I-III research quality requirements. The guide also includes summaries of studies about the effects of school-sponsored extracurricular programs, comparison group studies that did not meet Tiers I-III but could provide evidence at Tier IV, and studies of programs that combine afterschool and summer learning,

BOX 1.1

ESSA programs supporting afterschool activities and their evidence requirements

ESSA FORMULA GRANT ACTIVITIES REQUIRING EVIDENCE OF EFFECTIVENESS

Most of Title I, Part A funds are distributed to states through federal formula and can be used for afterschool activities consistent with the purposes of the Title. Under ESSA Title I, Part A, Section 1003, states must set aside 7 percent of those funds to help support school improvement plans, which ESSA requires for schools designated by states as needing improvement. Each school improvement plan must include evidence-based interventions (i.e., Tiers I-IV). For those schools that receive funds from the 7 percent set-aside for low-performing schools, at least one intervention in each school's improvement plan must be supported by evidence at Tiers I-III.

ESSA FORMULA GRANT ACTIVITIES THAT ENCOURAGE EVIDENCE OF EFFECTIVENESS

21st Century Community Learning Centers (Title IV, Part B). The primary federal source for afterschool funding, the 21st Century Community Learning Centers program, is authorized under Title IV, Part B of ESSA. Funding for this program is distributed to states through a formula. States then use the bulk of this to fund afterschool providers through a competitive process. Providers can be local education agencies, community organizations, or other public or private entities. Authorized activities include academic enrichment, including tutoring, and a broad array of other services and activities, such as arts, music, health, social and emotional development, physical fitness, nutrition, substance abuse prevention, career and technical programs, and internship or apprenticeship programs, among others.

Direct Student Services funded under Title I (Section 1003A). ESSA allows states to set aside up to 3 percent of Title I funds for Direct Student Services, including “high-quality academic tutoring” that meets criteria outlined in ESSA (Section 1003(e)). Under No Child Left Behind (NCLB), a previous authorization of the law, these services were known as Supplemental Educational Services (SES); Title I schools that had not made Adequate Yearly Progress for three years were required to offer this option to their low-income students. Some of these SES programs fit our definition of an afterschool program, and several studies of SES are included in this review. ESSA made such services optional, and states may elect to require that program models have evidence of effectiveness particularly as part of

examining their demonstrated record of success during the state approval process for providers.

Title I, Part A formula grants for Targeted Assistance Programs (Section 1115). In targeted assistance schools, Title I funds may be used only to provide supplemental instructional services to students identified as having the greatest need for additional help. Such services may include out-of-school time and expanded learning time programs, including programs held before or after school and during the summer.

FEDERAL DISCRETIONARY GRANTS THAT REQUIRE EVIDENCE OF EFFECTIVENESS

ESSA authorizes seven discretionary grant programs that award competitive points for evidence of effectiveness, per ESSA requirements. Of these, four could be used for afterschool programming.

Promise Neighborhoods (Section 4624(b)). This program supports neighborhood-based, cross-sector, “cradle-to-career” solutions to support academic success.

Full-Service Community Schools (Section 4625(b)(3)). This program supports partnerships between education agencies and community-based and nonprofit organizations to provide comprehensive services for students and families.

Jacob K. Javitz Gifted and Talented Students Education Program (Section 4644(f)(2)). A key purpose of the program is to support programs and projects for identifying and serving gifted and talented students, including through out-of-school time programs.

Literacy for All, Results for the Nation (LEARN) grants (Section 2224(b)). These grants to states support development of comprehensive literacy plans and a range of other activities, including professional development for educators and stakeholder engagement. Funds unexpended for these key activities may be used to support services for students, including “connecting out-of-school learning opportunities to in-school learning to improve children’s literacy achievement.”

How We Conducted the Review

To conduct this review, we used the evidence framework outlined in the Every Student Succeeds Act of 2015 (ESSA) and the U.S. Department of Education’s evidence guidance (2016) to identify studies that are strong tests of afterschool program effectiveness. A thorough review of the effectiveness literature has three key elements: (1) a systematic and comprehensive literature search, guided by clear criteria for eligible programs, outcomes, and research designs; (2) ratings of eligible studies for evidence of cause-and-effect relationships against a set of standards; and (3) a summary of the findings about cause-and-effect as well as key components of the program, study participants, and context. This chapter provides more detail on the methods we used to conduct the review.

How we conducted a systematic and comprehensive literature search

Literature searches are described as *systematic* when they are guided by a review protocol that outlines all activities of the review and *comprehensive* when they seek to identify every study that falls within the protocol’s guidelines.

The review protocol

The review protocol, which is drafted before the review begins, specifies the characteristics that studies must have to be eligible for review and the standards used to determine whether studies show evidence of program effectiveness. The protocol also describes the type of information to be recorded about each

study, including the data needed to determine the ESSA evidence tier and details about the program components, study context, and students participating in the studies. An abbreviated protocol is provided in the Appendix.

To be included in this review, studies had to meet criteria for the type of program, outcomes, students, research designs, comparison groups, and publication dates and language.

ELIGIBLE AFTERSCHOOL PROGRAMS

This review summarizes studies of voluntary psychological, educational, or behavioral afterschool programs intended to improve student outcomes in one or more of the outcome domains described below (see Box 2.1 for key definitions). Programs eligible for this review could be offered in school or community locations but had to be delivered during the school year outside of regular school hours. In addition, programs had to be offered primarily after school in the afternoon or evening. This review does not include summer programs (which are the subject of a different review also funded by The Wallace Foundation), combined afterschool/summer learning programs where the effect of each cannot be estimated separately, extended school time programs that focus solely on increasing the length of the regular school day, or programs that combine afterschool activities with activities during the regular school day.

ELIGIBLE OUTCOMES

Eligible outcomes in this review were indicators of positive youth development, including mastery of academic knowledge and skills, engagement and persistence with schooling, and development of social and emotional competencies. Because the absence of risk-taking or delinquency do not indicate that youth are acquiring the skills and competencies to succeed as adults, we did not include those outcomes in this review.

BOX 2.1

Key definitions

FINDING: an impact on an *outcome measure* for a specific sample of students in a *study*.

OUTCOME MEASURE: an instrument or tool that measures student outcomes. Similar outcomes are categorized into *outcome domains*.

OUTCOME DOMAIN: a broad category of similar *outcome measures*. Examples are mathematics achievement or school attendance.

PROGRAM: a program is a set of afterschool activities carried out in a specific way and intended to improve one or more student

outcomes. A program may be a distinct, “branded” model (for example, Girls on the Run) or may be locally-developed.

STUDY: an investigation of the effectiveness of a *program* using a specific sample of students. When results are reported separately for different samples (for example, different grade levels), we review and count these samples as separate studies.

In sum, there may be multiple *studies* of the same program; some studies may contribute multiple *findings* to one *outcome domain*; and some *studies* may contribute *findings* to more than one *outcome domain*.

We examined the effectiveness of afterschool programs for 10 outcome domains, as follows:

- Attendance and Enrollment
- General Achievement
- Mathematics Achievement
- Physical Activity/Health
- Promotion and Graduation
- Reading/English Language Arts (ELA) Achievement
- School Engagement
- Science Achievement
- Social and Emotional Competencies
- Other Achievement

When achievement outcomes were not subject-specific (for example, when a study reported reading and math outcomes as a composite), we report these outcomes under the domain of General Achievement. Some other types of achievement—for example, social studies achievement—are summarized as Other Achievement. Box 2.2 presents detailed descriptions of these outcome domains, including eligible outcome measures.

ELIGIBLE STUDENTS

Eligible programs provided services or activities directly to school-age students in kindergarten through 12th grade.

ELIGIBLE RESEARCH DESIGNS

Two general types of research designs were eligible: (1) *experimental designs or randomized controlled trials* in which students or groups of students were randomly assigned to program and comparison groups and (2) *quasi-experimental designs* that compare outcomes for students in afterschool programs to other students, without using random assignment.

These are the two general types of research designs specified in ESSA.

ELIGIBLE COMPARISON GROUPS

Eligible studies compared the effects of at least one afterschool program to at least one control or comparison condition. We required that comparisons be “no treatment” or “business as usual” or any similar condition set up as a contrast to the program condition. Studies that compared an afterschool program to another specific program, whether delivered after school or during another time, were not eligible for this review. We made this decision because studies that compare one program to another answer different questions from studies with “business as usual” comparison groups. For this reason, the impacts from such studies cannot easily be compared to each other.

PUBLICATION CRITERIA

Studies had to be conducted in the United States and reported in 2000 or later, in English. Journal articles, theses and dissertations, and all types of technical reports were eligible for the review. Publications did not have to be peer reviewed.

BOX 2.2

Eligible outcomes and domains

ACADEMIC ACHIEVEMENT (five domains reported separately: **Mathematics, Reading/English Language Arts, Science, General Achievement, Other Achievement**). These outcome domains are defined as the extent to which students master academic content. General Achievement refers to composite outcomes that combined academic subjects. Other Achievement includes outcomes that were not commonly reported and could not be categorized into the four other achievement domains (for example, social studies achievement). Achievement could be measured using any standard measure, including standardized achievement tests, grades, GPAs, or teacher reports of grades or performance; student reports of their own achievement were not eligible.

ATTENDANCE AND ENROLLMENT. This outcome domain includes outcomes that measure enrollment, attendance, or absenteeism at school, including the number or percentage of days absent or present during a school term, excessive absenteeism, referrals for truancy, and the like. Objective measures of attendance, such as those from school administrative records, were required.

PHYSICAL ACTIVITY/HEALTH. This outcome domain includes physical activity and fitness, weight management, and nutrition. Positive indicators of physical activity included time spent in moderate or vigorous physical activity or other objective indicators of activity, such as daily step counts. Negative indicators of physical activity were amount of time spent in sedentary activities, such as watching TV or working on computers. Physical fitness was defined as any measure related to cardiovascular fitness (for example, step test, systolic blood pressure); skeletal health (for example, bone mineral density); and/or muscular strength (for example, number of sit-ups).

Measures of weight management included BMI, percent body fat, waist circumference, and skinfold thickness. Nutrition included calories consumed, consumption of fruits, vegetables, and whole grains (positive indicator), and consumption of sweetened beverages (negative indicator). Objective indicators of these outcomes were preferred, but self-reports were acceptable if more objective indicators were not available.

PROMOTION AND GRADUATION. This outcome domain includes completion of key education milestones, such as high school graduation, being on track for graduation, grade-to-grade promotion, and the like. Objective measures of attainment, such as school administrative records, were required.

SCHOOL ENGAGEMENT. This outcome domain includes attitudes toward school and subjects taught at school and school behaviors (for example, homework completion). Objective indicators of these outcomes were preferred, but self-reports were acceptable if more objective indicators were not available.

SOCIAL AND EMOTIONAL COMPETENCIES. This outcome domain includes intra-personal and/or inter-personal competencies. Eligible intra-personal competencies included conscientiousness, initiative, flexibility, emotional regulation, and grit. Eligible inter-personal competencies included skills needed to relate to other people, such as communication, collaboration, conflict resolution, and leadership. Objective indicators of these competencies, such as direct assessments or observations, were preferred, but behavior checklists or ratings forms completed by teachers, parents, or service providers were permitted. Checklists or ratings completed by students were permitted provided they contributed to composite score that included at least one rather other than the student.

The literature search

Our search strategy was designed to assemble virtually all credible evidence on the effectiveness of afterschool programs. The primary way we located potentially eligible studies was a comprehensive search of electronic bibliographic databases. We supplemented this search by examining relevant review articles and research syntheses for additional study citations. The full search strategy, including the search terms and databases used, is provided in the Appendix.

RESULTS OF THE SEARCH

Titles and abstracts of each study identified in the search were screened for relevance to the review. Using this process, we identified 690 studies that warranted a more in-depth review for relevance using the full set of eligibility criteria. This screening process resulted in 216 studies eligible for review against research quality standards. Of these 216 studies, we judged that 148 met ESSA evidence Tiers I-III (tier requirements are described in the next section) and 58 met our criteria for Tier IV. Figure 2.1 shows the flow of studies through this process.

How we rated studies against standards for evidence of cause-and-effect relationships

The second step of the process was to review the studies that met all eligibility criteria described above against a set of standards for cause-and-effect. A trained reviewer recorded two types of information about the study: (1) the study methods and procedures and (2) the outcomes and findings of the study, including statistical significance.

ESSA's evidence framework

ESSA specifies four tiers of evidence: *Strong* (Tier I), *Moderate* (Tier II), *Promising* (Tier III), and a fourth category (Tier IV) that has been titled *Demonstrates a Rationale* in guidance from the U.S. Department of Education. The law provides only a minimal description of the evidence required to meet each tier (see Box 2.3 for relevant provisions from ESSA). Specifically, the law requires that:

- **Programs with Tier I evidence** must be supported by at least one *experimental study*, the “gold standard” for establishing cause-and-effect relationships. In these studies, students are randomly assigned to experience a program or to the control group. The study must show that the program improved at least one outcome, and the improvement must be *statistically significant*, or unlikely to be the result of chance variation.

- **Programs with Tier II evidence** must be supported by at least one *quasi-experimental study* that compares outcomes for afterschool program participants to outcomes for a comparison group that is closely matched on important characteristics. As with Tier I evidence, the study must show that the program improved at least one outcome, and the improvement must be statistically significant.
- **Programs with Tier III evidence** must be supported by at least one study that the law describes as “correlational... with statistical controls for selection bias.” Although not specified in the law, the implication is that Tier II and Tier III studies have many similarities but program and comparison groups in Tier III studies are not as closely matched. For example, compared to Tier II studies, Tier III studies may have larger differences between the program and comparison groups on previous achievement, which raises more doubt about whether the study represents an “apples-to-apples” comparison.
- **Programs that meet Tier IV requirements** provide a rationale for why outcomes are likely to improve based on existing research described only as “high-quality” in the law and are undergoing evaluation of their effectiveness.

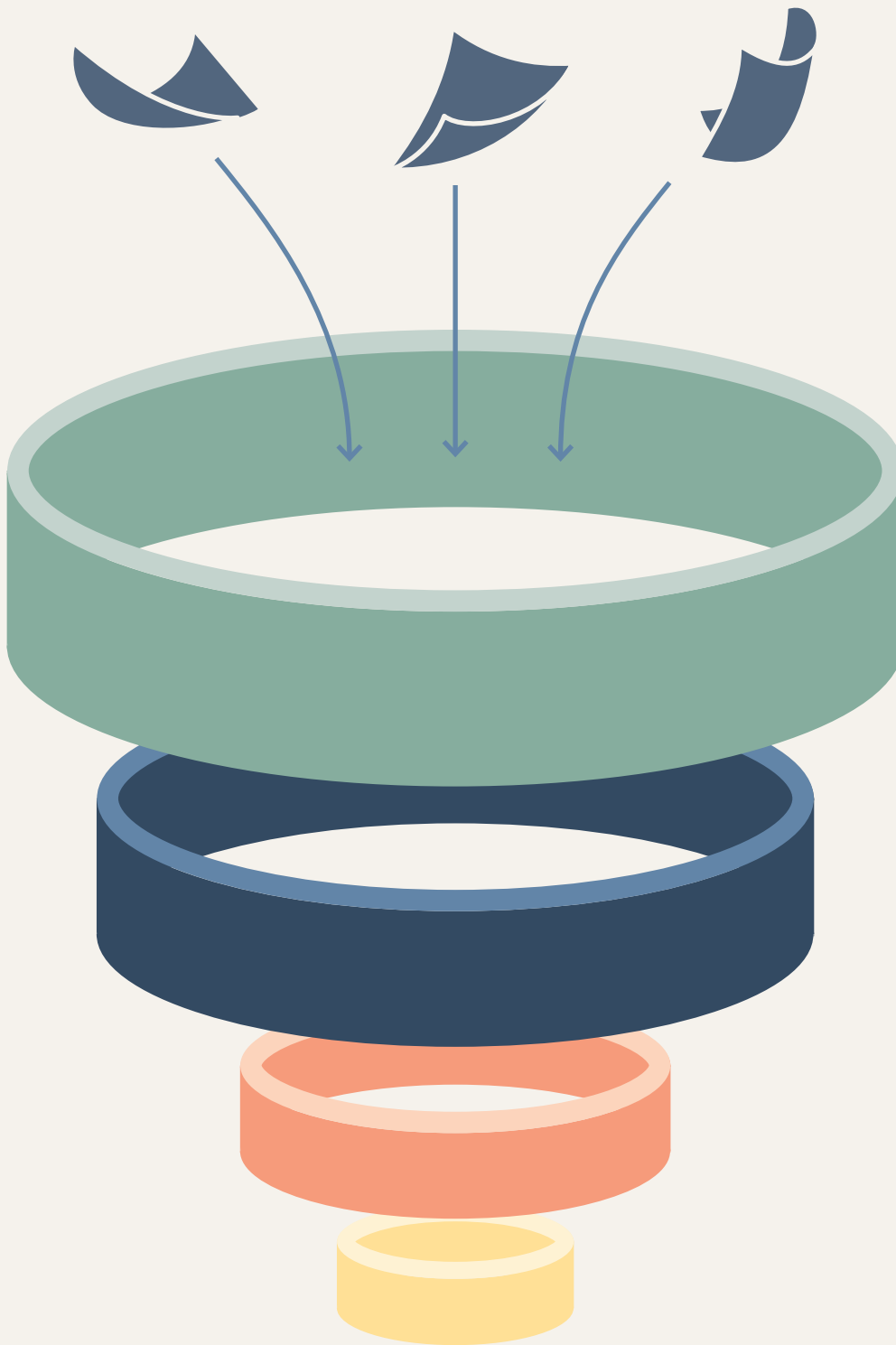
Evidence guidance clarifies some of ESSA's requirements and provides recommendations for evidence use in practice

To help states apply ESSA evidence tiers to activities funded under the law, in 2016 the U.S. Department of Education issued non-binding, non-regulatory guidance for using the evidence tiers in practice. The guidance references the standards of the What Works Clearinghouse (WWC), an initiative of the Department's Institute of Education Sciences that reviews and summarizes evidence of effectiveness of education interventions. For this review, the most relevant aspects of the Department's guidance are:

- **Definition of “well-designed and well-implemented.”** For Tiers I and II, the guidance further defines the meaning of the law's “well-designed and well-implemented” language. Well-designed and well-implemented Tier I evidence could meet the WWC's highest standard (meets standards *without* reservations), while well-designed and well-implemented Tier II evidence could meet the WWC's next-highest standard (meets standards *with* reservations).

FIGURE 2.1

Flow of manuscripts and studies through the review process



The Process

SEARCH

The literature search identified **10,153 manuscripts** to be screened for relevance.

INITIAL SCREENING

After an initial screen for relevance, **690 studies** received a full-text review to determine eligibility.

FULL-TEXT REVIEW

The full-text review identified **216 studies** eligible for review.

Studies were excluded at this stage because of an ineligible program type (158 studies), research design (185 studies), study sample or outcomes (86 studies), or language, location, or program year (45 studies).

ESSA EVIDENCE TIER RATING

148 studies (of 216 reviewed) met Tiers I-III Cause-and-Effect requirements, and **58 studies** met Tier IV Cause-and-Effect requirements. (One study did not meet Tiers I-IV, and nine studies had insufficient information to assign a rating.)

- **Broad application.** The Department recommends that Tier I and II evidence have broad application beyond a small group of students and a single site. Specifically, the Department suggests that evidence for Tiers I and II come from a study or a combination of studies involving at least 350 students and more than one site. Although guidance stops short of defining “site,” the Department has used “school district” as the definition of site in its own tiered-evidence grant competitions, and we use that definition in this report.
- **Relevance to the population and/or context where the program will be implemented.** The Department recommends that Tier I evidence come from studies that were conducted with a similar population *and* context to those where the program will be offered. Tier II evidence should come from a similar population *or* context. Determining whether a population or context is similar is a judgment call, and the Department does not define population and context further. In its own grant competitions, however, the Department has considered *population* to reference student characteristics and context to reference elements of the setting (for example, urban or rural, school or non-school location) that may be relevant to the effectiveness of the program.
- **Overall effectiveness and absence of harm.** The Department makes two recommendations that seek to steer decision-makers away from selectively choosing research results with the most favorable outcomes for a program. The first recommendation is to consider the overall effectiveness of the program by examining the full body of evidence from studies that are good tests of cause-and-effect. The second recommendation is to consider whether any improved outcomes are “overridden” by negative results—in other words, whether any negative results would cast doubt on the overall potential of the program to improve outcomes for students.

Table 2.1 summarizes ESSA’s requirements and the Department’s recommendations and introduces vocabulary that we use throughout this report to describe the extent to which studies and programs fulfill ESSA’s requirements and the Department’s recommendations. The first column presents the word or phrase that we use as shorthand for the requirement or recommendation, and subsequent columns explain the key question, source, and specifics of each requirement or recommendation, as well as the tiers to which it applies.

BOX 2.3

Definition of Evidence-Based in ESSA

In Title VIII, Sec. 8002(21)(A), ESSA defines evidence-based as:

“an activity, strategy, or intervention that

(i) demonstrates a statistically significant effect on improving student outcomes or other relevant outcomes based on—

(I) strong evidence from at least 1 well-designed and well-implemented experimental study;

(II) moderate evidence from at least 1 well-designed and well-implemented quasi-experimental study; or

(III) promising evidence from at least 1 well-designed and well-implemented correlational study with statistical controls for selection bias; or

(ii) *

(I) demonstrates a rationale based on high-quality research findings or positive evaluation that such activity, strategy, or intervention is likely to improve student outcomes or other relevant outcomes; and

(I) includes ongoing efforts to examine the effects of such activity, strategy, or intervention.”

* Part (ii) provides the definition of Tier IV (Demonstrates a Rationale).

TABLE 2.1

Summary of evidence requirements from ESSA and recommendations from U.S. Department of Education guidance

Shorthand Title	Key Question	Source	Requirement or Recommendation	Notes
WHAT WE PRESENT IN THIS REPORT				
Cause-and-Effect	Does the study provide a credible assessment of whether the program has an effect on student outcomes?	ESSA	Study must have internal validity appropriate to its tier.	This requirement applies to Tiers I-III only.
	Does the evidence meet What Works Clearinghouse (WWC) standards?	Guidance	Tier I-II studies should meet WWC standards with or without reservations.	This recommendation applies to Tiers I-II only.
Improved Outcome	Did the program improve any student outcomes or other relevant outcomes?	ESSA	There must be at least one statistically significant improvement.	This requirement applies to Tiers I-III only.
Broad Application	Has the program demonstrated its effectiveness with a sufficiently large group of students and in multiple places?	Guidance	A study (or studies) should involve a sample of at least 350 students and more than one site (school district).	This recommendation applies to Tiers I-II only.
ADDITIONAL DETERMINATIONS TO BE CONSIDERED BY DECISION-MAKERS				
Similarity	Has this program improved outcomes for similar students or in a similar context?	Guidance	Evidence should be from a similar population (of students) and/or context (for example, urban v. rural, type of education setting).	This recommendation applies to Tiers I-III only.
Overall Effectiveness	Considering all the evidence from strong cause-and-effect studies of this program, how effective is the program?	Guidance	Decision-makers should consider the overall body of evidence on the program.	This recommendation applies to Tiers I-III only.
Absence of Harm	Is there any evidence from strong cause-and-effect studies that this program harms students?	Guidance	There should be no negative findings that would cast doubt on the overall benefit of the program for students.	This recommendation applies to Tiers I-III only.

In this report, we provide information for whether a program has evidence that meets the first three criteria (Cause-and-Effect, Improved Outcome, and Broad Application). For the other three criteria (Similarity, Overall Effectiveness, and Absence of Harm), each of which requires professional judgment, we provide information that decision-makers can use to make their own determination. This information is available in a companion evidence guide.

To determine whether a study meets Cause-and-Effect criteria for Tiers I and II, we were able to follow the Department's guidance, which references an elaborated set of standards from the What Works Clearinghouse. However, because there are no such standards for Tiers III and IV, we had to further operationalize the guidance for this review (see Box 2.4 and the Appendix for more detail).

BOX 2.4

Further definition of Tiers III and IV

The Department of Education's evidence guidance provides additional clarification about which types of studies can meet Cause-and-Effect criteria for Tiers III and IV. However, the guidance does not provide a detailed roadmap or set of standards for these tiers. For this review, we developed a set of review criteria, guided by the practicalities of systematic review methods and what we believe is likely to be most useful to decision-makers.

HOW WE DEFINED TIER III

ESSA requires that Tier III evidence be from a "well-designed and well-implemented correlational study with statistical controls for selection bias." In this review, we require that Tier III studies use a quasi-experimental design, like the Tier II requirement. However, we allow baseline differences between treatment and comparison groups to be higher than those for Tier II if the study authors made appropriate adjustments. Outcome measures must meet the requirements specified in the What Works Clearinghouse Standards Handbook (Version 4.0). More detail about how study ratings are determined is provided in the Appendix.

HOW WE DEFINED TIER IV

Tier IV presented special challenges. The Department's evidence guidance recommends that a Tier IV rating is justified when a program has "a well-specified logic model... informed by research or an evaluation that suggests how the intervention is likely to

improve relevant outcomes." It was not feasible to locate and evaluate the large number of logic models that might be defined for the diverse set of programs included in our review. It was also difficult to imagine a straightforward way to translate such information into a useful evidence report.

Despite these challenges, we wanted to include in this review information about comparison group studies that could provide suggestive evidence at Tier IV that a program might have positive effects. This evidence could suggest that a program is worthy of implementation and study with a more rigorous design that could meet Tiers I-III Cause-and-Effect requirements. To characterize Tier IV studies, we decided to use standards that follow from the definitions of Tiers I-III. By our definition, Tier IV studies may have large differences in pre-treatment characteristics between treatment and comparison groups that are not statistically adjusted, and they may employ outcome measures without demonstrated reliability.

An additional purpose of the Tier IV category is to provide decision-makers with a comprehensive assessment of comparison group studies of afterschool programs that may come their way from colleagues or vendors. This compilation of Tier IV studies enables decision-makers to double check whether this review overlooked any comparison group studies.

How we report study findings and program information

In this document, we provide a list of the afterschool programs with at least one study meeting Tier I-III Cause-and-Effect requirements and showing at least one statistically significantly improved outcome (see Chapter Four, Table 4.1). A separate evidence guide includes detailed information about these programs, descriptions of studies meeting Tier IV criteria, a summary of studies of school-sponsored extracurricular programs, and a summary of studies to which we were unable to assign a tier rating.

Studies that provide evidence of Cause-and-Effect at Tiers I-III

Studies that meet Cause-and-Effect requirements for Tiers I-III are presented individually, organized by program, in a companion evidence guide. Because it is important for decision-makers to have a complete picture of the evidence, we present findings regardless of whether the study showed improved outcomes, negative outcomes, or no effect of the program. We also provide detailed information about each measure and comparison.

To help readers understand the overall effectiveness of a program, as recommended by the Department guidance, we provide summaries for each outcome domain using (1) a statement of the overall consistency and direction of effects and (2) an average effect size. This summary is particularly useful for outcome domains that include multiple outcome measures.

For each study, we summarize overall effects in each of the outcome domains studied with one of four descriptors. When there is more than one study of a program, we provide these descriptors for all studies of the program, combined. To contribute to an outcome domain descriptor, the outcome must meet Tier I-III Cause-and-Effect requirements. More detail about how these descriptors are determined appears in Box 2.5.

- **Positive Effect.** The study found at least one improved outcome and no overriding contrary evidence.
- **Mixed Effects.** The study found a mix of improved and null or negative outcomes.
- **No Effect.** The study found neither improved nor negative outcomes.
- **Negative Effect.** The study found at least one negative outcome and no overriding contrary evidence.

Studies of school-sponsored extracurricular programs at Tiers I-IV

In our search, we identified studies of school-sponsored extracurricular programs that met our eligibility criteria. These studies examine the effects of participation in any of a school's extracurricular offerings, rather than participation in a specific activity. Because these studies contain few details about the components of this extracurricular programming, we describe the studies in brief paragraphs in Appendix EG-1 of the companion evidence guide, rather than in the more detailed program summaries. Each description includes the study's ESSA tier rating.

Comparison group studies that meet Tier IV Cause-and-Effect requirements

Forty-eight studies of afterschool programs use comparison group designs and meet this review's Tier IV Cause-and-Effect requirements. These studies do not provide a strong test of the effect of afterschool programs, although they are stronger than studies without comparison groups. To provide a full accounting of comparison group studies of afterschool programs, we briefly summarize studies that meet Tier IV criteria in an annotated bibliography (Appendix EG-2 of the companion evidence guide). A program with a Tier IV study with positive effects could be implemented and studied for effectiveness using a research design that could meet Tiers I-III.

Studies that cannot be assigned an ESSA tier

In a few cases, studies met Tier I-III Cause-and-Effect requirements, but the study authors did not report all the information needed to assign an ESSA evidence tier. For example, some authors did not report information on the statistical significance of a finding or provide enough information that would allow others to compute the statistical significance. Without information on whether a finding is statistically significant, the tier cannot be determined. These studies are described briefly in Appendix EG-3 of the companion evidence guide.

BOX 2.5

Definitions of overall effectiveness

To summarize the overall effectiveness of programs for improving specific outcome domains, this review uses the following definitions. These definitions are adapted from the criteria used by the What Works Clearinghouse to determine effectiveness ratings (see *What Works Clearinghouse Procedures Handbook v. 4.0*, Table IV.3).

POSITIVE EFFECT. In this review, an outcome domain is described as demonstrating a positive effect when:

- There is only one outcome in the domain, and that outcome shows a statistically significant improvement, OR
- There are at least two statistically significant improved outcomes and no statistically significant negative outcomes.

MIXED EFFECTS. In this review, an outcome domain is described as demonstrating mixed effects when:

- There is at least one statistically significant improved outcome and at least one statistically significant negative outcome, but there are more improved than negative outcomes, OR

- There is at least one statistically significant improved outcome, but there are more “no effect” (not statistically significant) outcomes than improved outcomes.

NO EFFECT. In this review, an outcome domain is described as demonstrating no effect when:

- There are no statistically significant improved outcomes nor statistically significant negative outcomes.

NEGATIVE EFFECT. In this review, an outcome domain is described as demonstrating a negative effect when:

- There is at least one statistically significant negative outcome and no statistically significant improved outcomes, OR
- There is at least one statistically significant negative outcome and at least one statistically significant improved outcome, but there are more negative than improved outcomes.

The Landscape of Afterschool Studies

In this chapter, we provide a broad overview of the characteristics of studies that examine the effects of afterschool programs. We describe the types of afterschool programs that these studies examine, as well as the studies' outcomes, outcome measures, and sample and site characteristics. Although all studies were published in 2000 or later, we also present the study publication dates and the years when the programs were implemented to narrow down the timeframe of implementation and publication.

Having evidence of the effectiveness of afterschool programs requires both (1) effective programs and (2) studies of those programs that are strong tests of cause-and-effect. In addition, these studies must be able to detect any positive outcomes with some precision. For these reasons, a clear understanding of the universe of afterschool research can help the reader interpret the findings in Chapter Four, including findings for how many studies

meet the statutory requirements of ESSA and the Department of Education's recommendations for applying ESSA evidence tiers.

Chapter Two explained how, to produce this report, we conducted a comprehensive search for studies that investigate the effects of afterschool programs. We winnowed down an initial set of over 10,000 references to a smaller set of 216 studies that met initial screening criteria for research design, program characteristics, and relevant outcomes.¹ Of these studies, 148 meet the Cause-and-Effect criteria for ESSA Tiers I, II, or III. Because this report primarily focuses on studies that meet these criteria, this chapter reports data only for these 148 studies.² In addition, because the purpose of this chapter is to explore the characteristics of studies that meet research design requirements, we include studies regardless of whether they find statistically significantly improved outcomes.

¹ As we describe in Chapter Two, we define a study as an investigation of the effects of a program for a specific sample of students, where the effects for that sample are reported separately. For example, if a research report presents results for elementary and middle school students separately, we count those results as two separate studies: one for elementary students and one for middle grades students.

² We identified an additional 58 studies that meet Cause-and-Effect requirements for ESSA Tier IV. An annotated bibliography of these studies is presented in Appendix EG-2 of the evidence guide that accompanies this review.

We underscore that the focus of this chapter is studies—not programs. The 148 studies that meet Tiers I-III Cause-and-Effect requirements represent fewer than 148 programs because the effects of some programs have been examined in more than one study. In Chapter Four, we turn our attention to what we learned about programs.

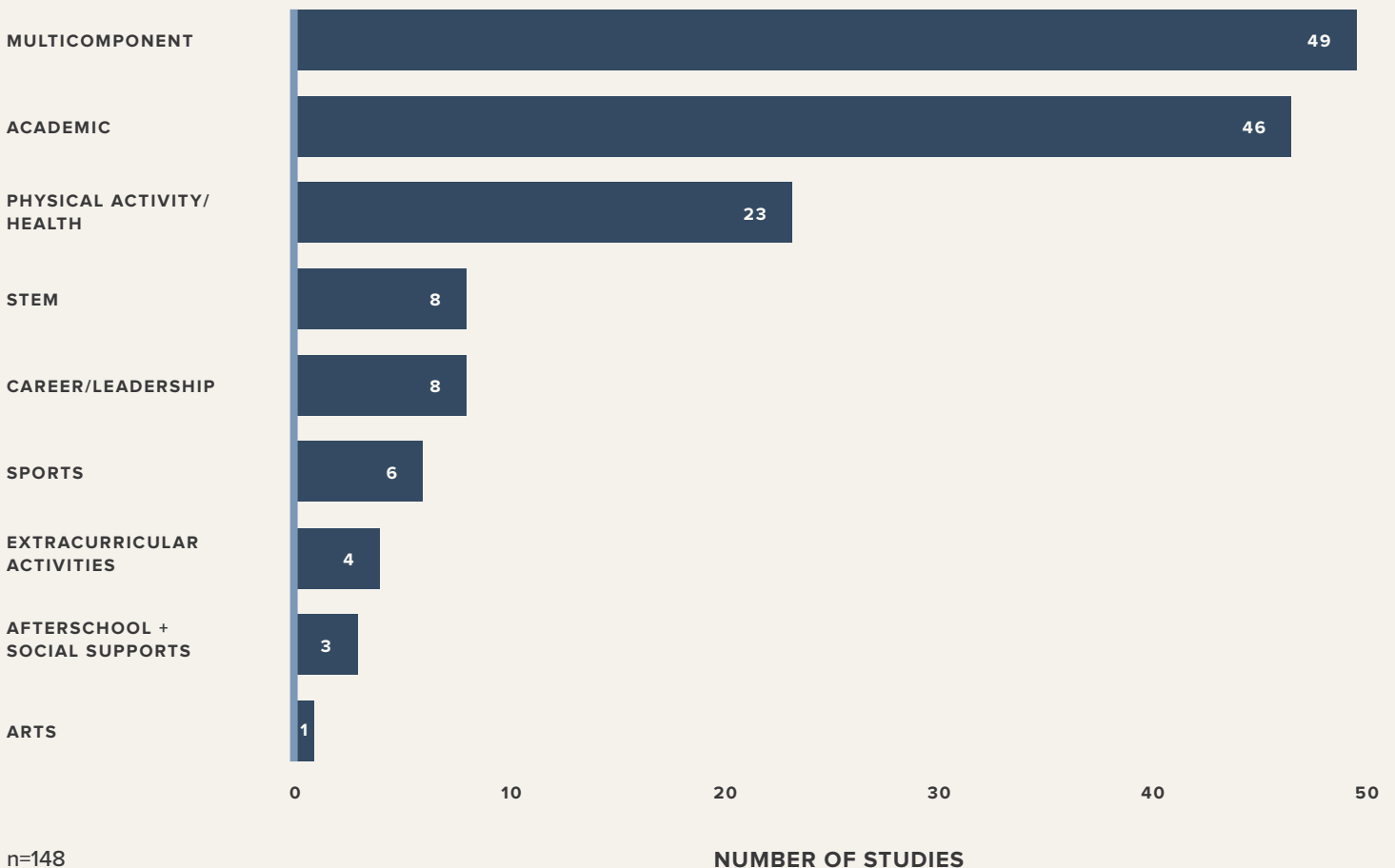
There are more studies of the effects of Multicomponent and Academic afterschool programs than of other types of programs.

We categorized the afterschool programs into nine types based on their stated purpose, primary activities, and the time that students spent on these activities (see Box 3.1 for a summary of these program types).

Forty-nine studies, or one-third of the total, are of Multicomponent programs (Figure 3.1). These afterschool programs offer multiple types of activities, typically including academic support, sports, and arts and crafts. An additional 46 studies examine the effectiveness of Academic programs, in which the majority of time is spent on tutoring or homework help, even though recreational or enrichment activities also may be offered. In all, then, almost two-thirds of the studies focus on one of these two types of programs.

FIGURE 3.1

Number of studies meeting Cause-and-Effect requirements for ESSA Tiers I-III, by program type



BOX 3.1

Afterschool program types

ACADEMIC

In these programs, students may receive tutoring in academic content areas and help with homework. Although the program may include other activities, such as recreation and enrichment, most of the time is spent on academic activities.

AFTERSCHOOL PLUS SOCIAL SUPPORTS (VARIANT OF MULTICOMPONENT)

These programs incorporate case management and more extensive family involvement in addition to a multicomponent program.

ARTS

These programs primarily focus on introducing students to one or more art forms (for example, visual arts, dance, theater, music) and/or developing their artistic skill.

CAREER/LEADERSHIP (VARIANT OF MULTICOMPONENT)

Career/Leadership programs are multicomponent programs that focus on career development, postsecondary readiness, and/or leadership development.

MULTICOMPONENT

Multicomponent programs offer multiple types of activities, and no type of activity dominates the time youth spend in the program. Multicomponent programs typically offer academic support, sports, and arts and crafts programming, and youth can experience multiple of these components.

PHYSICAL ACTIVITY/HEALTH

In these programs, most of the time is focused on healthy living and physical activity.

SCHOOL-SPONSORED EXTRACURRICULAR ACTIVITIES

There are numerous studies of the effects of any extracurricular participation, without assessing the effects of participation in any specific activity. We group these studies under the heading “extracurricular participation.” Studies of the effects of specific programs (for example, robotics teams) are assigned to other program types reflecting their focus.

SPORTS

Sports programs have a competitive component. Examples include running clubs and school-sponsored team sports.

STEM

In STEM programs, participants develop their interests in science, technology, engineering, and/or mathematics-related topics. Improving school academic performance is generally not the focus of these programs; instead, the programs seek to engage students in hands-on, active activities that nurture an interest in STEM fields. Examples include robotics teams and math clubs.

Less commonly studied are Physical Activity/Health programs, with 23 studies, or 15 percent of the total. There are fewer than 10 studies of programs in each of the remaining categories: STEM, Career/Leadership, Sports, Afterschool Plus Social Supports, Extracurricular Activities, and Arts.

We cannot tell from these data whether there are more studies of Multicomponent and Academic programs because these programs are more commonly offered than others. It is also possible that more Multicomponent and Academic programs had a requirement and resources for evaluations of effectiveness. We observe that many of these Multicomponent and Academic studies were of programs funded by the U.S. Department of Education's 21st Century Community Learning Centers program, which may mean that there were both more programs of these types and that more of them had evaluation requirements.

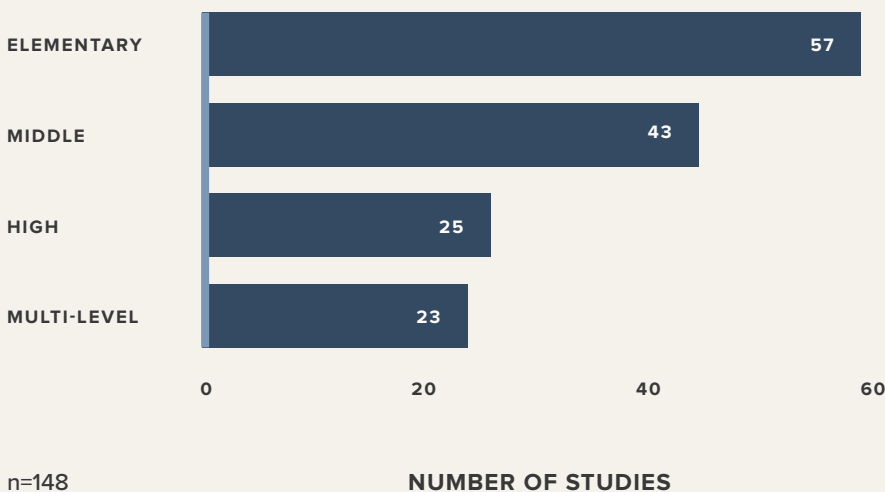
There are at least 25 studies at each grade level, though there are twice as many studies of elementary programs as high school programs.

We identified at least 25 studies at each grade level (elementary, middle, and high) that meet the Cause-and-Effect requirements for ESSA Tiers I-III (Figure 3.2). Studies of afterschool programs for the elementary grades are most common (57 studies, or 39 percent of the total). There are 43 studies in the middle grades (29 percent of the total). In sum, about two-thirds of the studies are of programs for elementary or middle grades students.

Twenty-three studies (16 percent) are of programs that served multiple grade levels (and did not estimate program effects separately for students at different grade levels). Grade level combinations include elementary-middle, middle-high school, and even elementary-high school. Studies of large statewide or districtwide afterschool initiatives are particularly likely to combine results across multiple grade levels.

FIGURE 3.2

Number of studies meeting Cause-and-Effect requirements for ESSA Tiers I-III across grade levels served



About 30 percent of studies have samples of fewer than 100 students, and over half include fewer than 350 students.

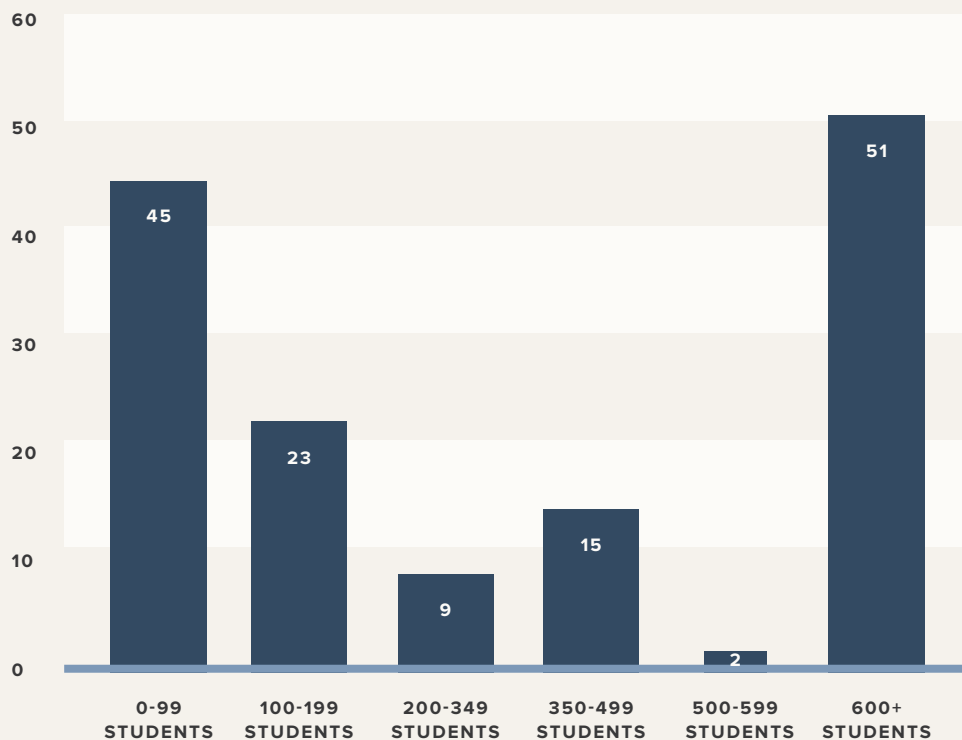
Among the studies meeting Cause-and-Effect criteria for ESSA Tiers I-III, sample sizes range from 11 students to over 200,000 students, with a mean of 6,604 and a median of 264. The large difference between the mean and median is due to several studies with very large samples of over 10,000 students. However, more than half (53 percent) of the studies with known sample sizes³ have fewer than 350 students. Fifty-three studies (37 percent) have sample sizes of 500 students or more, and 45 (31 percent) have sample sizes of 100 or fewer (Figure 3.3).

Sample sizes are important to keep in mind when reviewing findings of effectiveness of afterschool programs. In small studies, it is more difficult to detect a statistically significant program effect, even if the program being studied is an effective one. That is, a program may in fact be effective, but the statistical tests are not powerful enough to see the impact. This is because small samples produce less precise estimates. A study with a sample that is too small to detect effects of a given magnitude is described as *underpowered* and is more likely to find no differences between treatment and comparison groups even when differences are present (see Box 3.2 for more explanation of sample size and statistical power).

³ Three studies did not report sample sizes.

FIGURE 3.3

Sample sizes of studies meeting Cause-and-Effect requirements for ESSA Tiers I-III



n=145

NUMBER OF STUDIES AT GIVEN SAMPLE SIZES

To meet Tiers I-III, ESSA requires that a study not only meet Cause-and-Effect criteria but also show at least one statistically significant positive effect of the program. The small sample sizes of a substantial subset of afterschool studies suggest that these studies may be underpowered—that is, they would not be able to detect a statistically significant effect even if the program were effective. In addition, as described in Chapter Two, in guidance about how to apply the ESSA evidence tiers, the U.S. Department of Education recommends that evidence for Tiers I and II come from a study (or studies) with a total sample size of at least 350. As Figure 3.3 shows, less than half of the afterschool studies we identified meet that sample size recommendation.

Mathematics Achievement and Reading/English Language Arts (ELA) Achievement are by far the most common outcome domains studied.

Each study in this review examines the effects of afterschool programs in at least one student outcome domain, and most studies examine outcomes in more than one domain. *Outcome domains* classify related outcomes into more general categories (see Chapter Two for more detail about how we defined outcome domains). Like studies, outcome domains can be characterized as meeting Cause-and-Effect criteria for Tiers I-III. On average, each study in this review has two outcome domains meeting these criteria. Some studies include additional outcome domains that do not meet the criteria.

BOX 3.2

Understanding statistical power

Studies conduct tests of statistical significance to assess the likelihood that any differences observed are due to chance. By using statistical significance tests, studies are making inferences from their samples to a larger population from which the sample is drawn. *Statistical power* is the ability of a study to detect an effect of a given magnitude with a certain amount of precision.

Consider a simple example of a study conducted within one hypothetical school. The sample size (treatment and control students) needed to detect effects of the magnitude commonly found in studies of afterschool programs is as follows:

- To detect an effect of .15, the study would need a sample of about 425 students.
- To detect an effect of .20, the study would need a sample of about 240 students.
- To detect an effect size of .25, the study would need a sample of about 150 students.
- To detect an effect size of .30, the study would need a sample of about 100 students.

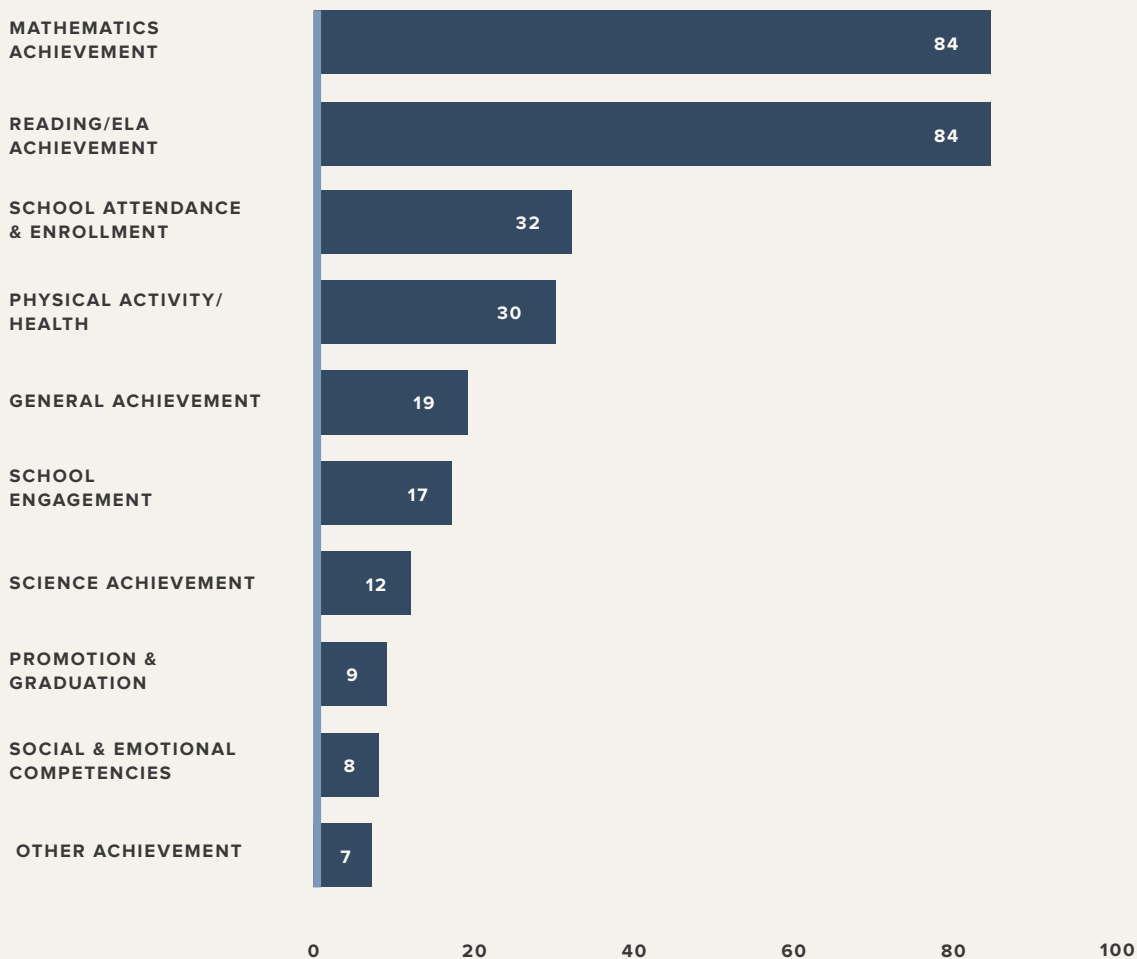
About 30 percent of the studies would have the statistical power to detect only effects that are considered large in education (effect sizes greater than .30).

Within each outcome domain, we looked at the number of studies with results meeting Cause-and-Effect criteria (Figure 3.4). Eighty-four studies (57 percent) investigate Mathematics Achievement, and the same number examine Reading/ELA Achievement. In separate analyses, we found that about 70 of the studies examining either Mathematics Achievement, Reading/ELA Achievement, or both are of Multicomponent or Academic programs, which often have improvement of academic outcomes as a primary goal.

Thirty-two studies (22 percent) examine Attendance and Enrollment, and almost that many studies investigate Physical Activity/Health outcomes (30 studies, or 20 percent). Fewer than 20 studies examine Science Achievement, School Engagement, Promotion and Graduation, Social and Emotional Competencies, or other types of achievement.

FIGURE 3.4

Number of studies with outcome domains meeting Cause-and-Effect requirements for ESSA Tiers I-III



State standardized tests were by far the most common outcome measures.

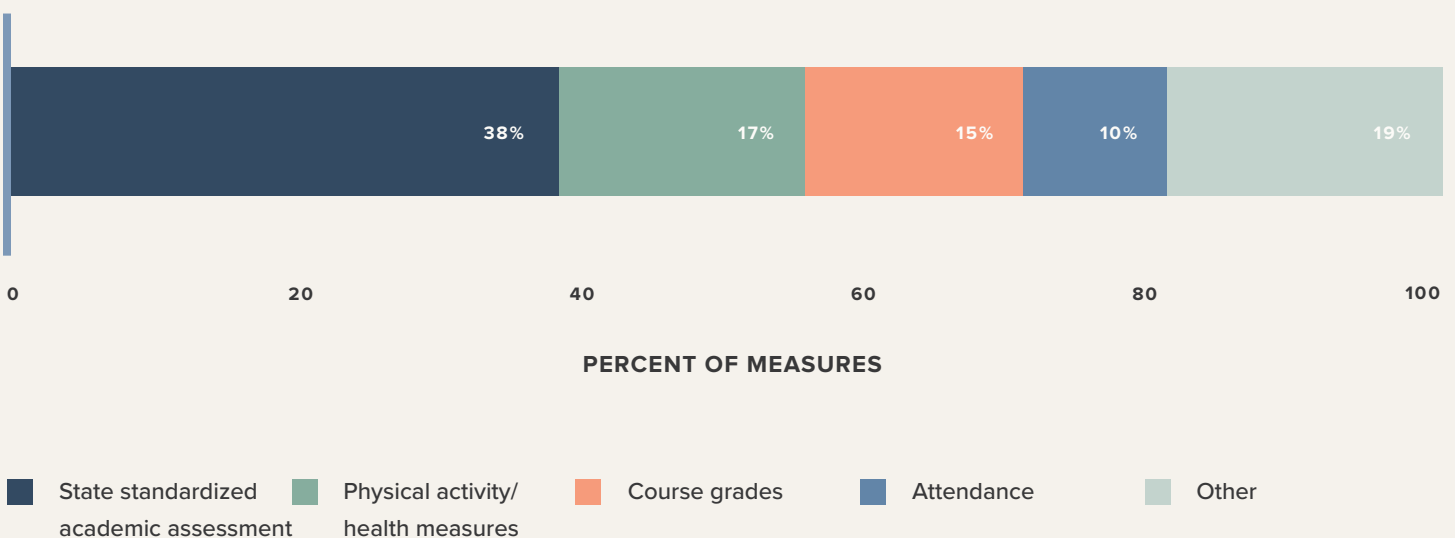
Within any outcome domain, a study can investigate multiple outcomes. For example, a study can examine two outcomes in the Mathematics Achievement domain: (1) a score on a mathematics achievement test and (2) a course grade in mathematics. To understand the most commonly used types of outcomes, we identified the specific measures used across the studies for any comparison meeting Cause-and-Effect criteria for Tiers I-III. We then summed the number of studies that use each of those measures and categorized the measures into types.⁴ We found that, on average, studies report on four measures.

Thirty-eight percent of the measures used are state standardized tests in reading, mathematics, science, or other content areas (Figure 3.5), making these the most commonly used type of measure. Physical activity/health measures were the next most common (17 percent), followed by course grades (15 percent) and attendance (10 percent). Measures that are easily available from school records—test scores, grades, and attendance, for example—make up at least two-thirds of all measures. Several other measures available from school data—Promotion and Graduation, for example—are included in “Other.” The “Other” category also includes measures of behavior and social and emotional competencies, among others.

⁴ For example, a study that uses the New Jersey state standardized test for reading and the New Jersey state standardized test for mathematics was counted as having two outcome measures. A second study, using the New Jersey state standardized test for math and course grades in math, would also be counted as having two measures. Together, these studies would contribute four measures to the total.

FIGURE 3.5

Categories of outcomes examined



NOTE: Percentages do not sum to 100% due to rounding.

The types of outcome measures used are important to keep in mind when reviewing findings of effectiveness of afterschool programs. This is because it may be more feasible for any positive effects of afterschool programs to manifest themselves in certain kinds of outcomes—for example, homework completion. Conversely, it is difficult to show impacts on some outcomes, such as standardized tests, which even intensive school-day academic interventions can struggle to change.

Outcome domains vary in the percentage of statistically significant improved outcomes.

We examined each outcome domain separately to understand how often studies find statistically significant improved outcomes. To do this, we considered only outcome domains meeting Cause-and-Effect criteria for Tiers I-III (Figure 3.6). The Promotion and Graduation outcome domain had the highest percentage of improved outcomes (five studies with improved outcomes, of 10 studies), followed by Physical Activity/Health (14 studies with improved outcomes, of 30 studies). One-third of the studies investigating Mathematics Achievement found a positive effect (28 of 84 studies), as did approximately one-quarter of the studies that examined Reading/ELA Achievement (22 of 84 studies).

As we have already observed in this chapter, finding a statistically significant result is related not only to the effectiveness of the program but also to the study sample size and the types of outcome measures used. We cannot say how these factors combined to produce the variation in improved outcomes across domains. What does seem clear is that afterschool studies more often find improvements in outcomes like graduation, physical activity/health, attendance, and engagement than academic achievement.

Average impacts of afterschool programs are positive across most outcome domains.

In addition to examining the percentage of studies with at least one statistically significant improved outcome (see Figure 3.6), we also looked at the size of program impacts within each outcome domain. These effect sizes give us a sense of how much afterschool programs impact student outcomes, on average. To conduct this analysis, we examined the average effect size and confidence interval⁵ for each outcome domain. The analysis includes only findings that meet Tier I-III Cause-and-Effect criteria.

⁵ The confidence interval is a measure of the precision with which we can estimate an average effect size. Smaller confidence intervals indicate greater precision. In addition, confidence intervals that include zero indicate that the average effect is statistically indistinguishable from zero (no effect).

Of the 10 outcome domains, seven have average effects that are positive and statistically significant (Figure 3.7). The domains without a positive, statistically significant average effect are Science Achievement, Social and Emotional Competencies, and Other Achievement. These domains also have fewer studies and findings than other domains, which contributes to the estimates being less precise.

Average effects of afterschool programs are largest in the domains of Attendance and Enrollment, School Engagement, and Promotion and Graduation, ranging from .12 to .18. The domains of Reading/ELA Achievement and Mathematics Achievement have smaller average effects (.05 to .06), but nevertheless are statistically significant. The smaller size of achievement outcomes relative to other domains in this review is not unexpected because of the many other influences on student achievement. Although there are no hard-and-fast rules about what constitutes a meaningful effect size, we suggest that effect sizes of .05 - .06 indicate a small effect and .12-.18 indicate a small-to-moderate effect.

FIGURE 3.6

Percentage of studies with an improved outcome, by outcome domain

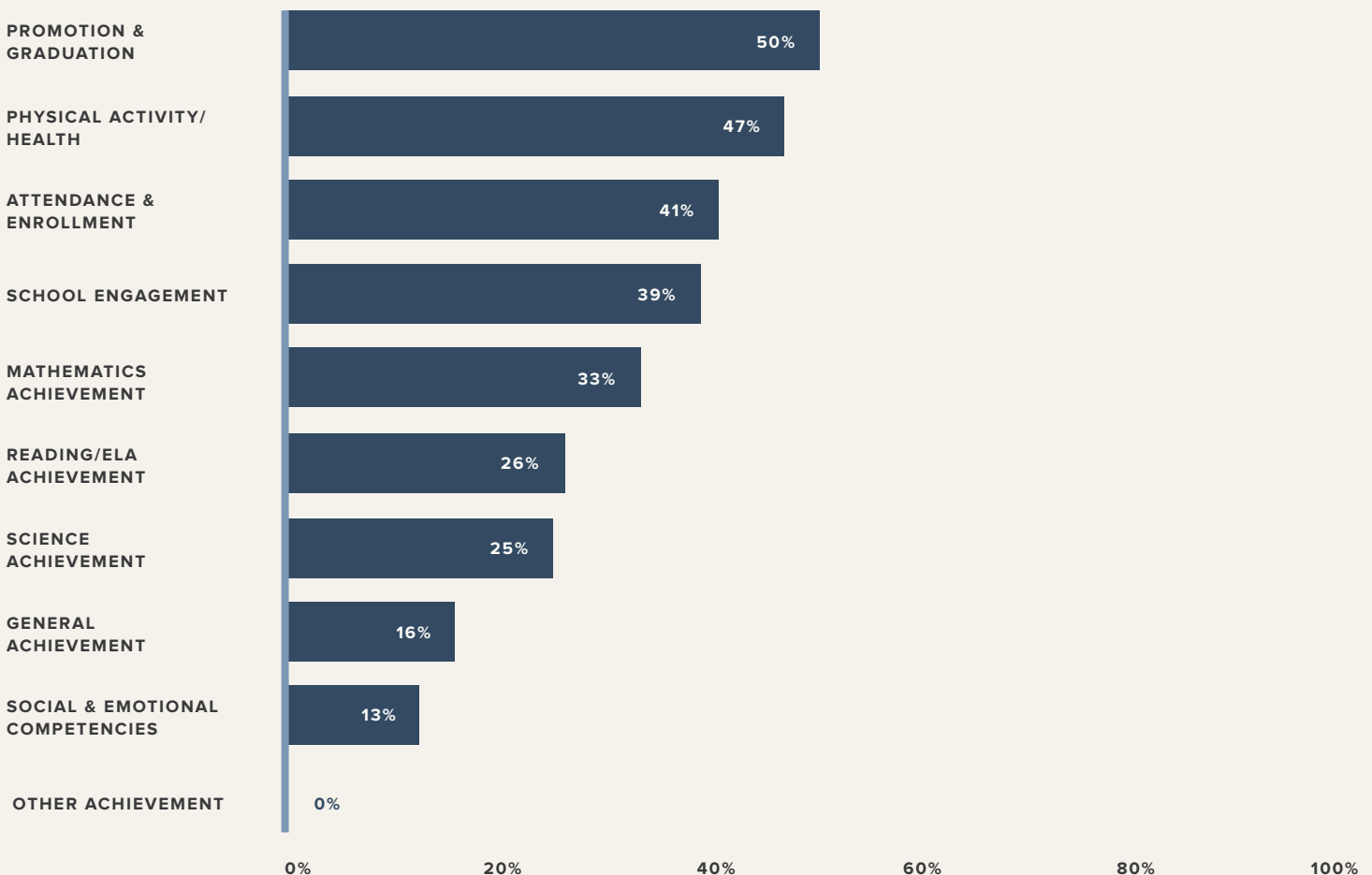
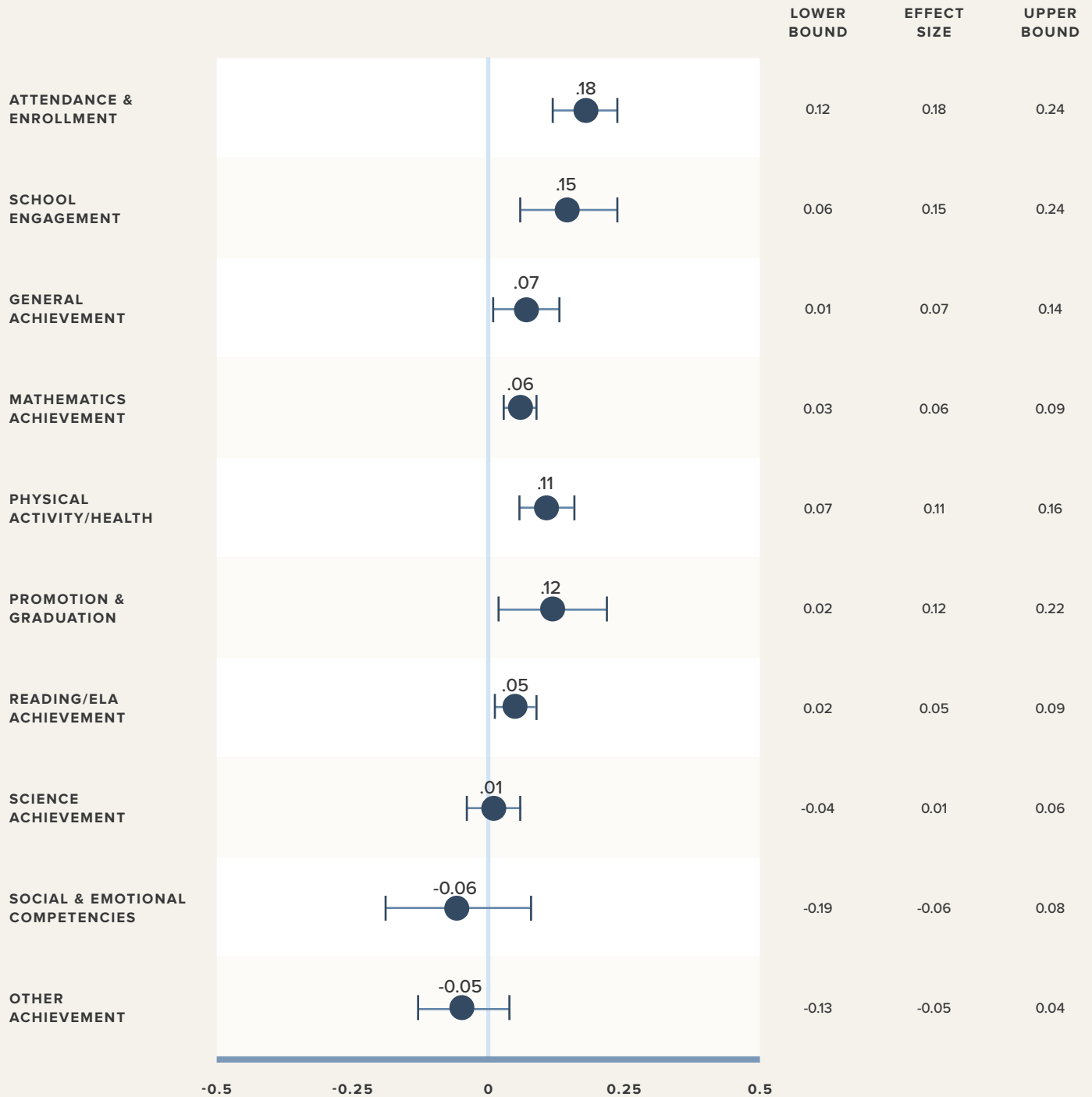


FIGURE 3.7

Average effect sizes and confidence intervals by outcome domain



Notes: Each mean effect size and confidence interval was computed with a multi-level weighted random effects model that accounts for the presence of multiple outcomes in the same domain within studies. Confidence intervals that do not cross 0 indicate that an effect size is statistically significant.

Most programs were implemented in 2006 or later, and most studies were published during that time frame.

This review was limited to research published in 2000 or later. However, to understand when the literature was published within that time frame, we examined the studies' publication dates. We also examined the years when the programs were implemented. It is important to look at the recency of this literature to assess its relevance to contemporary decision-makers.

Overall, we find that both publication dates and implementation years are relatively recent. The number of studies of afterschool effects grew from 27 in 2000-05 to 63 in 2011-15, an increase of about 130 percent (Table 3.1); as a result, more than half of the studies in this review were published in 2011 or later.⁶ Approximately one third of the studies were reported in dissertations, one third in journal articles, and one third in reports (primarily from school districts, states, or research firms) (Table 3.2).

TABLE 3.1

Study publication year

Study publication year	
YEARS	NUMBER OF STUDIES
2000-2005	27
2006-2010	44
2011-2015	63
2016-2017	12
Unknown	2
Total	148

TABLE 3.2

Study publication format

Study publication format	
FORMAT	NUMBER OF STUDIES
Dissertation	52
Journal	47
Report	48
Conference paper	1
Total	148

⁶ We can tell that the two studies with unknown dates were published in our review timeframe because they reference programs implemented in 2000 or later.

Publication dates lag program implementation years, so even though eligible studies were limited to those published in 2000 or later, some of the programs were implemented prior to 2000 (13, or about 10 percent of the studies that reported program implementation dates). About 70 percent of the studies with implementation dates were of programs offered in 2006 or later (Table 3.3). We note that studies of Physical Activity/Health programs were especially likely not to report implementation dates.

TABLE 3.3

Program implementation years

Program Implementation Years	
YEARS	NUMBER OF STUDIES
Before 2000	13
2000-05	25
2006-2010	51
2011-now	37
Not reported	22
Total	148

Summary

In this chapter, we summarized characteristics of studies that meet Cause-and- Effect requirements for ESSA Tiers I-III; these are the “well-designed and well-implemented” studies that provide good evidence of whether afterschool programs improve student outcomes. It is important to understand the characteristics of these studies because their designs, sample sizes, and outcomes affect what can be known about program effectiveness.

We found that studies of Academic and Multicomponent programs are most common, although a wide range of afterschool program types and outcomes have been studied. The most typical outcomes are Mathematics and Reading/ELA Achievement as measured by state standardized tests, which are some of the most difficult measures on which to show effects in education. Further, there is a substantial subset of studies with small sample sizes that are likely unable to detect effects of the magnitude that might be expected from an afterschool program.

In the next chapter, we present information about the effectiveness of programs. For some programs with more than one study of effectiveness, this involves summarizing effects across studies.

Evidence-Based Afterschool Programs under the Every Student Succeeds Act

In this chapter, we describe what we learned about the effectiveness of afterschool programs when we apply evidence requirements from ESSA and guidance from the U.S. Department of Education. The overarching purpose of this chapter is to identify programs with evidence at ESSA Tiers I-III so that afterschool providers, policymakers, funders, and others can understand the range of programs eligible for funding under ESSA at the three highest tiers and assess current programmatic approaches against the evidence.

The focus of this chapter is *programs*, while the previous chapter examined *studies*. Although we identified 148 studies meeting Cause-and-Effect criteria for ESSA Tiers I-III (Chapter Three), these studies represent 124 programs. This is because some

programs have been studied more than once. In this chapter, we show how many afterschool programs are at each evidence tier when different criteria are applied and present the availability of evidence-based options for different program types and grade levels.

This chapter includes at-a-glance tables describing the programs that meet the evidence criteria for Tiers I-III and had at least one improved outcome, organized by program type and grade. Each of these programs, along with others that have strong research designs but no evidence of a positive effect, is described in more detail in program summaries presented in a companion evidence guide.

BOX 4.1

Why we use the term “program” in this review

We use the term “program” to describe sets of afterschool activities because it is more commonly used by the afterschool community than other terms used by ESSA and the U.S. Department of Education to describe what must be evidence-based. Further, the term is broadly consistent with the language used by ESSA and U.S. Department of Education guidance and with the lineage of these terms in the What Works Clearinghouse. ESSA does not use the term “program” but instead refers to “an activity, strategy, or intervention” (Section 8101(21)(A)). U.S. Department of Education evidence guidance uses the shorthand term “intervention,” which is also how the Department’s What Works Clearinghouse describes a “program, product, practice, or policy aimed at improving student outcomes” (What Works Clearinghouse, 2018).

What is a “program?”

In this report, we use the term “program” to refer to a set of afterschool activities carried out in a particular way with the intention of improving one or more student outcomes. Summarizing the effects of afterschool programs is complicated by the fact that there are both branded and unbranded programs in the field.

Branded programs have a recognizable name or “brand,” are marketed by a program developer, and may involve specific activities, training, staffing, and/or materials. Examples of branded programs that we reviewed are Girls on the Run (www.girlsontherun.org) and Building Educated Leaders for Life (BELL) (www.experiencebell.org). **Unbranded programs** are developed by school districts, nonprofits, and other providers. These programs may have general similarities to other programs of a given type (for example, Multicomponent programs) and may even have the same funding source (for example, 21st Century Community Learning Centers), but their specific components, staffing, emphasis, and duration may differ.

The vast majority of studies that we identified are of unbranded programs. While some programs had general similarities, we did not feel confident that their approaches were similar enough to justify classifying them as a single program. For example, we identified numerous studies of Supplemental Educational Services, afterschool tutoring programs offered under No Child Left Behind to students in underperforming schools. However, these programs varied in their approaches to tutoring and, most challenging of all, not all studies clearly describe the approach used by the program.

For all these reasons, we examine and report the effects of each unbranded program separately. For example, each Supplemental Educational Services program that meets Cause-and-Effect and Positive Result criteria for ESSA Tiers I-III is reported separately in the at-a-glance tables in this chapter and in a detailed program summary in a companion evidence guide. In our summaries of the number of programs meeting certain criteria, we count each of these programs individually.

Likewise, we do not combine results for programs that are identified as having been funded under the federal 21st Century Community Learning Centers program (21CCLC). We made this choice for two reasons. First, 21CCLC is a funding stream that supports many kinds of programs with different target populations and components. Second, we are not confident that all studies of 21CCLC programs report the program’s funding source. While we do not combine separately reported 21CCLC studies, we do review and report on studies that estimate impacts for groups of 21CCLC programs (for example, statewide studies of 21CCLC programs).

We realize that some readers who would prefer an overall effectiveness rating for large categories of programs may find this approach unsatisfactory. Some readers may believe that certain programs that we report individually should be combined instead. For these readers, the study details that we provide in a companion evidence guide provide the building blocks for an analysis of the effectiveness of broader categories of programs, such as Supplemental Educational Services.

We identified 124 programs with at least one study that meets the Cause-and-Effect requirements at Tiers I-III. Of these, 15 programs have studies meeting Tier I Cause-and-Effect requirements, 42 have studies meeting Tier II requirements, and 67 have studies meeting Tier III requirements (Figure 4.1). It is important to note that, even though these programs have studies meeting research quality requirements for Tiers I-III, not all improved outcomes. Meeting the Cause-and-Effect requirement is an indication only that the program has at least one study that can provide good evidence of *whether* the program is effective—it does not necessarily provide evidence *that* the program is effective.

We also identified 58 comparison group studies that did not meet Cause-and-Effect requirements for Tiers I-III but did meet those for Tier IV. These studies are summarized in Appendix EG-2 of the companion evidence guide. For a program to meet Tier IV requirements, it must also be undergoing evaluation of its effectiveness. The implication is that this evaluation should provide evidence at Tiers I-III.

Of the 124 afterschool programs with evidence meeting Cause-and-Effect requirements for ESSA Tiers I-III, half have at least one statistically significant improved outcome.

Sixty-two afterschool programs have at least one study with an improved outcome. Of these, the majority (42 programs, or 68 percent) have evidence meeting Cause-and-Effect requirements for ESSA Tier III. Seventeen programs (27 percent) have evidence at Tier II, and three programs (5 percent) have evidence at Tier I (Figure 4.1).

The percentage of programs with an improved outcome was higher for those with Cause-and-Effect evidence at Tier III (42 of 67 programs, or 63 percent) than at Tier II (17 of 42 programs, or 40 percent) or Tier I (3 of 15 programs, or 20 percent).

Programs without a statistically significant improved outcome could show (1) no statistically significant effects (neither positive nor negative), (2) statistically significant negative results, or (3) a mixture of both. However, few statistically significant negative results were reported in the studies that we reviewed. Although we cannot definitively explain why there is a small number of negative results, we think a probable reason for this phenomenon is that afterschool programs are unlikely to harm students in the school-related, physical activity/health, and social and emotional domains that are the focus of this review.

Programs that meet Cause-and-Effect requirements for Tiers I-III and improved at least one outcome are reported individually, by program type, in Tables 4.1 – 4.6 at the end of this chapter.

Almost all program types have one or more programs with evidence meeting Cause-and-Effect requirements for ESSA Tiers I-III and at least one statistically significant improved outcome.

Among program types, the Academic and Multicomponent categories have the largest number of programs meeting the Cause-and-Effect and Improved Outcome criteria (Figure 4.2). Sixteen Academic programs (37 percent of the 43 Academic programs with Cause-and-Effect evidence) have a study showing a statistically significant improved outcome, and 22 Multicomponent programs (59 percent of the 37 Multicomponent programs with Cause-and-Effect evidence) show an improved outcome. All program types except for Afterschool Plus Social Supports and Arts have at least one program with an improved outcome.

Four program types (Academic, Physical Activity/Health, Afterschool Plus Social Supports, and Arts) each have fewer programs with an improved outcome than programs with no improved outcomes. In fact, two of these types (Afterschool Plus Social Supports and Arts) have no programs showing improved outcomes. Academic programs are more likely than other programs to have academic achievement as their sole outcome domains, and, as we showed in Chapter Three, there are fewer improved outcomes for these measures (Figure 3.6).

FIGURE 4.1

Number of programs with evidence meeting Cause-and-Effect requirements, by tier and improved outcome

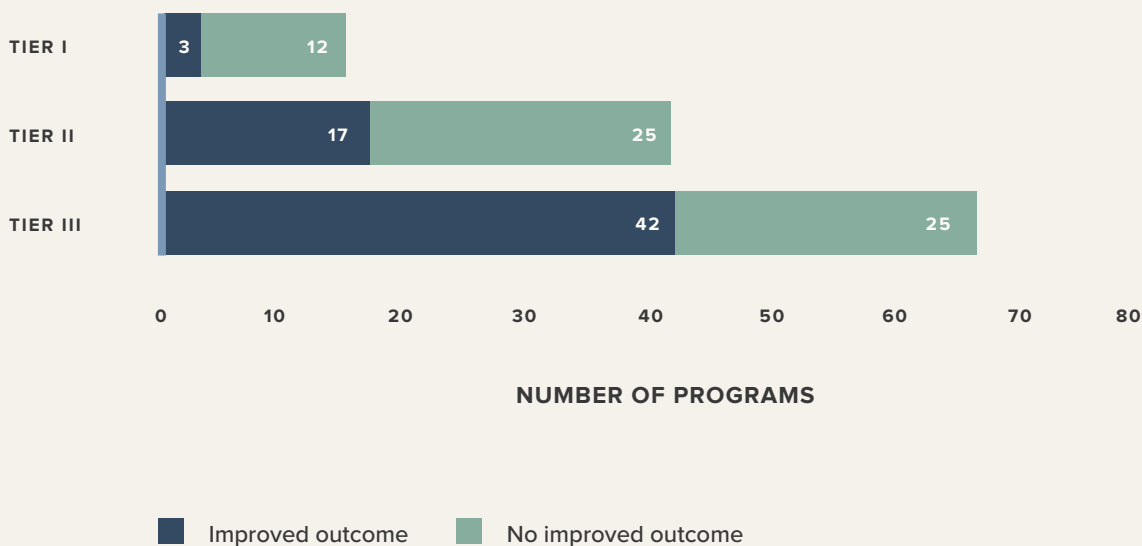
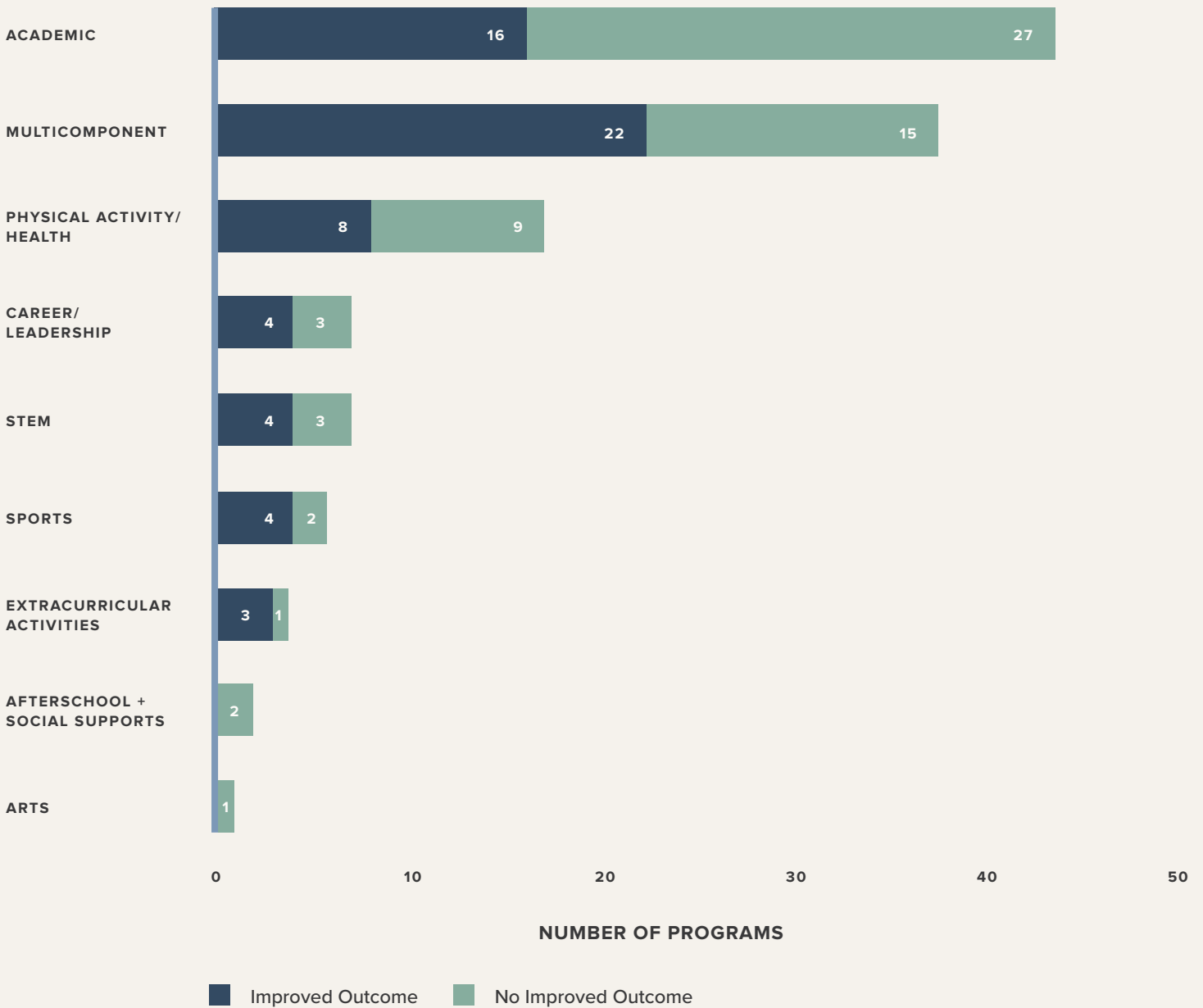


FIGURE 4.2

Number of programs with an improved outcome, by program type



Few afterschool programs meet Tiers I or II when Broad Application criteria (number of study sites and sample size) are applied.

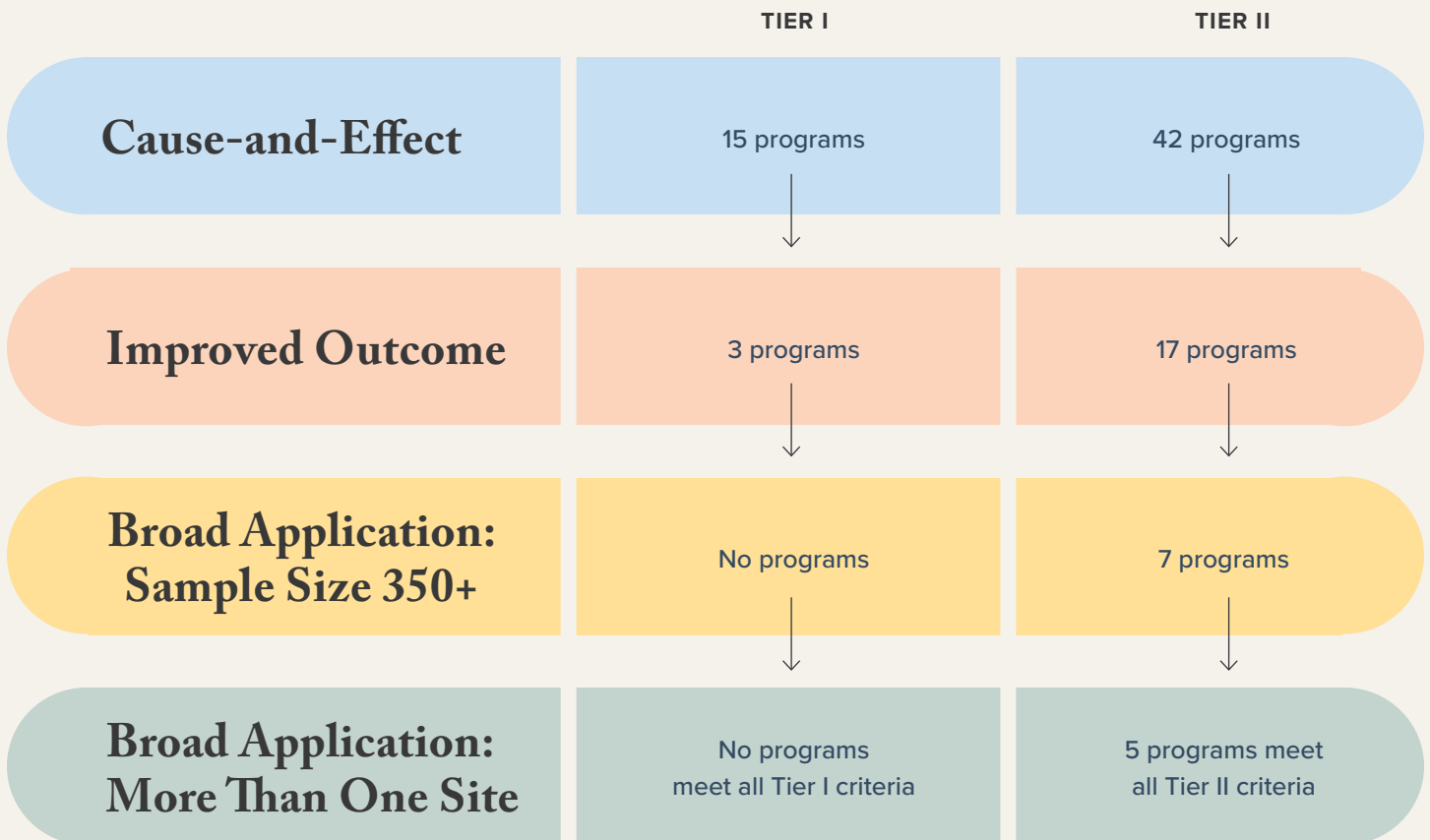
To encourage informed use of evidence, the U.S. Department of Education elaborated on ESSA evidence requirements in a guidance document released in September 2016 (see Chapter Two for more information on this guidance). The Department recommended two additional criteria for a program to be supported by Tier I or II evidence: (1) the study (or studies) showing the improved outcome should include at least 350 students, and (2) the study (or studies) should have been conducted in two or more sites. As we explain in Chapter Three, the Department has interpreted “site” to mean “school district” in its own K-12 discretionary grant competitions that require or incentivize applicants to submit evidence of effectiveness.

These sample size and site number recommendations are an additional check on the program’s evidence to see whether it has broader applicability beyond a handful of students in one location.⁷ For this reason, in this review, we have titled these recommendations the Broad Application criteria.

⁷ The recommendations correspond to the What Works Clearinghouse moderate-to-large extent of evidence definition (see page 26 of *What Works Clearinghouse™ Procedures Handbook v. 4.0*).

FIGURE 4.3

Programs meeting successive ESSA screens for Tiers I-II



When the Broad Application criteria—sample size and number of sites—are applied, no afterschool programs have Tier I evidence, and five programs have Tier II evidence (Figure 4.3). Of the three programs with studies meeting Tier I Cause-and-Effect and Improved Outcome criteria, none has a study with a sample size of 350 or more. Among the 17 programs with studies that meet Tier II Cause-and-Effect and Improved Outcome criteria, seven have studies sample sizes of 350 or more, and of these, five also have studies that were conducted in more than one site.

When Cause-and-Effect, Improved Outcome, and Broad Application criteria are applied, no afterschool programs meet Tier I requirements, five programs meet Tier II, and 59 programs meet Tier III.

The U.S. Department of Education recommends that the two Broad Application criteria (sample size and number of sites) apply to Tiers I and II only. Therefore, all programs that meet Tier I or Tier II for Cause-and-Effect and Improved Outcome but do not meet Broad Application criteria become Tier III programs.

Two Physical Activity/Health programs meet the Tier II requirements (Georgia FitKid, Youth Fit for Life), as does one Multicomponent program (LA's BEST). Two studies of large sets of Multicomponent programs (21st Century Learning Centers [middle school] and the Texas Afterschool Initiative) meet the Tier II requirements.

The evidence supporting these programs illustrates the judgment calls that sometimes must be made in assessing whether the evidence meets a given ESSA tier. LA's BEST, for example, was offered in one school district only, the Los Angeles Unified School District. However, the sheer size of the district, coupled with the fact that studies of LA's BEST cover multiple cohorts of students, convinces us that the program meets the spirit of the Broad Application criteria.

For any of these programs, education decision-makers should make additional judgment calls about the evidence.

The U.S. Department of Education recommends that, for programs to have evidence at Tiers I-III, improved outcomes should not be “overridden by statistically significant and negative... evidence on the same intervention” (U.S. Department of Education, 2016, p. 8) from studies that meet Cause-and-Effect requirements. This recommendation is a reminder not to selectively choose one improved outcome from a study but rather to examine the totality of the evidence for a program. In many cases, whether an improved outcome is overridden by negative evidence is a judgment call. In this report, we do not make that call for decision-makers, though we provide detailed information in the program summaries (available in a companion evidence guide) that enables users to make this determination.

In addition, the Department recommends that decision-makers assess whether the evidence for a program's improved outcome was from an “overlapping”—that is, similar—population of students and/or from a similar context as those where the program might be implemented. Because there are no hard-and-fast rules for what constitutes a similar student population or context, we recommend that decision-makers make common-sense judgments about whether the relevance of evidence stretches credulity. For example, it does not seem sensible to use evidence for a high school robotics program as justification for a science-based afterschool program for students in the early elementary grades. In program summaries presented in a companion evidence guide, we provide information about the student populations and contexts of studies.

Summary

In this chapter, we identified 62 afterschool programs that meet the Cause-and-Effect and Improved Outcome criteria for Tiers I-III. The majority of the programs meeting these criteria are Academic or Multicomponent programs, but most program types have at least two options meeting these criteria. In addition, programs with Tier I-III evidence and improved outcomes can be found across the grades. Tables 4.1 through 4.6 summarize these programs in at-a-glance tables, and a companion evidence guide provides more detailed information to help decision-makers determine whether a programmatic approach is relevant to their needs.

We note that there is a substantial number of afterschool programs (57) that have been the focus of the most rigorous and well-implemented effectiveness studies—that is, they meet Cause-and-Effect requirements for Tiers I-II. This suggests that although it can be challenging to conduct rigorous effectiveness studies of afterschool programs, including with randomized designs, it is indeed possible to carry out such studies. We identified 15 programs with at least one study that meets Cause-and-Effect requirements for Tier I; that is, we judge that these studies could meet What Works Clearinghouse standards without reservations. An additional 42 programs have studies that meet Cause-and-Effect requirements for Tier II and that we judge could meet What Works Clearinghouse standards with reservations. Further, we note that 20 of the programs studied with the most rigorous designs have at least one improved outcome.

TABLE 4.1

ACADEMIC

programs that meet criteria for ESSA Tiers I-III

Intervention	Grades	Domains with at least one improved outcome at Tiers I-III
ELEMENTARY GRADES		
21st Century Community Learning Centers (Pittsburgh Charter Schools - Elementary)	3-5	Mathematics Achievement, Reading/ELA Achievement
Academically-focused extended day program	K-5	Mathematics Achievement, Reading/ELA Achievement
Afterschool reading tutoring (Texas)	3	Reading/ELA Achievement
Afterschool tutoring program	4	General Achievement, Mathematics Achievement, Reading/ELA Achievement
Building Educated Leaders for Life (BELL)	3-6	Mathematics Achievement
ELA Extra and Math Mania	4	Mathematics Achievement
Raising Education Attainment Challenge	3-4	Reading/ELA Achievement
MIDDLE GRADES		
Middle School Academic Intervention Program	6-8	Reading/ELA Achievement
Supplemental Educational Services (Los Angeles- Middle Grades)	6-8	Reading/ELA Achievement
Warrior After School	Middle Grades	Mathematics Achievement, Reading/ELA Achievement
ELEMENTARY AND MIDDLE GRADES		
Supplemental Educational Services (Chicago)	3-8	Mathematics Achievement, Reading/ELA Achievement
Supplemental Educational Services (Dallas)	3-8	Mathematics Achievement, Reading/ELA Achievement
Supplemental Educational Services (Florida)	4-10	Mathematics Achievement
Supplemental Educational Services (Minnesota)	3-8	Mathematics Achievement, Reading/ELA Achievement
HIGH SCHOOL		
Academic Volunteer Mentor Service program	9-12	Attendance & Enrollment, General Achievement, Reading/ELA Achievement
Supplemental Educational Services (Los Angeles-High School)	9-12	Reading/ELA Achievement

TABLE 4.2

CAREER/LEADERSHIP

programs that meet criteria for ESSA Tiers I-III

Intervention	Grades	Domains with at least one improved outcome at Tiers I-III
MIDDLE GRADES		
Citizen Schools	5-8	Attendance & Enrollment, Mathematics Achievement, Promotion & Graduation, Reading/ELA Achievement School Engagement
Stay-in-School for College and Career Opportunities (SISCO)	6-8	Attendance & Enrollment, Promotion & Graduation
HIGH SCHOOL		
After School Matters	9-12	School Engagement
Hispanic Youth Leadership Program	9-12	Attendance & Enrollment, Promotion & Graduation

TABLE 4.3

MULTICOMPONENT

programs that meet criteria for ESSA Tiers I-III

Intervention	Grades	Domains with at least one improved outcome at Tiers I-III
ELEMENTARY GRADES		
21st Century Community Learning Centers (Fresno)	3-6	Mathematics Achievement
21st Century Community Learning Centers (Philadelphia)	1-5	Attendance & Enrollment, Mathematics Achievement, Reading/ELA Achievement
After School Education and Safety (California – Elementary)	3-5	Attendance & Enrollment, Physical Activity/Health
Baltimore Community Schools (Elementary)	K-5	Attendance & Enrollment
LA's Better Educated Students for Tomorrow (LA's BEST)	2-5	Mathematics Achievement
Multicomponent program (Northeast)	1-3	Physical Activity/Health, Social & Emotional Competencies
Tutoring and enrichment program	3-5	Mathematics Achievement

TABLE 4.3 (CONTINUED)

Intervention	Grades	Domains with at least one improved outcome at Tiers I-III
MIDDLE GRADES		
21st Century Community Learning Centers (Marietta Boys and Girls Club)	7-8	Mathematics Achievement, Reading/ELA Achievement
21st Century Community Learning Centers (Middle Grades)	6-8	Attendance & Enrollment, School Engagement
21st Century Community Learning Centers (Philadelphia)	6-8	Mathematics Achievement, Reading/ELA Achievement
After School Education and Safety (California – Middle Grades)	6-8	Attendance & Enrollment, Mathematics Achievement, Physical Activity/Health
The After-School Corporation (Middle Grades)	6-8	Promotion & graduation
AfterZone	6-8	Attendance & Enrollment, Physical Activity/Health
Baltimore Community Schools (Middle Grades)	6-8	Attendance & Enrollment
Cooke Middle School Afterschool Recreation Program	5-8	Physical Activity/Health, School Engagement
Texas After School Initiative	6-8	Attendance & Enrollment
ELEMENTARY AND MIDDLE GRADES		
21st Century Community Learning Centers (St. Louis, Missouri)	Elementary and middle	Mathematics Achievement
The After-School Corporation (Pre-K – Grade 8)	3-8	Attendance & Enrollment, Mathematics Achievement
Cool Girls	4-8	Physical Activity/Health
HIGH SCHOOL		
21st Century Community Learning Centers (Philadelphia)	9-12	Promotion & Graduation
After School Safety and Enrichment for Teens	9-11	Attendance & Enrollment, Mathematics Achievement, Reading/ELA Achievement
The After-School Corporation (High School)	9-12	Attendance & Enrollment

TABLE 4.4

PHYSICAL ACTIVITY/HEALTH

programs that meet criteria for ESSA Tiers I-III

Intervention	Grades	Domains with at least one improved outcome at Tiers I-III
ELEMENTARY GRADES		
America SCORES	4-5	Physical Activity/Health
FITKids	Elementary	Physical Activity/Health
Georgia Prevention Institute Physical Activity Program	3-5	Physical Activity/Health
Girls in the Game	3-5	Physical Activity/Health
LA Sprouts	4-5	Physical Activity/Health
Youth Fit For Life	Ages 9-12	Physical Activity/Health
MIDDLE GRADES		
Fitness-focused afterschool programs (California)	5-7	Physical Activity/Health
Girls on the Move	5-8	Mathematics Achievement
Student-centered physical activity program	6	Physical Activity/Health

TABLE 4.5

STEM

programs that meet criteria for ESSA Tiers I-III

Intervention	Grades	Domains with at least one improved outcome at Tiers I-III
ELEMENTARY GRADES		
4-H Robotics Program	Ages 9-11	Science Achievement
Bringing Up Girls in Science (BUGS)	4-5	Science Achievement
MIDDLE GRADES		
The Investigators Club	7	School Engagement, Science Achievement
HIGH SCHOOL		
FIRST Robotics	9-12	Attendance & Enrollment

TABLE 4.6

SPORTS

programs that meet criteria for ESSA Tiers I-III

Intervention	Grades	Domains with at least one improved outcome at Tiers I-III
MIDDLE GRADES		
School-sponsored sports (Texas)	8	School Engagement
HIGH SCHOOL		
High school interscholastic sports	10-12	Mathematics Achievement, Reading/ELA Achievement
School-sponsored sports (Miami)	9-12	Mathematics Achievement, Reading/ELA Achievement
School-sponsored sports (South Texas)	11	Attendance & Enrollment, General Achievement, Mathematics Achievement, Reading/ELA Achievement

Key Findings and Recommendations

This chapter summarizes major findings from our review of the research on the effectiveness of afterschool programs. Building on these findings, we also provide several recommendations for program providers, state and federal policymakers, and researchers who study afterschool programs.

Key Findings

There is a substantial set of afterschool programs with evidence of effectiveness meeting ESSA Tiers I-III, including branded and unbranded programs.

We identified more than 60 programs with research evidence at Tiers I-III showing one or more improved outcomes for students. As we showed in Chapter Four, these programs included a mix of branded programs or well-developed models (for example,

Building Educated Leaders for Life [BELL], AfterZone, and Chicago After School Matters) and unbranded, “homegrown” programs (for example, 21st Century Community Learning Centers programs in Philadelphia). Taken together, the programs improved outcomes in a variety of domains, including Mathematics Achievement, Reading/ELA Achievement, Science Achievement, Physical Activity/Health, Attendance and Enrollment, Promotion and Graduation, and Social and Emotional Competencies. At the same time, we found relatively few instances of statistically significant negative outcomes for students.

Detailed information about the goals, components, and studies of these programs, as well as of other programs with similar research quality that did not have any positive outcomes, is presented in a companion evidence guide.

Effective afterschool programs can be found at each grade level and within almost every program type.

There are more programs with evidence of effectiveness for students in the elementary and middle grades, but afterschool programs meeting ESSA Tier I-III criteria can be found across the grade levels and for almost all program types. Academic programs, which focus on improvement of academic skills, and Multicomponent programs, which offer a wide range of activities, have the largest number of effective program options. There are more studies of these programs than others, which may be a function of their being more commonly offered.

There are more afterschool program options for improving Mathematics and Reading/ELA Achievement than improving other outcomes.

Mathematics and Reading/ELA Achievement, as measured by standardized test scores, are the most commonly studied outcomes, by far. For this reason, even though the percentage of positive findings was lower for Mathematics Achievement and Reading/ELA Achievement than for outcome domains like Attendance and Enrollment or Physical Activity/Health (Figure 3.6), there are more programs that improved outcomes in the Mathematics Achievement and Reading/ELA Achievement domains. Because improving academic outcomes is a key goal of the 21st Century Community Learning Centers program, it is important to note that numerous programs have demonstrated effects on Mathematics and Reading/ELA Achievement, including improvements in standardized test scores.

Recommendations for afterschool program providers

This set of recommendations is geared to those who make decisions about which afterschool programming to offer. This includes district and school personnel as well as community groups, nonprofits, and others who provide afterschool programs.

Think of research evidence as a tool to help you make the best bets about what will work for your students.

Research evidence should complement, not replace, program providers' craft knowledge and good judgment. Research studies, even the most rigorous ones, provide information only about whether a program has worked in a place (or places) where it has been tried in the past. For this reason, the best way to think about evidence is as a tool that helps program providers increase the likelihood of offering effective programs. The more studies of a program, the larger the sample sizes, and the more consistent the results, the more confident we can be about the amount of evidence for a program. Larger amounts of evidence provide more assurance that an approach reliably produces certain results, and such evidence should be taken especially seriously, even if it challenges one's beliefs and assumptions.

Most importantly, even an afterschool program with strong evidence of effectiveness is unlikely to generate benefits if its implementation is not high quality. An evidence-based program must be replicated with high fidelity to the model to have a reasonable chance of achieving its intended impacts.

Use information about program components, staffing, and dosage to assess whether a program might be a good fit for local needs and resources.

One aspect of using good judgment in applying evidence to afterschool programming is to determine whether the program seems like a good fit for a given location, including whether resources are in place to support quality implementation. The companion evidence guide provides a high-level summary of this information, but decision-makers should see this

information as a starting place for their investigations. More nuanced descriptions can be found in many of the reports, and fortunately, studies with the most extensive documentation of components and implementation conditions tend to be freely available for web download. Decision-makers may want to contact the providers of branded programs for information about cost; we found that program cost information was infrequently reported in studies.

Programs with evidence of effectiveness may need to be adapted to fit new contexts. For example, a science program with evidence at Tiers I-III may have used university student-volunteers to staff the program, but an afterschool provider considering a similar approach may not be located near a university. In a case like this, the provider should consider what qualities the university students likely brought to the program (for example, extra hands to help with young students? Specialized knowledge and skills?) and determine whether suitable staffing alternatives can be used for this key component of the model.

It is common for programs to be effective in some domains, but not others, so consider overall evidence of effectiveness and pay special attention to specific outcomes that are priorities for improvement.

The good news is that relatively few programs in this review resulted in negative outcomes—that is, harmed youth. Still, it is important to use good judgment when applying evidence by looking at the overall effectiveness of a program, rather than focusing selectively on positive findings. For example, a program might have positive benefits in the Physical Activity/Health domain but not in Mathematics Achievement. If improving both types of outcomes is a high priority for a provider or community, it must find a program that seems likely to improve outcomes in both domains or else choose to implement more than one type of afterschool program.

In addition, for any program that is of initial interest, we recommend examining the specific outcomes and measures used in studies of the program. Even within a given outcome domain, some outcomes may be more global than others. For example, within the Reading/ELA Achievement domain, measures may be as specific as vocabulary acquisition and as broad as a state standardized reading/ELA assessment. Moreover, the size of the effects might vary across these seemingly similar outcome measures. For example, a program may have evidence that student participation has a large effect on vocabulary acquisition and a very small, or no, effect on a state standardized reading/ELA assessment. It is important to think about the specifics of the outcome you are seeking to achieve as you consider the evidence presented in this report.

To meet ESSA sample and site requirements for Tiers I and II, combine evidence from multiple studies of similar programs.

As we showed in Chapter Four, numerous studies of afterschool programs meet Cause-and-Effect and Improved Outcome criteria for Tiers I and II but do not meet the minimum sample sizes or site numbers that programs should have, as recommended by the U.S. Department of Education’s evidence guidance. However, the minimum sample size and site number recommendations can be achieved with multiple studies showing improved outcomes that, together, meet these criteria. This would require making a compelling argument that programs classified as distinct in this report are, in fact, examples of a single type of program. This is likely to require gathering additional information on the programs highlighted in the companion evidence guide by looking at original sources, contacting study authors, and/or talking to program representatives.

This approach can also be used going forward as new studies are conducted. For example, a program with one study that meets Cause-and-Effect requirements for Tier I and a sample size of fewer than 350 students is currently classified as Tier III. With another study of equal quality and positive effects, and a combined sample size across the two studies of at least 350, the program could meet Tier I criteria.

ESSA does not require that states limit afterschool programs to using evidence from the highest tiers. However, as described in Chapter One (Box 1.1) some U.S. Department of Education discretionary grant programs require or incentivize applicants to submit research evidence meeting criteria for these higher ESSA tiers.

Check the quality of research provided by vendors against the assessments provided in this report and accompanying volume.

Program providers can check the quality of any vendor-supplied evidence against study reviews presented in this report and the accompanying volume. Because this report is based on a thorough, comprehensive literature search, we believe that we have identified and given a tier rating to almost every study published from 2000 through 2017 examining the effects of afterschool programs on the outcomes we specified. It is possible, of course, that this review inadvertently overlooked a comparison group study, but vendor-supplied evidence that is not reviewed in this report is likely not to use a comparison group, which we consider an essential element for credible studies of afterschool program effects.

Recognize that studies answer different types of questions.

Most of the studies that we identified and reviewed focus closely on the effects of a single program or program model. These studies are useful for individual program providers who may be trying to decide on a specific approach to programming, including staffing, components, and program duration. Other studies, however, focus on whether programs of a general type, perhaps supported by a state or federal funding stream, improved outcomes for a whole state. These studies may include only the most general information about the content and format of the afterschool programs. These studies are likely of greatest value to state-level decision-makers who are trying to decide whether to invest in afterschool programming or who want to use these studies as a starting point for a deeper investigation of why an initiative might not have achieved hoped-for effects.

If existing afterschool programs with evidence at Tiers I-III do not fit with needs and resources, use the flexibility built in to ESSA Tier IV.

Tier IV of ESSA's evidence framework provides a door through which new afterschool program models can be introduced and evaluated for evidence of effectiveness. If no evidence exists at Tiers I-III, a program can meet Tier IV criteria if there is a well-specified logic model, informed by research, that provides reason to believe that the program will improve outcomes. Neither ESSA nor the Department's guidance specify the type or qualities of research that should inform the logic model. However, for program providers wishing to use the Tier IV option, a starting place is studies that meet Tier IV requirements and are presented in Appendix EG-2 of the companion evidence guide. As comparison group studies, they are superior to studies that use a pretest-posttest design without a comparison group.

Recommendations for states

States are justified in using a more liberal tier standard (Tiers I-IV) for afterschool programs until there are more programs with evidence meeting Tiers I and II.

With some exceptions, ESSA allows states full discretion about which level of evidence (Tiers I-IV) to require for activities funded under the law. States can select the level of evidence for the major source of afterschool funding under ESSA, the 21st Century Community Learning Centers program. Based on our review of the afterschool literature, we believe that states are justified in allowing programs to cite evidence at ESSA Tiers I-IV until there is more evidence at Tiers I-II. For example, although 20 afterschool programs meet Tier I-II research design requirements and improved at least one student outcome, only five programs meet these tiers when all the Department's recommendations, such as minimum study sample size, are applied.

To obtain larger sample sizes and multiple sites recommended by the U.S. Department of Education guidance, states could encourage and facilitate evaluations of similar programs across providers and districts.

As shown in Chapter Four, there are numerous studies that meet Cause-and-Effect and Improved Outcome criteria for Tiers I and II but do not have the sample sizes and number of sites recommended by the Department of Education's evidence guidance. State departments of education, through which federal afterschool funding flows to localities, have an important role to play in facilitating rigorous impact studies across districts and providers. We recommend that states work with districts to develop learning priorities and systematically design evaluations to address those priorities. For example, states and districts might prioritize learning about the effects of different afterschool tutoring models. Models could vary in terms of duration, intensity, materials used, and other dimensions, and their effects could be investigated using an experiment or quasi-experiment.

States should strongly encourage districts and providers to conduct evaluations designed to meet Tier I and Tier II criteria for research quality and should provide adequate evaluation funding.

This review identified 57 programs with research judged to be consistent with What Works Clearinghouse (WWC) quality standards for experiments or quasi-experiments. These findings tell us that rigorous studies of effectiveness can be done, and are being done, in the afterschool field. The WWC, an initiative of the Department of Education's Institute of Education Sciences, reviews studies of the effectiveness of education interventions, and its standards are a benchmark in the field. Designing and implementing studies to meet WWC standards is a reasonable and achievable goal.

Additional studies, particularly those designed to meet Tiers I-II, are needed to identify more afterschool programs that improve outcomes for students. States distribute 21st Century Community Learning Centers funding and are in a unique position to encourage systematic, rigorous research about afterschool programs. By incentivizing program providers to clarify their logic models, implement their programs with fidelity, and conduct rigorous studies of the effectiveness and implementation of their afterschool programs, states could increase the field's understanding of what works in afterschool programming and what it takes to achieve desired results.

The What Works Clearinghouse has made its training freely available online (<https://ies.ed.gov/ncee/wwc/OnlineTraining>), and there are many other resources available for evaluators who want to understand how to design and implement strong impact studies.

Recommendations for the federal government

The U.S. Department of Education’s 21st Century Community Learning Center (21CCLC) program should encourage program evaluations that investigate a range of student outcomes, including academic achievement measured other than by state standardized tests.

The U.S. Department of Education describes a broad vision for the 21st Century Community Learning Centers program, including providing opportunities for academic enrichment and a wide array of additional services, programs, and activities for students and families. Given this broad mandate and focus on enrichment, the indicators used by the federal government to judge the program’s performance may capture only a subset of potentially important outcomes. **We particularly note that using state standardized test scores as key performance measures may set unrealistic expectations for what afterschool programs can and should accomplish.** We recommend that the Department encourage states and their evaluators to use measures that are clearly aligned to the purposes of the programs and sensitive to changes that they realistically could produce.

Recommendations for evaluators

Thoroughly report program elements, implementation challenges, and cost.

To use research in practice, providers need to know what the program offered, how it offered these components, and what they cost. Providers can use information about a program’s implementation to assess whether the program is suitable for their setting, whether their staff have the training or capacity to implement the activities, what implementation challenges to expect, and the like. To respond to this need, we collected and summarized information available from the studies included in the review, but we found that too often studies provide minimal descriptions of the program. If research is to support decision-making, evaluators need to provide much more detail about the features of programs and settings they have studied.

Provide all information needed to conduct a study review for ESSA.

To determine the ESSA tier for each study that we reviewed, we used only the information that study authors provided about their findings. We did not contact authors for additional information if an important piece of technical detail, such as a standard deviation or sample attrition, was missing. In a few cases, the lack of this information meant that we could not rate the study, and in other cases, studies might have been rated at a lower tier than they would have been if we had all the information we needed.

The information we needed to review studies is described in the What Works Clearinghouse Reporting Guide for Study Authors (https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_gsa_v1.pdf), and we recommend that study authors check their reporting against this document. In addition to information about the intervention and the measures, authors should explain how the program and comparison groups were formed and include the sample size, mean, and standard deviation on baseline and post-intervention measures for both groups. For an excellent example of a brief study that provided all information needed for a review, see this report on Elevate Math by the Regional Educational Laboratory West at WestEd: https://ies.ed.gov/ncee/edlabs/regions/west/pdf/REL_2015096.pdf/.

Register effectiveness studies and report all confirmatory analyses, including negative impacts.

The field of education research increasingly is taking steps to ensure transparency and completeness in reporting about findings from effectiveness studies, to ensure a full accounting of an intervention’s effects—negative, positive, and null. Registering studies as they begin, including identifying planned confirmatory analyses, is one way of increasing this transparency. In the context of ESSA, these registries have another purpose: enabling states, school districts, and afterschool providers to see whether a study is underway of an afterschool model of interest. As we describe in Chapter Two, one of the requirements for ESSA Tier IV is that a study of the program be underway.

The Society for Research on Educational Effectiveness hosts a free study registry. To register a study, navigate to <https://www.sreereg.org/>.

References

- Acaira, E., Vile, J., & Reisner, E. R. (2010). *Citizen Schools: Achieving high school graduation: Citizen Schools' youth outcomes in Boston*. Boston, MA: Policy Studies Associates, Inc.
- Afterschool Alliance. (2014, February). *Taking a deeper dive into afterschool: Positive outcomes and positive practices*. Retrieved from http://afterschoolalliance.org/documents/Deeper_Dive_into_Afterschool.pdf
- Afterschool Alliance. (2018, April). *21st Century Community Learning Centers*. Retrieved from <http://afterschoolalliance.org/documents/21stCCLC-Overview-2018.pdf>.
- Afterschool Alliance. (2018b). Data on webpage: <https://www.afterschoolalliance.org/policy21stcccl.cfm>.
- Allen, M. L. (2008). *Assessing the effectiveness of supplemental education services in urban Florida school districts* (Doctoral dissertation). Retrieved from http://ufdcimages.uflib.ufl.edu/UF/E0/02/19/69/00001/allen_m.pdf
- American Institutes for Research. (2014, July 11). *The in-school and afterschool social and emotional learning connection: A planning tool*. Retrieved from <http://www.air.org/resource/school-and-afterschool-social-and-emotional-learning-connection-planning-tool>.
- Anderson-Butcher, D., & Cash, S. (n.d.) *Capital Kids evaluation: Overall and site specific results for 2003-2004*. Columbus, OH: Ohio State University.
- Annesi, J. J. (2006). Relations of physical self-concept and self-efficacy with frequency of voluntary physical activity in preadolescents: Implications for after-school care programming. *Journal of Psychosomatic Research*, 61(4), 515-520.
- Annesi, J. J., Faigenbaum, A. D., Westcott, W. L., & Smith, A. E. (2008). Relations of self-appraisal and mood changes with voluntary physical activity changes in African American preadolescents in an after-school care intervention. *Journal of Sports Science & Medicine*, 7(2), 260-268.
- Armstrong, B. L. P. (2007). *Evaluating the impact of the LeapTrack Assessment and Instruction System through an afterschool reading intervention class* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3282864)
- Assouline, S. G., Ihrig, L. M., & Mahatmya, D. (2017). Closing the excellence gap: Investigation of an expanded talent search model for student selection into an extracurricular STEM program in rural middle schools. *Gifted Child Quarterly*, 61, 250-261.
- Avery, C. (2013). *Evaluation of the College Possible program: Results from a randomized controlled trial* (NBER Working Paper No. 19562). Cambridge, MA: National Bureau of Economic Research.
- Barbeau, P., Johnson, M. H., Howe, C. A., Allison, J., Davis, C. L., Gutin, B., & Lemmon, C. R. (2007). Ten months of exercise improves general and visceral adiposity, bone, and fitness in black girls. *Obesity*, 15, 2077-2085.
- Barker, B. S., & Ansorge, J. (2007). Robotics as means to increase achievement scores in an informal learning environment. *Journal of Research on Technology in Education*, 39(3), 229-243.
- Barnhart, M. K. (2011). *The impact of participation in supplemental educational services (SES) on student achievement: 2009-10* (Publication No. 379). Los Angeles, CA: Los Angeles Unified School District Research Unit.
- Beckemeyer, P. (2009). *An intervention study of pre-teen girls to improve eating behavior and physical activity levels: Growing Optimally, Girls Improving Health and Raising Levels of Self-efficacy (GO-GIHRLS)* (Unpublished doctoral dissertation). Northern Illinois University, DeKalb, IL.
- Beller, J. M. (2013). *Positive youth development programs and high school girls' psychosocial behaviors* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3601435)

- Birmingham, J., & White, R. N. (2005). *Promoting positive youth development for high school students after school: Services and outcomes for high school youth in TASC programs*. Washington, DC: Policy Studies Associates, Inc.
- Blanchard, S., Judy, J., Muller, C., Crawford, R. H., Petrosino, A. J., White, C. K., ... & Wood, K. L. (2015). Beyond blackboards: *Engaging underserved middle school students in engineering*. *Journal of Pre-college Engineering Education Research*, 5(1), 1-14.
- Blowers, J. G. (2007). *Impact of an afterschool martial arts program on at-risk students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3257356)
- Bohnert, A. M., & Ward, A. K. (2013). Making a difference: Evaluating the girls in the game (GIG) after-school program. *Journal of Early Adolescence*, 33(1), 104-130.
- Boulden, W. T. (2006). Evaluation of the Kansas City LULAC National Education Service Center's Young Reader's Program. *Children & Schools*, 28(2), 107-114.
- Bridges, C. (2011). *The correlation between after school tutoring and reading scores of at-risk third graders* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3482150)
- Broh, B. A. (2002). Linking extracurricular programming to academic achievement: Who benefits and why? *Sociology of Education*, 75, 69-95.
- Burton, M. D. (2006). *The impact of a leadership development program on fifth-grade students' disciplinary referrals and leadership* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3206885)
- Callender, S. L. (2015). *Student participation in extracurricular activities and the impact on academic achievement, self-concept, and academic self-concept during the middle school years* (Doctoral dissertation,). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3663019)
- Carr, E. M. (2015). *Afterschool program interventions that support the academic achievement, behavior, and engagement of at-risk student populations* (Unpublished doctoral dissertation). University of Houston-Clear Lake, Houston, TX.
- Carroll, S. (2004). *The effect of after school tutoring and summer enrichment programs on the algebra I achievement of at-risk students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3122343)
- Carswell, S. B., Hanlon, T. E., Watts, A. M. & O'Grady, K. E. (2014). Prevention-related research targeting African American alternative education program students. *Education and Urban Society*, 46(4) 434-449.
- Chatterji, M., Kwon, Y. A., & Sng, C. (2006). Gathering evidence on an after-school supplemental instruction program: Design challenges and early findings in light of NCLB. *Education Policy Analysis Archives*, 14(12), 1-47.
- Chiefs for Change. (2016). *Expanding equity: Leveraging the Every Student Succeeds Act (ESSA) to provide Direct Student Services*. Washington, DC: Chiefs for Change. Retrieved from <http://chiefsforchange.org/wp-content/uploads/2016/04/Chiefs-for-Change-Direct-Student-Services-April-2016.pdf>.
- Chinen, M. H. (2010). *The effects of after-school program participation on mathematics achievement: The case of LA's BEST* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3450983)
- Collum, R. C., Jr. (2004). *An investigation of the impact of participation in an afterschool program on the academic achievement and social behavior of adolescent Black males defined as at-risk* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3098371)
- Cooper, T. M. (2007). *The effects of supplemental educational services on student achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3262811)

- Cosden, M., Morrison, G., Albanese, A. L., & Macias, S. (2001). When homework is not home work: After-school programs for homework assistance. *Educational Psychologist, 36*(3), 211-221.
- da Cruz, K. (2017). *Effects of a randomized trial after-school physical activity club on the math achievement and executive functioning of girls* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 10252421)
- Cross, A. B. (2009). *An experimental evaluation of after school program participation on problem behavior outcomes: Does pre-existing risk moderate the effects of program participation?* (Unpublished doctoral dissertation). University of Maryland, College Park, MD.
- Cross, A. B., Gottfredson, D. C., Wilson, D. M., Rorie, M., & Connell, N. (2009). The impact of after-school programs on the routine activities of middle school students: Results from a randomized, controlled trial. *Criminology and Public Policy, 8*, 391-412.
- Dallman, R. (2004). *Students' bond to their school and interscholastic athletics in a consolidated school* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3163140)
- Darling, N., Caldwell, L. L., & Smith, R. (2005). Participation in school-based extracurricular activities and adolescent adjustment. *Journal of Leisure Research, 37*(1), 51-76.
- Davis, C. L., Tkacz, J., Gregoski, M., Boyle, C. A., & Lovrekovic, G. (2006). Aerobic exercise and snoring in overweight children: A randomized controlled trial. *Obesity, 14*(11), 1985-1991.
- Davis, J. N., Ventura, E. E., Cook, L. T., Gyllenhammer, L. E., & Gatto, L. M. (2011). LA Sprouts: A gardening, nutrition, and cooking intervention for Latino youth improves diet and reduces obesity. *Journal of the American Dietetic Association, 111*, 1224-1230.
- de Heer, H. (2009). *Project L.E.A.N.: An after-school health and exercise program for elementary school children in El Paso, Texas* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3358902)
- de Heer, H. (2009). *A path analysis of an after-school health and exercise program for elementary school children in El Paso, Texas* (Master's thesis). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 1462432)
- de Heer, H. D., Koehly, L., Pederson, R., & Morera, O. (2011). Effectiveness and spillover of an after-school health promotion program for Hispanic elementary school children. *American Journal of Public Health, 101*, 1907-1913.
- Deeb-Westervelt, W. (2003). *The effects of an after -school academic intervention services math program on the grade four New York state mathematics assessment: A quasi -experimental case study*. (Doctoral dissertation, Hofstra University). Retrieved from <https://search.proquest.com/docview/305328443?accountid=14541>
- Delucci, G. R. (2010). *An evaluation of an after-school program for low income elementary and middle school students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3420956)
- Dreyer, K. J. (2010). *An examination of academic outcomes for students who attend a school-based afterschool program* (Doctoral dissertation, University of Pittsburgh). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3417442)
- Duncan, G., & Murnane, R, eds. (2011). *Wither opportunity? Rising inequality, schools, and children's life chances*. New York: Russell Sage Foundation.
- Dynarski, M., Moore, M., Mullens, J., Gleason, P., James-Burdumy, S., Rosenberg, L., . . . , Deke, J. (2003). *When schools stay open late: The national evaluation of the 21st Century Community Learning Centers program. First year findings* (Report No. ED473230). Princeton, NJ: Mathematica Policy Research.
- Dynarski, M., Moore, M., James-Burdumy, S., Rosenberg, L., Deke, J., & Mansfield, W. (2004). *When schools*

stay open late: *The national evaluation of the 21st Century Community Learning Centers program. New findings* (MPR Report No. 8658-800/8720-700). Princeton, NJ: Mathematica Policy Research.

Earle, A. (2009). *Roadmap to afterschool for all: Examining current investments and mapping future needs*, Washington, D.C.: Afterschool Alliance. Retrieved from: <http://www.afterschoolalliance.org/documents/roadmap/RoadmapToAfterschoolForAllReport.pdf>.

Educational Research Consultants. (2016). *EduCare Foundation: After school program report card for 2014-2015*. Retrieved from <http://www.ercdata.com>.

Evans, S. W., Langberg, J. M., Schultz, B. K., Vaughn, A., Altaye, M., Marshall, S. A., & Zoromski, A. K. (2016). Evaluation of a school-based treatment program for young adolescents with ADHD. *Journal of Consulting and Clinical Psychology, 84*, 15-30.

Every Student Succeeds Act of 2015, Pub. L. No. 114-95 (2015).

Felix, E. E. (2009). *The impact of a supplemental education program on academic performance and attendance of ninth grade students at-risk for school failure* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3351551)

Fink, B. L. (2010). *The effect of a seventh grade after school leadership program on the developmental assets, academic achievement, and behavior of non-thriving students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3450199)

Ford, J., Harrison, L., Neergaard, L. Qian, H., Zoblotsky, T., McKay, D., & Ross, S. M. (2009). *Supplemental educational services in the Commonwealth of Virginia: 2007-2008* (CREP Report). Memphis, TN: Center for Research in Educational Policy, University of Memphis.

Fountain, A. R., Gamse, B., Velez, M., Hillard, M., & Cropper, P. (2016). *Evaluation of Citizen Schools' expanded learning time model: Final report*. Cambridge, MA: Abt Associates.

Fulmer, A. D. (2013). *Investigating relationships between academic achievement and after school programs in a high school environment* (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=3378&context=etd>

Gallagher, B., & Zoblotsky, T. (2009). *Supplemental educational services in the state of North Carolina: 2007-2008 student achievement analysis* (CREP Report). Memphis, TN: Center for Research in Educational Policy, University of Memphis.

Gao, J., Hallar, B., & Hartmann, T. A. (2014). *A snapshot of OST programs in Philadelphia: An evaluation of eleven 21st Century Community Learning Center grantees*. Philadelphia, PA: Research for Action.

Gardner, L. T. (2014). *The impact of afterschool tutoring on reading scores of elementary students* (Unpublished doctoral dissertation). Walden University, Minneapolis, MN.

Gatto, N. M., Ventura, E. E., Cook, L. T., Gyllenhammer, L. E., & Davis, J. N. (2012). LA Sprouts: A garden-based nutrition intervention pilot program influences motivation and preferences for fruits and vegetables in Latino youth. *Journal of the Academy of Nutrition and Dietetics, 112*, 913-920.

Gearhart, S. (2016). *Reading comprehension through incidental learning: Efficacy of an after-school literacy program utilizing the Barton Reading and Spelling System®* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 10240947)

Gentilcore, J. (2002). *The effect of an after-school academic intervention service on a New York state eighth grade English language arts assessment: A case study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3057258)

Gill, M. J. (2006). *Effective literacy and math instruction, better student achievement: An evaluative study of after-school programming* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3210266)

Goerge, R., Cusick, G. R., Wasserman, M., & Gladden, R. M. (2007). *After-school programs and academic impact: A study of Chicago's After School Matters* (Issue Brief #112). Chicago, IL: Chapin Hall Center for Children, University of Chicago.

Goldschmidt, P., & Huang, D. (2007). *The long-term effects of after-school programming on educational adjustment and juvenile crime: A study of the LA's BEST after-school program* (CRESST Report). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.

Gooding, L. F. (2010). *The effect of a music therapy-based social skills training program on social competence in children and adolescents with social skills deficits* (Doctoral dissertation). Retrieved from <http://diginole.lib.fsu.edu/islandora/object/fsu%3A254290>

Gorke, G. K. (2006). *Academic intervention instructors: Effectiveness of graduate-level college students as intervention tutors in an afterschool program* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3227047)

Gottfredson, D. C., Cross, A. B., Wilson, D. M., Rorie, M., & Connell, N. (2010a). An experimental evaluation of the All Stars Prevention Curriculum in a community after school setting. *Prevention Science, 11*, 142-154.

Gottfredson, D. C., Cross, A. B., Wilson, D. M., Rorie, M., & Connell, N. (2010b). Effects of participation in after-school programs for middle school students: A randomized trial. *Journal of Research on Educational Effectiveness, 3*, 282-313. Graham, S., Taylor, A., & Hudley, C. (2015). A motivational intervention for African American boys labeled as aggressive. *Urban Education, 50*(2), 194-224.

Grant, S., Hamilton, L., Wrabel, S., Gomez, C., Whitaker, A., Tamargo, J.,..., Ramos, A. (2017). *Social and emotional learning interventions under the Every Student Succeeds Act: Evidence review*. Santa Monica, CA: RAND Corporation.

Grassi, A. M. (2012) *Quality after-school programming and its relationship to achievement-related behaviors and academic performance* (Doctoral dissertation). Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=csu1354646131&disposition=inline

Green, H. K. (2010). *The impact of an academic sports-mentoring afterschool program on academic outcomes in at-risk youth* (Unpublished doctoral dissertation). Philadelphia, PA: Drexel University.

Grolnick, W. S., Farkas, M. S., Sohmer, R., Michaels, S., & Valsiner, J. (2007). Facilitating motivation in young adolescents: Effects of an after-school program. *Journal of Applied Developmental Psychology, 28*, 332-344.

Gross, C. (2016). *The effect of an evidence based physical education curriculum and/or an after school activity program on cardiovascular endurance and fundamental motor skill proficiency in grade 5* (Doctoral dissertation). Retrieved from <http://digitalcommons.plymouth.edu/etd/104>

Gutin, B., Yin, Z., Johnson, M., & Barbeau, P. (2008). Preliminary findings of the effect of a 3-year after-school physical activity intervention on fitness and body fat: The Medical College of Georgia FitKid Project. *International Journal of Pediatric Obesity, 3*, 3-9.

Hailey, L. A. (2014). *The effects of an after-school program: Changing academic performance and promoting success* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3618240)

Hall, M. K., Foutz, S., & Mayhew, M. A. (2012). Design and impacts of youth-directed Café Scientifique program. *International Journal of Science Education, Part B: Communication and Public Engagement, 1*-24.

Hammett, L. A. (2004). *A study of the effects of an after-school program on the discipline and the academic achievement of selected seventh grade students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3257307)

- Hanlon, T. E., Simon, B. D., O'Grady, K. E., Carswell, S. B., & Callaman, J. M. (2009). The effectiveness of an after-school program targeting urban African American youth. *Education and Urban Society, 42*(1), 96–118.
- Hartmann, T. A., Gao, J., Kumar, A., & Edmunds, K. (2013). *A snapshot of OST programming in Philadelphia: An evaluation of six 21st Century Community Learning Center grantees*. Philadelphia, PA: Research for Action.
- Hartmann, T., Good, D., & Edmunds, K. (2011). *Éxito: Keeping high-risk youth on track to graduation through out-of-school time supports*. Philadelphia: Research for Action.
- Heinrich, C. J., Burch, P., Good, A., Acosta, R., Cheng, H., Dillender, M., . . . , Stewart, M. (2014). Improving the implementation and effectiveness of out-of-school-time tutoring. *Journal of Policy Analysis and Management, 33*(2), 471-494.
- Herman, R., Gates, S., Arifkhanova, A., Bega, A., Chavez-Herrerias, E., Han, E., ..., Wrabel, S. (2017). *School leadership interventions under the Every Student Succeeds Act: Evidence review: Updated and expanded*. Santa Monica, Calif.: RAND Corporation.
- Higgins, J. D. (2016). *Get in the game: A study of the relationship between participation in extracurricular activities and academic achievement in the Shawnee Mission School District* (Doctoral dissertation,). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 10112403)
- Hillman, C. H., Pontifex, M. B., Castelli, D. M., Khan, N. A., Raine, L. B., Scudder, M. R., . . . , Kamijo, K. (2014). Effects of the FITKids randomized controlled trial on executive control and brain function. *Pediatrics, 134*, e1063-e1071.
- Hirsch, B. J., Hedges, L. V., Stawicki, J. & Mekinda, M. A. (2011). *After-school programs for high school students: An evaluation of after school matters*. Retrieved from <https://www.sesp.northwestern.edu/docs/publications/1070224029553e7f678c09f.pdf>
- Hobbs, C. L. (2012). *Effects of an afterschool program on elementary and middle school math achievement in Georgia schools* (Unpublished doctoral dissertation). Liberty University, Lynchburg, VA.
- Holochwost, S. J., Propper, C. B., Wolf, D. P., Willoughby, M. T., Fisher, K. R., Kolacz, J., ... & Jaffee, S. R. (2017). Music education, academic achievement, and executive functions. *Psychology of Aesthetics, Creativity, and the Arts, 11*(2), 147-166.
- Howe, C. A., Harris, R. A., Gutin, B. (2011). A 10-month physical activity intervention improves body composition in young black boys. *Journal of Obesity, 2001*, 1-8.
- Huang, C., Gao, Z., Hannon, J. C., Schultz, B., Newton, M., & Jenson, W. (2012). Impact of an after-school physical activity program on youth's physical activity correlates and behavior. *ICHPER-SD Journal of Research, 7*(2), 18-23.
- Huang, D., Kim, K. S., Marshall, A., & Pérez, P. (2005). *Keeping kids in school: An LA's BEST example. A study examining the long-term impact of LA's BEST on students' dropout rates* (CRESST Report). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.
- Huang, D., La Torre, D., Leon, S., Duong, N., & Hodson, C. (2011). *Supporting student success in middle schools: Examining the relationship between elementary afterschool program participation and subsequent middle school attainments* (CRESST Report). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.
- Huang, D., Leon, S., Harven, A. M., La Torre, D., & Mostafavi, S. (2009). *Exploring the relationship between LA's BEST program attendance and cognitive gains of LA's BEST students* (CRESST Report No. 757). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.

Huang, D., Leon, S., La Torre, D., & Mostafavi, S. (2008). *Examining the relationship between LA's BEST program attendance and academic achievement of LA's BEST students* (CRESST Report No. 749). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.

Huang, D., Silver, D., Cheung, M., Duong, M., Gualpa, A., Hodson, C., . . . , Vazquez, V. (2011). *Independent statewide evaluation of after school programs* (CRESST Report No. 789). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.

Huang, D., & Wang, J. (2012). *Independent statewide evaluation of ASES and 21st CCLC afterschool programs: May 1, 2008-December 31, 2011* (Report No. CDE4/CN077738/2011). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing, Center for the Study of Evaluation, Graduate School of Education and Information Studies, University of California, Los Angeles.

Hughes, J. N., Cao, Q., & Kwok, O. M. (2016). Indirect effects of extracurricular participation on academic adjustment via perceived friends' prosocial norms. *Journal of Youth and Adolescence*, 45(11), 2260-2277.

Institute of Education Sciences. (2018). *What Works Clearinghouse™ glossary of terms*. Retrieved from November 13, 2018 from <https://ies.ed.gov/ncee/wwc/Glossary>.

Institute of Education Sciences. (n.d.) *What Works Clearinghouse™ Procedures Handbook, v. 4.0*. Retrieved November 13, 2018 from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf.

Isbell, A. M. (2014). *Analysis of the relationships between after-school programs/activities and competencies* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3367008)

Israel, J. M. (2013). *Student extracurricular participation, student achievement, and school perception: An elementary school perspective* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3561144)

JacAngelo, N. P. (2003). *The relation of sports participation to academic performance of high school students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3126425)

James, S. M. (2007). *Identity and science learning in African American students in informal science education contexts* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3300822)

James-Burdumy, S., Dynarski, M., & Deke, J. (2007). When elementary schools stay open late: Results from the national evaluation of the 21st Century Community Learning Centers program. *Educational Evaluation and Policy Analysis*, 29(4), 296-318.

James-Burdumy, S., Dynarski, M., & Deke, J. (2008). After-school program effects on behavior: Results from the 21st Century Community Learning Centers program national evaluation. *Economic Inquiry*, 46(1), 13-18.

Jarratt, K. (2014). *Mathematics achievement outcomes for middle school students attending school-based afterschool mathematics programs* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3582791)

Jenner, E., & Jenner, L. W. (2007). Results from a first-year evaluation of academic impacts of an after-school program for at-risk students. *Journal of Education for Students Placed at Risk*, 12, 213-237.

Jones, B. R. (2014). *An after-school program and its effect on the math and reading performance levels of the standardized testing and reporting (STAR) for identified at-risk students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3609435)

- Jones, C. J. (2009). *The 2008 supplemental educational services program: Year 5 summative evaluation* (Unpublished report). Chicago, IL: Chicago Public Schools Office of Extended Learning Opportunities.
- Jones, S., & Workman, E. (2016). *ESSA's well-rounded education*. Washington, D.C.: Education Commission of the States. Retrieved from <http://www.ecs.org/wp-content/uploads/ESSAs-Well-Rounded-Education-1.pdf>
- Kaufmann, A. M. (2002). *Interscholastic sports participation as a predictor of academic success for high school students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3047511)
- Kauh, T. J., (2011). *AfterZone: Outcomes for youth participating in Providence's citywide after-school system*. Philadelphia, PA: Public/Private Ventures.
- Koch, M., Georges, A., Gorges, T., & Fujii, R. (2010). Engaging youth with STEM professionals in afterschool programs. *Meridian Middle School Computer Technologies Journal*, 13(1), 1-10.
- Koumoullou, M. (2013). *The academic differences between students involved in school-based robotics programs and students not involved in school-based robotics programs* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3573701)
- Kuperminc, G. P., Thomason, J., DiMeo, M., & Broomfield-Massey, K. (2011). Cool Girls, inc.: Promoting the positive development of urban preadolescent and early adolescent girls. *Journal of Primary Prevention*, 32(3-4), 171-183.
- Lacy, S. S. (2011). *The impact of afterschool tutoring on reading achievement of elementary students in a Mississippi rural school setting* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3450309)
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76(2), 275-313.
- Laughlin, S. R. (2013). *Robotics: Assessing its role in improving mathematics skills for grades 4 to 5* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3557417)
- Lauver, S. C. (2002). *Assessing the benefits of an after -school program for urban youth: An impact and process evaluation* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3043903)
- Le, T. N., Arifuku, I., Vuong, L., Tran, G., Lustig, D. F., & Zimring, F. (2011). Community mobilization and community-based participatory research to prevent youth violence among Asian and immigrant populations. *American Journal of Community Psychology*, 48(1-2), 77-88.
- Li, Y. (2016). *Three essays on economic evaluation of health intervention programs and health policy* (Doctoral dissertation). Retrieved from <https://oaktrust.library.tamu.edu/handle/1969.1/158142>
- Lightner, P. A. (2010). *Expelled middle school students: A study of the effects of a short-term, after-school reading intervention program* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3423819)
- Lipscomb, S. (2007). Secondary school extracurricular involvement and academic achievement: A fixed effects approach. *Economics of Education Review*, 26(4), 463-472.
- Lipton, C. M. (2007). *A domain-specific approach to giftedness: An empirical study of the effects of an enrichment program on students' leadership, social skills, and attention* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3269342)
- Little, K. N. (2009). *Effects of an intervention after-school program on academic achievement among middle school students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3342460)

- London, R. A., & Gurantz, O. (2013). Afterschool program participation, youth physical fitness, and overweight. *American Journal of Preventive Medicine, 44*(s3), S200-S207.
- London, R., Gurantz, O., & Norman, J. (2011). The effect of afterschool program participation on English language acquisition. *Afterschool Matters, Spring, 22-29*.
- Lovell, C. (2006). *The relationship between participation in extended-day programs and mathematics achievement of Title-I eligible participants* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3226843)
- Ludwig, M., Boyle, A., & Lindsay, J. (2017). *Review of evidence: Arts integration research through the lens of the Every Student Succeeds Act (ESSA)*. Washington, D.C.: American Institutes for Research, 2017. Retrieved from <http://www.wallacefoundation.org/knowledge-center/Pages/ESSA-Arts-Evidence-Review-Report.aspx>
- Madsen, K., Thompson, H., Adkins, A., & Crawford, Y. (2013). School-community partnerships a cluster-randomized trial of an after-school soccer program. *JAMA Pediatrics, 167*(4), 321-326.
- Mahoney, J. L., Lord, H., & Carryl, E. (2005). Afterschool program participation and the development of child obesity and peer acceptance. *Applied Developmental Science, 9*, 202-215.
- Manlove, K. J. (2013). *The impact of extracurricular athletic activities on academic achievement, disciplinary referrals, and school attendance among Hispanic female 11th grade students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3566593)
- Marquez, S. A., & Kovacic, C. B. (2007). *Supplemental educational services evaluation report, 2006-2007* (Unpublished report). Santa Fe, NM: New Mexico Public Education Department.
- Martin, C. J. (2015). *Student involvement in extracurricular activities and post-secondary education placement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 10030342)
- Mason, M. J., & Chuang, S. (2001). Culturally-based after-school arts programming for low-income urban children: Adaptive and preventive effects. *Journal of Primary Prevention, 22*(1), 45-54.
- Matvienko, O. (2007). Impact of a nutrition education curriculum on snack choices of children ages six and seven years. *Journal of Nutrition Education and Behavior, 39*(5), 281-285.
- Maxfield, M., Schirm, A., & Rodriguez-Planas, N. (2003). *The Quantum Opportunity Program demonstration: Implementation and short-term impacts* (MPR Report No. 8279-093). Washington, DC: Mathematica Policy Research.
- Maxfield, M., Castner, L., Maralani, V., & Vencill, M. (2003). *The Quantum Opportunity Program demonstration: Implementation findings* (MPR Report No. 8279-080). Washington, DC: Mathematica Policy Research.
- Maxwell, J. B. (2010). *The impact of supplemental education services program on middle school at-risk learners' math and reading scores* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3403821)
- McHale, J. P., Vinden, P. G., Bush, L., Richer, D., Shaw, D., & Smith, B. (2005). Patterns of personal and social adjustment among sport-involved and noninvolved urban middle-school children. *Sociology of Sport Journal, 22*(2), 119-136.
- McKay, D., Paek, J., Harrison, L., Qian, H., Zoblotsky, T., Ross, S. M., Fedde, F., & Ford, J. (2008). *Supplemental education services in the state of Virginia: 2006-2007* (CREP Report). Memphis, TN: Center for Research in Educational Policy, University of Memphis.
- Melnyk, B. M., Small, L., Morrison-Beedy, D., Strasser, A., Spath, L., Kreipe, R., . . . O'Haver, J. (2007). The COPE healthy lifestyles teen program: Feasibility, preliminary efficacy, and lessons learned from an after school group intervention with overweight adolescents. *Journal of Pediatric Health Care, 21*(5), 315-322. <http://doi.org/10.1016/j.pedhc.2007.02.009>

- Miller, J. S. (2004). *An examination of academic achievement and the development of personal attributes of students participating in extracurricular activities in two rural Idaho high schools* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3123843)
- Moldow, E. (2007). *After-school program activities and academic achievement: A study in one urban k–8 school* (Doctoral dissertation, Brandeis University). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3278159)
- Molina, B. S. G., Flory, K., Bukstein, O. G., Greiner, A. R., Baker, J. L., Krug, V., & Evans, S. W. (2008). Feasibility and preliminary efficacy of an after-school program for middle schoolers with ADHD: A randomized trial in a large public middle school. *Journal of Attention Disorders, 12*(3), 207-217.
- Monti, J. M., Hillman, C. H., & Cohen, N. J. (2012). Aerobic fitness enhances relational memory in preadolescent children: The FITKids randomized control trial. *Hippocampus, 22*(9), 1876-1882.
- Moran, R. A. (2009). *Extracurricular activities: Investigating the effects of participation-nonparticipation on the Georgia high school graduation test* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3350419)
- Muñoz, M. A., Chang, F., & Ross, S. M. (2012). No Child Left Behind and tutoring in reading and mathematics: Impact of supplemental educational services on large scale assessment. *Journal of Education for Students Placed at Risk, 17*(3), 186-200.
- Nance, E. E., Moore, D. H., & Lewis, C. (1999). 21st Century Community Learning Centers: Do they affect student achievement? *Community Education Journal, 27*(1-2), 7-11.
- Nears, K. (2007). *The achievement gap: Effects of a resilience-based after school program on indicators of academic achievement*. (Doctoral dissertation). Retrieved from <https://repository.lib.ncsu.edu/handle/1840.16/3402>
- Niederdeppe, N. N. (2009). *The impact of a theatre arts intervention on underachieving gifted Latinos* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3351755)
- Office of Research, Evaluation, and Accountability. (2007). *SES tutoring programs: An evaluation of year 3 in the Chicago Public Schools* (Unpublished report). Chicago, IL: Chicago Public Schools Office of Extended Learning Opportunities.
- Ogden, C. (2008). *Measuring the effectiveness of after-school programs via participants' pre and posttest performance levels on the Georgia Criterion Referenced Competency Test* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3297561)
- O'Grady, S. T. (2013). *Analyzing the effects of in-season sports participation on at-risk student-athletes* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3602814)
- Olson, L. S. (2014). *A first look at community schools in Baltimore*. Baltimore, MD: Baltimore Education Research Consortium.
- Olson, L. S., Connolly, F., & Kommajesula, A. H. (2013). *Family League 2011-2012 out of school time programs in Baltimore city*. Baltimore, MD: Baltimore Education Research Consortium.
- Oyserman, D., Terry, K., & Bybee, D. (2002). A possible selves intervention to enhance school involvement. *Journal of Adolescence, 25*, 313-326.
- Pate, R. R., Saunders, R. P., Ward, D. S., Felton, G., Trost, S. G., & Dowda, M. (2003). Evaluation of a community-based Intervention to promote physical activity in youth: Lessons from active winners. *American Journal of Health Promotion, 17*(3), 171–182.
- Patterson, P. W. (2007). *The benefits of an after-school program on African American students in two rural schools* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3361406)

- Perry, J. C., & Shannon, L. (2017). How vocational psychologists can make a difference in K-12 education. *International Journal of Educational and Vocational Guidance, 17*, 97-115.
- Pettway, D. D. (2005). *An evaluation of SmartLinks* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3173506)
- Phillips, T. D. (2004). *The effectiveness of writing voice in an afterschool writing program for fourth-graders* (Master's thesis). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 1423877)
- Prenovost, J. K. E. (2001). *A first-year evaluation of after-school learning programs in four urban middle schools in the Santa Ana Unified School District* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3007008)
- Price, S. A. (2014). *Academic outcomes of elementary after-school programs: A matched-pair study within high-poverty, predominantly Hispanic student populations* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3702868)
- Reardon, S. (2011). The widening academic achievement gap between rich and poor: New evidence and possible explanations. In G. Duncan and R. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91-115). New York, NY: Russell Sage Foundation.
- Reisner, E. R., White, R. N., Russell, C. A., & Birmingham, J. (2004). *Building quality, scale, and effectiveness in after-school programs: Summary report of the TASC evaluation*. Washington, DC: Policy Studies Associates, Inc.
- Rhea, A. (2013). *Academic Achievement Academy (AAA), 2011-12* (D & A Report No. 13.03). Cary, NC: Wake County Public School System, Data and Accountability Department.
- Ritchie, L. D., Sharma, S., Ikeda, J. P., Mitchell, R. A., Raman, A., Green, B. S., ... & Fleming, S. E. (2010). Taking Action Together: A YMCA-based protocol to prevent type-2 diabetes in high-BMI inner-city African American children. *Trials, 11*(1), 60.
- Roberson, I. (2006). *After school education and safety program* (Final Evaluation Report. San Francisco, CA: San Francisco Unified School District, Program Evaluation and Research Department.
- Roberts, G. A. (2008). *Effect of an after-school remediation program on the academic achievement of students with disabilities* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3324724)
- Robinson, T. N., Killen, J. D., Kraemer, H. C., Wilson, D. M., Matheson, D. M., Haskell, W. L., . . . , Varady, A. (2003). Dance and reducing television viewing to prevent weight gain in African-American girls: The Stanford GEMS pilot study. *Ethnicity and Disease, 13*, S1-65-S1-77.
- Rosenkranz, R. R., Behrens, T. K., & Dziewaltowski, D. A. (2010). A group-randomized controlled trial for health promotion in girl scouts: Healthier troops in a SNAP (Scouting Nutrition & Activity Program). *BMC Public Health, 10*(1), 81-93.
- Rothman, T., & Henderson, M. (2011). Do school-based tutoring programs significantly improve student performance on standardized tests? *RMLE Online: Research in Middle Level Education, 34*(6), 1-10.
- Roukema, R. A. (2005). *The impact of the Support Our Students (SOS) after-school program on the achievement of middle-grade students at risk of academic failure* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3195186)
- Rumph, M. K. (2011). *Evaluation of CATCH Kids Club after school program: A nutrition and physical fitness intervention for third, fourth, and fifth graders* (Master's thesis). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 1500125)
- Russell, C. A., Mielke, M. B., Miller, T. D., & Johnson, J. C. (2007). *After-school programs and high school success: Analysis of post-program educational patterns of former middle-grades TASC participants*. Washington, DC: Policy Studies Associates, Inc.

- Sanford, E. E. (2010). *Examining the impact of complementary assistance learning on student CRCT scores* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3397116)
- Scarpati, J. (2017). *Are 21st Century after school programs an effective academic intervention for elementary school students attending high poverty schools?* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 10273768)
- Scarpello, J. P. (2014). *Examination of the relationship between academic and demographic variables at Hillsborough High School, NJ* (Doctoral dissertation). Retrieved from <http://delaware.contentdm.oclc.org/cdm/compoundobject/collection/p15323coll5/id/26596/rec/1>
- Schirm, A., Rodriguez-Planas, N., Maxfield, M., & Tuttle, C. (2003). *The Quantum Opportunity Program demonstration: Short-term impacts* (MPR Report No. 8279-093). Washington, DC: Mathematica Policy Research.
- Schirm, A., Stuart, E., & McKie, A. (2006). *The Quantum Opportunity Program demonstration: Final impacts* (MRP Report No. 8279-932). Washington, DC: Mathematica Policy Research.
- Scott, V. (2014). *The impact of an after-school tutorial program on at-risk ninth grade students' academic performance in an Urban Midlands School District* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3662524)
- Sebastian, J. (2013). *The impact of an after-school intervention program on academic achievement among middle school students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3587666)
- Shapley, K., Vicknair, K., Sheehan, D., Pieper, A., Jepson, D., & Sturges, K. (2004). *Texas study of students at risk: Efficacy of grants supporting academic success from elementary through high school*. Austin TX: Texas Center for Educational Research.
- Smart, V. J. (2016). *Improving student achievement in after school programs through technology: An action research study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 10144298)
- Smith, C. F. (2015). *Transforming attitudes and lives: Liberating African-American elementary and middle school students in out-of-school time STEM education* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3684470)
- Smith, T. A. (2005). *An examination of the effectiveness of a specific after-school academic intervention program on the success of at risk students* (Doctoral dissertation). Retrieved from <https://scholarship.shu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1203&context=dissertations>
- Son, D. (2010). *Racial differences in the effectiveness of after school programs* (Unpublished master's thesis). University of Maryland, College Park, MD.
- Speroni, K. G., Earley, C., & Atherton, M. (2007). Evaluating the effectiveness of the Kids Living Fit™ program: A comparative study. *Journal of School Nursing, 23*(6), 329-336.
- St. Pierre, T. L., Mark, M. M., Kaltreider, D. L., & Campbell, B. (2001). Boys & Girls Clubs and school collaborations: A longitudinal study of a multicomponent substance abuse prevention program for high-risk elementary school children. *Journal of Community Psychology, 29*(2), 87-106.
- Starks, F. I., Jr. (2002). *Mentoring at-risk youth: An intervention for academic achievement* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3053015)
- Stencel, J. R. (2005). *A study of the relationship between interscholastic athletic participation and academic achievement for a group of Tennessee high school students* (Doctoral Dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3197287)
- Stephens, L. J., & Schaben, L. A. (2002). The effect of interscholastic sports participation on academic achievement of middle level school students. *NASSP Bulletin, 86*(630), 34-41.

- Story, M., Sherwood, N. E., Himes, J. H., Davis, M., Jacobs, D. R., Jr., Cartwright, Y., Smyth, M., & Rochon, J. (2003). An after-school obesity prevention program for African-American girls: The Minnesota GEMS pilot study. *Ethnicity and Disease, 13*, S1-54-S1-64.
- Tarchichi, R. (2017). *The impact of mathematics, science and language arts writing to learn strategies on the culture of learning in primary and secondary students* (Doctoral dissertation). Retrieved from <https://rdw.rowan.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=3361&context=etd>
- Tom, D. M. (2012). *Impact on achievement with ST math after school instruction* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3530680)
- Townsend, D. R. (2007). *The academic vocabulary development of middle school English learners: An intervention study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3273870)
- Trousdale, D. (2000). First-year evaluation of an after-school program for middle school youth. *ERS Spectrum, 18*(3), 3-11.
- Tyler-Wood, T., Ellison, A., Lim, O., & Periathiruvadi, S. (2012). Bringing Up Girls in Science (BUGS): The effectiveness of an afterschool environmental science program for increasing female students' interest in science careers. *Journal of Science Education and Technology, 21*(1), 46-55.
- U.S. Department of Education. (2016, September). *Non-Regulatory guidance: Using evidence to strengthen education investments*. Washington, D.C.: Author. Retrieved from <https://www2.ed.gov/policy/elsec/leg/essa/guidanceusesinvestment.pdf>
- U.S. Department of Education. (2017). *21st Century Community Learning Centers (21st CCLC) analytic support for evaluation and program monitoring: An overview of the 21st CCLC performance data: 2015–16* (12th report). Retrieved from <https://www2.ed.gov/programs/21stcclc/performance.html>
- Venzen, M. A. (2011). *Impact of the Twenty-first Century afterschool program on student achievement in mathematics and language arts* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3487238)
- Wahlstrom, K., Sheldon, T., & Lewis, A. (2004). *Final evaluation report: 21st Century Community Learning Centers Pathways to Progress, Saint Paul Public Schools*. St. Paul: University of Minnesota Center for Applied Research and Educational Improvement.
- Waldron, A. (2006). *A mixed methods evaluation of an afterschool science club's influence on adolescent girls' attitudes toward science* (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.
- Walker, C. S., Kronick, R., & Diambra, J. F. (2007). Assessment of a full-service school, after-hours tutoring and enrichment program. *Journal of At-Risk Issues, 13*(2), 21-27.
- Watkins, T. R. (2016). *The influence of literacy intervention on mathematics* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 10241100)
- Watson, N. (2003). *The effects of the ninth grade success initiative at Pearland High School* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3094018)
- Welch, A., & Huffman, D. (2011). The effect of robotics competitions on high school students' attitudes toward science. *School Science and Mathematics, 111*(8), 416-424.
- White, L., & Hixson N. (2014). *21st Century Community Learning Centers 2013: A quasi-experimental investigation of program impacts on student achievement in mathematics and reading/language arts*. Charleston, WV: West Virginia Department of Education, Division of Teaching and Learning, Office of Research.
- White, L. J. (2014). *21st Century Community Learning Centers: A descriptive evaluation for 2012-2013*. Charleston, WV: West Virginia Department of Education, Division of Teaching and Learning, Office of Research.

White, L. J., Hammer, P. C., & Whisman, A. (2015). *21st Century Community Learning Centers, 2014-2015: A descriptive evaluation*. Charleston, WV: West Virginia Department of Education, Division of Teaching and Learning, Office of Assessment.

White, L. J., & Whisman, A. (2015). *21st Century Community Learning Centers 2014: A quasi-experimental investigation of program impacts on student achievement in mathematics and reading/language arts*. Charleston, WV: West Virginia Department of Education, Division of Teaching and Learning, Office of Research.

White, R. N., Reisner, E. R., Welsh, M., & Russell, C. A. (2001). *Patterns of student-level change linked to TASC participation, based on TASC projects in year 2*. Washington, DC: Policy Studies Associates, Inc.

Wilcox, M. L. (2012). *The impact of extracurricular activities on academic performance for rural secondary students in Indiana* (Doctoral dissertation). Retrieved from <http://scholars.indstate.edu/handle/10484/4006>

Wilson, D. K., Evans, A. E., Williams, J., Mixon, G., Sirard, J. R., & Pate, R. (2005). A preliminary test of a student-centered intervention on increasing physical activity in underserved adolescents. *Annals of Behavioral Medicine, 30*, 119-124.

Wilson, R. W. (2016). *The effect of school-based physical activity on fifth grade cardiovascular fitness* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 10156429)

Wright, G. B. (2002). *The impact of after school programs on students' academic achievement in Title I schools* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database. (UMI No. 3361423)

Wyatt, J. M. (2002). *An outcome evaluation of an after-school program for children with emotional and behavioral disorders* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3055298)

Yin, Z., Gutin, B., Johnson, M. H., Hanes, J., Moore, J. B., Cavnar, M., . . . , Barbeau, P. (2005). An environmental approach to obesity prevention in children: Medical College of Georgia FitKid Project year 1 results. *Obesity Research, 13*, 2153-2160.

Yin, Z., Hanes, J., Moore, J. B., Humbles, P., Barbeau, P., & Gutin, B. (2005). An after-school physical activity program for obesity prevention in children: The Medical College of Georgia FitKid Project. *Evaluation and the Health Professions, 28*, 67-89.

Yin, Z., Moore, J. B., Johnson, M. H., Barbeau, P., Cavnar, M., Thornburg, J., & Gutin, B. (2005). The Medical College of Georgia FitKit Project: The relations between program attendance and changes in outcomes in year 1. *International Journal of Obesity, 29*, S40-S45.

Yokley-Busby, S. (2013). *The impact of attendance longevity in an after school program, designed to build intentional relationships and support academic success, on urban elementary students' achievement, attendance, and school awards* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3601426)

Zief, S. G. (2005). *A mixed-methods study of the impacts and processes of an after-school program for urban elementary youth* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses Database (UMI No. 3165777)

Zosky, D. L., & Crawford, L. A. (2003). No Child Left Behind: An assessment of an after-school program on academic performance among low-income at-risk students. *School Social Work Journal, 27*, 18-31.

Review Methods

I. Literature Search Strategy

Eight electronic databases were accessed through the library of Abt Associates, Inc., using two blocks of terms designed to capture the universe of potentially eligible studies of afterschool programs. The databases included Academic Search Complete, EconLit, ERIC, Education Research Complete, and JSTOR searched using the EBSCO interface, PsycInfo, ProQuest Dissertations and Theses, and the Web of Science. The text of the search string was modified to suit the conventions of each database (for example, the PsycInfo interface we used does not permit wildcard characters within quotes), but the same strategy was used across all databases. The searches were restricted to publication dates in 2000 and forward and, where possible, studies conducted in the U.S. and written in English. The searches were conducted in 2017 during July and August.

Block A: Methods Terms

“control group*” OR “control condition*” OR random* OR “comparison group*” OR “comparison condition*” OR “regression discontinuity” OR RDD OR “matched group*” OR baseline OR treatment* OR experiment* OR trial OR intervention* OR empirical OR evaluation* OR “research study” OR impact* OR effectiveness OR causal* OR posttest* OR “post-test*” OR “follow-up” or “follow up” OR pretest* OR “pre-test*” OR “pre-post” OR “pretest-posttest” OR QED* OR RCT* OR “propensity score*” OR “quasi-experiment*” OR “research synthesis” OR “meta-analysis” OR “systematic review”

AND

Block B: Program Terms

After-school* OR afterschool* OR “after school” OR “Out of school” OR “enrichment program” OR “extended service school*” OR “boys and girls club*” OR “boys & girls club*” OR BGCA OR “21st Century Community Learning Center*” OR “21st CCLC*” OR “extended school day*” OR “Girls Inc” OR “girl scout*” OR “boy scout*” OR “police athletic league*” OR “police activit* league*” OR “higher achievement” OR “wings for kids” OR extracurricul* OR extra-curricul* OR “extra curricular” OR “youth development” OR “youth serving organization*” OR “beacon center” OR “supplementary education” OR “supplemental instruction”

Search results were loaded into an EndNote library and deduplicated. To screen titles and abstracts for relevance, we used the free, open source application abstrackr (<http://abstrackr.cebm.brown.edu>).

II. The Review Process

Fourteen trained reviewers conducted the reviews for this project. After training on the review protocol and standards, studies were reviewed by a single reviewer. All reviews were quality checked by a senior member of the review team.

III. Additional Details About How the ESSA Tiers Were Operationalized

All studies were reviewed against evidence standards based on ESSA Tiers I-IV, defined as follows:

Tier I (Strong Evidence): this tier corresponds to a rating of Meets Group Design Standards without Reservations using version 4.0 of the What Works Clearinghouse Procedures and Standards Handbook.

Tier II (Moderate Evidence): this Tier corresponds to a rating of Meets Group Design Standards with Reservations using version 4.0 of the What Works Clearinghouse Procedures and Standards Handbook.

When studies were required to adjust for baseline imbalance to meet the moderate evidence tier, the list of approved adjustment strategies provided below was used.

Tier III (Promising Evidence): to meet the promising evidence requirements, studies must have used a randomized controlled trial design or a quasi-experimental design with a comparison group. Studies met the promising evidence requirements under the following circumstances:

- RCTs with high attrition or quasi-experimental designs that did not meet WWC baseline equivalence requirements because the relevant baseline effect size was indeterminate or outside the adjustment range ($>.25$) could meet the promising requirement if they controlled for selection bias using one of the approved adjustment strategies provided below.
- Quasi-experimental designs that did not meet WWC baseline equivalence requirements because the relevant baseline effect size was outside the adjustment range or indeterminate could meet the promising requirement without using an approved adjustment strategy if a propensity score matching procedure was used to construct the comparison group.

- RCTs with high attrition or quasi-experimental designs for which baseline information was provided but not on the analytic sample and either established baseline equivalence or used an approved adjustment method or both could meet the promising requirement.
- RCTs or quasi-experimental designs that had a temporal confound (for example, they used a prior-year comparison group). These studies must have also met any baseline equivalence requirements defined for the promising evidence tier.
- Outcome measures must have met all the requirements for outcome measures described below.

Tier IV (Emerging Evidence): Studies must have used a randomized controlled trial design or a quasi-experimental design with a comparison group to meet the requirements for emerging causal evidence. Studies met the emerging evidence requirements under the following circumstances:

- RCTs with high attrition or QEDs for which baseline equivalence is outside the adjustment range and impact estimates are unadjusted or do not use an approved adjustment strategy;
- RCTs with high attrition or QEDs for which baseline equivalence is indeterminate and impact estimates are unadjusted or do not use an approved adjustment strategy;
- RCTs or quasi-experimental designs that have provider or administrative unit (i.e., $n=1$) confounds;
- Outcomes must not be overaligned, must be collected consistently across conditions, and must be face valid, but may be missing information about reliability. Studies using outcome measures with known low reliability cannot meet emerging causal evidence standards.

Sample Attrition

This review used the WWC liberal boundary for attrition. We selected this boundary based on the assumption that most attrition in studies of afterschool interventions is likely due to factors that are not strongly related to intervention status.

Baseline Equivalence

If the study design was a randomized controlled trial with high levels of attrition or a quasi-experimental design, the study was required to demonstrate baseline equivalence of the intervention and comparison groups for each analytic sample and (if necessary) use an approved adjustment strategy to meet the requirements for strong or moderate evidence. To meet baseline equivalence requirements for the promising evidence tier, the requirement that baseline equivalence be demonstrated on the exact analytic sample was waived.

If demonstration of baseline equivalence was required for a study, the following pre-intervention (or baseline) characteristics were preferred:

- A pre-intervention measure of the outcome (i.e., a pretest), or
- A measure in the same domain as the outcome (i.e., a close proxy for a pretest).

If a pretest or pretest proxy was not available, the study was required to demonstrate baseline equivalence on at least two of the sociodemographic characteristics on the following list:

- Student socioeconomic status [SES] (for example, family income, free- or reduced-price lunch status, parent education levels)
- Student race/ethnicity
- Student grade level or age
- Student gender

If the calculated difference of a specified baseline characteristic was greater than .25 standard deviations in absolute value, the study could meet the Tier III evidence requirements provided acceptable adjustments were made, as described below.

For differences in the specified baseline characteristics that were between .05 and .25 standard deviations, the analysis was required to include an acceptable statistical adjustment for the baseline characteristics to meet the baseline equivalence requirement. In this case the study could meet the Tier II evidence requirements provided acceptable adjustments were made, as described below.

Differences of less than or equal to .05 standard deviations required no statistical adjustment and the study could meet Tier II requirements.

Acceptable statistical adjustments:

- Ordinary or multi-level regression model that includes the covariate at issue as an independent variable
- Analysis of covariance that includes the covariate at issue
- Repeated measures ANOVA (note that the effect size computation for repeated measures ANOVAs typically requires the group by time interaction term)
- Simple gain scores
- Difference-in-differences adjustment
- Fixed effects for individual students

Procedures for Statistical Adjustment for Studies with Baseline Covariate Imbalance

These procedures applied to all studies for which baseline equivalence was required to be demonstrated (i.e., RCTs with high attrition and all quasi-experimental studies). If a pretest was available for an outcome and the difference between conditions at baseline was shown to be within the range that required statistical adjustment, the statistical adjustment was only needed for that outcome. For example, if vocabulary, reading comprehension, and reading fluency were available as pre- and post-intervention measures, and the pre-intervention difference in reading comprehension required statistical adjustment, only the analysis of reading comprehension needed to adjust for baseline differences in reading comprehension.

For outcomes that did not have a pretest or close proxy, if the difference between conditions at baseline on one of the required covariates was shown to be within the range that required statistical adjustment, then adjustment was required only for the covariate in the adjustment range, but was required for all outcomes that did not have pretests or close proxies.

For example, if academic achievement was judged to be within the range that required statistical adjustment and SES was very closely balanced (i.e., it was not in the adjustment range), then all outcomes without pretests needed to adjust for the measure of academic achievement; adjustment for baseline SES was not required.

Requirements for Outcomes

The outcome standards for this review were aligned with those in use by the WWC. Specifically, there are four outcome standards:

- Face validity
- Reliability
- Lack of over-alignment
- Consistency of measurement between treatment and comparison groups

Face Validity. To satisfy the criterion for face validity, there needed to be a sufficient description of the outcome measure for the reviewer to determine that the measure was clearly defined, had a direct interpretation, and measured the construct it was designed to measure.

Reliability. To satisfy the reliability criteria, the outcome measure was required to either (A) be a standardized measure, or (B) meet one or more of the following criteria for reliability: internal consistency (such as Cronbach's alpha) of .50 or higher, test-retest reliability of .40 or higher, or inter-rater reliability (percentage agreement, correlation, or kappa) of .50 or higher. Standardized tests relevant to this review were assumed to meet the eligibility criteria, where standardized is defined as a test designed to be administered, scored, and interpreted in the same way no matter when and where it is administered.

Over-Alignment. To satisfy the criterion related to over-alignment, the outcome could not be closely aligned or tailored to the intervention being tested. This typically occurred when an outcome measure was created by a researcher or intervention developers. Evidence of over-alignment could include an outcome measure that assessed respondents using some of the same materials that are part of the intervention, which could give the intervention group an unfair advantage over the comparison group. Standardized measures or measures that were in general use outside of the particular study being reviewed were unlikely to be deemed as over-aligned.

Consistency of Measurement between Treatment and Comparison Groups. The standard for consistency of measurement required that:

- Measures are constructed the same way for both treatment and comparison groups
- The data collectors and data collection modes for data collected from treatment and comparison groups are either the same or are different in ways that would not be expected to affect the measures
- The time between pre-test (baseline) and post-test (outcome) measures does not does not systematically differ between treatment and comparison groups

Confounding. For this review, confounds were defined as in the WWC Handbook. A confound occurs when a component of the study design or the circumstances under which the intervention was implemented is perfectly aligned, or confounded, with either the intervention or comparison group. That is, some factor is present for members of only one group and absent for all members in the other group.

There are two common types of confounds, which are addressed differently across the evidence tiers:

- **Time confound:** a time confound occurs in quasi-experimental studies when the comparison group is constructed from a different time than the intervention group. For example, if the intervention group is comprised of 4th grade students from 2015 and the comparison group is constructed of 4th graders from the previous school year, a time confound is present.
- **Provider or administrative unit confound (also referred to as an n=1 confound):** This type of confounding occurs when the intervention or comparison group contains a single study unit—for example, when all the intervention students are taught by one teacher, all the comparison classrooms are from one school, or all the intervention group schools are from a single school district. In these situations, there is no way to distinguish between the effect of the intervention and that unit. For example, if all students who use a mathematics intervention are taught by a single teacher, then any subsequent differences between the outcomes of students who use the mathematics intervention and those who do not may be due to the intervention, the teacher, or both.

In addition, when the treatment group is comprised of individuals or units that were offered and accepted treatment and all of the comparison group is comprised of individuals or units that were known to have been offered and refused treatment, the design has a confound. Studies with this type of confound do not meet the requirements for any evidence tier.

IV. Effect Sizes

The standardized mean difference effect size, specifically Hedges' g , was used as the effect size metric for all findings in this review (Hedges, 1981). Effect sizes from binary outcomes were translated to the standardized mean difference metric using the Cox transformation (Sanchez-Meca etc.). Common formulae for computing effect sizes from other statistics were used when means, adjusted means, or adjusted mean differences (as from a regression model) and standard deviations were not given (cf. Lipsey & Wilson, 2001).

