Texas 21st Century Community Learning Centers

2014–15
Evaluation Report

Prepared for: Texas Education Agency

May 2016

Copyright © Notice: The materials are copyrighted © and trademarked ™ as the property of the Texas Education Agency (TEA) and may not be reproduced without the express written permission of TEA, except under the following conditions:

1) Texas public school districts, charter schools, and Education Service Centers may reproduce and use copies of the Materials and Related Materials for the districts' and schools' educational use without obtaining permission from TEA.

2) Residents of the state of Texas may reproduce and use copies of the Materials and Related Materials for individual personal use only without obtaining written permission of TEA.

3) Any portion reproduced must be reproduced in its entirety and remain unedited, unaltered and unchanged in any way.

4) No monetary charge can be made for the reproduced materials or any document containing them; however, a reasonable charge to cover only the cost of reproduction and distribution may be charged.

Private entities or persons located in Texas that are not Texas public school districts, Texas Education Service Centers, or Texas charter schools or any entity, whether public or private, educational or non-educational, located outside the state of Texas MUST obtain written approval from TEA and will be required to enter into a license agreement that may involve the payment of a licensing fee or a royalty.

For information contact: Office of Copyrights, Trademarks, License Agreements, and Royalties, Texas Education Agency, 1701 N. Congress Ave., Austin, TX 78701-1494; e-mail: copyrights@tea.texas.gov.

Texas Assessment of Knowledge and Skills™ (TAKS™) and State of Texas Assessments of Academic Readiness® (STAAR®) are registered trademarks of TEA. Other product and company names mentioned in this report may be the trademarks of their respective owners.

Table of Contents

Highlights of the 2014–15 Evaluation i
Executive Summary iv
Background and Context iv
Statewide Evaluation: Year 1 (2010–11) and Year 2 (2011–12) v
Statewide Evaluation: Year 3 (2012–13) and Year 4 (2013–14) vi
Statewide Evaluation: Year 5 (2014–15) viii
A Summary of Findings ix
2014–15 Evaluation Conclusions x
Broad Conclusions About the Effectiveness of the ACE Program xiii
Recommendations xv
Chapter 1: Introduction 1
Statewide Evaluation: Year 1 (2010–11) and Year 2 (2011–12) 4
Statewide Evaluation: Year 3 (2012–13) and Year 4 (2013–14) 6
Statewide Evaluation: Year 5 (2014–15) 10
Organization of the Report 13
Chapter 2: Connections Between Program Quality and Youth Outcomes 15
Previous Explorations of Quality, 2010–2014 15
RQ 1.1—Relationship Between Quality and Student Outcomes 18
RQ 1.2—Relationship Between Quality and Student Outcomes for SPP and ACE-Only Students 20
RQ 1.3—Mediating Effect of Student Engagement on Quality and Outcomes 22
RQ 1.4—Mediating Effect of Youth Mindsets and Behaviors on Outcomes 26
Conclusions and Key Findings 32
Chapter 3: Cost Effectiveness of the SPP Program 34
Methods 34
Sample 35
RQ 2.1—SPP Cost per Student Relative to School-Related Outcomes 36
RQ 2.2—Relationship Between Funding for Specific Activities and Student Outcomes 40
RQ 2.3—Return on Investment for SPP Programs 42

RQ 2.4—Impact of SPP Implementation on Organizational and Instructional Practice 43
Conclusions and Key Findings 54
Chapter 4: Impact of the ACE Program 57
RQ 3.1—Impact of ACE Programming by Attendance and SPP Status 58
RQ 3.2—The Added Benefit of Participating in SPP and ACE Programming 69
RQ 3.3—Examining Program Typology 72
Conclusions and Key Findings 75
Chapter 5: Conclusions and Recommendations 77
2014–15 Evaluation Conclusions 77
Broad Conclusions about the Effectiveness of the ACE Program 80
Recommendations 84
References 86
Appendices 90
Appendix A. ACE Program's Critical Success Model 91
Appendix B. Study Methods and Data Sources 92
Appendix C. Survey Constructs 96
Appendix D. Analysis Methods 98
Appendix E. RQ 2—Cost Analysis Study Detail 109
Appendix F. RQ 3—Impact Analyses Data Table Detail 113
Appendix G. Program Typology Development Process 115

List of Tables

Table 1. Student Sample for RQ 1.2 20

Table 2. Higher Versus Lower Quality SPP and ACE Centers on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14) 21

Table 3. Higher Versus Lower Quality SPP and ACE Centers on STAAR Test Performance, 30+ Day Attendees (2013–14) 22

Table 4. Effects of Program Quality on STAAR Test Performance via Engagement, 30+ Day Attendees (2013–14) 25

Table 5. Effects of Program Quality via Engagement on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14) 25

Table 6. Frequency Distribution of Construct Category by Time Point (2013–14) 27

Table 7. Pre- to Post-Change on Mindsets and Behaviors Survey by Construct, 30+ Days Attendees (2013–14) 28

Table 8. Relationship Between Program Participation and Mindsets and Behaviors, 30+ Day Attendees (2013–14) 29

Table 9. Effects of Program Participation on Achievement via Youth Mindsets and Behaviors, 30+ Day Attendees (2013–14) 30

Table 10. Effects of Program Participation on School-Day Discipline Incidents and Absences via Youth Mindsets and Behaviors, 30+ Day Attendees (2013–14) 31

Table 11. Distribution of Students Participating in SPP Activities for 30+ Day, by Grade Level (2013–14) 36

Table 12. Percentage of SPP Students Meeting or Exceeding Standard for STAAR-Mathematics and Reading, for 30+ Day Attendees by Grade Level (2012–14) 38

Table 13. Estimated Association Between per-Student Program Expenditures and Student Mathematics and Reading Performance, for Students in Grades 4–9, 30+ Day Attendees (2013–14) 40

Table 14. Estimates of the Association Between per-Student Program Spending by Mode of Instruction and STAAR-Mathematics and Reading Performance, for Students in Grades 4–9, 30+ Day Attendees (2013–14) 42

Table 15. Number of ACE Student Cases Analyzed by Grade and Participation Level (2013–14) 60

Table 16. Participant and Nonparticipant Demographics for 30+ and 60+ Day Attendees (2013–14) 61

Table 17. Effect of ACE Program Participation on STAAR-Reading and English I and II EOC Exams for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14) 62

Table 18. Effect of ACE Program Participation on STAAR-Mathematics and Algebra I EOC for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14) 63
Table 19. Effect of ACE Program Participation on School-Day Discipline Incidents for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14) 64
Table 20. Effect of ACE Program Participation on School-Day Absences for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14) 65
Table 21. Effect of ACE Program Participation on Grade-Level Promotion for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14) 66
Table 22. Student Sample for Analysis of Impact of SPP Program (2013–14) 67
Table 23. Program Participants Versus Nonparticipants on Test Performance, 30+ Day Attendees (2013–14) 68
Table 24. Program Participants Versus Nonparticipants on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14) 69
Table 25. SPP+ACE Versus ACE-Only on STAAR Test Performance, 30+ Day Attendees (2013–14) 71
Table 26. SPP+ACE Versus ACE-Only on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14) 71
Table 27. SPP+ACE Versus ACE-Only on Grade-Level Promotion, 30+ Day Attendees (2013–14) 72
Table 28. SPP+ACE versus ACE-Only on STAAR Test Performance, 30+ Day Attendees (2013–14) 73
Table 29. SPP+ACE Versus ACE-Only on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14) 74
Table C-1. Student Engagement Survey Items (2013–14) 96
Table C-2. Grades 4–12 Survey Items Related to Effort and Persistence and Learner Behaviors (2013–14) 96
Table C-3. Survey Items Related to Sense of Competence as a Learner (2013–14) 97
Table D-1. PQA Subscales 100
Table D-2. Average PQA Scale Scores by Subscale and Center Type (2013–14) 101
Table E-1. Association Between SPP Program Total per Student Expenditures and Student Performance on STARR and EOC Mathematics and Reading (2013–14) 109
Table E-2. Association Between SPP Program Computer-Based and Face-to-Face Instruction per-Student Expenditures and Student Performance on STARR and EOC Mathematics and Reading (2013–14) 111
Table F-1. Effect of ACE Program Participation on School-Day Discipline Incidents for 30+ and 60+ Day Attendees, Relative to Nonparticipants (2013–14) 113

Table F-2. Effect of ACE Program Participation on School-Day Absences for 30+ and 60+ Day Attendees, Relative to Nonparticipants (2013–14)..... 113

Table F-3. Effect of ACE Program Participation on Grade-Level Promotion for 30+ and 60+ Day Attendees, Relative to Nonparticipants (2013–14)..... 114

List of Figures

Figure ES-1. SPP Theory of Change vii
Figure 1. Theory of Youth Change Related to Afterschool Program Participation 5
Figure 2. SPP Theory of Change 7
Figure 3. SPP Theory of Change, Implementation and Student Experiences 17
Figure 4. Percentage of SPP Students Meeting or Exceeding the Standard on STAAR-Mathematics and Reading by Instruction Type, 30+ Day Attendees (2013–14) 39
Figure 5. SPP Return on Investment, Relationship Between Spending and Rate of Meeting STAAR-Mathematics and STAAR-Reading Standards[1] 43
Figure 6. Theory of Youth Change Related to Afterschool Program Participation 82
Figure 7. SPP Theory of Change 82
Figure D-1. Differences Between Higher and Lower Quality Clusters on Key Indicators, 2013–14 102
Figure D-2. Path Analysis Variables for RQ 1.3 103
Figure D-3. Path Analysis Variables for RQ 1.4 104

List of Acronyms

21st CCLC	21st Century Community Learning Centers
ACE	Afterschool Centers on Education
AIR	American Institutes for Research
APT-O	Assessment of Program Practices Tool–Observation
EOC	end-of-course
ELA	English language arts
ESEA	Elementary and Secondary Education Act
ESSA	Every Student Succeeds Act
HLM	hierarchical linear modeling
OCES	Observation of Child Engagement Scale
OST	out-of-school time
PBL	project-based learning
PEIMS	Public Education Information Management System
POS	point-of-service
PQA	Program Quality Assessment
PSM	propensity score matching
RFP	request for proposal
RQ	research question
SPP	STAAR Pilot Project
SE	standard error
STAAR	State of Texas Assessments of Academic Readiness
TAKS	Texas Assessment of Knowledge and Skills
TEA	Texas Education Agency
TEKS	Texas Essential Knowledge and Skills
TX21st	TX21st CCLC Student Tracking System

Texas 21st Century Community Learning Centers

2014–15 Evaluation Report

May 2016

Prepared for:

Texas Education Agency

1120 East Diehl Road, Suite 200
Naperville, IL 60563-1486
800-356-2735 • 630-649-6500
www.air.org

CREDITS

American Institutes for Research
1120 East Diehl Road, Suite 200
Naperville, IL 60563-1486
Phone: 630-649-6500
www.air.org

Gibson Consulting Group
1221 South MOPAC Expressway, Suite 355
Austin, TX 78746
Phone: 512-328-0884
www.gibsonconsult.com

Contributing Authors

American Institutes for Research
Elizabeth Devaney
Neil Naftzger
Feng Liu
Samantha Sniegowski

Gibson Consulting Group
Joseph Shields
Eric Booth

Submitted to

Texas Education Agency
1701 North Congress Avenue
Austin, Texas 78701-1494
Phone: 512-463-9734

Highlights of the 2014–15 Evaluation

The 2014–15 evaluation is organized around three key research questions that support the overarching goals of examining program implementation and impact. Highlights from analyses related to these questions include:

Exploring the Relationship Between Program Quality, Youth Engagement, and Outcomes[1]

- *Participating in high-quality programs may lead to improved outcomes under certain conditions.* Participation in high-quality Afterschool Centers on Education (ACE) was associated with fewer absences. Likewise, participation in high-quality State of Texas Assessments of Academic Readiness (STAAR) Pilot Project (SPP) programming was associated with fewer absences and improved mathematics achievement.
- *High-quality and high-academic content activities can lead to positive outcomes despite low student engagement.* Although the evaluation team saw a negative relationship between quality and engagement and between quality and outcomes when examined separately, there was a positive relationship between high-quality programs and outcomes when viewed through the lens of engagement. This indicates that it is still possible to see improvements in student outcomes when they participate in high-quality programs even if their engagement is low.
- *No relationship was found between improvement in youth-reported mindsets and behaviors and school-related outcomes.* Although it is possible that mindsets and behaviors have no relationship to participation and outcomes, the evaluation team believes it is more likely that the youth report survey used to measure mindsets and behaviors is imperfect or that only a subset of students see improvements on certain skills and belief areas and that these improvements may be difficult to detect when examining students in aggregate.

Exploring the Cost Effectiveness and Sustainability of the SPP Program

- *There is no significant relationship between per-pupil SPP program spending and academic performance, regardless of activity type.* Analyses did not reveal any relationship between per-student spending and performance on the STAAR-

[1] Program quality is determined, for the purposes of this evaluation, by observations of program activities using the Youth Program Quality Assessment (YPQA), Observation of Child Engagement Scale (OCES), and sections of the Assessment of Program Practices Tool (APT). Broadly, these tools ask raters to examine the extent to which staff are able to create a safe and supportive environment, the interactions among staff and youth, how staff engage youth, how youth demonstrate engagement in activities, and the explicit academic content of the programs. More details on these tools and their constructs are included in Appendix B of the main report.

Reading and STAAR-Mathematics exams overall or when the mode of delivery (i.e., Computer-Based or Face-to-Face) was taken into account.

- *The implementation of SPP programming contributed to changes in organizational and instructional practice in many centers.* Most project directors indicated that their experience with the SPP program changed their philosophy toward afterschool program delivery, making them more focused on academic content, alignment to the regular school-day curriculum, and building meaningful partnerships with school leaders and regular school-day staff. Many programs are sustaining components of the SPP program in their centers.

Exploring the Impact of ACE Programming on a Range of Student Outcomes

- *Students who participated in ACE for 60 or more days showed improved STAAR mathematics performance.* Analyses comparing ACE participants to similar nonparticipants found a small relationship between participation and improved mathematics performance. Findings were largest for students in Grade 9.
- *Students who participated in ACE for 60 or more days were more likely than nonparticipants to have a decrease in school-day disciplinary incidents and absences.* Overall, there appears to be a strong relationship between high levels of participation in ACE and decreases in problematic school-related behaviors, particularly for high school students.
- *There appears to be an added benefit to participating in SPP+ACE programming rather than ACE-only programming.* Analysis examining SPP participants compared to nonparticipants—those not enrolled in any ACE or SPP programming—did not find a positive relationship between participation and academic performance. However, when looking at the added benefit of participating in SPP+ACE programming versus ACE-only programming, analyses showed that SPP+ACE was associated with improved performance on STAAR mathematics. This suggests that participating in SPP programming alone may not lead to improved outcomes, but participating in SPP+ACE may contribute to improved outcomes.
- *The Learning Strategies approach and Face-to-Face mode of delivery may be associated with improved outcomes.* Analyses showed evidence of a relationship between the Learning Strategies approach and Face-to-Face mode of delivery and improved STAAR mathematics performance as well as decreased school-day disciplinary incidents.[2]

[2] A *Learning Strategies* approach focuses on learning how to learn and gathering skills applicable to many different content areas rather than building skills in one specific content area.

Recommendations

- *Continue to support the growth and development of ACE programs across the state.* The ACE program as a whole is having an impact on student outcomes, which indicates it may be a worthwhile investment.
- *Continue to emphasize and support quality programming through continuous improvement.* Program quality appears to play an important role in youth outcomes, indicating that providing support for continuous program improvement for ACE program staff is a critical step.
- *Encourage use of the Learning Strategies approach.* Use of the approach appears to be related to improved outcomes. Training for centers on this approach may be valuable.
- *When considering intensive academic interventions, always couple them with traditional ACE enrichment.* SPP activities alone were associated with low levels of engagement and may not be effective on their own. However, when paired with ACE programming, participants appeared to experience improved outcomes.

Executive Summary

Background and Context

A large body of research has shown that afterschool programs can have a positive impact on the young people who attend them, particularly young people from low-income communities. In fact, studies have shown that participating in high-quality programs on a regular basis can contribute to improved academic and social and emotional outcomes, including attendance, discipline referrals, achievement tests, and critical thinking and self-management skills (Auger, Pierce, & Vandell, 2013; Durlak & Weissberg, 2007; Kauh, 2011; Miller, 2003; Naftzger et al., 2013).

The 21st Century Community Learning Centers (21st CCLC) program, originally authorized under Title IV, Part B, of the Elementary and Secondary Education Act (ESEA), as amended by the No Child Left Behind Act of 2001, provides academic enrichment opportunities during nonschool hours for children, particularly students who attend high-poverty and/or low-performing schools.[3] The federal formula grants are awarded to state education agencies, which, in turn, make competitive grant awards to eligible entities to support afterschool and summer learning programs.[4] In July 2002, the federal government awarded the Texas Education Agency (TEA) $24.5 million to fund TEA's first cohort of 21st CCLC grantees for the 2003–04 school year. Since then, Texas has received annual awards that have been used to fund eight grant cycles, and the ninth cycle will begin in the 2016–17 school year.[5] All centers funded by the Texas 21st CCLC program, known in Texas as the *Afterschool Centers on Education* (ACE),[6] are expected to provide programs and services designed to support student performance in the following areas: academic performance, school attendance, school behavior, promotion rates, and graduation rates.

---

[3] For more information, review the authorizing legislation as part of the Elementary and Secondary Education Act (2001), Title IV, Part B at http://www2.ed.gov/policy/elsec/leg/esea02/pg55.html. ESEA was replaced in December, 2015, by the Every Student Succeeds Act (ESSA), which continues funding for the 21st CCLC program. For more details on ESSA, see https://www.whitehouse.gove/sites/whitehouse.gov/files/documents/ESSA_Progress_Report.pdf.

[4] Grantees include local education agencies, nonprofit organizations, for-profit organizations, institutions of higher education, and city or county government agencies.

[5] A cycle represents a cohort of grantees that receive funding for five years. Cycle 5, for example, represents the fifth such cohort to receive funding since TEA began funding for this grant.

[6] In Texas, the 21st CCLC program has its own unique brand that communicates the characteristics of the program and creates statewide awareness so that all Texas centers can identify themselves as part of a bigger picture. Although 21st CCLC is the federal funding source, the programs in Texas are referred to as *Afterschool Centers on Education*, or *Texas ACE*. The term *ACE* will be used throughout the report to refer to the programs in Texas unless reference is made to the federal funding source, in which case the term *21st CCLC* will be used.

The ACE programs have been a long-established resource for providing Texas students with academic support and enrichment opportunities. In 2012, TEA sought to utilize the ACE programs to provide more focused academic support to Texas students who were at risk of academic failure, as measured by the state's assessment of student learning, the State of Texas Assessments of Academic Readiness (STAAR) that was new at that time. STAAR is administered to students in Grades 3–8, and the STAAR end-of-course (EOC) assessments are currently administered after completion of various high school courses.[7] Although overall pass rates for the state were approximately 50% for mathematics and 67% for English language arts (ELA) in 2013, students identified as economically disadvantaged passed at much lower rates than their more advantaged peers (by an average of 26 percentage points) (Center for Public Policy Priorities, 2013).

To provide more focused academic support to academically at-risk youth, TEA created a supplemental grant program as part of its ACE program initiative, with funding beginning in the 2012–13 school year and continuing through the 2013–14 school year. This program, called the *STAAR Pilot Project* (SPP), provided additional funding for ACE programs to increase their academic support in core subjects using evidence-based interventions. Fifteen grantees were awarded competitive grants to establish SPP programming in selected centers where they already had established ACE-only programs.[8]

Statewide Evaluation: Year 1 (2010–11) and Year 2 (2011–12)

Beginning in fall 2010, TEA contracted with American Institutes for Research (AIR) and its partners at Gibson Consulting Group and the David P. Weikart Center for Youth Program Quality to conduct a statewide evaluation of the ACE program, geared toward two primary research objectives that TEA established for the project:

---

[7] STAAR includes annual assessments in reading and mathematics for Grades 3–8, writing at Grades 4 and 7, science at Grades 5 and 8, social studies at Grade 8, and EOC assessments for English I, English II, Algebra I, biology, and U.S. history.

[8] *Grantee* refers to the organization that serves as the fiscal agent on the 21st CCLC grant, and *the center* refers to the physical location where grant-funded services take place. The centers have defined hours of operation and a dedicated staff; they are required to have a position akin to a site coordinator. Each ACE grantee must have at least one center and may have as many as 20 centers. With regard to SPP centers, there were not dedicated SPP centers specifically, even though programs were sometimes referred to that way. Rather, SPP centers were ACE centers that included targeted interventions as the part of the programming that students were recruited to attend. Moreover, there were also ACE-only students and traditional ACE programming at a given SPP center that was available to all students enrolled in the center. The key distinction is that only certain ACE centers received specific supplemental grant funding to include SPP programming, either alongside other traditional ACE programming or as stand-alone programming.

- **Research Objective 1:** Identify and describe innovative strategies and approaches implemented by successful 21st CCLC programs.
- **Research Objective 2:** Conduct a statewide assessment of 21st CCLC programs, operations, participation in the program, and student achievement outcomes.

The results of the initial years of the evaluation are presented in two reports (Naftzger, Manzeske, Nistler, & Swanlund, 2012; Naftzger et al., 2013). Key findings were as follows:

- Three instructional approaches were found to be associated with high levels of student engagement: clarity of purpose, intentional use of time, and an active and interactive activity leader.
- Organizational practices associated with high levels of quality included intentional program design, staff development and collaboration, methods to monitor improvement, linkages to the school day, and community connections.
- There was some evidence of a connection between high-quality programs and high levels of participation.
- Higher levels of attendance (60+ days) in 21st CCLC–funded programs were associated with higher levels of Texas Assessment of Knowledge and Skills (TAKS)-Reading/ELA and mathematics performance, reduced school-day disciplinary incidents and school-day absences, and supported grade-level promotion for participants versus nonparticipants. In addition, a high level of point-of-service (POS) quality was associated with fewer school-day disciplinary incidents and a greater likelihood of grade-level promotion.[9]
- For high school students, participation in an ACE program increased the likelihood of being promoted to the next grade level by 97%. There were similar findings for elementary and middle school students, but the magnitude was much smaller.

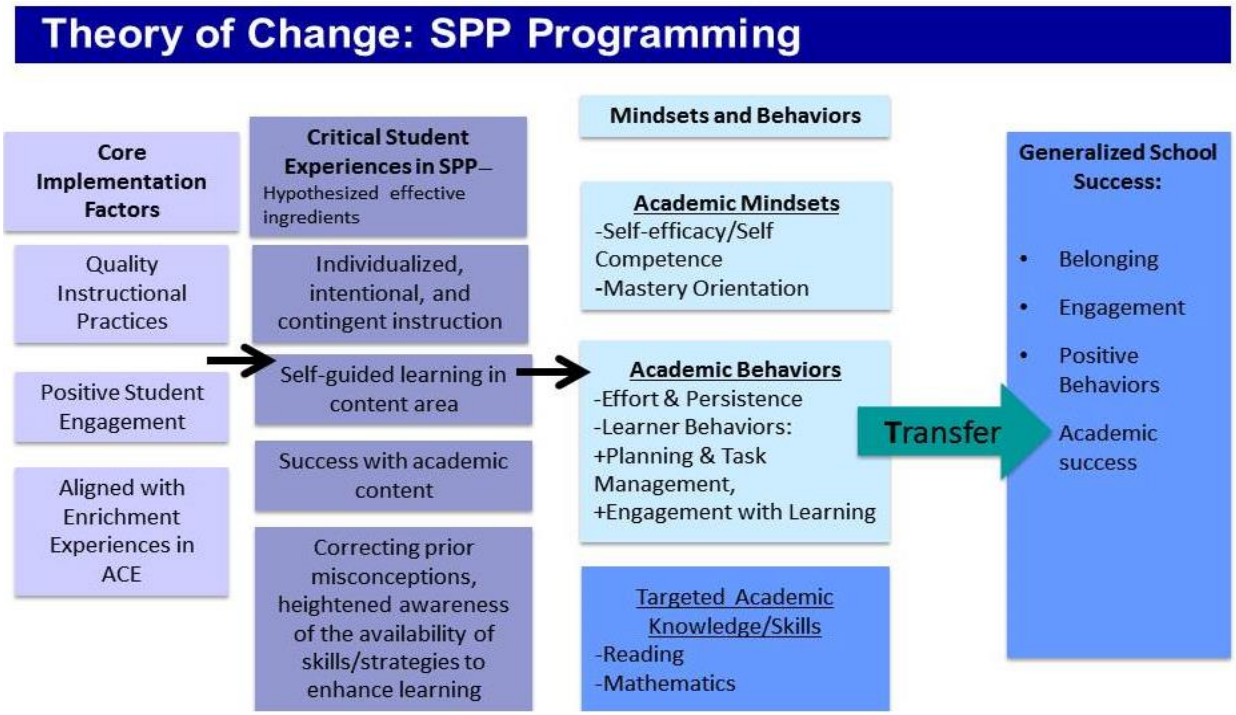
Statewide Evaluation: Year 3 (2012–13) and Year 4 (2013–14)

The introduction of the SPP program led to a change in the evaluation focus, although the overall scope remained the same. For the 2012–13 and 2013–14 evaluations, AIR focused its activities and questions specifically on the SPP program as it compared with traditional ACE programming. In order to make that shift in focus, the evaluation team worked with TEA to develop the SPP theory of change, depicted in Figure ES-1. This theory of change articulates the key facets of SPP implementation that contribute to the

[9] The point of service is where adults and students interact, such as the during program activities.

experiences youth may have in the program. The hypothesis is that a sequence of high-quality, engaging experiences across time will lead to students developing key beliefs and skills, both social-emotional and academic in nature. Improving these key mindsets and behaviors will, in turn, affect youth performance on key metrics during the school day.

Figure ES-1. SPP Theory of Change



The results of the 2012–13 and 2013–14 evaluations are presented in a combined report (Devaney et al., 2015). Key findings from those two years were as follows:

- SPP programs were more academic in nature than were ACE-only programs. This finding was expected given the purpose of the funding for SPP programming and the stated and explicit intention that SPP programs help students at risk for academic failure improve their skills. More specifically, SPP programs hired more certified teachers; SPP students spent more time in academic activities; and SPP activities used smaller groups and longer activities to support academic learning.
- Students participating in SPP programming tended to be more academically at risk and less proficient in key academic mindsets and behaviors than were their ACE-only peers at program onset. This finding is important because it indicates that the SPP program was successful in recruiting the types of students it intended to serve—that is, those at risk for academic failure.

- SPP programs and ACE-only programs had roughly the same level of average overall quality based on a sample of program observations conducted by the evaluation team. Although overall there was not a significant difference between the quality of ACE-only and SPP programs, youth-reported engagement was lower in SPP programs than in ACE-only programs.
- SPP activities that used a Learning Strategies approach, a combination of Computer-Based and Face-to-Face delivery, and a low staff-to-youth ratio, were the most engaging to young people. In addition, the Learning Strategies approach was associated with higher levels of quality than was any other approach.
- There was a small but positive impact of both SPP programming and ACE-only programming on many of the measured academic mindsets and behaviors in the 2012–13 school year.

Statewide Evaluation: Year 5 (2014–15)

The evaluation of the ACE program has concluded its fifth year and was designed to bring all four years of evaluation together to examine the program as a whole and its impact statewide as well as continue exploring the validity of the theory of change. The final year of the statewide evaluation of the ACE program was designed to answer three key research questions, all related to the two overarching objectives of the evaluation to identify innovative strategies and to examine overall program impact. The three research questions (RQ) and associated subquestions are as follows:

- RQ 1—What key practices, strategies, and approaches are especially related to quality programming that leads to improved youth outcomes?
  - RQ 1.1—What is the relationship between program quality[10] and student outcomes?
  - RQ 1.2—How does the relationship between quality and student outcomes differ for SPP students and ACE-only students?
  - RQ 1.3—How does the level of youth engagement mediate the relationship between quality and outcomes?
  - RQ 1.4—How do changes in youth mindsets and behaviors mediate the relationship between participation and outcomes?

[10] Program quality is determined, for the purposes of this evaluation, by observations of program activities using the Youth Program Quality Assessment (YPQA), Observation of Child Engagement Scale (OCES), and sections of the Assessment of Program Practices Tool (APT). Broadly, these tools ask raters to examine the extent to which staff are able to create a safe and supportive environment, the interactions among staff and youth, how staff engage youth, how youth demonstrate engagement in activities, and the explicit academic content of the programs. More details on these tools and their constructs are included in Appendix B of the main report.

- RQ 2—What is the cost effectiveness and sustainability of the SPP program?
  - RQ 2.1—What is the per-student cost of the SPP program relative to school-related student outcomes?
  - RQ 2.2—What is the relationship between the allocation of grant spending to specific activities and student outcomes?
  - RQ 2.3—What is the return on investment for SPP programs?
  - RQ 2.4—In what ways has the SPP experience impacted the work of project directors in how they organize and deliver afterschool programs, and which organizational or instructional components have they incorporated into 2014–15 afterschool programming?
- RQ 3—What is the impact of ACE programming on a range of student outcomes?
  - RQ 3.1—How does the impact of programming differ by attendance level for both the overall ACE program as well as students specifically enrolled in SPP activities?
  - RQ 3.2—What impact does participation in SPP+ACE have on youth outcomes compared to similar youth that participated only in ACE activities?
  - RQ 3.3—If SPP+ACE is found to have a significant impact on youth outcomes, what program typologies are associated with larger program effects?

RQ 1 and RQ 2 serve as a continuation of the 2012–13 and 2013–14 evaluations, concluding analysis of the SPP theory of change. With RQ 3, the evaluation team revisits the analysis done in the 2010–11 and 2012–13 evaluations, looking at the overall impact of the ACE program and attempting to better understand subsets of the participating population (e.g., students participating in SPP activities, students participating in activities employing certain typologies). The 2014–15 evaluation serves as a culminating analysis of the ACE initiative over the past five years.

A Summary of Findings

The 2014–15 evaluation is organized around three key research questions that support the overarching goals of examining program implementation and impact. This summary outlines each research question and the findings associated with it, then provides some overarching conclusions based on five years' worth of evaluation findings, and finally concludes with a series of recommendations for the ACE program as a whole going forward.

2014–15 Evaluation Conclusions

RQ 1—What key practices, strategies, and approaches are especially related to quality programming that leads to improved youth outcomes?

Analyses undertaken to explore this research question revealed that **there is a positive relationship between quality and outcomes**, but that relationship and **the mediating effect of youth-reported engagement on it is not clear, potentially because of the presence of explicit academic content in SPP programming.** In addition, **youth mindsets and behaviors do not appear to have a relationship to participation and outcomes**, although more research is needed to understand these complicated relationships. The findings that contribute to this overall conclusion include:

- *High-quality programs may lead to improved outcomes under certain conditions.* The evaluation team found evidence of a relationship between quality and outcomes when ACE and SPP programs were examined separately. Participation in both high-quality ACE and SPP programs was associated with fewer absences, and participation in high-quality SPP programs was associated with improved mathematics achievement.
- *High-quality and high-academic content activities can lead to positive outcomes despite low engagement.* Although the evaluation team saw a negative relationship between quality and engagement and between quality and outcomes when examined separately, there was still a positive relationship between high-quality programs and outcomes when viewed through the lens of engagement. This indicates that it is still possible for students to improve outcomes when participating in high-quality programs even if their engagement is low. It is the evaluation team's hypothesis that programs with explicit academic content may not be as engaging to youth while still being high quality. The majority of activities observed and included in the analyses, both SPP and ACE-only, this year were academic in nature, which may have led to the results. However, analysis to test this hypothesis was beyond the scope of this evaluation.
- *No relationship was found between improvement in youth-reported mindsets and behaviors and school-related outcomes.* The evaluation team did not find a mediating effect of mindsets and behaviors on the relationship between participation and outcomes. Although it is possible that mindsets and behaviors have no relationship to participation and outcomes, the evaluation team believes it is more likely that the youth report survey used to measure mindsets and behaviors is imperfect or that only a subset of youth see improvements on certain skills and belief areas and that it may be difficult to detect these improvements when examining students in aggregate. Sample size considerations and project

resources did not afford the evaluation team the opportunity to consider all of these options.

RQ 2—What is the cost effectiveness and sustainability of the SPP program?

By conducting a thorough return on investment and cost per student analysis as well as interviews with project directors, the evaluation team was able to determine both that **the amount of funds that centers dedicated to SPP was not related to student outcomes** and that **the SPP program did lead to changes in staff operational and instructional practices** in most centers. This combination of findings may indicate that intensive academic interventions such as SPP can lead to sustained change in practice regardless of how much funding is dedicated to the process. More specifically, the evaluation team found the following:

- *There is no significant relationship between per-pupil SPP program spending and academic performance, regardless of activity type.* Analyses did not reveal any relationship between per-student spending and performing on the STAAR-Reading and STAAR-Mathematics exams overall and when taking the delivery mode (i.e., Computer-Based or Face-to-Face) into account.
- *Implementation of SPP programming contributed to changes in organizational and instructional practice in many centers.* Most project directors indicated that their experience with implementing the SPP program changed their philosophy toward afterschool program delivery, making them more focused on academic content, alignment to the regular school-day curriculum, and building meaningful partnerships with school leaders and regular school-day staff. This change in philosophy has resulted in the incorporation of SPP-related organizational and instructional approaches into their current afterschool work, including those related to staffing and staff training, program monitoring, and recruiting students. Many programs are sustaining components of SPP in their programs despite the conclusion of the funding cycle.

RQ 3—What is the impact of ACE programming on a range of student outcomes?

By conducting rigorous impact analyses on the ACE programming overall as well as a variety of correlational analyses looking at the impact of subsets of students and programs, the evaluation team was able to determine a **relationship between participation and improved STAAR mathematics performance as well as participation and reduced school-day disciplinary incidents and absences**. Findings appear to be **strongest among students who participate at high levels (60+ days) and who are in Grades 9–12**. Specific findings that contribute to these overall conclusions include:

- *Students who participated in ACE for 60 or more days showed improved STAAR mathematics performance.* Analyses comparing ACE participants to similar nonparticipants found a small relationship between participation and improved mathematics performance. Findings were largest for students in Grade 9. Although small, the results were similar to those found in other statewide 21st CCLC evaluations, including the 2011–12 evaluation of ACE programming.
- *Students who participated in ACE for 60 or more days were more likely than nonparticipants to have a decrease in school-day disciplinary incidents and absences.* Here again, there were larger effects for high school students than for elementary and middle school students, but overall, there appears to be a strong relationship between high levels of participation in ACE and decreases in problematic school-related behaviors.
- *Although there does not appear to be an impact on academic performance when comparing SPP participants to nonparticipants, there does appear to be an added benefit to participating in SPP+ACE programming rather than ACE-only programming.* Analyses looking at SPP participants compared to those who did not participate did not find a positive relationship between participation and academic performance. However, when looking at the added benefit of participating in SPP+ACE programming versus ACE-only programming, analyses showed that SPP+ACE was associated with improved performance on STAAR mathematics. This suggests that participating in SPP programming alone may not lead to improved outcomes, but participating in SPP+ACE may contribute to improved outcomes.
- *The Learning Strategies and Face-to-Face approaches may be associated with improved mathematics performance and decreased school-day disciplinary incidents.*[11] In conducting an analysis of the various program typologies, the evaluation team found evidence of a relationship between Learning Strategies and Face-to-Face approaches and improved STAAR mathematics performance. In addition, although both the Learning Strategies and Skills-Based approaches were associated with a decrease in school-day disciplinary incidents, the Learning Strategies approach had a larger effect. Finally, the Face-to-Face approach was also associated with a decrease in school-day disciplinary incidents. Although these findings were simply correlational and cannot definitively point to these two approaches as superior, the analysis results are consistent with the 2012–13 evaluation results that found a Learning Strategies approach was associated with higher quality programming.

[11] *Learning Strategies* approach focuses on learning how to learn and gathering skills applicable to many different content areas rather than building skills in one specific content area.

Broad Conclusions About the Effectiveness of the ACE Program

Conclusions About Impact

- Students participating in ACE programming at high levels (60+ days per year) do see improvements on key school-related indicators. The largest impacts were consistently related to on-time grade-level promotion followed by fewer absences and decreased school-day disciplinary incidents. The program has also had a small impact on mathematics achievement.
- The largest impacts were felt by students in Grades 9–12, especially for on-time grade-level promotion. These impacts were very high in the 2011–12 evaluation and somewhat smaller but still substantial in the 2014–15 evaluation.
- The results related to the SPP program were somewhat less conclusive. Participating in SPP, as compared to not participating in ACE programming at all, was associated with substantial decreases in school-day disciplinary incidents and somewhat smaller but still strong decreases in absences. It was also associated with slight declines in both mathematics and reading performance on the STAAR exam.
- Participating in SPP+ACE rather than ACE-only programming was associated with significantly improved STAAR mathematics performance. Participation in this combination of programming was also associated with decreased school-day disciplinary incidents and absences. These findings suggest that an intensive academic intervention paired with ACE enrichment may provide the right combination of programming to support positive outcomes.
- A *Learning Strategies* approach may be the most effective in supporting student learning. In 2012–13, the *Learning Strategies* approach was associated with higher quality and greater student engagement. In 2014–15, analyses suggested that a relationship may exist between participation in programs using a *Learning Strategies* approach and decreased school-day disciplinary incidents and improved STAAR mathematics performance for students experiencing both SPP and ACE programming.

Conclusions About Quality and Engagement

Conclusions about program quality and engagement as a result of the five-year evaluation are somewhat more complicated than those related to impact. They are further complicated by the introduction of SPP programming to the evaluation starting in 2012–13. Collectively, the five-year evaluation has told us the following about the relationship between quality and engagement:

- **Program quality is related to cognitive engagement.** In 2010–11 and 2012–13/2013–14, the evaluation team looked at the relationship between quality and engagement at the activity level. In both sets of analyses, AIR was able to closely link Program Quality Assessment (PQA) quality scores with youth-reported engagement scores for the same activity and found significant and positive relationships. In 2014–15, the evaluation team did not find this same level of relationship. This might be due to the fact that the analysis was done at the center level rather than the activity level, where there was a more direct connection between what was observed and what youth experienced as a result.
- **Explicit academic content may depress engagement**. In 2013–14, the evaluation team conducted analyses of the relationship between quality and engagement. Analyses were conducted at the activity level and revealed a strong positive relationship between quality and youth-reported engagement in most cases, while SPP–funded activities and activities with explicit academic content were negatively related to engagement. The evaluation team hypothesized at the time that the explicit academic content present in SPP activities may contribute to decreased engagement. This hypothesis could also explain the findings from the 2014–15 analysis that found no relationship between quality and youth-reported engagement because most activities observed, both SPP and ACE-only, were academically oriented. Investigation of this notion was beyond the scope of this evaluation and is an area for further study.
- **Program quality can be related to improved outcomes, even when engagement is low, if academic content is explicit.** Findings from the 2011–12 and 2014–15 evaluations found a relationship between quality and outcomes, albeit through somewhat different paths. In 2011–12, analyses found a relationship between higher quality programs and improved student outcomes when the programs provided intensive academic activities and had high levels of observed engagement. In 2014–15, analyses found a relationship between higher quality programs and improved student outcomes when youth-reported engagement was low. Although these findings seem contradictory, it is important to note that the relationship between quality and outcomes is consistent, and that the measure of engagement varied in the two years. The evaluation team hypothesizes that it is the presence of SPP programming that accounts for this discrepancy.

In sum, the five-year evaluation has demonstrated that there does appear to be relationships among program quality, engagement, and student outcomes for youth participating in ACE programming at high levels.

Recommendations

Based on findings from the specific 2014–15 evaluation and the overarching five-year evaluation, AIR would recommend the following key next steps for the ACE program and the Texas Education Agency.

- ***Continue to support the growth and development of ACE programs across the state.*** The ACE program as a whole is having an impact on student outcomes and may be a worthwhile investment that is helping schools in their ultimate mission to support student success.
- ***Continue to emphasize and support quality programming through continuous improvement.*** Program quality appears to play an important role in youth outcomes. Therefore, it stands to reason that providing support for continuous program improvement for the ACE centers is a critical step. It may be beneficial to provide training on specific practices outlined in the PQA related to creating a supportive environment, supporting strong interactions between adults and youth in programs, and engaging youth in activities.
- ***Encourage use of the Learning Strategies approach.*** The evaluation revealed that a *Learning Strategies* instructional approach may be more effective than the *Skills-Based* approach in engaging youth and contributing to youth outcomes. A *Learning Strategies* approach focuses more on learning how to learn and gathering skills applicable to many different content areas rather than learning specific skills associated with one content area. TEA may consider working with its technical assistance provider to develop training related to this approach.
- ***When considering intensive academic interventions, always couple them with traditional ACE enrichment.*** There was some evidence that an intensive academic intervention such as SPP can be successful within an out-of-school time (OST) setting when students experience it alongside traditional ACE programming. SPP activities alone were associated with low levels of engagement and may not have been effective on their own. However, when paired with ACE programming, they appeared to have positive impacts on participants. There may be value to supporting intensive academic sessions, but students still need the fun and engaging enrichment opportunities that ACE provides.

Chapter 1: Introduction

A large body of research has shown that afterschool programs can have a positive impact on the young people who attend them, particularly young people from low-income communities. In fact, studies have shown that participating in high-quality programs on a regular basis can contribute to improved academic and social and emotional outcomes, including attendance, discipline referrals, achievement tests, and critical thinking and self-management skills (Auger, Pierce, & Vandell, 2013; Durlak & Weissberg, 2007; Kauh, 2011; Miller, 2003; Naftzger et al., 2013). Despite this research, across the United States, at least 11 million children are left alone and unsupervised each weekday afternoon (Afterschool Alliance, 2014). Although a multitude of youth programs are available, existing programs cannot meet the demand, especially for students living in high-poverty communities (Afterschool Alliance, 2015). In Texas alone, there were more than 5 million (5,232,065) students enrolled in public schools during the 2014–15 school year (Texas Education Agency [TEA], 2016). Of those students, 59% (just over 3 million) are identified as economically disadvantaged (TEA, 2016). Young people living in poverty are less likely to graduate than their higher income peers and are more likely to have lower school achievement levels (The Equity and Excellence Commission, 2013). Afterschool opportunities have been shown to correct this imbalance between lower and higher income youth (Auger et al., 2013).

The 21st Century Community Learning Centers (21st CCLC) program, originally authorized under Title IV, Part B, of the Elementary and Secondary Education Act (ESEA), as amended by the No Child Left Behind Act of 2001, attempts to correct this lack of access to critical afterschool opportunities for young people through the creation of community learning centers that provide academic enrichment opportunities during nonschool hours for children, particularly students who attend high-poverty and/or low-performing schools.[12] The federal formula grants are awarded to state education agencies, which, in turn, make competitive grant awards to eligible entities to support afterschool and summer learning programs.[13] In July 2002, the federal government awarded TEA $24.5 million to fund TEA's first cohort of 21st CCLC grantees for the 2003–04 school year. Since then, Texas has received annual awards that have been used to fund eight grant cycles, and the ninth cycle will begin in the 2016–17 school

[12] For more information, review the authorizing legislation as part of the Elementary and Secondary Education Act (2001), Title IV, Part B at http://www2.ed.gov/policy/elsec/leg/esea02/pg55.html. ESEA was replaced in December, 2015, by the Every Student Succeeds Act (ESSA), which continues funding for the 21st CCLC program. For more details on ESSA, see https://www.whitehouse.gove/sites/whitehouse.gov/files/documents/ESSA_Progress_Report.pdf.
[13] Grantees include local education agencies, nonprofit organizations, for-profit organizations, institutions of higher education, and city or county government agencies.

year.[14] All centers funded by the Texas 21st CCLC program, known in Texas as the *Afterschool Centers on Education* (ACE),[15] are expected to provide programs and services designed to support student performance in the following areas: academic performance, school attendance, school behavior, promotion rates, and graduation rates. Currently, the program serves approximately 130,000 students across the state, or about 4% of students identified as economically disadvantaged in the state.

The ACE programs have been a long-established resource for providing Texas students with academic support and enrichment opportunities. In 2012, TEA sought to utilize the ACE programs to provide more focused academic support to Texas students who were at risk of academic failure, as measured by the state's assessment of student learning, the State of Texas Assessments of Academic Readiness (STAAR) that was new at that time. STAAR is a more rigorous assessment than the Texas Assessment of Knowledge and Skills (TAKS), which it replaced. STAAR is administered to students in Grades 3–8, and the STAAR end-of-course (EOC) assessments are currently administered after completion of various high school courses.[16] Although overall pass rates for the state were approximately 50% for mathematics and 67% for English language arts (ELA) in 2013, students identified as economically disadvantaged passed at much lower rates than their more advantaged peers (by an average of 26 percentage points) (Center for Public Policy Priorities, 2013). This lower pass rate is notable given the fact that the federal 21st CCLC program is designed to target students attending schools serving students identified as economically disadvantaged and so have an opportunity to help those students perform better on state tests.

To provide more focused academic support to academically at-risk youth, TEA created a supplemental grant program as part of its ACE program , with funding beginning in the 2012–13 school year and continuing through the 2013–14 school year. In the request for proposal (RFP) issued on October 3, 2012, the STAAR Pilot Project (SPP) was described as a "pilot initiative which will provide current 21st CCLC grantees with additional funding to increase their use of evidence-based interventions to support additional academic assistance in core subjects to help the students meet or exceed

---

[14] A cycle represents a cohort of grantees that receive funding for five years. Cycle 5, for example, represents the fifth such cohort to receive funding since TEA began funding for this grant.

[15] In Texas, the 21st CCLC program has its own unique brand that communicates the characteristics of the program and creates statewide awareness so that all Texas centers can identify themselves as part of a bigger picture. Although 21st CCLC is the federal funding source, the programs in Texas are referred to as *Afterschool Centers on Education*, or *Texas ACE*. The term *ACE* will be used throughout the report to refer to the programs in Texas unless reference is made to the federal funding source, in which case the term *21st CCLC* will be used.

[16] STAAR includes annual assessments in reading and mathematics for Grades 3–8, writing at Grades 4 and 7, science at Grades 5 and 8, social studies at Grade 8, and EOC assessments for English I, English II, Algebra I, biology, and U.S. history.

standards established by the new STAAR assessments." Grantees were awarded competitive grants to establish SPP programming in selected centers where they already had established ACE-only programs.[17]

Eligibility for SPP funding in 2012–13 was limited to grantees that had received 21st CCLC funds in Cycles 5, 6, and 7 and already had demonstrated success with respect to adherence to program requirements and the implementation of practices for struggling students. Eligibility also was limited to grantees that had met certain program implementation criteria (i.e., requiring low to medium technical assistance, having no noncompliance incidents related to grant and program operation requirements, having met student participation requirements in all ACE programs in 2011–12, and having spent 85% or more of their fiscal year 2012 grant award).

Consistent with the requirements for ACE programs, the grantees were required to develop SPP programs aligned with the ACE program's Critical Success Model (see Appendix A). Specifically mentioned in the RFP were these items:

- Innovative instructional techniques based on research and best practices be implemented
- Adult advocates support student involvement in school
- Preassessment and postassessment data be used to identify student needs and provide targeted interventions
- Professional development be provided to staff to increase their effectiveness

In November 2012, 15 21st CCLC grantee agencies were awarded SPP supplemental grants through a competitive process and then established SPP programs in 46 centers where they also were implementing ACE programs. That means that in 2012–13, approximately 5% (46 out of the 864) of all Texas ACE centers had SPP programming. The grantee agencies included 10 school districts (Austin, Fort Worth Cycles 6 and 7, Harlingen, Richardson, Sherman, Snyder, Temple, Taylor, and Valley View); one regional education service center (ESC 13); two community-based organizations (Communities in Schools, one in San Antonio and the other in southeast Harris County);

[17] *Grantee* refers to the organization that serves as the fiscal agent on the 21st CCLC grant, and *the center* refers to the physical location where grant-funded services take place. The centers have defined hours of operation and a dedicated staff; they are required to have a position akin to a site coordinator. Each ACE grantee must have at least one center and may have as many as 20 centers. With regard to SPP centers, there were not dedicated SPP centers specifically, even though programs were sometimes referred to that way. Rather, SPP centers were ACE centers that included targeted interventions as the part of the programming that students were recruited to attend. Moreover, there were also ACE-only students and traditional ACE programming at a given SPP center that was available to all students enrolled in the center. The key distinction is that only certain ACE centers received specific supplemental grant funding to include SPP programming, either alongside other traditional ACE programming or as stand-alone programming.

one county department of education (Harris County); and one charter school (NYOS). The SPP grants were continued in the 2013–14 school year, but two Cycle 5 grantees (Richardson and Communities in Schools in southeast Harris County) ended their programs at the conclusion of the 2012–13 school year. Therefore, 13 grantees operating 94 centers, 37 of which had SPP funding, were included in the 2013–14 and 2014–15 evaluations.

Statewide Evaluation: Year 1 (2010–11) and Year 2 (2011–12)

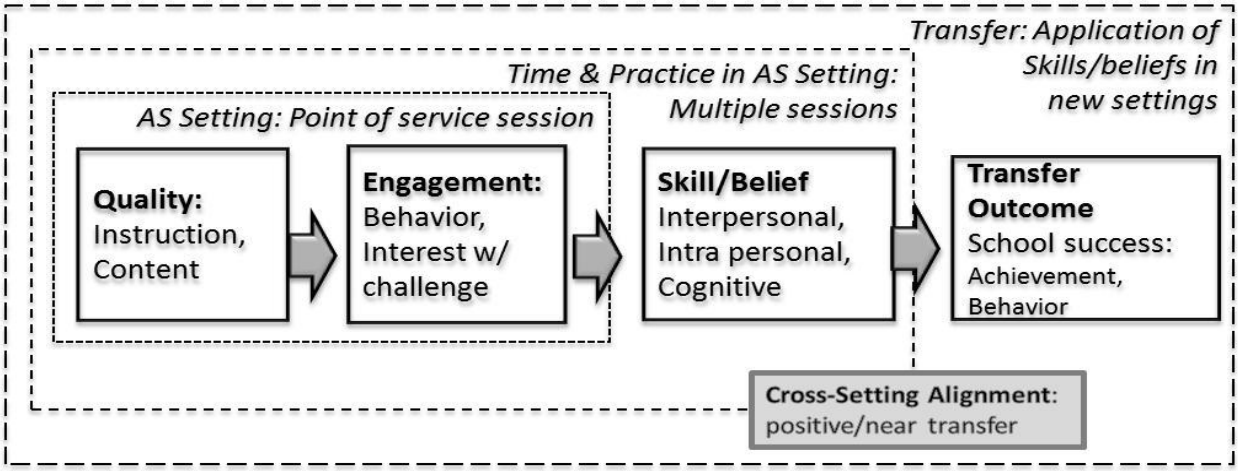
Beginning in fall 2010, TEA contracted with American Institutes for Research (AIR) and its partners at Gibson Consulting Group and the David P. Weikart Center for Youth Program Quality to conduct a statewide evaluation of the ACE program, geared toward two primary research objectives TEA established for the project:

- **Research Objective 1:** Identify and describe innovative strategies and approaches implemented by successful 21st CCLC programs.
- **Research Objective 2:** Conduct a statewide assessment of 21st CCLC programs, operations, participation in the program, and student achievement outcomes.

In order to accomplish these two objectives and to guide the overall evaluation, AIR identified the theory of child-level change outlined in Figure 1, developed in partnership with the David P. Weikart Center for Youth Program Quality. Although not specific to the ACE program, it has been used by multiple afterschool programs around the country and provided an overarching frame for the five-year evaluation effort. In particular, the theory outlines how the ACE program might have an impact on school success (i.e., TAKS or STAAR performance in reading and mathematics, school-day attendance, grade-level promotion) by first establishing high-quality instruction and content, engaging youth in programming, and developing key mindsets and behaviors in the afterschool program that can then be transferred to new settings. This theory is supported by a 2007 meta-analysis examining the connection between developing personal and social skills in afterschool settings and a range of outcomes, including academic achievement (Durlak & Weissberg, 2007). The study found that afterschool programs employing what the authors dubbed the *S.A.F.E. features* (for sequenced, active, focused, and explicit) had significant benefits for youth on a wide range of outcomes. Programs missing these four features showed no significant outcomes. Since the release of this study, additional research has emerged on the importance of point-of-service (POS) quality in afterschool and expanded learning programs. Such research has found that programs characterized by a breadth of offerings and substantive levels of participation by youth across time can achieve a variety of social-emotional and academic outcomes (Auger et al., 2013; Kataoka & Vandell, 2013; Kauh,

2011). More specifically, youth who report high levels of engagement in their afterschool programs have better prosocial skills, task persistence, and work habits (Kataoka & Vandell, 2013).

Figure 1. Theory of Youth Change Related to Afterschool Program Participation



AS = afterschool

For the first two years of the evaluation, AIR began to explore the validity of this theory of change by answering the following evaluation questions associated with TEA's broad evaluation objectives:

- To what extent do students participating in services and activities funded by ACE demonstrate better achievement (along with other student outcomes) compared with similar students not participating in the program (separate estimates were calculated for youth attending one day or more, 30 days or more, and 60 days or more)?
- To what extent do students who have higher participation rates (60 days or more) demonstrate better achievement (along with other student outcomes) compared with similar students who participate in ACE at lower levels (30 days to 59 days)?
- Does ACE's impact on student outcomes vary by relevant program characteristics, particularly program quality?
- Are particular instructional and organizational practices and approaches more likely to produce positive changes in students' engagement in ACE programming?

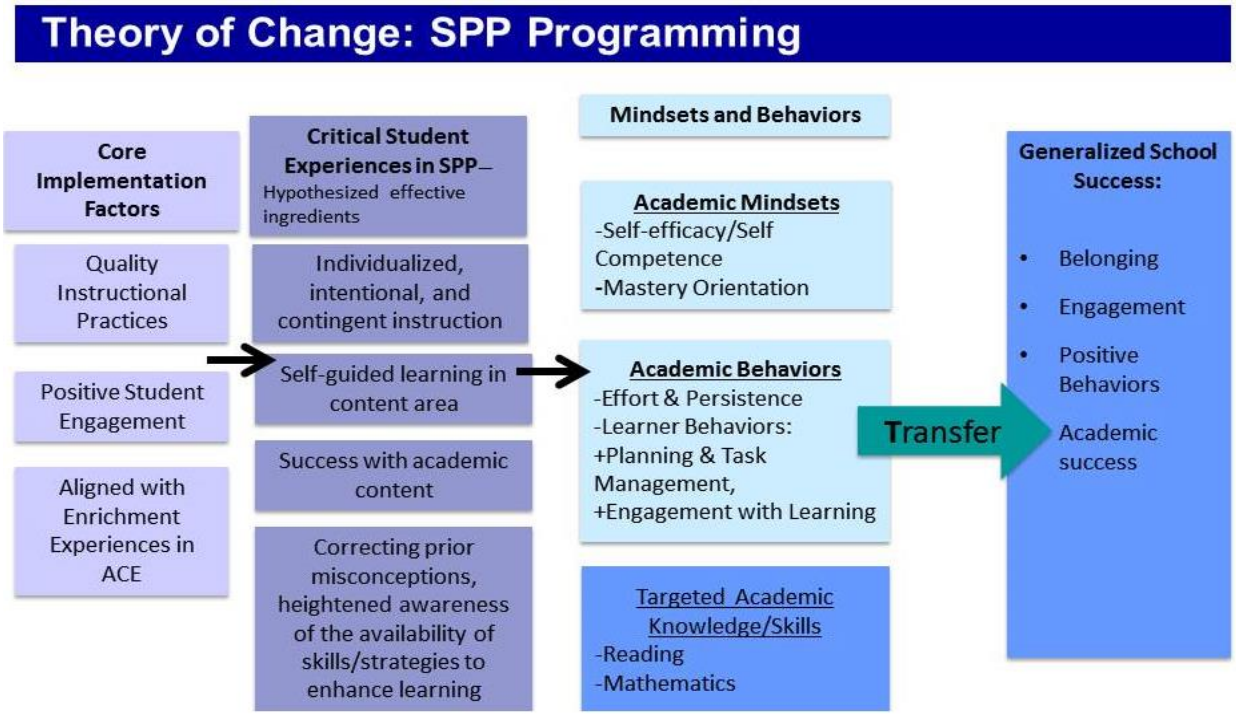
The results of the initial years of the evaluation are presented in two reports (Naftzger, Manzeske, Nistler, & Swanlund, 2012; Naftzger et al., 2013). Key findings were as follows:

- Three instructional approaches were found to be associated with high levels of student engagement: clarity of purpose, intentional use of time, and an active and interactive activity leader.
- Organizational practices associated with high levels of quality included intentional program design, staff development and collaboration, methods to monitor improvement, linkages to the school day, and community connections.
- There was some evidence of a connection between high-quality programs and high levels of participation.
- Higher levels of attendance (60+ days) in ACE programs were associated with higher levels of TAKS-Reading/ELA and mathematics performance, reduced school-day disciplinary incidents and school-day absences, and supported grade-level promotion. In addition, a high level of POS quality was associated with fewer school-day disciplinary incidents and a greater likelihood of grade-level promotion.
- For high school students, participation in an ACE program increased the likelihood of being promoted to the next grade level by 97%. There were similar findings for elementary and middle school students, but the magnitude was much smaller.

Statewide Evaluation: Year 3 (2012–13) and Year 4 (2013–14)

The introduction of the SPP program led to a change in the evaluation focus, although the overall scope remained the same. For the 2012–13 and 2013–14 evaluations, AIR focused its activities and questions specifically on the SPP program as it compared with traditional ACE programming. In order to make that shift in focus, the evaluation team worked with TEA to adapt the generic theory of change presented earlier to fit the specific SPP model. The SPP theory of change, depicted in Figure 2, articulates the key facets of SPP implementation that contribute to the experiences youth may have in the program. As with the generic theory of change, the evaluation team hypothesized that a sequence of high-quality, engaging experiences across time would lead to students developing key beliefs and skills, both social-emotional and academic in nature. Improving these key mindsets and behaviors would, in turn, affect youth performance on key metrics during the school day. The specifics of the theory of change follow Figure 2.

Figure 2. SPP Theory of Change



SPP Theory of Change Components

Core Implementation Factors

Quality Instructional Practices. As shown in Figure 2, high-quality instruction is one of the core implementation factors that leads to the second column of critical student experiences in SPP programming. Quality instructional practices are conceptualized as a series of practices and approaches that result in the creation of developmentally appropriate, high-quality settings for youth at the POS. These practices and approaches include the following:

- *Organizational-Level Practices.* Practices, structures, and approaches adopted by the organization as a whole, such as how staff are hired, oriented, trained, and evaluated; how partners are selected and integrated into program operations; and how the program creates developmentally appropriate settings aligned with core youth development principles and meaningful and relevant academic content
- *Activity Leader-Level Practices.* Practices and approaches directly adopted by activity leaders to make activity sessions supportive, interactive, engaging, and cognitively stimulating at the POS.

- Data related to program quality are important because they make it possible to explore the relationship between program quality and youth outcomes.

Positive Student Engagement. As the model posits, students need to engage positively in activities in order to benefit from participation. Thus, for the evaluation of the SPP programs, data on youth engagement are essential.

Aligned Enrichment Experiences. A third core implementation factor is aligned enrichment experiences that allow youth to engage in fun, hands-on learning that extends their SPP academic activities and allows for continued skill development.

Critical Student Experiences in SPP

The second column in Figure 2 indicates the critical experiences students must have in the SPP programming in order to develop the mindsets and behaviors as intended. These experiences represent high-quality practices, such as individualized instruction, self-guided learning, and correcting knowledge, so that students can have success with the content.

Mindsets and Behaviors

The third column in Figure 2 hypothesizes that if youth sustain engagement in high-quality activities throughout multiple sessions, they will change their mindsets and behaviors as a result of their participation. These mindsets and behaviors include a *Sense of Self-Competence* and *Self-Efficacy*, an orientation toward mastery, an ability to engage with learning, and a willingness to persist. With the change in academically relevant mindsets and behaviors, specific, content-related skills in areas such as reading and mathematics may be strengthened, particularly as a result of individualized academic support to students.

Generalized School Success

Finally, the last column in Figure 2 depicts learning transferred to the school setting. Context-specific mastery experiences support longer term skill development and skill transfer to the school day, ultimately leading to improved academic outcomes, such as improved grades and scores on state achievement tests, greater sense of attachment to school, and reduced dropout rates.

Research Questions

For the third and fourth year of the evaluation of ACE programs, and specifically those with funding to implement SPP programs, AIR focused on six research questions (RQs) that continued to support the broader statewide evaluation objectives of impact and

implementation and were designed to explore the validity of the SPP theory of change. The six RQs that guided the 2012–13 and 2013–14 evaluation activities were as follows:

- How does SPP programming compare with ACE-only programming in centers administered by the same grantee?
- How do students participating in SPP programming differ from students who participate in ACE-only programming?
- How does the quality of delivery differ between SPP and ACE-only programming?
- What instructional and administrative practices lead to high student engagement?
- How engaged are young people in SPP and ACE-only programming? What is the relationship between quality of delivery and student engagement?
- What is the impact of SPP programming on students' academic mindsets and behaviors? How does this compare with ACE-only students?

The results of the 2012–13 and 2013–14 evaluations are presented in a combined report (Devaney et al., 2015). Key findings from those two years were as follows:

- SPP programs were more academic in nature than were ACE-only programs. This finding was expected given the purpose of the funding for SPP programming and the stated and explicit intention that SPP programs help students at risk for academic failure improve their skills. More specifically, SPP programs hired more certified teachers; SPP students spent more time in academic activities; and SPP activities used smaller groups and longer activities to support academic learning.
- Students participating in SPP programming tended to be more academically at risk and less proficient in key academic mindsets and behaviors than their ACE-only peers at program onset. This finding is important because it indicates that the SPP program was successful in recruiting the types of students it intended to serve—that is, those at risk for academic failure.
- SPP programs and ACE-only programs had roughly the same level of average overall quality based on a sample of program observations conducted by the evaluation team. Although overall there was not a significant difference between the quality of ACE-only and SPP programs, youth-reported engagement was lower in SPP programs than in ACE-only programs.[18]

[18] Throughout this report, statistically significant differences are defined as those with a *p* value of 0.05 or below, which means that there is a 5% probability (or less) of randomly observing a difference of this size

- SPP activities that used a *Learning Strategies* approach, a combination of *Computer-Based* and *Face-to-Face* delivery, and a low staff-to-youth ratio, were the most engaging to young people. In addition, the *Learning Strategies* approach was associated with higher levels of quality than any other approach.
- There was a small but positive impact of both SPP programming and ACE-only programming on many of the measured academic mindsets and behaviors in the 2012–13 school year.

Statewide Evaluation: Year 5 (2014–15)

The fifth year of the ACE program evaluation was designed to bring all four years of evaluation together to examine the program as a whole and its impact statewide as well as continue exploring the validity of the generic theory of change outlined in Figure 1 and the SPP theory of change presented in Figure 2. The remainder of this report is dedicated to describing findings from the fifth year of the evaluation.

Research Questions

The final year of the statewide evaluation of the ACE program was designed to answer three key research questions, all related to the two overarching objectives of the evaluation to identify innovative strategies and to examine overall program impact. The three research questions and associated subquestions are as follows:

- RQ 1—What key practices, strategies, and approaches are especially related to quality programming that leads to improved youth outcomes?
  - RQ 1.1—What is the relationship between program quality and student outcomes?
  - RQ 1.2—How does the relationship between quality and student outcomes differ for SPP students and ACE-only students?
  - RQ 1.3—How does the level of youth engagement mediate the relationship between quality and outcomes?
  - RQ 1.4—How do changes in youth mindsets and behaviors mediate the relationship between participation and outcomes?

---

or greater if no difference exists. A moderately significant difference would yield a *p* value less than 0.1, where there would be a 10% probability of observing a difference of this size by chance. Essentially, these two terms help to illustrate the degree of confidence the research team has in the findings. Generally, the evaluation team presents only findings that are statistically significant, unless otherwise noted. In a few places, when it is relevant to the discussion, nonsignificant or moderately significant findings are shared to clarify a point or provide context for follow-up analysis.

- RQ 2—What is the cost effectiveness and sustainability of the SPP program?
  - RQ 2.1—What is the per-student cost of the SPP program relative to school-related student outcomes?
  - RQ 2.2—What is the relationship between the allocation of grant spending to specific activities and student outcomes?
  - RQ 2.3—What is the return on investment for SPP programs?
  - RQ 2.4—In what ways has the SPP experience impacted the work of project directors in how they organize and deliver afterschool programs, and which organizational or instructional components have they incorporated into 2014–15 afterschool programming?
- RQ 3—What is the impact of ACE programming on a range of student outcomes?
  - RQ 3.1—How does the impact of programming differ by attendance level for both the overall ACE program as well as students specifically enrolled in SPP activities?
  - RQ 3.2—What impact does participation in SPP+ACE have on youth outcomes compared to similar youth that participated only in ACE activities?
  - RQ 3.3—If SPP+ACE is found to have a significant impact on youth outcomes, what program typologies are associated with larger program effects?

RQ 1 and RQ 2 serve as a continuation of the 2012–13 and 2013–14 evaluations, concluding analysis of the SPP theory of change. With RQ 3, the evaluation team revisits the analysis done in the 2010–11 and 2012–13 evaluations, looking at the overall impact of the ACE program and attempting to better understand subsets of the participating population (e.g., students participating in SPP activities, students participating in activities employing certain typologies). Together, the 2014–15 evaluation serves as a culminating analysis of the ACE initiative over the past five years. The chapters that follow highlight specific analyses related to each research question and then conclude with overall findings and recommendations that tie the five-year study together.

Data Sources and Methods

A comprehensive set of data sources and methods have informed the findings presented in this report. Although analyzed in 2014–15, data that provide the basis for the report were collected during the 2013–14 school year and include data collected from the SPP program staff through interviews and surveys, student pre-post surveys measuring key mindsets and behaviors, data collected through observations of program

activities and student engagement surveys, and Texas administrative and student data. The data sources and methods are described in more detail in Appendix B.

Limitations

There are several important limitations to keep in mind when reviewing the data and findings in this report.

1. ***Intervention duration.*** The SPP program was a short-term grant program to test the concept of a more intensive academic intervention within the traditional ACE program. The analyses related to the SPP program presented in this report should be considered exploratory given the fact that the program may not have become fully operational and adherent to the model after only 1.5 years of implementation.
2. ***Propensity score matching (PSM) limitations.*** In addition, for the impact analyses presented in Chapter 4, it is important to keep in mind that the PSM approach that was employed to carry out these analyses should not be considered equivalent to an experimental study, which has stronger internal validity. The goal of a PSM approach is to create a comparison group that is as similar to the intervention or treatment group as possible and therefore minimize the impact of differences between the two groups on the estimates of a program's impact. However, the process cannot control for all bias. To the extent that other variables exist (not available for this analysis) that predict student attendance in the program and are also related to student outcomes, these analyses may be limited. Therefore, these analyses should be viewed as initial evidence about the impact of ACE on academic achievement and other outcomes.
3. ***Small sample size.*** For some of the analyses in this report, PSM was deemed not to be viable given limited sample sizes for the creation of a comparison group, requiring the evaluation team to use a less rigorous correlational approach.[19] The implementation window (1.5 years) and the relatively small number of grantees (as compared to the ACE program) for the program mean that the number of students participating was limited. When the evaluation team tried to limit the SPP sample to certain types of students (e.g., those participating in highly academic programming or enrolled in specific grades), the sample was limited even further.
4. ***Inferences about quality.*** The evaluation team conducted site visits at only a subset of centers offering SPP and ACE programming and within those centers, observed only some of the activities being offered. The evaluation team made inferences about overall program quality based on those site visits and observations. It is possible that these inferences were not accurate given the relatively small amount of programming that was observed.

[19] Analyses that use a correlational approach are specified in the individual report sections.

5. ***The varying nature of the SPP and ACE programs.*** There are a number of other limitations related to the nature of the SPP and ACE programs that are worth noting here. One is that neither SPP nor ACE implementation fidelity were measured as part of this evaluation, so it is possible that implementation varied widely across centers. In addition, each SPP grantee was able to select an intervention that best met their local STAAR performance improvement needs. Therefore, SPP grantees generally used different interventions and selected different student populations to participate. It is important to note that these interventions may vary in their implementation as well as in their individual effectiveness. Further, some ACE centers with SPP programming divided the two groups of students so that SPP students participated in activities only with other SPP students, even when they were engaged in enrichment versus academic intervention. In other centers, students stayed in SPP groups during academic interventions only some of the days after school (e.g., Tuesday/Thursday) and were placed into ACE activities when they were not explicitly participating in SPP. In addition to this, some grantees had begun to incorporate aspects of SPP into their overall ACE programming in the second year of the SPP grant. Although the evaluation team tried to identify centers where that was happening, it is possible aspects were implemented without being identified. Finally, SPP and ACE students both participated in school-day programming, activities, and curriculum that were not measured here and could influence the results.

Organization of the Report

This report contains a description of findings and recommendations from the 2014–15 evaluation and analyses associated with each of the research questions presented previously. Each of the next three chapters addresses a different research question. The report then concludes with a summary of findings and recommendations. Finally, the report contains extensive appendices that provide more detail about the evaluation methods, site visits, and instrumentation. The content of the remainder of the report is as follows:

- **Chapter 2:** This chapter shares an overview of findings related to exploration of the SPP theory of change. In particular, the chapter addresses findings related to the relationships among the quality of programming, youth engagement, changes in youth mindsets and behaviors, and youth outcomes.
- **Chapter 3:** This chapter presents analyses of cost per student as related to student outcomes for SPP programs; grant spending and specific student outcomes to show how spending may be related to outcomes; and return on investment. It also presents qualitative interview findings related to sustainability

including how the SPP program has changed the way they operate their programs.

- **Chapter 4:** This chapter provides the reader with an understanding of the ACE program's overall impact on youth outcomes as well as the impact of the SPP program in particular. The chapter explores how impacts differ by level of program attendance as well as whether there is an added benefit to students who participated in the SPP program in addition to ACE programming. Finally, this chapter discusses the association between program typology and outcomes for SPP programs.
- **Chapter 5:** This final chapter provides the reader with evaluative summary and recommendations for how findings can be used to improve programming.
- References
- Appendices

Chapter 2: Connections Between Program Quality and Youth Outcomes

The generic theory of change for afterschool programs across the country presented in the introductory chapter of this report contends that when afterschool programs provide young people with high-quality programming, it can lead to high levels of engagement and participation, which in turn leads to improvements in key mindsets and behaviors and eventually school-related outcomes. The five-year evaluation of the ACE program has consistently looked at the connections between participation, quality, and engagement. In this final year of the evaluation, AIR has attempted to complete its examination of the full theory of change. This chapter briefly describes previous efforts to explore the importance of quality and then outlines analyses conducted during 2014–15 exploring the full theory of change. In particular, this chapter examines the connections among quality programming, student engagement, key mindsets and behaviors and desired outcomes to answer the following research questions:

- RQ 1—What key practices, strategies, and approaches are especially related to quality programming that leads to improved youth outcomes?
  - RQ 1.1—What is the relationship between program quality and student outcomes?
  - RQ 1.2—How does the relationship between quality and student outcomes differ for SPP students and ACE-only students?
  - RQ 1.3—How does the level of youth engagement mediate the relationship between quality and outcomes?
  - RQ 1.4—How do changes in youth mindsets and behaviors mediate the relationship between participation and outcomes?

Previous Explorations of Quality, 2010–2014

In previous years, the evaluation of the ACE program has looked at quality in a number of ways. In the first two years, the evaluation team measured the presence of instructional practices in ACE activities likely to support youth interest, engagement, and skill development through direct observation of programming by using quality assessment tools commonly used in the out-of-school time (OST) field.[20] These tools included the Program Quality Assessment (PQA), which focuses on *process* quality (e.g., the extent to which staff provide supports and opportunities that result in the creation of a supportive, interactive, and engaging environment) and the *Academic Skill Building* section of the Assessment of Program Practices Tool–Observation (APT-O),

[20] At this point in the evaluation, SPP did not exist, and only ACE activities were included in the observations.

which outlines the types of supports and opportunities OST staff can provide to support skill development in particular *content* areas (see the Observation Measures section in Appendix B for more detail on these two instruments).[21] In addition, the evaluation team measured youth engagement in the program using two approaches: (1) a youth survey administered to students in Grades 4 and above at the conclusion of an observation asking students to report on their levels of interest and engagement during the observed session (see Appendix C, Table C-1, for the survey items) and (2) use of an observation-based behavioral measure of youth engagement, the Observation of Child Engagement Scale (OCES) (see the Observation Measures section in Appendix B for more information on the OCES). The evaluation team hypothesized that greater implementation of the instructional practices outlined in the PQA and APT-O would be associated positively with both youth engagement in programming as measured by the youth survey and the OCES and the cultivation of the types of youth outcomes desired from participation in OST offerings.

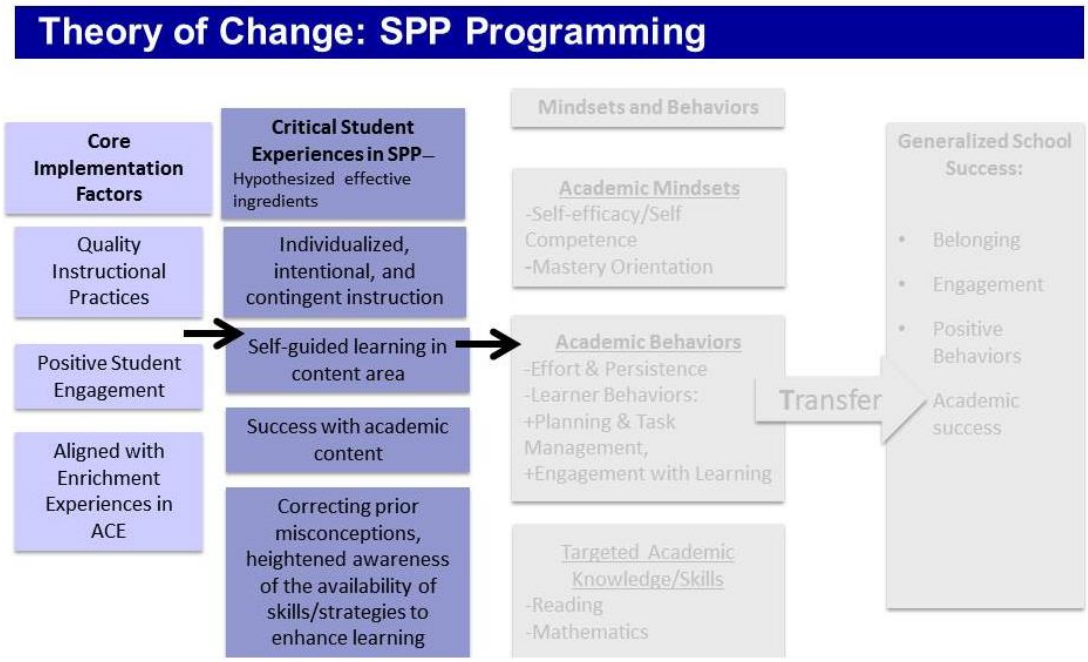
In 2010–11, a significant and positive relationship was found between the total PQA score and youth-reported engagement on the youth engagement survey. This analysis was conducted at the activity level so that the evaluation team could closely link quality data and engagement data from the same activity.[22] The finding suggested that the more staff adopted practices and created opportunities for students on the basis of the criteria outlined in the PQA, the higher the engagement students reported on the survey. In 2011–12, the evaluation team explored the relationship between program quality in ACE-funded activities and school-related outcomes. The goal of these analyses was to answer the following question: Does the impact on student outcomes vary by relevant ACE program characteristics, including center quality? These analyses differed from the 2010–11 analyses in two important ways. First, the analyses took place at the center rather than at the activity level, meaning that the connection between quality and engagement was less tightly coupled. The evaluation team used activity observation data to create an aggregate level of quality for the center.[23] Quality was also defined differently than in 2010–11. Instead of using the total PQA score, the team used a cluster analysis approach that placed centers into either a higher or lower quality group. A variety of data were used in this approach, including the instructional PQA score (which included three out of the four major domains on the PQA), OCES data, and APT-O data. The analyses revealed that centers that were (a) observed providing extensive academic content and

[21] Two versions of the PQA were used. For programs serving students in Grades 6–12, the Youth PQA was used. For those serving students in Grades K–5, the School Age PQA was used.
[22] Findings were based only on a subset of offerings provided within a relatively small sample of programs operating in Texas. This limits the generalizability of findings, although the sample was selected using both a stratified random sampling approach (40 centers) and a purposeful sampling approach (40 centers) based on criteria believed to be indicative of program quality.
[23] There were 238 observations over 80 centers, or approximately three observations per center.

supportive, interactive, and engaging activities and (b) characterized by higher levels of observed youth engagement demonstrated more of an impact on reducing school-day disciplinary incidents and on increasing student grade-level promotion.[24] In addition, centers demonstrating less of a capacity to provide activities with these characteristics demonstrated less of an impact on supporting student performance on the TAKS-Reading/ELA assessment.

Taken together, results from the first two years of the evaluation supported the general theory of change presented in the introductory chapter of this report. These findings led the evaluation team to conduct an additional set of analyses in the 2013–14 evaluation to test the hypothesis that higher quality programming is associated with higher student engagement in ACE programs and to determine whether the findings from 2010–11 were replicated when the SPP program was incorporated into the evaluation. Specifically, the evaluation examined the first two columns of the SPP theory of change, presented in Figure 3, to better understand how quality instructional practices and student engagement lead to critical student experiences in SPP.

Figure 3. SPP Theory of Change, Implementation and Student Experiences



[24] In conducting these analyses, the evaluation team compared ACE program participants with similar youth not participating in the program but attending the same schools as the ACE participants. Program quality variables were based on PQA, APT-O, and OCES scores obtained from the activities observed by the evaluation team. The team used correlational analyses to assess the relationship between a center's quality status and each of the center-level effect size estimates. This approach, as compared to the PSM techniques used in other sections of this report, is not a method that attempts to model causality. In this sense, we can say only that a relationship exists between program quality and youth outcomes.

The 2013–14 analyses used a similar method to the one used in 2010–11. That is, the evaluation team conducted analyses at the activity level and closely paired youth engagement survey data with quality data, again using only the total PQA score to define quality. For this analysis, the site visits included both ACE centers with SPP funding and a sample of similar ACE-only centers within the same grantee that did not have SPP funding.

The main finding from the 2013–14 analysis was that practices outlined in the PQA were associated with higher levels of youth engagement. More specifically, in two of the three models exploring the relationship between PQA scores and youth engagement (both observed and self-reported), greater adoption of practices represented in the PQA was associated with higher levels of youth-reported engagement in programming. A second finding from this set of analyses was that youth-reported engagement was lower in SPP programs, even if quality was high. In three of the four models exploring main effects, the SPP status of the center was related negatively to youth engagement in activities. Although it is not possible to ascertain definitively why this was the case, the evaluation team hypothesized that the lower level of youth-reported engagement may be related to the overt academic content covered in SPP activity sessions that largely targeted students who were especially academically at risk.

The evaluation team concluded that the results of the 2010–14 analyses supported, overall, the theory of change and indicated that further study to explore the second part of the theory—that quality and engagement in turn lead to improved mindsets and behaviors and eventually school-related outcomes—was a worthwhile endeavor.

RQ 1.1—Relationship Between Quality and Student Outcomes

To continue exploring the validity of the theory of change, in 2014–15, the evaluation team conducted analyses to determine the relationship between overall program quality and student outcomes, essentially replicating the analysis conducted during the 2011–12 evaluation. The first step in this process was to develop criteria and assign centers to either a higher or a lower quality group. As in 2011–12, the evaluation team did this using a technique called hierarchical cluster analysis that took into account variables related to how supportive, interactive, and engaging observed offerings were found to be (PQA scores), academic climate scores from the PQA, and youth engagement (OCES). This analysis resulted in 14 centers designated as lower quality and 12 centers designated as higher quality. Again, it is important to note that for this analyses, quality was examined at the center level. The evaluation team used observational data from a subset of program activities to create an aggregate quality rating for the whole center. More details on the cluster analysis approach are contained in Appendix D.

Next, the evaluation team conducted a PSM process to match youth participating in higher quality programs with a sample of similar youth participating in lower quality programs based on the center-level quality designations described previously. Only students with 30 or more days of participation at both types of centers and those in Grades 4–8 were included.[25] Limiting the grade level resulted in the elimination of several centers that served only students in Grades K–3. As a result, the final center sample included 10 SPP centers and 11 ACE-only centers. At the centers with SPP programming, only students who participated in SPP activities were included in the analyses. At the ACE-only centers, all students meeting the participation threshold were included in the analyses. The resulting sample included 667 students in higher quality programs (in 10 centers) and 564 in lower quality programs (in 11 centers). The goal for the PSM analysis was to reduce the likelihood that youth attending higher and lower quality programs would be significantly different on preexisting characteristics that may be related to the school-related outcomes examined as part of the analysis. Again, details on the PSM approach are included in Appendix D.

Once youth attending higher and lower quality programs were appropriately matched, the evaluation team conducted hierarchical linear modeling (HLM) analyses to determine the relationship between participation in lower and higher quality centers and student outcomes (see Appendix D for more detail on the HLM process). The hypothesis was that participating at high levels in high-quality programs would contribute to improved youth outcomes (e.g., improved achievement, lower discipline and absence rates, and greater on-time grade-level promotion). Overall, the results did not support the theory of change. That is, the evaluation team found no significant relationship between program quality and student outcomes. It is possible that the small sample size contributed to the nonsignificant findings. It is also possible that, although there were statistically significant differences between the two quality clusters on each of the measures, the cluster analysis grouping programs into the higher and lower quality groups may not have produced a sharp enough contrast between the two groups. For example, a center at the very high end of the lower quality group and the very low end of the higher quality group may not actually be all that different in their overall quality. Figure D-1 in Appendix D shows the differences between the two clusters in more detail. Finally, it is possible that the structural relationship between program quality and student outcomes is different for SPP centers than it is for ACE-only centers. This first set of analyses did not take center type into account. Analyses presented later in the report address this possibility.

[25] Grades 4–8 were included because in order to create a suitable comparison group, the evaluation team needed to use prior year achievement test scores. This eliminated students in Grades K–3, who would not have a prior year achievement test score because testing begins in Grade 3. It also eliminated students in Grades 9–12 because students in those grades take a different assessment (STAAR EOC exams), so the tests would not have been comparable for matching achievement levels.

RQ 1.2—Relationship Between Quality and Student Outcomes for SPP and ACE-Only Students

Although there were no significant findings for the overall sample of students in relation to quality and outcomes, the evaluation team continued its analysis to determine whether there was a difference between higher and lower quality centers when the SPP status of the student was taken into account. For the second part of RQ 1, the evaluation team conducted analyses to determine whether there was a relationship between quality and outcomes for SPP students and again for ACE-only students. The student sample was the same for this question as for RQ 1.1, but this time the students were broken into SPP and ACE-only groupings as defined in Table 1. Students were designated as SPP if they participated in at least one SPP activity at a given center and as ACE-only if they did not participate in SPP activities, regardless of what type of center they were attending.[26]

Table 1. Student Sample for RQ 1.2

Quality	SPP	ACE-Only
High Quality	66 (3 centers)	601 (7 centers)
Low Quality	301 (7 centers)	263 (4 centers)

Students enrolled in each program type (SPP and ACE-only) were analyzed separately, meaning that ACE-only and SPP were not compared with one another but rather each was considered its own treatment group. Because of the small sample size, the evaluation team used a descriptive approach using a multiple regression model without matching, examining only the correlations between two things (when controlling some student- and school-level covariates) rather than the PSM approach that was used in the previous research question, which gets closer to determining whether the treatment caused an outcome. Hence, the analyses used for RQ 1.2 can tell us only if there is a relationship between quality and improved outcomes for the two treatment groups (SPP or ACE-only).

For this set of analyses, the evaluation team did find some significant results. For example, the team found a statistically significant relationship between youth participation in higher quality SPP programs and fewer school-day absences compared to SPP youth enrolled in lower quality programs. More specifically, SPP students

[26] To determine SPP status, the evaluation team used the same techniques used in the 2012–13 and 2013–14 evaluations, identifying a student as SPP if he or she attended at least one SPP activity at a given center. At this stage in the analysis, SPP dosage is not taken into account. Any student with SPP activity participation at any level was included in the SPP analysis.

participating in higher quality programs were found to have a school-day absence rate that was 36% lower than their peers who participated in lower quality programs. ACE-only students participating in higher quality programs were even more likely to have a lower absence rate than their peers in lower quality programs, with 80% fewer school-day absences (or 1.607 fewer absences on average). The evaluation team also found a relationship between participation in a higher quality SPP programs and higher STAAR-Mathematics achievement as compared to youth participating in lower quality SPP programs. The effects associated with STAAR-Mathematics achievement and fewer school-day absences for ACE students were only moderately significant. Again, it is important to note that these results only suggest that a relationship exists between youth participation in higher and lower quality programs and STAAR-Mathematics achievement and school-day absences but do not prove that quality level is the cause. No significant differences were found between higher and lower quality programs on STAAR-Reading achievement or school-day disciplinary incidents. Tables 2 and 3 show the analysis results.

Table 2. Higher Versus Lower Quality SPP and ACE Centers on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14)

Group	School-Day Discipline Incidents				School-Day Absences			
	Rate Ratio[1]	Effect Size[2]	S.E.[3]	*p*	Rate Ratio[1]	Effect Size[2]	S.E.[3]	*p*
SPP	Model did not converge[4]				-36%	-0.443	0.225	0.049**
ACE-Only	-84%	-1.846	1.491	0.216	-80%	-1.607	0.882	0.068*

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Rate ratio is the percentage difference between participating in high-quality and low-quality programs.

[2] Unstandardized regression coefficient

[3] Standard error

[4] This means the analysis could not be conducted due to the small variance in the outcome measure.

Note: This analysis includes Grades 4–8.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; quality observations at site visits including Program Quality Assessment, Observation of Child Engagement Scale, and Assessment **of** Program Practices Tool scores, 2013–14.

Table 3. Higher Versus Lower Quality SPP and ACE Centers on STAAR Test Performance, 30+ Day Attendees (2013–14)

Group	STAAR-Reading			STAAR-Mathematics		
	Effect Size[1]	S.E.[2]	*p*	Effect Size[1]	S.E.[2]	*p*
SPP	0.163	0.322	0.614	0.537	0.309	0.084*
ACE-Only	0.212	0.677	0.755	0.418	0.706	0.554

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[2] Standard error

Note: This analysis includes Grades 4–8.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; quality observations at site visits including Program Quality Assessment, Observation of Child Engagement Scale, and Assessment of Program Practices Tool scores, 2013–14.

RQ 1.3—Mediating Effect of Student Engagement on Quality and Outcomes

The results from RQs 1.1 and 1.2 suggest that there may be a relationship between quality and outcomes, but it may work differently for youth enrolled in ACE-only versus SPP programming. Given the findings from RQs 1.1 and 1.2, the evaluation team decided to continue to explore the theory of change by examining RQ 1.3, which added engagement into the analysis. In particular, the evaluation team conducted a path analysis to determine what mediating role youth-reported engagement might play on the relationship between quality and outcomes. An explanation of the path analysis as well as the variables included in the analysis and how they were theorized to be related to one another can be found in Appendix D, Figure D-2.

Again, the student sample for RQ 1.3 drew from students attending the SPP and ACE-only comparison centers where the evaluation team conducted observations. As with the previously described analyses, only students in Grades 4–8 with 30 or more days of participation were included. In addition, to examine the connections between quality (still defined in these analyses as either higher or lower based on the cluster analysis method described earlier), youth-reported engagement, and outcomes, the evaluation team had to limit the analysis to centers with a large enough number of youth surveys to make the sample viable. PSM was employed but failed to find a matched comparison group that was equivalent with the treatment group because of the limited sample size.

The first step in the mediation analysis was to look at the relationship between quality and youth-reported engagement in the given sample. Contrary to the theory of change, the analyses revealed a very large and highly significant negative relationship between quality and engagement. As Table 4 shows, students in higher quality programs were less likely to be engaged than students in lower quality programs. This finding is counterintuitive and on its face seems to contradict findings from the 2010–11 and 2013–14 evaluations. However, it is important to note again that these analyses took place at the center level rather than the activity level, where we could more closely couple quality observation scores with youth-reported engagement for the same activity.

Some indicators from the 2013–14 evaluation may help explain these findings. First, it is important to note that the vast majority of activities observed in the spring of 2014 were characterized by an explicit focus on academic skill building. The SPP activities observed were explicitly academic in nature. Because of this, and in order to ensure that the comparison site observations were as similar as possible to the SPP observations, the evaluation team asked to observe activities in ACE centers that supported academic skill building. In some cases, the evaluation team was able to observe only nonacademic enrichment, and those activities were left out of this analysis. Further, activities were attended by youth that had demonstrated a propensity to struggle academically (i.e., analyses of grantee characteristics during the 2012–13 evaluation showed that SPP activities in particular targeted lower performing students, and the ACE program as a whole was geared toward students in need of additional support). Second, it is also possible that there is a connection between youth self-concept (i.e., how they view themselves), interest, and engagement. Although not assessed within the confines of this evaluation, it is plausible to argue that youth attending these offerings were more likely to experience a lower self-concept in relation to the academic content being delivered. We know that interest, a key component of engagement, tends to be related to a student's self-concept of ability (Durik, Hulleman, & Harackiewicz, 2015). In this sense, youth engagement could be lower in activities that focused on academic content where youth came into the activity with a lower self-concept of ability given past or current struggles in that content area.

Findings from the 2013–14 evaluation support the viability of this hypothesis. Although a positive relationship was found between PQA scores and levels of reported youth engagement, both the SPP status of the activity and the presence of explicit levels of academic content were negatively associated with engagement. In fact, when exploring correlations between PQA scores and youth-reported engagement, a significant and positive relationship was only found when considering activities not classified as having explicit reading or mathematics content ($r = .438$, $p < .05$).

Next, the evaluation team examined the relationship between youth-reported engagement and student outcomes, again at the center level. These analyses also revealed significant findings that contradict the theory of change. As Tables 4 and 5 show, high levels of youth-reported engagement were associated with lower levels of STAAR-Reading and STAAR-Mathematics achievement and more school-day absences.

Finally, the evaluation team assessed the mediating role of engagement to determine the relationship between quality and outcomes when taking engagement into consideration. Here, the analysis revealed the expected effect—higher quality was associated with statistically significant improvements in STAAR-Reading and STAAR-Mathematics performance and fewer school-day absences when engagement was included as a mediator in the analysis. In all three cases, the effect sizes were large as compared to similar evaluations.[27] Tables 4 and 5 detail these results. In sum, the results of RQ 1.3 analyses revealed that youth participating in high-quality programs were less engaged but have improved outcomes.

[27] Kane (2004) argues that the expected impact of afterschool programs on annual assessment results in reading and mathematics is somewhere in the range of .05 to .07 student-level standard deviations, given the amount of time youth spend in these programs during the span of the school year (60+ days of attendance equals approximately 120 hours of participation). In contrast, Hill, Bloom, Black, and Lipsey (2008) found that, on average, the effect of a whole year of learning—both in school and out of school—on assessment results in reading and mathematics for youth enrolled in Grades 4–8 averaged .31 standard deviations for reading and .42 standard deviations for mathematics. AIR's statewide evaluations of the 21st CCLC program conducted between 2009 and 2013 in five states have yielded results consistent with Kane's predictions, with effects on statewide mathematics and reading scores ranging from 0.031 to 0.063 (Naftzger et al., 2013; Naftzger, Vinson, Liu, Zhu, & Foley, 2014; Naftzger, Vinson, & Lui, 2013; Vinson, Marchand, Sparr, & Moroney, 2013; Moroney et al., 2013; Naftzger, Vinson, Swanlund, & Sparr, 2012; Naftzger, Vinson, Manzeske, & Gibbs, 2011). As with the results presented here, significant program effects were more consistently found for mathematics as opposed to reading, with an average effect size of .043.

Table 4. Effects of Program Quality on STAAR Test Performance via Engagement, 30+ Day Attendees (2013–14)

Path	Engagement			Reading			Mathematics		
	Effect Size[1]	S.E.[2]	*p*	Effect Size[3]	S.E.[2]	*p*	Effect Size[3]	S.E.	*p*
Program Quality	-10.974	2.994	0.000***						
Engagement				-0.019	0.009	0.030**	-0.016	0.004	0.000***
Program Quality via Engagement				0.206	0.111	0.064*	0.175	0.063	0.006**

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Unstandardized regression coefficient

[2] Standard error

[3] Standardized regression coefficient. Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

Note: This analysis includes Grades 4–8.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; quality observations at site visits including Program Quality Assessment, Observation of Child Engagement Scale, and Assessment of Program Practices Tool scores, 2013–14.

Table 5. Effects of Program Quality via Engagement on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14)

Path	Engagement			School-Day Discipline Incidents			School-Day Absences		
	Effect Size[1]	S.E.[2]	*p*	Effect Size[1]	S.E.[2]	*p*	Effect Size[1]	S.E.[2]	*p*
Program Quality	-10.974	2.994	0.000***						
Engagement				0.035	0.026	0.174	0.167	0.067	0.013**
Program Quality via Engagement				-0.387	0.276	0.161	-1.832	0.788	0.020**

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Unstandardized regression coefficient

[2] Standard error

Note: Model did not converge for grade promotion outcome. This means the analysis could not be conducted due to the small variance in the outcome measure.

Note: This analysis includes Grades 4–8.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; quality observations at site visits including Program Quality Assessment, Observation of Child Engagement Scale, and Assessment of Program Practices Tool scores, 2013–14.

As already noted, these findings were not consistent with what was theorized. To explain this outcome, the evaluation team's hypothesis is that low self-concept of ability combined with participation in explicitly academic activities (as noted, the majority of activities observed and included in this analysis, both SPP and ACE-only, were explicitly academic in nature) served to depress student interest in the activities being offered and therefore lowered youth engagement. It is important to note that this is simply conjecture at this point because it was beyond the scope of this evaluation to measure self-concept for students attending observed activities. However, although youth engagement was lower compared to youth participation in nonacademic offerings, students in these offerings still benefitted from participation in higher quality activities as demonstrated by the variety of positive outcomes detailed in Tables 4 and 5. In this sense, although the activities youth were participating in may not have sparked their interest, the opportunity to practice and rehearse content-related skills in a high-quality environment appeared to support the types of outcomes sought by the SPP and ACE program.

RQ 1.4—Mediating Effect of Youth Mindsets and Behaviors on Outcomes

For RQ 1.4, the evaluation team was interested in exploring the mediating effect of youth mindsets and behaviors on participation and outcomes. In other words, do students with more pre- to post-change on the Youth Mindsets and Behaviors Survey (see Appendix C, Tables C-2 and C-3 for constructs) have better outcomes when taking participation into consideration? Originally, the goal was to trace the entire theory of change across one set of students and centers to explore the connections between quality, engagement, mindsets and behaviors, and outcomes. The previous research questions looked at connections between quality and outcomes and quality and engagement. For this analysis, there were not enough students in those centers that the team visited in 2013–14 with matched pre- and post-Mindsets and Behaviors Survey data to create a sufficient comparison group. In addition, students in the comparison centers (ACE-only centers) visited during the site visits did not take the youth Mindsets and Behaviors Survey, so data on those students were limited. Therefore, the student sample for RQ 1.4 differs substantially from the sample used in the other research questions in this chapter. For RQ 1.4, in order to have enough students to allow for analysis, the sample included any student in Grades 4–8 attending a center with SPP programming (whether the student attended SPP programming or not) who had 30 or more days of participation, had matched pre- and post-Mindsets and Behaviors Survey data, and had room to improve on at least one of the five constructs on the presurvey (e.g., they fell into the lowest two response categories—not at all like me and sort of like me—on at least one construct). There were 764 students who met this criteria. The sample was then further limited to only students who participated in activities with

explicit academic content because those students were most likely to see growth on the sense of competence scales and it ensured that the SPP and ACE-only students in the sample were more similar. The evaluation team used a variable identified by site visitors during observations of the ACE-only and SPP activities to classify activities as having explicit academic content[28]. Limiting in this way reduced the final sample to 529 students. Table 6 displays the frequency distribution of construct category by time point.

Table 6. Frequency Distribution of Construct Category by Time Point (2013–14)

Construct	Time Point	Not At All Like Me	Sort of Like Me	A Lot Like Me
Sense of Competence as a Learner	Pre	4.5%	95.5%	0.0%
	Post	4.7%	68.5%	26.8%
Sense of Competence as a Reader	Pre	21.9%	78.1%	0.0%
	Post	22.4%	57.0%	20.6%
Sense of Competence in Mathematics	Pre	26.6%	73.4%	0.0%
	Post	17.1%	61.0%	21.8%
Effort and Persistence	Pre	10.3%	89.7%	0.0%
	Post	5.3%	67.5%	27.2%
Learning Behaviors and Engagement	Pre	4.6%	95.4%	0.0%
	Post	4.1%	65.4%	30.4%

SOURCE: Youth Mindsets and Behaviors Survey Responses (Grades 4–12), 2013–14.

The evaluation team started by reexamining the results of the pre- to post-change analysis on the Mindsets and Behaviors Survey data conducted for the 2013–14 evaluation (see Appendix D for details on the Rasch analysis methods used). That set of analyses, which did not take participation levels into account, found that there was no change or a decline in mindsets and behaviors among participants. This year, the evaluation team limited the sample to students who had participated for 30 or more days.[29] When limited in this way, the results were quite different. In fact, there was significant and positive pre- to post-change on all constructs, as shown in Table 7. Given this finding, the evaluation team felt it was viable to continue with RQ 1.4 to explore the mediating effect of mindsets and behaviors on outcomes.

[28] Observers requested academic content activities for observation. The APT-O observation instrument allowed the team to confirm if an activity was academic enrichment or non-academic enrichment.
[29] 30 or more days of attendance is considered to be "regular" attendance by the U.S. Department of Education for the 21st CCLC program and has been used by AIR in past Texas as well as other statewide 21st CCLC evaluations as a cutoff point for regular attendance. In this case, it was hypothesized that limiting the sample to regular attendees might yield pre- to post-change when not limiting in the previous year's evaluation found no pre- to post-change.

Table 7. Pre- to Post-Change on Mindsets and Behaviors Survey by Construct, 30+ Days Attendees (2013–14)

Construct	Mean Pretest	Mean Posttest	Pre- to Post-Change	*p*
Sense of Competence as a Learner	44.57	47.27	2.70	0.000***
Sense of Competence as a Reader	45.47	47.67	2.20	0.000***
Sense of Competence in Mathematics	43.15	46.63	3.48	0.000***
Effort and Persistence	42.52	47.17	4.65	0.000***
Learning Behaviors and Engagement	43.50	46.97	3.46	0.000***
***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level				

SOURCE: Youth Mindsets and Behaviors Survey Responses (Grades 4–12), 2013–14.

To conduct the mediation analysis (see Appendix D, Figure D-3, for more detail on the mediation analysis), first, the evaluation team looked at the relationship between participation and the five constructs on the Mindsets and Behaviors Survey—*Sense of Competence in Mathematics, Sense of Competence in Reading, Sense of Competence as a Learner, Effort and Persistence,* and *Learning Behaviors and Engagement*. Despite the positive pre- to post-change found in the descriptive analyses of the data, the team did not find any significant relationships between higher levels of program participation and improvement on the five constructs. That is, among students who participated at least 30 days in the program, participating increasingly more days was not associated with improvement on the five constructs. Table 8 shows the results of these analyses.

Table 8. Relationship Between Program Participation and Mindsets and Behaviors, 30+ Day Attendees (2013–14)

	LC[1]		RC[1]		MC[1]		EP[1]		BE[1]	
	Effect Size[2] (S.E.[2])	*p*	Effect Size[2] (S.E.[3])	*p*	Effect Size[2] (S.E.[3])	*p*	Effect Size[2] (S.E.[3])	*p*	Effect Size[2] (S.E.[3])	*p*
Higher levels of Participation	-0.020 (0.020)	0.324	0.007 (0.012)	0.597	-0.001 (0.019)	0.978	-0.024 (0.029)	0.407	-0.021 (0.021)	0.312

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] LC = *Sense of Competence as a Learner;* RC = *Sense of Competence as a Reader*; MC = *Sense of Competence in Mathematics*; EP = *Effort and Persistence*; BE = *Learner Behaviors and Engagement*

[2] Unstandardized regression coefficient

[3] Standard error

Note: This analysis includes Grades 4–8.

SOURCE: TX21st CCLC Student Tracking System, 2013–14; Youth Mindsets and Behaviors Survey data (Grades 4–12), 2013–14

Next, the evaluation team looked at the relationship between performance on the five constructs and youth outcomes (the top half of Tables 9 and 10) and then how performance on the five constructs might have a mediating effect on the relationship between participation and outcomes (the bottom half of Tables 9 and 10). For all of these analyses, the evaluation team found very few significant effects, and even the nonsignificant effects were very small (less than .005 in most cases), indicating that the evaluation team could not detect a mediating effect of mindsets and behaviors on the relationship between participation and outcomes. Tables 9 and 10 show the results of these analyses.

Table 9. Effects of Program Participation on Achievement via Youth Mindsets and Behaviors, 30+ Day Attendees (2013–14)

Path	STAAR-Reading			STAAR-Mathematics		
	Effect Size[1]	S.E.[2]	*p*	Effect Size[1]	S.E.[2]	*p*
Sense of Competence as a Learner	-0.002	0.004	0.651	0.005	0.004	0.239
Sense of Competence as a Reader	0.000	0.003	0.960	-0.001	0.003	0.754
Sense of Competence in Mathematics	0.006	0.003	0.030**	0.002	0.003	0.348
Effort and Persistence	0.002	0.005	0.606	0.002	0.002	0.405
Learner Behaviors and Engagement	0.001	0.004	0.691	-0.002	0.003	0.468
Participation via LC[3]	0.00004	0.0000	0.698	-0.0001	0.0002	0.428
Participation via RC[3]	0.000	0.000	0.960	0.000	0.000	0.776
Participation via MC[3]	0.000	0.000	0.978	-0.000	0.000	0.978
Participation via EP[3]	-0.00005	0.000	0.677	-0.00005	0.000	0.578
Participation via BE[3]	-0.00002	0.000	0.720	-0.00004	0.000	0.556

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[2] Standard error

[3] LC = *Sense of Competence as a Learner;* RC = *Sense of Competence as a Reader*; MC = *Sense of Competence in Mathematics;* EP = *Effort and Persistence*; BE = *Learner Behaviors and Engagement*

Note: This analysis includes Grades 4–8.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; Youth Mindsets and Behaviors Survey data (Grades 4–12), 2013–14.

Table 10. Effects of Program Participation on School-Day Discipline Incidents and Absences via Youth Mindsets and Behaviors, 30+ Day Attendees (2013–14)

Path	School-Day Discipline Incidents			School-Day Absences		
	Effect Size[1]	S.E.[2]	*p*	Effect Size[1]	S.E.[2]	*p*
Sense of Competence as a Learner	0.006	0.003	0.049**	0.013	0.017	0.438
Sense of Competence as a Reader	0.001	0.004	0.732	-0.007	0.012	0.565
Sense of Competence in Mathematics	0.001	0.002	0.754	0.041	0.021	0.048**
Effort and Persistence	-0.005	0.022	0.832	0.007	0.035	0.848
Learner Behaviors and Engagement	0.002	0.003	0.645	-0.022	0.018	0.208
Participation via LC[3]	-0.00012	0.000	0.405	-0.0003	0.000	0.5111
Participation via RC[3]	0.000	0.000	0.779	-0.00005	0.000	0.639
Participation via MC[3]	0.000	0.000	0.979	-0.00004	0.001	0.978
Participation via EP[3]	0.00012	0.000	0.803	-0.00017	0.001	0.869
Participation via BE[3]	-0.00004	0.000	0.660	0.00046	0.001	0.486

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1]Unstandardized regression coefficient

[2] Standard error

[3] LC = *Sense of Competence as a Learner;* RC = *Sense of Competence as a Reader*; MC = *Sense of Competence in Mathematics;* EP = *Effort and Persistence*; BE = *Learner Behaviors and Engagement*

Note: This analysis includes Grades 4–8.

Note: Model did not converge for grade-level promotion outcome. This means the analysis could not be conducted due to the small variance in the outcome measure.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; Youth Mindsets and Behaviors Survey data (Grades 4–12), 2013–14.

There are a variety of hypotheses for why no effects were found for this analysis. One, of course, is the possibility that mindsets and behaviors do not have any relationship to student outcomes regardless of participation. Another possibility, however, is that the youth report survey on mindsets and behaviors is an imperfect measure, and/or the sample is too diverse to detect impacts. For example, it may be the case that by grouping all students together, negative effects on one group of students may cancel out the positive effects on another, resulting in no effects. It is possible that, if the sample size were larger, the evaluation team could examine student groups for specific constructs (e.g., looking at only students participating in activities focused on explicit

mathematics instruction and looking only at sense of competence in mathematics) and see different results that are more consistent with the theory of change. Because of the small sample size, however, that level of analysis was not possible.

Conclusions and Key Findings

The overarching goal of the analyses discussed in this chapter was to examine further the theory of change presented earlier in this report and look at the remaining relationships between quality, engagement, youth mindsets and behaviors, and school-related outcomes that had not yet been explored in the previous years of the evaluation. Specifically, this chapter set out to answer the following questions:

- RQ 1—What key practices, strategies, and approaches are especially related to quality programming that leads to improved youth outcomes?
  - RQ 1.1—What is the relationship between program quality and student outcomes?
  - RQ 1.2—How does the relationship between quality and student outcomes differ for SPP students and ACE-only students?
  - RQ 1.3—How does the level of youth engagement mediate the relationship between quality and outcomes?
  - RQ 1.4—How do changes in youth mindsets and behaviors mediate the relationship between participation and outcomes?

In the concluding chapter of this report, we summarize the findings from this year's evaluation and put them in context with the previous four years of evaluation. In the section that follows, we summarize only the findings from analyses related to RQ 1 and its subquestions. Those findings are as follows:

- **High-quality programs may lead to improved outcomes under certain conditions.** Although the evaluation team did not find a relationship between quality and outcomes when all students were analyzed together, there was evidence of a relationship when ACE and SPP programs were examined separately. Both high-quality ACE and SPP programs were associated with decreased school-day absences, and high-quality SPP programs were associated with improved mathematics achievement.
- **Quality and engagement may not be related in activities with explicit academic content.** Analyses associated with RQ 1.3 revealed a negative relationship between quality and engagement, contrary to the theory of change. In the 2013–14 evaluation of the ACE program, the evaluation team found a negative relationship between SPP programs and engagement. It is the evaluation team's hypothesis that programs with explicit academic content may

not be as engaging to youth while still being high quality. The majority of activities observed and included in the analyses this year, both SPP activities and ACE-only comparison activities, were academic in nature, which may have led to the results.

- **High-quality and explicit academic content activities lead to positive outcomes**. Although the evaluation team saw a negative relationship between quality and engagement and between quality and outcomes when examined separately, there was still a positive relationship between high-quality programs and outcomes when viewed through the lens of engagement. This indicates that it is still possible for students to improve outcomes when participating in high-quality programs even if engagement is low. Again, the evaluation team hypothesizes that in activities with explicit academic content, even though cognitive engagement may be low, if quality is high, students may see improvements to key school-related outcomes.
- **No relationship was found between improvement in youth-reported mindsets and behaviors and school-related outcomes**. Although there was pre- to post-improvement on the Youth Mindsets and Behaviors Survey for youth who participated in programming more than 30 days, the evaluation team did not see a relationship between youth mindsets and behaviors and school-related outcomes. Nor did the team find a mediating effect of mindsets and behaviors on the relationship between participation and outcomes. Although it is possible that mindsets and behaviors have no relationship to participation and outcomes, the evaluation team believes it is more likely that the youth report survey is imperfect or that it is too difficult to measure effects when taking all students and all constructs on the survey into account. Further analyses in the future on specific subsets of students (e.g., those participating in mathematics activities) and specific constructs (e.g., sense of competence in mathematics) may prove more revealing.

Taken together, the findings discussed in this chapter indicate that there is a relationship between quality and outcomes and that the presence of explicit academic content may complicate the relationship between quality, outcomes, and engagement that emerged during the first four years of the evaluation.

Chapter 3: Cost Effectiveness of the SPP Program

Although the SPP program has ended, understanding its cost effectiveness and the extent to which it provided a reasonable return on investment is still relevant for TEA as it considers investing in other focused interventions to promote student achievement in OST settings. Therefore, the evaluation team included the following research questions and subquestions in the 2014–15 evaluation, all of which are addressed in this chapter:

- RQ 2—What is the cost effectiveness and sustainability of the SPP program?
  - RQ 2.1—What is the per-student cost of the SPP program relative to school-related student outcomes?
  - RQ 2.2—What is the relationship between the allocation of SPP grant funding to specific activities and student outcomes?
  - RQ 2.3—What is the return on investment for SPP programs?
  - RQ 2.4—In what ways has the SPP experience impacted the work of project directors and the delivery of afterschool programs, and which organizational or instructional components have they incorporated into 2014–15 afterschool programming?

To answer these questions, the evaluation team explored relationships between center-level expenditures and student outcomes (e.g., results on STAAR-Reading and Mathematics). The goal of these analyses is to provide insight into an important element of SPP programs—that is, which SPP program elements and features were associated with the most substantial returns in terms of student outcomes.

Methods

Budgets contained in 2013–14 SPP grant proposals were reported at the grantee level, not the center level; therefore, it was necessary to collect supplemental financial data about SPP expenditures by center.[30] Evaluators collected these data, disaggregated them by key programmatic functions and payroll functions, and then analyzed them descriptively. The team also assessed the ratio of center-level spending to student enrollment in SPP activities (for students attending SPP activities for 30+ days) to provide a measure of per-pupil spending by center.[31]

[30] SPP grants were awarded at the grantee level, and then grantees distributed the funds to one or more of their centers for program implementation.

[31] Throughout the evaluation, analyses include only students who attended 30+ days because that is the point after which research shows that program effects are most likely to appear (Kauh, 2011; Naftzger et al., 2012). In addition, for this particular analysis, excluding students with less than 30 days of attendance reduced the risk of including students who may have initially enrolled but did not actively participate in SPP activities.

To collect center-level spending data systematically, the evaluation team created standardized data collection forms that requested budget and expenditure data from each grantee. The evaluation team sent each project director a data collection form, along with detailed instructions for its completion and a request to complete one form for each center receiving SPP funding. It is important to note that a total of 32 centers representing 11 grantees returned surveys. Although this was a strong response rate, it is still a small number of grantees and centers, which could have affected the findings. Specific data elements collected for this report included budget and expenditures by expenditure category (e.g., payroll, professional services, contracted services) and by mode of delivery (e.g., Computer-Based academic interventions, Face-to-Face academic interventions). Evaluators requested that the financial information cover all expenditures related to SPP programming that occurred between September 1, 2013, and May 20, 2014, making it possible to explore how SPP program funds were spent during the 2013–14 academic year. These supplemental financial data were submitted to the evaluation team in June and July 2014. The evaluation team also provided technical assistance on completing the forms through emails and phone calls with individual project directors.

Sample

For the analyses presented in this chapter, only students who participated in SPP activities for 30+ days were included as “SPP students.” As Table 11 shows, a total of 1,546 students met this criterion.

In addition, Table 11 shows the distribution of SPP students across grade levels. Approximately 54% of all participants in SPP activities were in elementary school (EL) grades (i.e., Grades 3–5), 25% were in traditional middle school (MS) grades (i.e., Grades 6–8), and 21% of the SPP students were in Grades 9–12.

Table 11. Distribution of Students Participating in SPP Activities for 30+ Day, by Grade Level (2013–14)

Grade Level	Number of SPP Students	Percentage of SPP Students
3	367	24%
4	360	23%
5	113	7%
6	75	5%
7	160	10%
8	142	9%
9	99	6%
10	128	8%
11	94	6%
12	8	0.5%
Total	1,546	100%

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

RQ 2.1—SPP Cost per Student Relative to School-Related Outcomes

To address RQ 2.1, the evaluation team used a variety of analytic techniques to explore how per-pupil program costs and spending on program components (e.g., mode of instructional delivery) were associated with student performance on STAAR exams, while holding other student characteristics (i.e., school-day attendance rate, race/ethnicity, economically disadvantaged status) constant.[32] The evaluation team relied upon center-level expenditure data reported by SPP project directors following the 2013–14 school year, SPP student enrollment data submitted through TX21st, and 2012–13 and 2013–14 student-level STAAR data for reading and mathematics. As noted earlier, analyses were limited to programs with SPP funding that operated in 2013–14 and students who attended SPP programming 30 or more days. In addition, students in Grade 3 were excluded from the analyses because they did not have a prior-year STAAR score, which means the evaluation team could not create a suitable comparison group.[33] Please see Appendix E for more information about the specific statistical models used in this analysis.

[32] This analysis does not compare outcomes with students at centers without any SPP program spending—those schools are omitted from the analyses.
[33] Accordingly, one center serving only students in Grades K–2 was dropped from the analysis.

Two pieces of information were important for conducting analyses related to RQ 2.1: (1) the per-student cost of SPP programming and (2) student performance on STAAR exams. First, the evaluation team calculated per-student SPP grant expenditures for each of the centers that reported having students participating in SPP activities during the 2013–14 school year.[34] SPP expenditures for 2013–14 varied widely, from a high of $6,672 per student to a low of $326 per student, based on the reported number of students served in the program. The median per-student SPP expenditure was $737, and the mean was $1,339.

Next, the evaluation team examined SPP student performance for 2012–13 and 2013–14 on the STAAR-Reading and Mathematics exams (Grades 4–8) and the English I and Algebra I EOC exams (Grade 9).[35] Table 12 summarizes the percentage of SPP students in each grade who met the state standard on the mathematics and reading tests.[36] Overall, the percentage of SPP students who met or exceeded the standard was lower than the overall average for non-SPP students in our dataset (STAAR-Mathematics was 66% and STAAR-Reading was 66% in 2013–14). In addition, the percentage of SPP students meeting or exceeding the standard for STAAR-Mathematics and STAAR-Reading was generally higher in 2013–14 than in 2012–13 for both tests across grades.

[34] A total of 32 centers with SPP funding were included in the expenditure analyses.
[35] Although the EOC exams are content- and not grade-specific, and therefore can be taken by students in Grades 10-12 as well, only Grade 9 students were included in these analyses to avoid including repeat test takers in the analysis.
[36] Throughout, meeting or exceeding the standard means students scored at or above the Level II Phase-in 1 standard.

Table 12. Percentage of SPP Students Meeting or Exceeding Standard for STAAR-Mathematics and Reading, for 30+ Day Attendees by Grade Level (2012–14)

Grade Level[1,2]	STAAR-Mathematics 2012–13 [3,4]	STAAR-Mathematics 2013–14[3,4]	STAAR – Reading 2012–13 [3,4]	STAAR-Reading 2013–14[3,4]
4	27%	58%	26%	48%
5	41%	55%	25%	44%
6	52%	52%	41%	36%
7	33%	39%	26%	59%
8	30%	59 %	45%	62%
9	59%	64%	69%	74%
Total	35%	54%	34%	53%

[1] Grade 3 students omitted from the analyses because they did not have prior STAAR score (i.e., 2012–13 results).

[2] Grades 4–7 reflect grade-level STAAR results, Grade 8 results reflect grade-level STAAR results or Algebra I EOC exams (used Algebra I if they had both Algebra I and STAAR), and Grade 9 reflects results from the Algebra I and English I EOC exams.

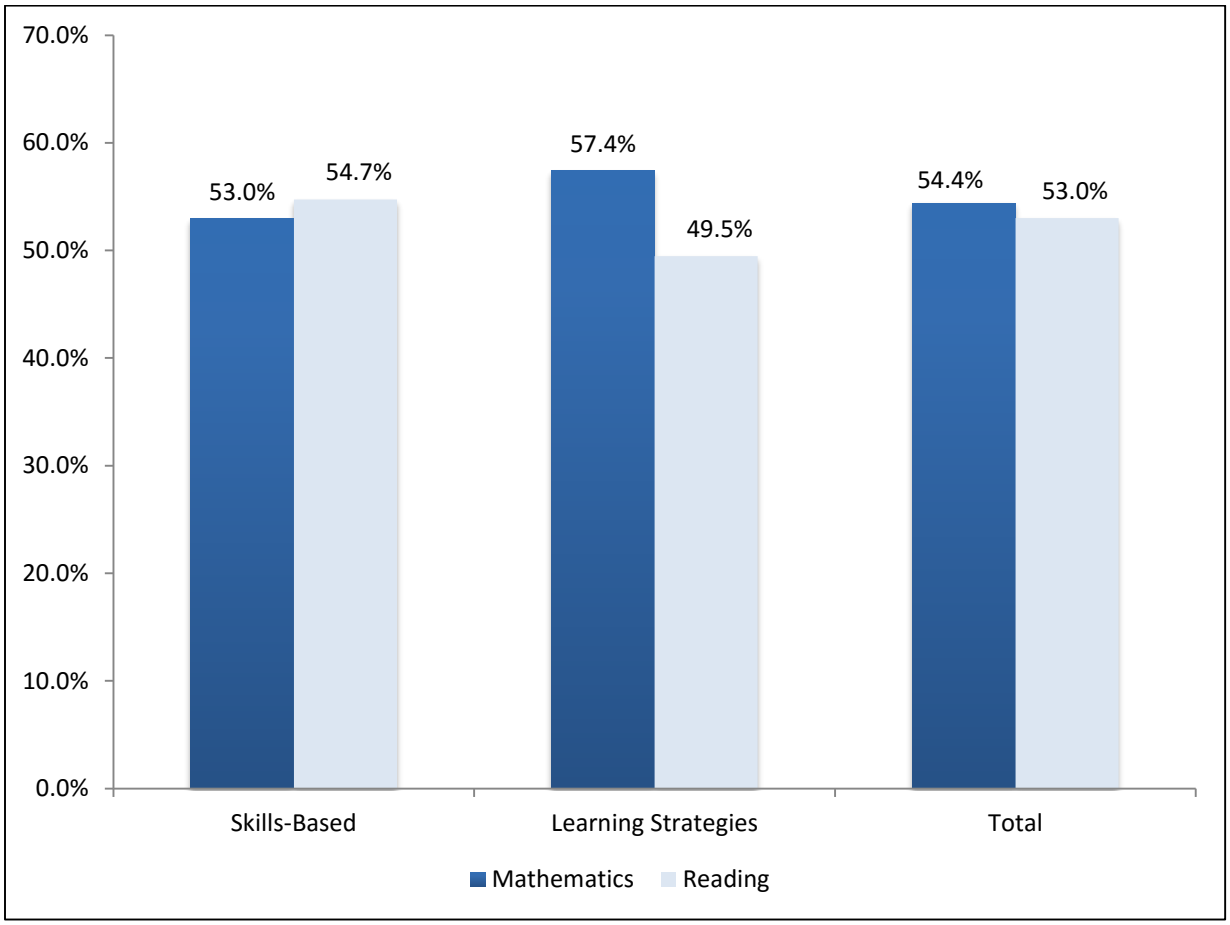
[3] Students who scored higher than the Level II Phase-in 1 standard were considered to have met or exceeded the standard.

[4] Pass rates include only SPP students who took the exam in both 2012–13 and 2013–14 (Mathematics *N* = 861; Reading *N* = 878).

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; and Gibson Consulting Group Spring 2014 Financial Data Collection, 2013–14.

Finally, the evaluation team looked at the type of instruction taking place at each center and its relationship to STAAR achievement prior to beginning analyses. Of the 32 centers in the study sample, 21 centers employed *Skills-Based* instruction, while 11 used a *Learning Strategies* approach. Figure 4 shows the percentage of students meeting or exceeding the standard for STAAR-Mathematics and STAAR-Reading by instruction type. A slightly higher percentage of students exposed to *Learning Strategies* instruction (57%) met or exceeded the standard for the 2013–14 STAAR-Mathematics test compared to students exposed to *Skills-Based* instruction (53.0%) in their SPP activities. However, the opposite was found for reading, where 50% of students exposed to *Learning Strategies* instruction in their SPP activities met or exceeded the standard for the STAAR-Reading exam compared to 55% of students provided with *Skills-Based* instruction in their SPP activities.

Figure 4. Percentage of SPP Students Meeting or Exceeding the Standard on STAAR-Mathematics and Reading by Instruction Type, 30+ Day Attendees (2013–14)



SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; and Gibson Consulting Group Spring 2014 Financial Data Collection, 2013–14.

Findings

In order to examine the association between per-student program expenditures and student performance, the evaluation team used a multilevel model to predict student performance, accounting for program expenditure measured at the center level. The model controls for individual student characteristics and the effect sizes for these controls are included in Appendix E, Table E-1. Table 13 shows that center-level per-student program spending was not related to student performance in STAAR-Reading and Mathematics except for Grade 9 students on the English I EOC. For Grade 9 students, although the substantive effect is small, every $1 increase in per-student program spending is associated with a .2% decrease in the odds of meeting or exceeding the standard for the English I EOC. This finding contradicts the evaluation team's hypothesis that an increase in spending would lead to greater achievement.

Table 13. Estimated Association Between per-Student Program Expenditures and Student Mathematics and Reading Performance, for Students in Grades 4–9, 30+ Day Attendees (2013–14)

	STAAR-Mathematics/Algebra I EOC[1]				STAAR-Reading/English I EOC[2]			
	Effect Size[3]	Percent Change in Odds of Meeting Standard	S.E.[4]	*p*	Effect Size[3]	Percent Change in Odds of Meeting Standard	S.E.[4]	*p*
Grade 4–8								
Per-student expenditures	-0.000	-0.1%	.000	0.517	-0.000	-0.1%	.000	0.073*
Grade 9								
Per-student expenditures	0.000	0.0%	.000	0.930	-0.002	-0.2%	.001	0.040**

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] *N*=760 for Grades 4–8 STAAR-Mathematics; *N*=69 for Grade 9 Algebra I EOC

[2] *N*=768 for Grades 4–8 STAAR-Reading; *N*=74 for Grade 9 English I EOC

[3] Standardized regression coefficient. Effect sizes < .001 are reported as .000 to avoid conveying false precision.

[4] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; and Gibson Consulting Group Spring 2014 Financial Data Collection, 2013–14.

RQ 2.2—Relationship Between Funding for Specific Activities and Student Outcomes

To answer RQ 2.2, the evaluation team looked at the student performance and per-student spending data used in RQ 2.1 as well as information on how much funding was used for different modes of instruction (i.e., Face-to-Face instruction or Computer-Based instruction). Overall, the average per-student expenditure for Face-to-Face instruction was $644 (Minimum = $0; Maximum = $2,706; Median = $415), and the average per-student expenditure level for Computer-Based instruction was $119 (Minimum = $0; Maximum = $1,985; Median = $73).

Findings

As with RQ 2.1, to examine the association between per-student program expenditures by instructional type and student performance, the evaluation team used a multilevel

model to predict student performance, accounting for program expenditures at the center level. Table 14 shows that center-level per-student program spending for Computer-Based and Face-to-Face instruction did not predict student performance on STAAR-Reading and STAAR-Mathematics, with one exception. For Grades 4–8, the odds of meeting the standard on the STAAR-Reading exam decreased as per-student Computer-Based instruction expenditures increased. Although the substantive effect is small, every $1 increase in per-student Computer-Based instruction spending is associated with a .3% decrease in the odds of meeting the standard on the STAAR-Reading test, suggesting a weak, negative relationship between reading performance and Computer-Based instruction.

Table 14. Estimates of the Association Between per-Student Program Spending by Mode of Instruction and STAAR-Mathematics and Reading Performance, for Students in Grades 4–9, 30+ Day Attendees (2013–14)

	STAAR-Mathematics[1]				STAAR-Reading[2]			
	Effect Size[3]	Percent Change in Odds of Meeting Standard	S.E.[4]	*p*	Effect Size[3]	Percent Change in Odds of Meeting Standard	S.E.[4]	*p*
Grades 4–8								
Per-student expenditures Computer-Based instruction	-0.000	-0.1%	0.000	0.421	-0.002	-0.3%	.001	0.043**
Per-student expenditures Face-to-Face instruction	-0.000	-0.1%	0.000	0.969	-0.000	-0.1%	.000	0.233
Grade 9								
Per-student expenditures Computer-Based instruction	0.001	0%	0.008	0.880	-0.008	-0.1%	.008	0.272
Per-student expenditures Face-to-Face instruction	-0.000	-0.1%	0.003	0.933	-0.001	-0.2%	.003	0.670

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] *N*=760 for Grades 4–8 STAAR-Mathematics; *N*=69 for Grade 9 Algebra I EOC

[2] *N*=768 for Grades 4–8 STAAR-Reading; *N*=74 for Grade 9 English I EOC

[3] Standardized regression coefficient. Effect sizes < .001 are reported as .000 to avoid conveying false precision.

[4] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; and Gibson Consulting Group Spring 2014 Financial Data Collection, 2013–14.

RQ 2.3—Return on Investment for SPP Programs

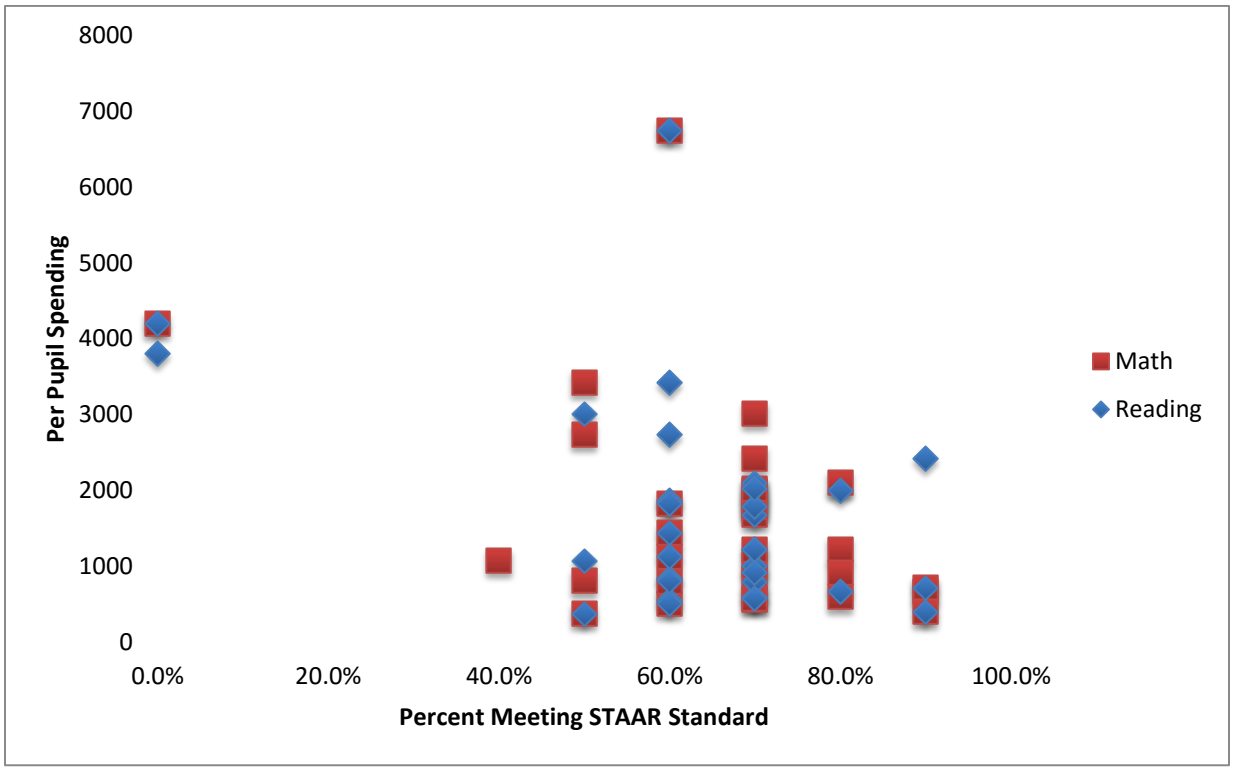
To investigate RQ 2.3, the evaluation team examined the results of RQ 2.1 and plotted the relationship between 2013–14 center-level per-student SPP expenditures and the

2013–14 rate of meeting the standard for STAAR-Mathematics and STAAR-Reading for the center that the majority of ACE participants in a given program attended.

Findings

As Figure 5 illustrates, aggregate rates for meeting STAAR standards are not markedly different across per-pupil spending levels, suggesting a lack of relationship between the two indicators. These analyses do not account for many other factors, such as regular school-day academic programs, instructor quality, SPP implementation fidelity, and other funds infused into afterschool programs beyond SPP grant funds, which may account for variation in student outcomes.

Figure 5. SPP Return on Investment, Relationship Between Spending and Rate of Meeting STAAR-Mathematics and STAAR-Reading Standards[1]



[1] A total of 32 centers were included the return on investment analysis.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14; and Gibson Consulting Group Spring 2014 Financial Data Collection, 2013–14.

RQ 2.4—Impact of SPP Implementation on Organizational and Instructional Practice

To examine RQ 2.4, the evaluation team conducted interviews with 11 project directors in January 2015. The purpose of these interviews was to learn more about which

aspects of the SPP program were particularly effective and which were difficult to implement as well as to learn about the organizational or instructional strategies project directors carried forward into 2014–15 (regardless of how programming was funded). Project director interviews covered the following topics:

- Effective aspects of the SPP program (organizational and instructional)
- Challenges to Implementing the SPP program
- Sustainability

Upon completion of interviews, the evaluation team analyzed transcripts for key themes and subthemes across centers, as well as for divergent opinions about SPP program implementation, including the following:

- Effective organizational strategies employed in the SPP program
- Effective instructional strategies employed in the SPP program
- Barriers encountered when implementing the SPP program
- How they would change their approach to the SPP program if implementing in 2014–15
- Approaches for sustaining SPP-related activities
- Organizational approaches carried forward into current afterschool programming
- Instructional or curricular approaches carried forward into current afterschool programming
- Funding for current afterschool activities and continuation of SPP-related activities
- How the SPP program has changed their philosophy on afterschool programming

The first analyses presented in the following section explores aspects of SPP programming that project directors found to be particularly effective and those they found difficult to implement or that represented barriers to successful implementation. The evaluation team analyzed these data separately for organizational aspects (e.g., staffing, support from host school, program monitoring, student recruitment, staff training) and instructional aspects (e.g., group Face-to-Face instruction, Computer-Based mathematics and reading programs, project-based learning, student interactions, enrichment activities) of the SPP program.

Next, the evaluation team analyzed interview data related to program sustainability approaches and the ways in which the SPP program experience may have impacted project directors' philosophical approaches to afterschool programming. Along with

these assessments, the team conducted analyses of the organizational and instructional features of the SPP program that carried over into the 2014–15 afterschool programming. The evaluation team also analyzed data related to funding sources and approaches for 2014–15 programming.

All analyses are descriptive in nature and represent the perspectives of SPP project directors as they reflected on their experiences with the program during the spring 2013 to summer 2014 period.

Findings

Effective Organizational Strategies Employed in the SPP Program

When SPP project directors were asked to elaborate on which organizational strategies they found to be effective between spring 2013 and summer 2014, SPP leaders indicated that their approaches to SPP staffing and staff training, SPP program monitoring, and recruiting students for the SPP were effective.

Program Staffing. According to SPP project directors, the primary shift in the staffing approach as a result of SPP funding was greater use of certified, school-day teachers to deliver afterschool programming as compared to ACE-only programming. This mirrors findings from analysis of staffing types done as part of the 2012–13 and 2013–14 evaluations of ACE and SPP programming that found centers with SPP programming employed more certified teachers than ACE-only centers (Devaney et al., 2015). One center implemented a master teacher approach, where a certified teacher from the district served as a monitor and advisor to afterschool instructors delivering the afterschool academic content in reading and mathematics. Comments from project directors related to effective staffing approaches include:[37]

> *"Teachers, because the teachers here were just involved in a lot of professional development and didn't necessarily want to put forth that extra time to work for the ACE program, but the SPP program was so closely aligned with the school-day curriculum and it had so much support from the school admin that we had a lot more buy-in for that program from the day-school teachers, and therefore a lot more of them were interested in working for our program."*
>
> *"One thing that we did that year that we do a little different this year is the teachers who helped us in our [intervention]. We had some extra duty—teachers came in and did some extra duty for us. We had school-day teachers that were helping us monitor that program."*

[37] Throughout this chapter, quotations have been edited for grammatical correctness and clarity, but the substance of the interviewer comments has not been altered in any way.

"All of those [teachers] were highly-qualified staff. Those were the same teachers that had the kids during the day. They already knew what the kids' strengths and weaknesses were. So when they transitioned to the afterschool program, you already had knowledge of the kids' backgrounds. Every kid also had an individual sheet where we had all their scores, from benchmarks to six-grade teacher recommendations, you name it."

Program Monitoring. Seven of the 11 SPP project directors felt that how they monitored their SPP programs was effective. Of the seven project directors who noted effective program monitoring for their SPP program:

- Three hired a dedicated program monitor for the SPP program.
- Two specifically mentioned more detailed tracking of grades and student progress.
- Two noted improved communications with regular school staff about student progress.
- Two mentioned an improved teacher observation system.
- Two mentioned monitoring support from TEA's technical assistance provider.

Project director comments regarding program monitoring include:

"For SPP, we developed an observation form that was very specific for the model and what it should look like, and I think that was really an important organizational lesson that we learned, was that thinking out observations ahead of time, and that's something that we're continuing to do now. You know we've now crafted two observation forms. One for the academic parts of our program and one for the enrichment part of our program that are really specific to the sort of model that we want to see in every classroom when we go in. You know, there's things that need to be happening for best practices every time."

"I think the program, the way it was set up to where we actually had the opportunity to work with the administration and school-day staff to select the students and basically target a particular population within the school, and also keeping that close tie to the regular school-day teachers, and informing them, and keeping track of those students' grades, text course, and everything that needed to be monitored during their program. That's a big plus."

"I think it taught us a lot, the ACE leadership team a lot, about observations and the importance of really thinking about the observation process."

Student Recruitment. Eight of the 11 SPP project directors shared that they modified their approach to recruiting students when they implemented the SPP program and that

this approach was effective. Six project directors noted that they were more focused than they had been in the past on recruiting students for the program based on prior performance on STAAR and district benchmark tests. Two directors indicated that their recruitment efforts were "definitely a lot more targeted and very specific" and designed to "more thoroughly identify students who need help academically." Project directors also spoke about being more data driven in their approach to identifying students: "Whenever we, who were running the program, actually were the ones that went in and looked at the data and tried pulling the correct students to be in the program, it seemed to be a little bit more effective." Another project director shared how they targeted English language learners for the SPP program:

> *"We recruited the ELLs—the kids who were the English-speaking strugglers, especially the recent immigrants—the ones that we needed to make sure they were caught up as quickly as possible. The test they were going to take was going to affect their scores for the campus. They were the ones that were targeted—the ones at the very bottom."*

Staff Training. Four of the 11 SPP project directors explained that their approach to staff training was particularly effective. One director shared the following about the training model:

> *"I feel like our training model even has evolved, and so instead of meeting with the instructors once a week, which made it a little difficult for them to work at their campus, we now just train with them once a month but we're able to—even though the training is less frequent, I feel like it's a more efficient model, and, you know, we realize how confident our coaches can go out and feel empowered, they'll leave the lesson and maintain the fidelity and maintain their objective even if the kids aren't feeling it that day."*

Another project director discussed how they have expanded their academic staff training to include enrichment staff, which has helped to improve the quality of programming and instructional delivery:

> *"And we've taken some parts of that and kind of tried to train our youth workers to think that way for any activity because a lot of the cognitively guided instructions is about getting the kids to think for themselves and come up with their own strategy and really acting as a guide to the students rather than teaching them. So we've been thinking about that a lot with our training for our staff for any activity that they lead as something they should implement."*

Effective Instructional or Curricular Strategies Employed in the SPP Program

Face-to-Face, Small-Group Instruction. Almost all SPP project directors (10 of 11) indicated that Face-to-Face, small-group instruction in reading and/or mathematics was an effective component of the SPP program during the spring 2013 to summer 2014 period. This was by far the most commonly noted instructional or curricular aspect of the SPP program viewed as effective by project directors.

Two directors reported effective direct teaching in both mathematics and reading; one director reported effective Face-to-Face instructional delivery in mathematics only; and the remaining seven project directors indicated that effective direct teaching sessions exclusively involved reading. Project directors shared that "having those low-student numbers to the one teacher boosted the students' confidence," that "Face-to-Face interaction helped a lot, even with the online piece," and that they "were able to keep our class sizes really small compared to other ACE enrichment classes," resulting in "classes where the instructors could really work small group or one-on-one with their kids."

Teacher-Student Interactions. Five of the 11 project directors interviewed indicated that the nature of teacher-student interactions was a positive product of their SPP program. The following quote was offered by a project director regarding effective teacher-student interactions facilitated by SPP interventions:

> *"It's called a read-back, so when the child is ready for that portion on the computer, they go to the teacher at her computer and they have some interaction with the teacher. In fact this year, we've done it a little bit differently. The older kids mainly do some journal writing. They do a little review of the story while they're sitting by their computer, and then they bring their notebook up to their teacher when they do their read-back. So they're getting kind of a double dose of everything."*

Computer-Based Reading Programs. Four of 11 directors felt their Computer-Based reading interventions were effective. Each of the project directors discussed the positive effect of different commercially purchased Computer-Based reading programs implemented in their SPP program.[38] One project director shared the following about the system:

> *"The computer system allowed us to have each kid have tutorials linked specifically to them, and then increasing linkage between the people that were*

[38] One project director noted two effective Computer-Based reading programs in place in their SPP program.

teaching the afterschool class and individual class to know who is exactly struggling with what."

Another director spoke about the reporting advantages of their Computer-Based reading program:

> *"This particular program that we use has a really good way of reporting progress, which really helps and also the classroom teachers whenever they need to do the RTI meetings or conferences with the parents."*

Project-Based Learning Interventions. Four of 11 directors shared that project-based learning (PBL) approaches were effective in their SPP programs. Project directors noted that PBL strategies were effective because it allowed them to be "able to apply the kind of intangible, theoretical things that they were learning in a project" and that the approach was something that they could "carry out through our other classes as we continue to develop courses." Another project director shared the following experience with PBL in their SPP program:

> *"We had one of our lead counselors write the curriculum for it and it was very thought out and well planned. We're doing a lot of project-based learning in our program because of some of the benefits we saw of that, of the student choice part of it."*

Other Instructional or Curricular Approaches. Other effective instructional strategies shared by project directors included Computer-Based mathematics programs (*n*=2), social services programs (*n*=1), the use of an outside vendor (*n*=1), and rotating students among work stations (*n*=3). Project directors shared the following thoughts about rotating students among work stations in their SPPs:

> *"It's really important and then allowing the students to rotate through workstations and have that independence…. You know and we really found that, that method worked very well in afterschool."*

> *"You had two kids who were with direct instruction with the teacher. You had two kids doing independent practice…. You have the daily practice, independent practice, and direct instruction. So they rotated. You'd have about 20 minutes each or longer."*

Challenges to SPP Program Implementation

Barriers. Project directors were asked to share the primary barriers they encountered when implementing the SPP program at funded centers. Student recruitment was the most commonly noted barrier, cited by 5 of 11 project directors. Other barriers cited by three or more project directors included the following:

- Program staffing (*n*=4)
- SPP program reporting requirements (*n*=4)
- Student attendance (*n*=3)
- Technology issues (*n*=3)

Administration/school buy-in to the SPP program, integrating enrichment into the program, and trying to implement too many strategies or trying to address too many Texas Essential Knowledge and Skills (TEKS) student expectations were each noted by two project directors as barriers to effective program implementation.[39]

Proposed Changes. Six of the 11 project directors indicated that if the program had continued, they would have modified their approach to how it was implemented. Responses covered a variety of topics, including instructional approaches, curriculum, and organizational methods. For example, one project director shared that she "would have been more focused from the beginning on a narrower set of skills within the ELA TEKS, such as vocabulary. And so really paring down that intervention to something that you can affect." Another director indicated that "we would use a blended online and Face-to-Face approach for project based learning with a preset curriculum that teachers could just come in and deliver." A third project director indicated that she would use one of their instructional approaches but would not use an outside vendor for the delivery of reading and writing curriculum in the future because of quality of instruction and cost issues.

The remaining five project directors were satisfied with their academic interventions and would continue or have continued their use, sometimes with minor adjustments.

Program Sustainability

Because funding for the SPP program concluded in August 2014, it is important to understand the extent to which SPP activities have been sustained at ACE centers that had SPP funding, how they have been sustained, and if the experience with the SPP program has had any meaningful impacts on how afterschool leaders approach organizational and instructional aspects of OST programming.

Funding for Current Afterschool Activities and Continuation of SPP-Related Activities. Although five project directors noted that their current afterschool programs were supported with 21st CCLC funds, only two of the 11 interviewees reported that they were funded exclusively with 21st CCLC funds.

[39] TEKS are the state standards for what students should know and be able to do.

The remaining nine project directors reported receiving funding from one or more funding sources in addition to 21st CCLC funds. More specifically, seven of the nine project directors shared that district funds were being used to support current afterschool programming; two indicated that they had received grants from a local foundation to support their afterschool programs; three noted that parent fee-for-service programs were in place; three directors reported receiving grant funding from other sources; and two directors indicated that they were currently applying for grant funding to help support their afterschool programs.

Maintenance of SPP Activities. Project directors shared the different approaches they are using to continue SPP-related organizational and instructional practices even though SPP funding has concluded. One project director shared that they have continued SPP programming at the originally funded centers and have expanded the program to new ACE centers so that the grantee now has a total of 15 centers with SPP programming. At those 15 centers, the SPP program still functions as a separate program for targeted students alongside the traditional ACE program. Four other project directors indicated that most or all of the SPP program components have been rolled into either a 21st CCLC–funded ACE program or a district-funded program. Two directors reported that some SPP components have been integrated into their current afterschool programs.[40] Three project directors indicated that SPP activities have not been folded into their current afterschool work. Therefore, it appears that the majority of programs have continued aspects of SPP programming in some form.

Organizational Approaches Carried Forward Into Current Afterschool Programming. Project directors noted that they have carried forward a number of different organizational strategies (learned during their experience with the SPP program) to their current afterschool work. Six of the 11 project directors shared that they have changed their approach toward working with school leaders to improve campus buy-in for the afterschool program. One project director noted that “what we’re doing different is pushing more tutorials and just meeting more with the principals and working more with their improvement plans and working alongside the regular school day more so than we were before.” Another project director indicated that “we are also meeting more frequently with the administrator that supports 21st Century on that school.”

Five of the project directors indicated that they have adjusted their approach toward recruiting students for their afterschool program—typically being more focused on academic need than they had been in the past—and that they are using regular, school-

[40] At one grantee, two centers have continued to offer a before school program for K–2 students through a local 21st CCLC grant. The program functions just as it did when funded through the SPP grant in spring 2013 and the 2013–14 school year.

day staff to assist with the identification of students for the program. Project directors noted that they are "more focused in targeting academically and behaviorally challenged students for the program" and that they "target the students based on the need and recruit those students for the program." One project director shared the following recruitment approach:

> *"We have principals and counselors and teachers identify students in each of the campuses that could use the extra help. What we did was we set up guidelines this year. Those students who didn't pass their STAAR test, those students who are failing their core subjects, and those students who may have fluency and comprehension difficulties."*

Five of the 11 project directors also noted that the SPP program experience has changed how they currently approach staffing for their afterschool program—typically focusing more on certified teachers to deliver academic interventions in the afterschool environment. Project directors indicated that they are now "working a lot more closely with core-day teachers," that "academic liaisons are working more closely with the core-day teaching staff to identify what needs are," and that they make a greater effort to "engage with campus leaders."

Four of the 11 project directors interviewed indicated that they have adopted SPP program monitoring strategies in their current afterschool programming. One interviewee shared: "One thing very definitely is I have used a program monitor in the SPP program because I am just not able to get to every place all the time and I hired somebody to come in…and monitor those programs."

Two project directors shared that they have continued the staff training approach employed in their SPP programs, and two project directors indicated that they have continued their SPP enrichment activities with their current afterschool programs.

Instructional or Curricular Approaches Carried Forward Into Current Afterschool Programming. Project directors were also asked to elaborate on any SPP-related instructional or curricular approaches that they continued into their current afterschool programming. Eight of the 11 interviewees reported that they are now more focused on small-group, Face-to-Face activities in their current afterschool programs. This includes incorporating more small-group, direct teaching time into primarily Computer-Based learning interventions for mathematics and/or reading. Comments made by project directors about their continuation of this instructional approach include:

> *"We did independent reading time last year in conjunction with the [the intervention] to make up the hour and half that the students were getting. We continued that. We do reading activities. We have kids read books, read chapter*

books as a group. They write poems. They journal. They letter write. Things like that."

"But as far as like the academic classes, the academic enrichment, they also contain more with a small group…. As a lesson we learned from SPP is you know having those, you know identifying those kids based on their academic performance and standardized test performance and then you know in the afterschool environment you're offering lessons that are targeting those skill gaps. So in a 1-to-10 ratio you know those lesson plans that are written by an academic liaison are being delivered to the kids in need. And so they are building you know on them continuously."

Other academic approaches that project directors have continued in their current afterschool work include the following:

- Computer-Based reading programs (*n*=4)
- PBL approaches (*n*=4)
- Rotating students to different work stations (*n*=3)
- Outside vendor programs (*n*=3)
- SPP enrichment sessions (*n*=3)
- Tutorials (*n*=3)
- Computer-Based mathematics programs (*n*=2)
- Homework help (*n*=2)
- Smartboard mathematics games/iPod reading and mathematics games (*n*=2)
- Social services for at-risk youth (*n*=1)
- Social-emotional learning (*n*=1)

SPP Program Impact on Project Director Philosophy. The majority of the SPP project directors (10 of 11) shared that their experience with the pilot program changed their philosophy toward afterschool programming, with the remaining director indicating that the SPP program reinforced her beliefs. The most common philosophical change was related to a more focused approach to academic content delivery, which was noted by five different project directors. Project directors shared that they are "focusing more on academics now" and are "doing much more individualized [instruction] and then using some of the computer types of tutorials paired with the kids working individually or in groups with the teacher." Another project director discussed how academic delivery has improved in the afterschool program after participating in SPP:

> *"I think our academic delivery has improved because of the implementation of SPP. I think SPP opened our eyes to many of the needs that students face academically. When we implemented SPP, we actually dug deeper as to what the academic needs of the students were."*

Three of the 11 project directors shared that their experience with the SPP program made them more focused on aligning afterschool content with the regular school-day curriculum. One project director shared: "The academic content delivery—yes because much like what the guys were getting from the day as far as you know like reporting, we are heavily focusing on school day alignment. We have implemented an academic [component] in every program that we currently have."

In a related response, two interviewees shared that the SPP program helped them better understand the importance of cultivating meaningful partnerships and relationships with school principals and other school leaders to improve buy-in for afterschool activities. One director shared the following experience:

> *"Well, again, I've tried to improve my relationship and enhance discussions with all of my principals. We encourage them to come to our monthly meetings and our academic liaison and our site coordinator and my family engagement specialist and I. And then hopefully, the principal—or at the very least, an assistant principal that has been designated to liaise and speak directly to the principal, that's usually an assistant, an AP—will come in and we'll talk on a monthly basis. We're telling our principals that if you can give us your very best and brightest teachers to help us with lesson plans and activity plans, we'll try to offer real substance in our afterschool programming. Our principals are far less likely this year to talk about the afterschool program as a babysitting program or something like that than they might have been before."*

Conclusions and Key Findings

Understanding the cost effectiveness and return on investment for the SPP program is an important part of TEA's determining the value of funding similar intensive academic support programs in the future. In service to that, this chapter attempted to answer the following research questions:

- RQ 2—What is the cost effectiveness and sustainability of the SPP Program?
  - RQ 2.1—What is the per-student cost of the SPP program relative to school-related student outcomes?
  - RQ 2.2—What is the relationship between the allocation of SPP grant funding to specific activities and student outcomes?
  - RQ 2.3—What is the return on investment for SPP programs?

- RQ 2.4—In what ways has the SPP experience impacted the work of project directors and the delivery of afterschool programs, and which organizational or instructional components have they incorporated into 2014–15 afterschool programming?

Exploratory analyses related to RQs 2.1–2.3 had the following key findings:

- There is no significant relationship between per-pupil SPP program spending and academic performance, regardless of activity type. The evaluation team's exploration of the relationship between per-student spending levels and rates of meeting STAAR-Reading and STAAR-Mathematics standards did not reveal any significant findings. That is, the amount spent per student attending 30 or more days of SPP programming was not associated with meeting STAAR standards. Similarly, the mode of instructional delivery (i.e., Computer-Based or Face-to Face) did not appear to have an effect on rates of meeting STAAR standards. It is important to note that the lack of significant findings may be related to small sample sizes (32 centers with SPP funding from just 11 grantees) and the nature of the financial data reported by grantees.[41]
- **SPP programming is being sustained across participating afterschool centers.** Despite the lack of significant findings from the statistical analysis, post-project interviews with SPP project directors indicated that the program did have an impact on how centers think about and implement their afterschool programming and how components of SPP programming are being sustained. Specifically, the majority of SPP project directors reported that they have made changes to their organizational and administrative practices related to staffing and staff training, program monitoring, and recruiting students as a result of the SPP program.
- **Participation in SPP programming had an impact on the philosophy of program directors.** In addition, and perhaps most notably, the vast majority of project directors indicated that their experience with the SPP program changed their philosophy toward afterschool program delivery, making them more focused on academic content, alignment to the regular school-day curriculum, and building meaningful partnerships with school leaders and regular school-day staff. This change in philosophy has resulted in the incorporation of SPP-related organizational and instructional approaches into their current afterschool work, regardless of the method of program funding (e.g., 21st CCLC, district funding, and foundation grant funding).

[41] Data reported by grantees represent estimates of how grant funds were expended across centers and across different types of program activities. Because SPP grantees did not necessarily account for expenditures in the manner requested, it is possible that some error in the estimated program-related expenditures may have occurred.

These findings indicate that although the amount of funds that centers dedicated to SPP programming was not related to student outcomes, the SPP program did lead to changes in staff operational and instructional practices in most centers. This finding combined with other analyses may provide greater insight into the value of specialized and intensive academic instruction in traditional ACE centers.

Chapter 4: Impact of the ACE Program

As noted in the introduction to this report, in 2011–12, AIR conducted an evaluation of the ACE program. That evaluation found significant program effects for students who participated in the program 60 or more days. These students demonstrated higher mathematics and reading assessment test scores on the TAKS assessment, fewer school-day disciplinary incidents and absences from school, and greater on-time grade-level promotion.[42] The primary goal of impact analyses summarized in this chapter was to determine whether similar impacts were found based on youth participation in programming during the 2013–14 school year and to explore whether participation in SPP programming may have provided an added benefit for those who participated on a variety of outcomes.

In addition to examining the overall impacts of the ACE program, the evaluation team conducted analyses of the impact of the SPP program. The purpose of this analysis is not to directly compare student participation in ACE-only and SPP programming but rather to look at how SPP participants fare compared to nonparticipants. More specifically, the goal of this set of analyses is to understand if participation in a supplemental program with a focus on intensive academic support contributes to improved student outcomes.

Next, the evaluation team examined the added benefit of participating in SPP activities—that is, did students who participated in SPP activities in combination with ACE program activities see greater improvements than those who participated only in ACE activities. Again, the purpose of this analysis was not to directly compare ACE and SPP but rather to determine whether there is an added benefit to participating in intensive academic enrichment in addition to general ACE programming.

Finally, the evaluation team wanted to look at which type of approach to instruction might be associated with the greatest benefits. In 2012–13, the evaluation identified several program typologies (e.g., *Learning Strategies* versus *Skills-Based* instructional approaches and Face-to-Face versus Computer-Based modes of delivery) and examined the level of quality in each. That analysis revealed that the *Learning Strategies* approach was associated with higher quality. The evaluation team wanted to learn whether these approaches were also associated with greater youth outcomes.

[42] It is important to note, however, that at that time the state was still using the TAKS assessment, so direct comparisons between the 2011–12 and 2014–15 assessment test impacts are not appropriate. The 2011–12 report can be viewed here.

Collectively, this chapter examines the following research questions:

- RQ 3—What is the impact of ACE programming on a range of student outcomes?
  - RQ 3.1—How does the impact of programming differ by attendance level for both the overall ACE program as well as students specifically enrolled in SPP activities?
  - RQ 3.2—What impact does participation in SPP+ACE have on youth outcomes compared to similar youth that participated only in ACE activities?
  - RQ 3.3—If SPP+ACE is found to have a significant impact on youth outcomes, what is the relationship between program typology and outcomes?

RQ 3.1—Impact of ACE Programming by Attendance and SPP Status

The evaluation team examined the impact of participation in ACE programming on a variety of youth outcomes, including academic performance, school-day attendance, school-day disciplinary incidents, and on-time grade-level promotion. Students were broken into categories by level of attendance and then by SPP status. The following sections outline the results of these analyses.

Impact of Overall ACE Programming by Attendance Groups

First, the evaluation team wanted to examine the overall impact of the ACE program by attendance level in an attempt to replicate the 2011–12 evaluation.

Sample

In any evaluation of a program where participants are not randomly assigned to participate in the program, it is important to consider how students select into the program or intervention under study. We know that it is likely that students who participate in 21st CCLC programming are different from those who do not attend. In general, because of the way the funding is structured, students who attend ACE programs tend to be lower achieving students than those who do not, prior to the start of the current academic year. Students participating in SPP programs tend to have even lower achievement than ACE-only students because that program specifically targeted the lowest performing students in most funded centers. These differences can bias estimates of program effectiveness because they make it difficult to disentangle preexisting differences between students who attended the program and those who did not from the effect of attending the program.

The evaluation team used a PSM approach like the one described earlier in this report to help mitigate that existing bias in program effect when examining the effect of

participating in ACE programming on students' STAAR-Reading and Mathematics achievement, grade-level promotion, number of school-day disciplinary incidents, and number of days absent. Specifically, the study compared the performance of students who participated in ACE to similar students who did not participate. Participation was defined two different ways for the purpose of the analysis. First, students who attended at least 30 days were compared to students who did not attend programming at all. Second, students who attended at least 60 days were compared to students who had no program attendance. These "treatment" definitions ensured that the comparison of program effect was based on students who received a significant dose of ACE programming. For more details on the PSM approach, see Appendix D.

For the first part of RQ 3.1, examining the overall impact of ACE programming, the sample includes all students participating in ACE programming (including SPP students) during the 2013–14 school year who had at least 30 days of attendance compared to similar nonparticipants. Using the PSM approach, the evaluation team was able to create a matched comparison group that was equivalent to the treatment group on a list of student-level variables such as prior achievement on STAAR-Reading and Mathematics, prior attendance, gender, and ethnicity, and on school-level variables such as school type (e.g., urban, rural), total enrollment, and student race/ethnicity composition.[43] Tables 15 and 16 show the total number of students included in the analysis by participation level and the demographic details of students included in the sample.

[43] For school-level variables, the evaluation team used the feeder school that the majority of students in a given ACE program attended. In most cases, the feeder school and the host school for the program were the same.

Table 15. Number of ACE Student Cases Analyzed by Grade and Participation Level (2013–14)

Grade Level	30+ Days of Participation		60+ Days of Participation	
	Treatment	Comparison[1]	Treatment	Comparison[1]
Grade 4	16,921	26,698	16,907	26,658
Grade 5	15,945	27,803	15,930	27,755
Grade 6	16,746	31,327	16,734	31,262
Grade 7	16,444	34,113	16,426	34,035
Grade 8	15,914	33,825	15,890	33,715
Grade 9	8,817	30,418	8,799	30,205
Grade 10	8,454	25,905	8,427	25,719
Grade 11	7,688	21,417	7,656	21,163
Grade 12[2]	NA	NA	NA	NA
Total	**106,929**	**231,506**	**106,769**	**230,512**

[1] Treatment and comparison groups differ in size substantially because the groups were initially very different in size. The comparison group consists of all students who attend feeder schools where the majority of students in a given ACE program attend. During the PSM matching process, the comparison sample was reduced in size to include only those students who were equivalent to the treatment group on a list of covariates but still remained larger than the treatment group because of its original large size.

[2] Grade 12 was not included in the analysis because of the small number of students in Grade 12 who met the 30+ day attendance benchmark.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

Table 16. Participant and Nonparticipant Demographics for 30+ and 60+ Day Attendees (2013–14)

	Participants (*n*=106,929 30+ Day Attendees *n*=106,769 60+ Day Attendees)[1]	Nonparticipants (*n*=231,506 30+ Day Attendees *n*=230,512 60+ Day Attendees)
Female	50%	51%
Male	50%	49%
White	62%	64%
Black	20%	15%
Asian	2%	3%
Pacific Islander	0%	0%
American Indian	19%	21%
Hispanic	65%	68%
Special Education Status	6%	6%
English Language Learners	22%	25%
At risk[2]	50%	50%
Economic	15.76	18.94

[1] The percentages for 30+ and 60+ day attendees were identical in all categories because of the small difference in sample size for the two groups. Therefore, the evaluation team decided to include the percentages only once.

[2] As defined by Texas Education Code §29.081(d) (2015).

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

STAAR-Mathematics and Reading Performance

Table 17 shows the impact of participation in ACE programming on STAAR-Reading performance for 30+ day attendees and 60+ day attendees. The analyses revealed a significant and negative relationship between participation in ACE and STAAR-Reading performance at the Grade 4–5 level. That means that students who participated in ACE had lower reading performance than their nonparticipating peers. At the Grade 6–8 level, there is a similar finding, but this time it is only moderately significant, indicating that the relationship between participation and STAAR-reading performance is not as strong. Overall, the effect sizes are quite small but suggest a very slight negative relationship between ACE participation and STAAR-Reading performance.

Table 17. Effect of ACE Program Participation on STAAR-Reading and English I and II EOC Exams for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14)

Group[1]	30+ Day Attendees			60+ Day Attendees		
	Effect Size[4]	S.E.[5]	*p*-value	Effect Size[4]	S.E.[5]	*p*-value
Grades 4–5—STAAR-Reading	-0.041	0.005	<0.001***	-0.039	0.006	<0.001***
Grades 6–8[2]—STAAR-Reading	-0.007	0.004	0.100	-0.011	0.006	0.071*
English I EOC[3]	-0.005	0.011	0.654	-0.016	0.018	0.367
English II EOC[3]	0.004	0.013	0.724	0.024	0.019	0.215

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Results are not aggregated across all grades because the nature of the tests for each grade level grouping are different.

[2] Only includes Grade 8 students who took the STAAR-Reading exam and not those who took the English I EOC exam.

[3] EOC refers to end-of-course assessments, in which students taking the courses described were assessed at the end of the course. This analysis includes only students in Grade 9 for English I and Grade 10 for English II. Grades 10–12 were excluded from the English I EOC analysis and Grades 11–12 were removed from the English II EOC analysis in an effort to limit the sample to those students who were likely taking the exam for the first time.

[4] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[5] Standard error

SOURCE: State of Texas Assessments of Academic Readiness, 2013–14.

Table 18 shows the impact of participation in ACE programming on STAAR-Mathematics performance for 30+ day attendees and 60+ day attendees. In contrast to the analysis of STAAR-Reading performance, this analysis showed positive and significant findings for almost all grade levels and attendance groupings, meaning that the analyses demonstrated a relationship between participation in ACE and better performance on STAAR-Mathematics. The effects were larger for students who attended 60 or more days and for students in Grade 9. Although the effects are small, they are typical of, if not slightly lower than, what the evaluation team has found in similar statewide evaluations of 21st CCLC programs in other states and in prior years in Texas and are commensurate with what would be expected for a program of the scope and duration of typical 21st CCLC programs.[44] Overall, these findings suggest a positive relationship between participating in ACE programming at high levels and better STAAR-Mathematics performance.

[44] See Footnote 27 about effect sizes in comparable evaluations.

Table 18. Effect of ACE Program Participation on STAAR-Mathematics and Algebra I EOC for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14)

Group[1]	30+ Day Attendees			60+ Day Attendees		
	Effect Size[4]	S.E.[5]	*p*-value	Effect Size[4]	S.E.[5]	*p*-value
Grades 4–5 STAAR-Mathematics	-0.002	0.005	0.631	0.011	0.006	0.041**
Grades 6–8 Mathematics[2]	0.009	0.004	0.046**	0.021	0.006	<0.001***
Algebra I EOC[3]	0.041	0.014	0.004**	0.072	0.023	0.002**

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Results are not aggregated across all grades because the nature of the tests for each grade level grouping are different.

[2] Includes only Grade 8 students who took the STAAR-Mathematics exam and not those Grade 8 students who took the Algebra I EOC exam.

[3] EOC refers to end-of-course assessments, in which students taking the courses described were assessed at the end of the course. This analysis includes only students in Grade 9 for Algebra I EOC. Grades 10–12 were excluded from analysis in an effort to limit the sample to those students who were likely taking the exam for the first time.

[4] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[5] Standard error

SOURCE: State of Texas Assessments of Academic Readiness, 2013–14.

Nonacademic Outcomes

In addition to examining mathematics and reading achievement that may have improved as a result of afterschool programming, the evaluation team examined a variety of nonacademic outcomes that are also critical for student success, including school-day attendance, school-day disciplinary incidents, and on-time grade-level promotion. As noted in the introductory sections of this report, research has shown that afterschool programs can have an influence on these types of indicators. Table 19 shows the impact of participation in ACE programming on school-day discipline incidents for 30+ day attendees and 60+ day attendees. The analyses revealed a negative and significant relationship between almost all grade levels and attendance groups. In other words, analysis revealed that participation in afterschool programming, especially at 60+ days and for students in Grades 9–11, is associated with fewer school-day disciplinary incidents. The rate ratio column indicates a percentage to help in understanding the results presented in Table 19. For example, students participating at high levels in Grades 9–11 had a school-day disciplinary incidence rate that was 23% lower than

those who did not participate in ACE programming. See Appendix F, Table F-1, for a breakdown of these findings by individual grade level.

Table 19. Effect of ACE Program Participation on School-Day Discipline Incidents for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14)

Group	School-Day Discipline Incidents— 30+ Day Attendees				School-Day Discipline Incidents— 60+ Day Attendees			
	Rate Ratio[1]	Effect Size[2]	S.E.	*p*	Rate Ratio[1]	Effect Size[2]	S.E.	*p*
Grades 4–11[3]	-4%	-0.046	0.006	<0.001***	-16%	-0.169	0.009	<0.001***
Grades 4–5	-3%	-0.027	0.018	0.142	-11%	-0.121	0.022	<0.001***
Grades 6–8	-2%	-0.021	0.008	0.006**	-15%	-0.163	0.011	<0.001***
Grades 9–11	-13%	-0.141	0.014	<0.001***	-23%	-0.260	0.024	<0.001***

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Rate ratio is the percentage difference between participants and nonparticipants in each attendance category.

[2] Unstandardized regression coefficient

[3] Grade 12 was excluded from analysis because of the small sample size.

SOURCE: Public Education Information Management System data, 2013–14.

Table 20 shows the impact of participation in ACE programming on school-day absences for 30+ day attendees and 60+ day attendees. The results show that participation at all grade and attendance levels was significantly associated with a decrease in school-day absences. That means that there appears to be a relationship between participation in ACE programming and fewer school-day absences. The results are particularly strong for students who participated 60 or more days, especially in Grades 6–11, where youth participating in the program had an absence rate that was more than 20% lower than their peers not participating in programming. See Appendix F, Table F-2, for a breakdown of these findings by individual grade level.

Table 20. Effect of ACE Program Participation on School-Day Absences for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14)

Grade Level	School-Day Absences—30+ Day Attendees				School-Day Absences—60+ Day Attendees			
	Rate Ratio[1]	Effect Size[2]	S.E.[3]	*p*	Rate Ratio[1]	Effect Size[2]	S.E.[3]	*p*
4–11[4]	-14%	-0.149	0.000	<0.001***	-19%	-0.214	0.000	<0.001***
Grades 4–5	-11%	-0.121	0.000	<0.001***	-15%	-0.159	0.000	<0.001***
Grades 6–8	-16%	-0.169	0.000	<0.001***	-22%	-0.257	0.000	<0.001***
Grades 9–11	-15%	-0.145	0.000	<0.001***	-21%	-0.230	0.000	<0.001***

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Rate ratio is the percentage difference between participants and nonparticipants in each attendance category.

[2] Unstandardized regression coefficient

[3] Standard error

[4] Grade 12 was excluded from analysis because of the small sample size.

SOURCE: Public Education Information Management System data, 2013–14.

Table 21 shows the impact of participation in ACE programming on grade-level promotion for 30+ day attendees and 60+ day attendees. As with school-day absences, the results from this analysis show that participation in ACE programming at all grades and attendance levels has a significant impact on on-time grade-level promotion. Specifically, youth participating in the program demonstrated a greater likelihood of being promoted to the next grade level than similar students not participating in the program. These findings are particularly large for those who participate 60 or more days and those in Grades 9–11, even at lower participation levels. For students in Grades 9–11, participating 60 or more days in ACE programming is associated with being 55% more likely to be promoted to the next grade level on time. See Appendix F, Table F-3, for a breakdown of these findings by individual grade level.

Table 21. Effect of ACE Program Participation on Grade-Level Promotion for 30+ and 60+ Day Attendees Relative to Nonparticipants (2013–14)

Group	Grade-Level Promotion—30+ Day Attendees				Grade-Level Promotion—60+ Day Attendees			
	Odds Ratio[1]	Effect Size[2]	S.E.[3]	*p*	Odds Ratio[1]	Effect Size[2]	S.E.[3]	*p*
Grades 4–11[4]	+23%	0.203	0.002	<0.001***	+54%	0.432	0.000	<0.001***
Grades 4–5	+17%	0.160	0.005	<0.001***	+24%	0.219	0.007	<0.001***
Grades 6–8	+15%	0.140	0.006	<0.001***	+21%	0.191	0.013	<0.001***
Grades 9–11	+41%	0.341	0.008	<0.001***	+55%	0.437	0.000	<0.001***

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Odds ratio indicates what the odds of being promoted are for the treatment group versus the comparison group.

[2] Unstandardized regression coefficient

[3] Standard error

[4] Grade 12 was excluded from analysis because of the small sample size.

SOURCE: Public Education Information Management System data, 2013–14.

Taken together, the analyses presented in this section suggest that participating in ACE programming can have an impact on key student outcomes. With a few exceptions, effects are most likely to take place when students participate at high levels and are largest for students who are in Grades 9–11. It is possible that there are some preexisting characteristics of high school students who enroll and participate in ACE programming that makes them more likely to experience positive outcomes. Although impacts on STAAR performance were not strong, and in the case of STAAR-Reading were even negative, participation in ACE programming was associated with moderate improvements on outcomes that are critical for success in school such as school-day attendance, discipline incidents, and grade-level promotion.

Impact of ACE Programming by SPP Status

In addition to examining the benefits of ACE programming overall on student outcomes, the evaluation team looked at a similar set of outcomes for students who participated in SPP programming. As noted earlier, this was not done as a direct comparison between SPP and ACE-only participation but rather as a comparison between participating in SPP programming and not participating in any kind of programming. The goal was to better understand the value of participating in an intensive academic intervention like SPP. For these analyses, the evaluation team included any student who was designated as an SPP participant and attended programming activities 30 or more days,

regardless of the how often they engaged in SPP activities specifically.[45] That is, students included in this analysis could have participated in SPP a little, some, or all of their time after school and may also have participated in regular ACE programming when they were not in SPP activities.[46] The comparison group for this set of analyses were matched students who were enrolled in the same schools that the ACE participants attended and who did not participate in either ACE or SPP programming. In the next section, the evaluation team examined whether SPP was an added benefit to participating in ACE programming. That analysis, as the next section explains, takes dosage in SPP-specific activities into account. Table 22 contains information about the student sample for this research question.

Table 22. Student Sample for Analysis of Impact of SPP Program (2013–14)

Grade Level	30+ Days of Participation	
	Treatment	Comparison[1]
Grades 4–5	546	48,385
Grade 6–8	410	94,216
Grades 9–11[2]	312	73,454
Total	**1,268**	**216,055**

[1] Treatment and comparison groups differ in size substantially because the groups were initially very different in size. The comparison group consists of all students who attend feeder schools where the majority of students in a given ACE program attend. During the PSM matching process, the comparison sample was reduced in size to include only those students who were equivalent to the treatment group on a list of covariates but still remained larger than the treatment group because of its original large size.
[2] Grade 12 was excluded from this analysis because of small sample size.

SOURCE: Public Education Information Management System data, 2013–14.

As Table 23 shows, the evaluation team did not find evidence of a positive relationship between SPP participation and STAAR performance as compared to similar students not participating in the program. In fact, at the elementary level, a negative relationship was found between SPP participation and both STAAR-Reading and Mathematics performance, although the latter finding was only moderately significant. In other words, students in Grade 4 and Grade 5 enrolled in SPP performed worse on reading and mathematics assessments than their nonparticipating peers. There were no significant

[45] The evaluation team looked only at the 30+ day attendees and not the 60+ day attendees as with the previous research questions because the sample size of SPP students was smaller than for the ACE program as a whole. Further, for this research question, the evaluation team was interested in learning about the impact of any level of SPP activity participation above the 30+ day benchmark.
[46] Some ACE centers with SPP programming separated the two groups of students so that SPP students participated in activities only with other SPP students, even when they were engaged in enrichment versus academic intervention. In other centers, students stayed in SPP groups during academic interventions only some of the days after school (e.g., Tuesday/Thursday) and were placed into ACE activities when they were not explicitly participating in SPP.

findings at the middle and high school levels. A possible reason for this finding is that each SPP grantee was able to select an intervention that best met its local STAAR performance improvement needs. Therefore, SPP grantees generally used different interventions and selected different student populations for participation. It is important to note that these interventions may vary in their implementation as well as in their individual effectiveness. In addition, the SPP program specifically targeted low-performing students, and each SPP program director may have defined that a little differently, depending on the needs and challenges of their student population. Although the evaluation team took measures to control for student characteristics and prior academic performance, it was not possible to get a perfect match.[47] The difference between the two groups could account for these negative findings.

Table 23. Program Participants Versus Nonparticipants on Test Performance, 30+ Day Attendees (2013–14)

Group	STAAR-Reading			STAAR-Mathematics		
	Effect Size[3]	S.E.[4]	*p*	Effect Size[3]	S.E.[4]	*p*
Grades 4–5	-0.128	0.044	0.004**	-0.089	0.046	0.053*
Grades 6–8	0.047	0.038	0.223	-0.018	0.040	0.648
Grade 9[1]	0.011	0.078	0.885	-0.086	0.091	0.345
Grade 10[2]	-0.130	0.097	0.181	NA		

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] English I EOC, Algebra I EOC

[2] English II EOC

[3] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[4] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14.

Table 24 shows findings related to analyses of how SPP participants fared on school-day disciplinary incidents and absences compared to those students who did not participate in any programming. This analysis found that among students in Grades 6–11, SPP participation is associated with fewer school-day discipline incidents and absences. More specifically, students attending SPP had a rate of school-day disciplinary incidents that was 32% to 48% lower than similar youth not participating and an absence rate that was 14% to 20% lower. Again, given that these models were correlational, we can say only that a relationship exists between participating in SPP

[47] There are a few covariates on which the treatment and matched comparison groups were not well balanced (Standard Mean Difference>0.05 because of the comparatively small number of treatment students), and they were controlled for in the impact model.

programming and school-day discipline incidents and absenteeism. It is possible that other key preexisting differences between youth attending SPP programming and the nonparticipant comparison group are driving the lower rate of school-day disciplinary incidents and absences. That being said, in many studies of afterschool program impacts, reduced school-day disciplinary incidents and absences are common findings (Naftzger et al., 2013; Naftzger et al., 2014).

Table 24. Program Participants Versus Nonparticipants on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14)

Group	School-Day Discipline Incidents				School-Day Absences			
	Rate Ratio[1]	Effect Size[2]	S.E.[3]	*p*	Rate Ratio[1]	Effect Size[2]	S.E.[3]	*p*
Grades 4–5	+ 3%	0.031	0.238	0.897	+ 2%	0.018	0.036	0.622
Grades 6–8	- 32%	-0.389	0.061	0.000***	- 20%	-0.228	0.023	0.000***
Grades 9–11[4]	- 48%	-0.651	0.098	0.000***	- 14%	-0.157	0.025	0.000***

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Rate ratio is the percentage difference between participants and nonparticipants in each attendance category.

[2] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[3] Standard error

[4] Grade 12 was excluded from analysis because of the small sample size.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2012–13 and 2013–14.

Overall, these findings indicate that a relationship may exist between SPP participation and reduced school-day disciplinary incidents and absences for students in Grades 6–11.

RQ 3.2—The Added Benefit of Participating in SPP and ACE Programming

As noted earlier in this chapter, as part of understanding the overall impact of the ACE program and the added benefit of an intensive academic intervention like SPP programming, the evaluation team examined the added benefit of participating in SPP programming in addition to ACE programming as compared to participating in ACE-only programs on a variety of outcomes.

Sample

For RQ 3.2, the sample was more limited than for RQ 3.1. Here, the evaluation team was interested in looking at the added benefit of participating in SPP activities in

addition to ACE activities as opposed to participating in just ACE activities or just SPP activities. In order to develop the SPP+ACE treatment group, the evaluation team wanted to identify students who spent at least 25% and not more than 75% of their time in each type of activity. Therefore, students who spent 25% to 75% of their time in SPP activities were included in the treatment sample (*N*=397). Students with more than 75% of their time in SPP activities were considered SPP-only and therefore were not included in either the treatment or comparison group. Again, the goal was not to compare ACE and SPP, in which case the analysis would have included SPP-only students. Instead, the purpose of this set of analyses was to see how those students experiencing both SPP and ACE activities fared compared to students participating only in ACE programs without the additional intensive academic support. To identify a suitable comparison group, the evaluation team first looked at ACE-only students (meaning they had no time in SPP activities) from the same center as the SPP+ACE sample. This left too few students to ensure an accurate match for the PSM process. Therefore, the evaluation team expanded the comparison group to include ACE-only students from other centers operated by the same grantees as those included in the treatment group (*N*=7,786) under the assumption that these centers would share similar organizational practices and often used similar interventions and structures. This allowed for a better matching process.

As Table 25 shows, analysis revealed a moderately significant relationship between participating in SPP+ACE programming for 30 or more days and higher STAAR-Mathematics achievement as compared to participation in ACE-only. For this analysis, the evaluation team used a PSM process to match the treatment group (SPP+ACE) to a nontreatment comparison group (ACE-only), thus creating the analytic group for the impact analysis. The finding is noteworthy for a couple of reasons. First, the effect of SPP on STAAR-Mathematics achievement, although still small and only moderately significant, is larger than typically found when comparing 21st CCLC participants with nonparticipants (see footnote 27 and Naftzger et al., 2013; Naftzger et al., 2014). In addition, the effect appears at only 30 or more days of participation. As noted earlier in the chapter, for the overall ACE programming impacts, the evaluation team also found a statistically significant relationship between participation and STAAR-Mathematics, but the effect was smaller and was found only at the 60+ day participation level. This finding indicates that although the significance is moderate, there may be an added benefit on mathematics performance to participating in SPP+ACE programming.

Table 25. SPP+ACE Versus ACE-Only on STAAR Test Performance, 30+ Day Attendees (2013–14)

Subject[1]	Effect Size[2]	S.E.[3]	*p*
Grades 4–8 STAAR-Reading	0.032	0.059	0.591
Grades 4–8 STAAR-Mathematics	0.107	0.059	0.069*

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] These analyses did not include Grades 9–12 because the EOC tests given in those grades are substantively different from the Grades 4–8 STAAR exams so could not be grouped and the sample size was too small to allow individual grade-level analyses.

[2] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[3] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2012–13 and 2013–14.

Table 26 shows results from an examination of the relationship between SPP+ACE participation and school-day discipline incidents and school-day absences as compared to ACE-only participation. This analysis found a negative and significant relationship between SPP+ACE participation and both outcomes. This indicates that participating in SPP+ACE contributed to fewer school-day absences and disciplinary incidents. The effect on school-day discipline incidents was particularly strong, with SPP+ACE students showing a rate of school-day disciplinary incidents that was 58% lower than youth attending only ACE programming.

Table 26. SPP+ACE Versus ACE-Only on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14)

Outcome[1]	Rate Ratio[2]	Effect Size[3]	S.E.[4]	*p*
School-Day Discipline Incidents	-58%	-0.864	0.181	0.000***
School-Day Absences	-11%	-0.121	0.042	0.004**

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] This analysis includes Grades 4–8 because academic indicators (e.g., STAAR exams) used in the PSM process are substantively different at different grade levels preventing grouping and sample size was too small to allow individual grade-level analysis.

[2] Rate ratio is the percentage difference between participants and nonparticipants in each attendance category.

[3] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[4] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

Finally, the evaluation team examined SPP+ACE participation related to grade-level promotion. These analyses, presented in Table 27, did not find a significant relationship between participation in SPP+ACE programming and on-time grade-level promotion as compared to ACE-only participation.

Table 27. SPP+ACE Versus ACE-Only on Grade-Level Promotion, 30+ Day Attendees (2013–14)

Subject[1]	Odds Ratio[2]	Effect Size[3]	S.E.[4]	*p*
Grade-Level Promotion	+79%	0.582	0.608	0.338

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] This analysis includes Grades 4–8 because academic indicators (e.g., STAAR exams) used in the PSM process are substantively different at different grade levels preventing grouping and sample size was too small to allow individual grade-level analysis.

[2] Odds ratio indicates what the odds of being promoted are for the treatment group versus the comparison group.

[3] Unstandardized regression coefficient

[4] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

Taken together, these analyses suggest that participating in SPP+ACE programming contributed to improved school-related outcomes for students in Grades 4–8 as compared to participating in ACE-only programming. Specifically, participating in combined programming led to fewer school-day discipline incidents and fewer school-day absences and moderately significant improvements in STAAR-Mathematics achievement.

RQ 3.3—Examining Program Typology

As a next step, the evaluation team wanted to look more deeply into the program typology developed during the 2012–13 evaluation for SPP programs. That typology included two dimensions of programming: the mode of delivery (how content is delivered) and the program approach to improving skills. More specifically, the evaluation team identified two categories of typologies, mode of delivery (Face-to-Face, Computer-Based, or Combined Mode) and overall curricular approach (*Learning Strategies* or *Skills-Building*). In the *Learning Strategies* approach, instructors emphasize general learning strategies that are applicable across different content areas, while in the *Skills-Building* approach, instructors focus more on specific skills associated with a subject area. To read more about how these typologies were developed, see Appendix G and the report from the combined 2012–13 and 2013–14 evaluation (Devaney et al., 2015). In the 2012–13 evaluation, the evaluation team learned that the Face-to-Face and Combined Modes of delivery and the *Learning Strategies* approach were more

closely associated with student engagement and higher quality assessment scores on the PQA than Computer-Based and *Skills-Based* approaches.

To explore these findings further, for the 2014–15 evaluation, the evaluation team refined the SPP+ACE versus ACE-only participation analysis by examining the differences between participant groups by program typology. The sample size for this analysis was quite small because the sample was first filtered to include SPP+ACE students who met the inclusion criteria described earlier in this chapter and then was reduced further into the various typology groupings. Because of the small sample size, the analysis was correlational and not causal. Therefore, the findings presented here should be interpreted with caution as merely suggesting evidence of a relationship.

Table 28 shows the analysis of STAAR-Reading and Mathematics performance for SPP+ACE versus ACE-only participants broken down by typology. The evaluation team ran a separate model for each typology. This analysis shows a moderately significant relationship between STAAR-Mathematics performance and both the *Learning Strategies* approach and the Face-to-Face mode of delivery. The effect sizes related to these two moderately significant findings were larger than found in other analyses examining STAAR assessment results.

Table 28. SPP+ACE versus ACE-Only on STAAR Test Performance, 30+ Day Attendees (2013–14)

Typology[1,2]	STAAR-Reading			STAAR-Mathematics		
	Effect Size[3]	S.E.[4]	*p*	Effect Size[3]	S.E.[4]	*p*
Learning Strategies approach	-0.044	0.144	0.758	0.220	0.120	0.068*
Skills-Based approach	-0.098	0.062	0.113	-0.096	0.061	0.119
Face-to-Face mode	-0.046	0.148	0.756	0.218	0.121	0.070*
Computer-Based mode	-0.115	0.11	0.300	0.016	0.111	0.885
Combined mode	-0.096	0.077	0.216	-0.099	0.077	0.197

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] This analysis includes Grades 4–8 because academic indicators (e.g., STAAR exams) used in the PSM process are substantively different at different grade levels preventing grouping and sample size was too small to allow individual grade-level analysis.

[2] *N* sizes for each typology analysis: *Learning Strategies* approach=504, *Skills-Based* approach=2118, Combined mode=1614, Face-to-Face mode=452, Computer-Based mode=556

[3] Standardized regression coefficient or beta coefficient, which shows the change in the dependent variable in standard deviation along with one unit change of the independent variable. It should be between -1 and 1.

[4] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; 21st CCLC Evaluation observations of 22 activities, 2012–13.

Table 29 shows the analysis of SPP+ACE versus ACE-only participation in relation to school-day absences and disciplinary incidents. As this table indicates, there is a significant relationship between both the *Learning Strategies and Skills-Based* approaches and school-day discipline incidents. The *Learning Strategies* approach effect is larger. In addition, the Face-to-Face mode of delivery is associated with decreases in school-day disciplinary incidents. Finally, a moderately significant relationship was found between the Computer-Based mode of delivery and decreased school-day absences, but the effect size is smaller. Taken together, these findings suggest evidence that the *Learning Strategies* approach and Face-to-Face mode of delivery may be most effective in decreasing school-day discipline incidents, with the rate of school-day disciplinary incidents being more than 50% lower among students designated as SPP+ACE compared to those designated as ACE-only when considering programs adopting these approaches to service delivery.

Table 29. SPP+ACE Versus ACE-Only on School-Day Discipline Incidents and Absences, 30+ Day Attendees (2013–14)

	School-Day Discipline Incidents				School-Day Absences			
Typology[1,2]	**Rate Ratio[3]**	**Effect Size[4]**	**S.E.[5]**	***p***	**Rate Ratio[3]**	**Effect Size[4]**	**S.E.[5]**	***p***
Learning Strategies approach	-54%	-0.766	0.311	0.014**	+7%	0.063	0.088	0.473
Skills-Based approach	-33%	-0.394	0.148	0.008**	+6%	0.062	0.040	0.124
Face-to-Face mode	-56%	-0.828	0.344	0.016**	+7%	0.063	0.088	0.473
Computer-Based mode	-22%	-0.243	0.188	0.196	+10%	0.091	0.055	0.099*
Combined mode	-21%	-0.242	0.253	0.339	+1%	0.005	0.059	0.936

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] This analysis includes Grades 4–8 because academic indicators (e.g., STAAR exams) used in the PSM process are substantively different at different grade levels preventing grouping and sample size was too small to allow individual grade-level analysis.

[2] *N* sizes for each typology analysis: *Learning Strategies* approach=504, *Skills-Based* approach=2118, Combined mode=1614, Face-to-Face mode=452, Computer-Based mode=556

[3] Rate ratio is the percentage difference between participants and nonparticipants.

[4] Unstandardized regression coefficient

[5] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; 21st CCLC Evaluation observations of 22 activities, 2012–13.

Taken together, these findings suggest that of the various typologies, the *Learning Strategies* approach and Face-to-Face mode of delivery may have the strongest relationship to improved outcomes. Although it was beyond the scope of this evaluation to examine more closely why these typologies might be related to decreased school-day disciplinary incidents and improved mathematics performance, it is the evaluation team's hypothesis that the *Learning Strategies* approach and Face-to-Face mode of delivery allow for more hands-on interaction between instructors and students. The PQA highlights interaction and relationships as key quality practices. It may be the case that these approaches use higher quality practices that, in turn, contribute to improved outcomes like those theorized in the theory of change and partially evidenced in the analysis for RQ 1 presented earlier in the report. These analyses were more descriptive in nature and should be interpreted with caution. Although evidence of a relationship exists, the sample sizes were quite small. In addition, the evaluation team used observations of a handful of activities at each center to determine a center-level designation as belonging to one typology or another. Although most of the centers adopted universal strategies when implementing SPP programming and could therefore be reasonably classified as following one typology or another, it is possible that not all activities in a given center followed the same typology.

Conclusions and Key Findings

The goal of this chapter was to examine the overall impact of the ACE program as well as to look at specific student groups and approaches. The chapter addressed the following research questions:

- RQ 3—What is the impact of ACE programming on a range of student outcomes?
    - RQ 3.1—How does the impact of programming differ by attendance level for both the overall ACE program as well as students specifically enrolled in SPP activities?
    - RQ 3.2—What impact does participation in SPP+ACE have on youth outcomes compared to similar youth that participated only in ACE activities?
    - RQ 3.3—If SPP+ACE is found to have a significant impact on youth outcomes, what is the relationship between program typology and outcomes?

Data examining ACE participation, SPP+ACE participation, and the role of different approaches to instruction revealed the following key findings:

- **Participating in ACE programming at high levels (60+ days) is related to small improvements in STAAR-Mathematics performance, particularly for Grade 9 students.** Analyses revealed that students who participate in ACE programming at high levels see improvements in their STAAR-Mathematics

performance. Effects at the high-school level were more than three times larger than those at the elementary and middle school levels. Although the effect sizes were small, they bear some similarity to effects found in other 21st CCLC evaluations nationwide and in Texas in previous years.

- **Participating in ACE programming at high levels (60+ days) is related to fewer school-day disciplinary incidents and school-day absences and increased on-time grade-level promotion.** In addition to improvements in academic indicators, the analyses found even stronger effects on nonacademic, but school-related, outcomes. Again, at higher participation levels and at almost all grade levels, analyses revealed that participation in ACE leads to a decrease in school-day disciplinary incidents and absences and an increase in on-time grade-level promotion. Again, as found with STAAR-Mathematics and as was found in the 2011–12 evaluation of this program, these effects were particularly strong for high school students.
- **Participation in the SPP program is not associated with improved academic performance as compared to nonparticipation but does have a relationship with decreased school-day discipline incidents and absences.** When looking at SPP students as compared to students who did not participate in ACE programming at all, participation in SPP programming was not associated with improved academic performance. In fact, the evaluation team found a negative relationship between SPP participation and STAAR-Reading and Mathematics performance. SPP participation was associated with a decline in school-day disciplinary incidents and absences, even at lower (30+ days) participation levels.
- **Participation in SPP+ACE programming was associated with improved STAAR-Mathematics performance and decreased school-day disciplinary incidents and absences.** The effects were minimal when comparing SPP participants to nonparticipants, but when looking at the added value of SPP on top of ACE programming (i.e., SPP+ACE vs. ACE-only), the results were more promising. Although the analysis was correlational, and therefore only merely suggestive of a relationship, there appears to be an added benefit of SPP+ACE programming on STAAR-Mathematics performance, school-day disciplinary incidents, and school-day absences.
- **The *Learning Strategies* approach and Face-to-Face mode of delivery may be more likely to lead to improved outcomes.** Although the sample size was too small for the evaluation team to definitively examine the typologies, analyses suggest that the *Learning Strategies* approach and Face-to-Face mode of delivery were associated with greater gains in STAAR mathematics performance and decreased school-day disciplinary incidents for SPP+ACE participants as compared to ACE-only participants.

Chapter 5: Conclusions and Recommendations

As noted in the introductory chapter of this report, the 2014–15 evaluation of the ACE program was designed as a culmination of the previous four years of the program's evaluation. In particular, the evaluation outlined in this report aimed to reexamine the generic theory of change as well as look at the SPP theory of change, both presented in the report's introduction, to get a better understanding of both the overall ACE program and the effectiveness of supplemental funding programs for intensive academic support within OST settings, such as the SPP program. The report is organized around three key research questions that support the evaluation's overarching goals of examining program implementation and impact. This chapter first outlines each research question and the findings associated with it, then provides some overarching conclusions based on five years' worth of evaluation findings, and finally concludes with a series of recommendations for the ACE program as a whole going forward.

2014–15 Evaluation Conclusions

RQ 1—What key practices, strategies, and approaches are especially related to quality programming that leads to improved youth outcomes?

- RQ 1.1—What is the relationship between program quality and student outcomes?
- RQ 1.2—How does the relationship between quality and student outcomes differ for SPP students and ACE-only students?
- RQ 1.3—How does the level of youth engagement mediate the relationship between quality and outcomes?
- RQ 1.4—How do changes in youth mindsets and behaviors mediate the relationship between participation and outcomes?

Analyses undertaken to explore these various research questions revealed that **there is a positive relationship between quality and outcomes**, but that relationship and **the mediating effect of youth-reported engagement on it is not clear, potentially because of the presence of explicit academic content in SPP programming.** In addition, **youth mindsets and behaviors do not appear to have a relationship to participation and outcomes**, although more research is needed to understand these complicated relationships. The findings that contribute to this overall conclusion include:

- *High-quality programs may lead to improved outcomes under certain conditions.* The evaluation team found evidence of a relationship between quality and outcomes when ACE and SPP programs were examined separately. Both high-

quality ACE and SPP programs were associated with decreased school-day absences, and high-quality SPP programs were associated with improved mathematics achievement.

- *High-quality and high academic content activities can lead to positive outcomes despite low engagement.* Although the evaluation team saw a negative relationship between quality and engagement and between quality and outcomes when examined separately, there was still a positive relationship between high-quality programs and outcomes when viewed through the lens of engagement. This indicates that it is still possible for students to improve outcomes when participating in high-quality programs even if engagement is low. It is the evaluation team's hypothesis that programs with explicit academic content may not be as engaging to youth while still being high quality. The majority of activities observed and included in the analyses this year, both ACE-only and SPP, were academic in nature, which may have led to the results. However, analysis to test this hypothesis was beyond the scope of this evaluation.
- *No relationship was found between improvement in youth-reported mindsets and behaviors and school-related outcomes.* The evaluation team did not find a mediating effect of mindsets and behaviors on the relationship between participation and outcomes. Although it is possible that mindsets and behaviors have no relationship to participation and outcomes, the evaluation team believes it is more likely that the youth report survey is imperfect or that only a subset of youth see improvements on certain skills and belief areas and that these improvements impact some outcomes and not others. Sample size considerations and project resources did not afford the evaluation team the opportunity to consider all of these options.

RQ 2—What is the cost effectiveness and sustainability of the SPP program?

- RQ 2.1—What is the per-student cost of the SPP program relative to school-related student outcomes?
- RQ 2.2—What is the relationship between the allocation of grant spending to specific activities and student outcomes?
- RQ 2.3—What is the return on investment for SPP programs?
- RQ 2.4—In what ways has the SPP experience impacted the work of project directors in how they organize and deliver afterschool programs, and which organizational or instructional components have they incorporated into 2014–15 afterschool programming?

By conducting a thorough return-on-investment and cost-per-student analysis as well as interviews with project directors, the evaluation team was able to determine both that **the amount of funds that centers dedicated to SPP was not related to student outcomes** and that **the SPP program did lead to changes in staff operational and instructional practices** in most centers. This combination of findings may indicate that intensive academic interventions such as SPP can lead to sustained change in practice regardless of how much funding is dedicated to the process. More specifically, the evaluation team found the following:

- *There is no significant relationship between per-pupil SPP program spending and academic performance, regardless of activity type.* Analyses did not reveal any relationship between per-student spending and performing on the STAAR-Reading and STAAR-Mathematics exams overall and when taking the delivery mode (i.e., Computer-Based or Face-to-Face) into account.
- *SPP contributed to changes in organizational and instructional practice in many centers.* Most project directors indicated that their experience with SPP changed their philosophy toward afterschool program delivery, making them more focused on academic content, alignment to the regular school-day curriculum, and building meaningful partnerships with school leaders and regular school-day staff. This change in philosophy has resulted in the incorporation of SPP-related organizational and instructional approaches into their current afterschool work, including those related to staffing and staff training, program monitoring, and recruiting students. Many programs are sustaining components of SPP in their programs despite the conclusion of the funding cycle.

RQ 3—What is the impact of ACE programming on a range of student outcomes?

By conducting rigorous impact analyses on the ACE programming overall as well as a variety of correlation analyses looking at the impact of subsets of students and programs, the evaluation team was able to determine a **relationship between participation and improved STAAR-Mathematics performance as well as participation and reduced school-day disciplinary incidents and absences**. Findings appear to be **strongest among students who participate at high levels (60+ days) and who are in Grades 9–12**. Specific findings that contribute to these overall conclusions include:

- *Students who participated in ACE for 60+ days showed improved STAAR-Mathematics performance.* Analyses comparing ACE participants to similar nonparticipants found a small relationship between participation and improved STAAR-Mathematics performance. Findings were stronger for students in Grade

9 (.07 versus .01 and .02 for elementary and middle school levels, respectively). Although small, the results were similar to those found in other statewide 21st CCLC evaluations, including the 2011–12 evaluation of ACE programming.

- *Students who participated in ACE for 60+ days were more likely than nonparticipants to have a decrease in school-day disciplinary incidents and absences.* Here again, high school students had larger effects than elementary and middle school students, but overall, there appears to be a strong relationship between high levels of participation in ACE and decreases in problematic school-related behaviors.
- *Although there does not appear to be an impact on academic performance when comparing SPP participants to nonparticipants, there does appear to be an added benefit to participating in SPP+ACE programming rather than ACE-only programming.* Analysis looking at SPP participants compared to nonparticipants did not find a positive relationship between participation and academic performance. However, when looking at the added benefit of participating in SPP+ACE programming versus ACE-only programming, analyses showed that SPP+ACE was associated with improved performance on STAAR-Mathematics.
- *The Learning Strategies approach and Face-to-Face mode of delivery may be associated with improved mathematics performance and decreased school-day disciplinary incidents.* In conducting an analysis of the various program typologies, the evaluation team found evidence of a relationship between the *Learning Strategies* approach and Face-to-Face mode of delivery and improved STAAR-Mathematics performance. In addition, although both the *Learning Strategies* and Skills-Based approaches were associated with a decrease in school-day disciplinary incidents, the *Learning Strategies* approach had a larger effect. Finally, the Face-to-Face mode of delivery was also associated with a decrease in school-day disciplinary incidents. Although these findings were simply correlational and cannot definitely point to these two approaches as superior, the analysis results are consistent the 2012–13 evaluation results that found a *Learning Strategies* approach was associated with higher quality programming.

Broad Conclusions about the Effectiveness of the ACE Program

Conclusions about Impact

Overall, the five-year evaluation of the ACE program has revealed that students participating at high levels (60+ days per year) do see improvements on key school-related indicators. In both the 2011–12 and 2014–15 evaluations, the largest impacts were related to on-time grade-level promotion followed by decreased school-

day absences and decreased school-day disciplinary incidents. The program has also had a small impact on mathematics achievement as measured first by the TAKS assessment and later by the STAAR exam. In both years, the largest impacts were felt by students in Grades 9–12, especially for on-time grade-level promotion. Those impacts were very high in the 2011–12 evaluation and somewhat smaller, but still substantial, in the 2014–15 evaluation.

The results related to the SPP program were somewhat less conclusive. Although the evaluation found that participating in SPP was associated with substantial decreases in disciplinary incidents and somewhat smaller but still strong decreases in school-day absences, it was also associated with slight declines in both mathematics and reading performance on the STAAR exam. The latter finding was somewhat troubling given the program's specific purpose to help students improve their mathematics and reading competency. However, these results looked at participating in SPP versus not participating in any programming and did not take the varying interventions into account. When examining the added benefit of participating in SPP+ACE rather than ACE-only programming, the evaluation team found that SPP+ACE participation was associated with significantly improved STAAR-Mathematics performance. Participation in this combination of programming was also associated with decreased school-day disciplinary incidents and absences. These findings suggest that an intensive academic intervention paired with ACE enrichment may provide the right combination of programming to support positive outcomes.

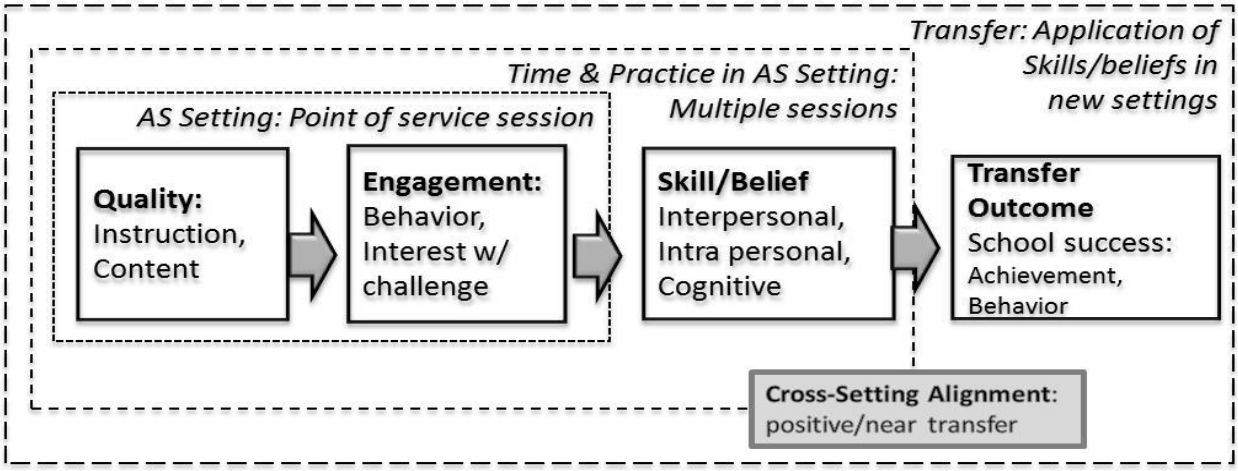
Finally, the 2012–13 and 2014–15 evaluations both suggest that employing a *Learning Strategies* approach to instruction may be the most effective way to support student learning. In 2012–13, the evaluation team developed the instructional typologies used in the evaluation and found that the *Learning Strategies* approach was associated with higher quality and greater student engagement. In 2014–15, the evaluation team conducted analyses to determine which typology might be most effective in improving student outcomes. Although correlational in nature, the analyses suggested a relationship may exist between participation in programs using a *Learning Strategies* approach and decreased school-day disciplinary incidents and improved STAAR-Mathematics performance for students experiencing both SPP and ACE programming.

Conclusions about Quality and Engagement

Conclusions about program quality and engagement as a result of the five-year evaluation are somewhat more complicated than those related to impact. They are further complicated by the introduction of SPP programming to the evaluation starting in 2012–13. AIR started the evaluation in 2010–11 with the generic theory of change

shown in Figure 6 as a guide. This theory of change was used in other programs across the country and has an evidence base to support it.

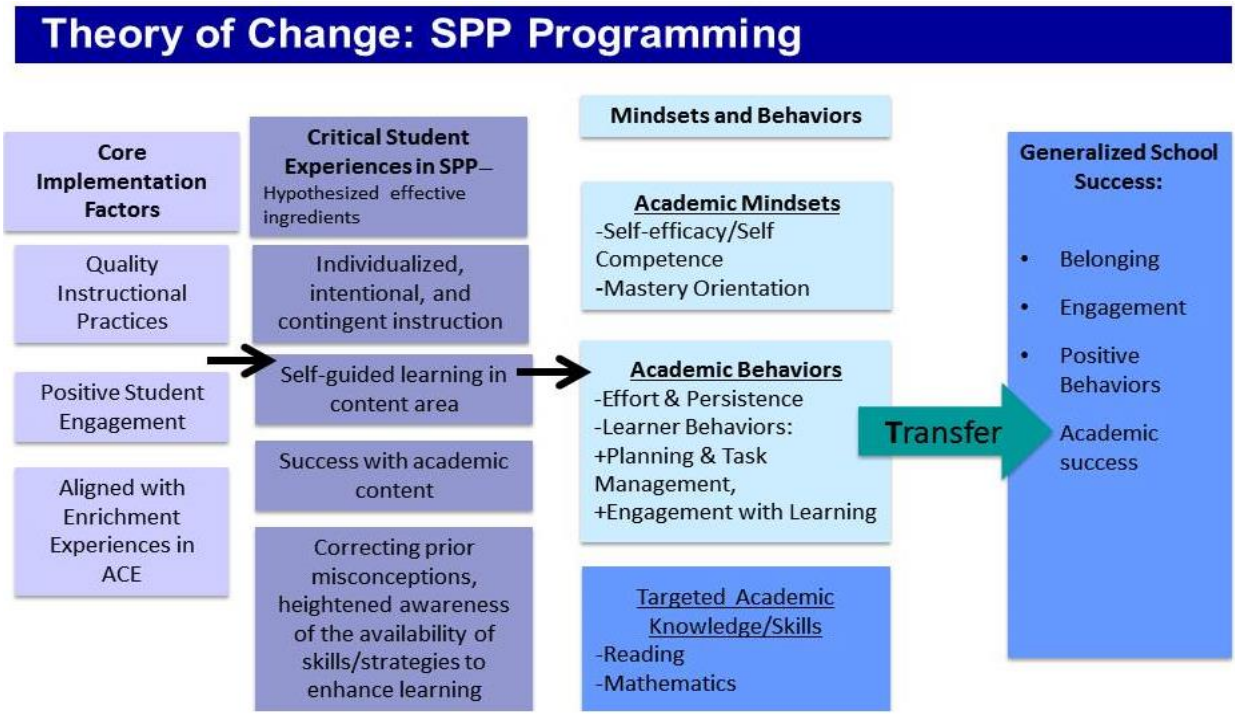
Figure 6. Theory of Youth Change Related to Afterschool Program Participation



AS = afterschool

The first two years of the evaluation explored this theory of change. Then in 2012–13, the SPP program was introduced as an intervention program within the ACE program, and the evaluation team developed a second theory of change specific to SPP. This theory of change is shown in Figure 7.

Figure 7. SPP Theory of Change



Both theories attempt to explain the relationships among quality, engagement, student experiences in programs, improved mindsets and behaviors, and outcomes. Through analyses during the course of the five years, the evaluation team has found aspects of both theories of change to be true and others to be inconclusive. Collectively, the five-year evaluation has told us the following about these key relationships:

- **Program quality is related to cognitive engagement.** In 2010–11 and 2012–13/2013–14, the evaluation team looked at the relationship between quality and engagement at the activity level. In both sets of analyses, we were able to closely link PQA quality scores with youth-reported engagement scores for the same activity and found significant and positive relationships. The evaluation team did not find this same level of relationship in the 2014–15 analysis. This might be due the fact that the analysis was done at the center level rather than the activity level, where there was a more direct connection between what was observed and what youth experienced as a result. For 2014–15, the evaluation team tried to extrapolate these in-the-moment, activity-level experiences to the center level across the span of the school year and for other youth attending the center not specifically enrolled in observed activities. This process of extrapolating to the center level introduced additional error into these calculations, which could have impacted the outcome of these analyses.
- **Explicit academic content may depress engagement**. In 2013–14, the evaluation team conducted analyses of the relationship between quality and engagement. Again, these analyses were conducted at the activity level and revealed a strong positive relationship between quality and youth-reported engagement in most cases, while SPP-funded activities and activities with explicit academic content were negatively related to engagement. The evaluation team hypothesized at the time that the explicit academic content present in SPP activities may contribute to decreased engagement. This hypothesis could also explain the findings from the 2014–15 analysis that found no relationship between quality and youth-reported engagement because most SPP and ACE-only activities observed were academically oriented. This notion was beyond the scope of this evaluation and is an area for further study.
- **Program quality can be related to improved outcomes, even when engagement is low, if academic content is high.** Findings from the 2011–12 and 2014–15 evaluation found a relationship between quality and outcomes, albeit through somewhat different paths. In 2011–12, analyses found a relationship between higher quality programs and improved student outcomes when the programs provided intensive academic activities and had high levels of observed engagement. In 2014–15, analyses found a relationship between higher quality programs and improved student outcomes when youth-reported

engagement was low. Although these findings seem contradictory, it is important to note that the relationship between quality and outcomes is consistent and that the measure of engagement varied in the two years—observed engagement in 2011–12 and youth-reported cognitive engagement in 2014–15. The evaluation team hypothesizes that it is the presence of SPP programming that accounts for this discrepancy. Although in both cases the team observed academic content in programming, the nature of how academics are presented in ACE activities is substantively different from that in SPP programs. ACE academic enrichment is typically project-based, hands-on learning, such as learning mathematics skills through a cooking class. SPP programming consisted of more traditional subject matter instruction, such as completing mathematics problems, where content was explicit and the focus of the activities was to improve STAAR performance. The SPP programs, therefore, were still high quality but may have been less engaging than academically oriented ACE programs observed in 2011–12.

In sum, the five-year evaluation has demonstrated that the theory of change holds value and warrants further exploration through future evaluations of ACE or of other similar programs across the country. Although the evaluation team was not able to prove the mediating role of youth mindsets and behaviors because of limitations in the data, and the introduction of explicit academic instruction via SPP programming complicated the path, there does appear to be relationships among program quality, engagement, and student outcomes for youth participating in ACE programming at high levels.

Recommendations

Based on findings from the specific 2014–15 evaluation and the overarching five-year evaluation, AIR recommends the following key next steps for the ACE program and the Texas Education Agency:

- ***Continue to support the growth and development of ACE programs across the state.*** The ACE program as a whole is having an impact on student outcomes and may be a worthwhile investment that is helping schools in their ultimate mission to support student success.
- ***Continue to emphasize and support quality programming through continuous improvement.*** Program quality appears to play an important role in youth outcomes. Therefore, it stands to reason that providing support for continuous program improvement for the ACE centers is a critical step. It may be beneficial to provide training on specific practices outlined in the PQA related to creating a supportive environment, supporting strong interactions between adults and youth in programs, and engaging youth in activities.

- ***Encourage use of the Learning Strategies approach.*** The evaluation revealed that a *Learning Strategies* instructional approach may be more effective than the *Skills-Based* approach in engaging youth and contributing to youth outcomes. A *Learning Strategies* approach focuses more on learning how to learn and gathering skills applicable to several different content areas rather than learning specific skills associated with one content area. Although the evaluation looked at this approach only in SPP activities, TEA may consider working with its technical assistance provider to develop training related to this approach for all of its ACE program staff.
- ***When considering intensive academic interventions, always couple them with traditional ACE enrichment.*** There was some evidence that an intensive academic intervention such as SPP can be successful within an OST setting when students experience it alongside traditional ACE programming. SPP activities alone were associated with low levels of engagement and may not be effective on their own. However, when paired with ACE programming, intensive academic interventions appeared to have positive impacts on participants. There may be value to supporting academics in intensive sessions, but students show more positive results when they also have access to the fun and engaging enrichment opportunities that ACE provides.

References

Afterschool Alliance. (2014). *America after 3pm: Afterschool programs in demand*. Washington, DC: Author. Retrieved from http://afterschoolalliance.org/documents/AA3PM-2014/AA3PM_National_Report.pdf

Afterschool Alliance (2015). *21st Century Community Learning Centers: Providing afterschool and summer learning support to communities nationwide.* Washington, DC: Author. Retrieved from http://www.afterschoolalliance.org/documents/challenge2014/21stCCLCOverview_FINAL.pdf

Auger, A., Pierce, K. M., & Vandell, D. L. (2013, April). *Participation in out-of-school settings and student academic and behavioral outcomes.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences.* Mahwah, NJ: Erlbaum.

Center for Public Policy Priorities. (2013). *Investing in our future: 2013 state of Texas children.* Austin, TX: Author. Retrieved from http://www.forabettertexas.org/images/CPPP13_KC-databook-v27.pdf

Devaney, E., Naftzger, N., Liu, F., Levings, J., Cristobal, S., Vinson, M., et al. (2015). *Texas 21st Century Community Learning Centers: 2012–13 and 2013–14 combined evaluation report*. Washington, DC: American Institutes for Research.

Durik, A. M., Hulleman, C. S., & Harackiewicz, J. M. (2015). One size fits some: Instructional enhancements to promote interest. In K. A. Renninger & M. Nieswandt (Eds.), *Interest, the self, and K–16 mathematics and science learning*. Washington, DC: American Educational Research Association.

Durlak, J. A., & Weissberg, R. P. (2007). *The impact of afterschool programs that promote personal and social skills.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning. Retrieved from http://www.uwex.edu/ces/4h/afterschool/partnerships/documents/ASP-Full.pdf

The Equity and Excellence Commission. (2013). *For each and every child—A strategy for education equity and excellence*. Washington, DC: Author. Retrieved from https://www2.ed.gov/about/bdscomm/list/eec/equity-excellence-commission-report.pdf

Hill, C. J., Bloom, H. S., Black, A. R. and Lipsey, M. W. (2008), Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.

Hong, J., & Hong, Y. (2009). Reading instruction time and heterogeneous grouping in kindergarten: An application of marginal mean weighting through stratification. *Educational Evaluation and Policy Analysis*, *31*(1), 54–81.

Kane, T. J. (2004). *The impact of afterschool programs: Interpreting the results of four recent evaluations* (Working Paper). New York, NY: William T. Grant Foundation. Retrieved from http://www.issuelab.org/click/download1/impact_of_after_school_programs_interpreting_the_results_of_four_recent_evaluations

Kataoka, S., & Vandell, D. L. (2013). Quality of afterschool activities and relative change in adolescent functioning over two years. *Applied Developmental Science, 17*(3), 123–134.

Kauh, T. J. (2011). *AfterZone: Outcomes for youth participating in Providence's citywide after-school system*. Philadelphia, PA: Public Private/Ventures. Retrieved from http://www.wallacefoundation.org/knowledge-center/after-school/evaluations/documents/afterzone-outcomes-youth-participating-providences-citywide-after-school-system.pdf

Kreft, G. G., & De Leeuw, J. (1998). *Introducing multilevel modeling, number 1 of introducing statistical methods*. London: SAGE.

Linacre, J. M. (2005) *WINSTEPS* [Computer program]. Chicago, IL: Winsteps.com.

Linacre, J. M., & Wright, B. D. (2004). Construction of measures from Many-Facet data. In E. V. Smith, Jr., & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 296–321). Maple Grove, MN: JAM Press.

Miller, B. (2003). *Critical hours: Afterschool programs and educational success.* Boston, MA: Nellie Mae Education Foundation. Retrieved from https://www.nmefoundation.org/getmedia/08b6e87b-69ff-4865-b44e-ad42f2596381/Critical-Hours?ext=.pdf

Moroney, D. A., Marchand, J., Naftzger, N., Liu, F., Vinson, M., & Sparr, M. (2013). *Intermediate report for the statewide program evaluation of 21st CCLC grants: Rhode Island Department of Education*. Chicago, IL: American Institutes for Research.

Naftzger, N., Manzeske, D., Nistler, M., & Swanlund, A. (2012). *Texas 21st Century Community Learning Centers: Interim evaluation report.* Chicago, IL: American Institutes for Research. Retrieved from http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147506331&libID=2147506324

Naftzger, N., Nistler, M., Manzeske, D., Swanlund, A., Shields, J., Rapaport, A., et al. (2013). *Texas 21st Century Community Learning Centers: Year two evaluation report.* Naperville, IL: American Institutes for Research. Retrieved from http://tea.texas.gov/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147506331&libID=25769804085

Naftzger, N., Vinson, M., & Liu, F. (2013). *New Jersey 21st Century Community Learning Centers year 4 evaluation report.* Washington, DC: American Institutes for Research.

Naftzger, N., Vinson, M., Liu, F., Zhu, B., & Foley, K. (2014). *Washington 21st Century Community Learning Centers program evaluation: Year 2.* Chicago, IL: American Institutes for Research.

Naftzger, N., Vinson, D., Manzeske, D., & Gibbs, C. (2011). *New Jersey 21st Century Community Learning Centers (21st CCLC) impact report 2009–10.* Naperville, IL: American Institutes for Research.

Naftzger, N., Vinson, M., Swanlund, A., & Sparr, M. (2012). *An evaluation of the Washington 21st Century Community Learning Centers (21st CCLC) program: Year one*. Naperville, IL: American Institutes for Research.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: SAGE.

Schafer, J. L., & Kang, J. D. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated case study. *Psychological Methods 13(*4), 279–313.

Shernoff, D. J., & Vandell, D. L. (2007). Engagement in afterschool program activities: Quality of experience from the perspective of students. *Journal of Youth and Adolescence*, *36*, 891–903.

Texas Education Agency. (2016). *Enrollment in Texas public schools, 2014–15* (Document No. GE16 601 09). Austin, TX: Author.

Texas Education Code. (2015). Retrieved April 28, 2016, from http://www.statutes.legis.state.tx.us/

Vinson, M., Marchand, J., Sparr, M. & Moroney, D. (2013). *Rhode Island 21st Century Community Learning Centers program evaluation: Evaluation report 2011–12.* Chicago, IL: American Institutes for Research.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Appendices

Appendix A. ACE Program's Critical Success Model

Appendix B. Study Methods and Data Sources

Appendix C. Survey Constructs

Appendix D. Analysis Methods

Appendix E. RQ 2—Cost Analysis Study Detail

Appendix F. RQ 3—Impact Analyses Data Table Detail

Appendix G. Program Typology Development Process

Appendix A. ACE Program's Critical Success Model

Texas ACE Critical Success Factors

Outcomes	Critical Success Factors (behaviors)	Critical Success Factors Performance Indicators	Milestones (grantees)	Milestone Performance Indicators (measure)
	Critical Success Factor #1: Student and Family Engagement			
• Improve Academic Performance • Improve Attendance	• Students and families actively participating and engaged in learning • Students and families displaying leadership roles, volunteering to participate and lead activities	• Increased student and family attendance in afterschool programs • Students mentoring other students • Students and families facilitating activities Measurement Tool • Instructor surveys/self assessment • Principal/Project Director survey • Observation/on-site visit	• Utilize innovative instructional techniques for academic and enrichment activities based on research and best practices	• Activity Tracking - TX21st (Three times per year- Summer, Fall, & Spring) • Curriculum/Lesson Plans
	Critical Success Factor #2: School Involvement			
• Improve Behavior	• Students increased sense of involvement in school	• Number of students participating in extracurricular activities • Increased number of mentors Measurement Tool • Student/Family surveys • Teacher surveys	• Provide adult advocates, based on student need and in accordance with best practices	• Number of meetings with students • Number of contacts made with Families, teachers, school day staff
	Critical Success Factor #3: Assessment Data			
• Increase Promotion Rates	• Use of assessment data to revise/ reevaluate student services	• Changes in student activities following re-assessment Measurement Tool • Document analysis of program files • Observation/on-site visits	• Conduct ongoing/continuous assessment to determine need and improve targeted services	• Methods of assessment: pre/ post tests, needs assessments, case plans, etc. • Use of PRIME Assessment
	Critical Success Factor #4: Professional Development Impact			
• Increase Graduation Rates	• Implementation of strategies learned through training • Noticeable difference in educational instruction (teaching methods)	• Changes in methods of instruction based on training Measurement Tool • Self assessments • Supervisor assessments	• Provide all required training opportunities for staff development	• Number of trainings • Schedule of trainings • Staff sign in sheets • Participant surveys • MyTexasACE Training Report

Appendix B. Study Methods and Data Sources

The evaluation methods for the 2014–15 study included an analysis of a variety of data collected primarily during the 2013–14 school year. These data include a combination of Texas administrative data, student and activity leader surveys, and site visits. The methods are presented in detail in this appendix.

Data Collection Methods

The majority of data used to conduct analyses for this 2014–15 evaluation report were collected during the 2013–14 school year as part of the fourth year of the overall evaluation. Each source of data used in this year's evaluation is described in detail in the following sections.

Texas Extant Data. AIR obtained a substantial amount of information housed in TEA's administrative and assessment data systems to support the evaluation. These data include the following:

TX21st. TX21st is a web-based data collection system developed and maintained by TEA to report required data on the program to the U.S. Department of Education. Throughout the program year, TX21st collects data directly from grantees on a broad array of ACE program characteristics, student demographics, and programs and activities. Data extracted from the tracking system were used to explore levels of enrollment and attendance in ACE– and SPP–funded activities and to assign students to the ACE-only or SPP+ACE treatment groups.

Additional TEA Data. The participant and impact analyses described in this report included variables on student demographics, school-day discipline incidents, school-day attendance, and grade-level promotion from the Public Education Information Management System (PEIMS). Student achievement data from the annual STAAR assessments also were included. Some of these variables were used for outcome variables in the impact analyses. Others, including prior year outcome variables, were used to support the matching of SPP and ACE youth and participating ACE students and nonparticipating students for the impact analyses. School-level performance data from the Texas Academic Performance Report[48] also were used to support matching efforts

Youth Mindsets and Behaviors Survey. A consistent goal of the evaluation has been to explore further the SPP theory of change, showing that high levels of engagement in high-quality afterschool programs can contribute to changes in academic mindsets and behaviors that then lead to generalized school success. In 2012–13, the evaluation

[48] Prior to 2012–13, the Texas Academic Performance Report was known as the *Academic Excellence Indicator System*.

team developed a measure, based on existing valid and reliable surveys, to assess student functioning associated with academic mindsets and behaviors. The surveys included items to measure the following constructs: *Effort and Persistence*, *Learner Behaviors*, *Sense of Competence as a Learner, Sense of Competence in Mathematics, and Sense of Competence as a Reader*. Students in Grades 4–12 completed the Mindsets and Behaviors Survey themselves, online or on paper. The survey for these students included items on all of the constructs.

For the 2014–15 evaluation, the evaluation team used the results of the survey that was implemented in 2013–14, first in November 2013 and again in April and May 2014. Youth surveys were administered at all centers with SPP programming, although only to a sample of SPP and ACE-only program participants. Using a stratified, randomized sampling approach, based on data housed in TX21st, AIR selected 50 SPP participants to take the survey. The evaluation team calculated the proportion of students in each grade and for each gender that matched the overall distribution of SPP students. A sampling of those students then was selected. A sample of alternate SPP students also was identified in the event that a student from the original sample either was not enrolled or had left the SPP program prior to data being collected.

In addition to developing a sample of SPP student survey respondents, the evaluation team asked ACE center staff to identify up to 30 ACE students who were not enrolled in the SPP programming. The evaluation team provided numbers of students and criteria for the 30 ACE-only students who centers could select to take the survey (Grades 4–12) or have an activity leader or teacher fill out a survey on their behalf (Grades K–3). The evaluation team asked that all centers survey 30 ACE-only students, even if the center had fewer than the desired 50 SPP students. The evaluation team provided the distribution of students in each grade and gender that matched the distribution of the SPP students and that added up to 30 total ACE-only students. In addition to the distribution of students, the evaluation team also included academic criteria for selecting the ACE-only students. To determine these criteria, the evaluation team surveyed project directors to determine how they identify students for the SPP program. Each center received a letter that included selection criteria along with the numerical distribution of students to guide the selection of ACE-only students to take the survey.

Site Visits. The evaluation team also conducted site visits to inform the evaluation at one center per grantee agency that had received SPP grants. In addition, the evaluation team identified a matched sample of ACE-only centers within the same grantee when possible for site visits, resulting in a total of 26 site visits to 13 SPP centers and 13 ACE-only centers. Project director interviews were used to assess the extent to which SPP activities were being offered in a uniform manner across participating centers and to provide recommendations for centers with strong management and effective implementation of the SPP program. Center selection was based on these

recommendations and the desire to visit centers serving a range of elementary, middle, and high school students.

Site visits took place over two days and included interviews with activity leaders and site coordinators, observations of program sessions, and administration of a student survey to Grade 4–12 students who participated in the sessions that were observed.

Observation Measures. The evaluation teams observed three to five activities at each center for a total of 89 observations. Observations were guided by three instruments:

- **PQA.** The primary observation tool employed was the PQA,[49] an instrument developed by the High Scope Education Foundation and now administered by the David P. Weikart Center for Youth Program Quality (http://cypq.org/downloadpqa). The tool frequently is used to assess the quality of extended learning programs at the POS.
- **APT-O.** This tool, specifically the section on academic skill building, was used to measure the types of supports and opportunities afterschool staff can provide to support skill development in particular content areas. The APT-O (http://www.niost.org/apt) is a comprehensive observation tool developed by the National Institute on Out-of-School Time for the Massachusetts Department of Elementary and Secondary Education. The tool was designed to support state efforts to improve 21st CCLC. The APT-O was selected to supplement the PQA because it includes measures of skill building in reading and mathematics.
- **OCES.** The OCES (developed by the University of Virginia Social Development Lab, http://www.socialdevelopmentlab.org/resources/measures/oces/), an adaptation of the National Institute of Child Health and Human Development Early Child Care Research Network Classroom Observation Scale, was used to measure the extent to which youth participating in a given activity were engaged in program activities, attentive, self-reliant in performing activity tasks, and compliant with requests and directives from activity leaders. During the observations, observers applied ratings for each of the constructs on the three measures as well as maintained a narrative description of the observed, describing activities, materials and resources, teacher and student interactions, student interactions, and student engagement. The measure consists of five items: engagement, attention, self-reliance, compliance, and disruptive behavior. Each item was rated on a four-point Likert-type scale.

Youth Surveys. In addition to the observation measures, the evaluation team disseminated a youth engagement survey to students in Grades 4 and above at the conclusion of the activity that was being observed. The survey is similar to one

[49] As noted earlier, two versions of the PQA were used to support observations conducted at centers with SPP funding. The School-Age PQA was used in centers serving elementary students. The Youth PQA was employed in centers serving middle and high school students.

employed by Shernoff and Vandell (2007) and includes eight questions on students' concentration, enjoyment, and interest during the session (see Appendix C for the specific survey items). The survey assesses a cognitive, as opposed to a behavioral, definition of engagement and relies on self-report rather than an observer scanning for and recording engagement levels of participating youth. In total, 401 student surveys were collected across the 89 activities that were observed.

Appendix C. Survey Constructs

Student Engagement Survey Construct

Table C-1. Student Engagement Survey Items (2013–14)

Student Engagement Survey Items
Thinking about today's activity that just ended, please answer the following questions: Was today's activity interesting? Was this activity important to you? Did you really have to focus to do the activity? Did you enjoy what you were doing during this activity? Was it easy to pay attention during today's activity? Was the activity something you were good at doing? Did you wish you were doing something else? Did you feel like you had a say in what you did during the activity?
Response Options: not at all, sort of, and very much

Academic Mindsets and Behaviors Survey Constructs

Table C-2. Grades 4–12 Survey Items Related to Effort and Persistence and Learner Behaviors (2013–14)

***Effort and Persistence* (6 Items)**	***Learner Behaviors* (7 Items)**
I try to do my schoolwork even when it looks really hard. When I am taught something that doesn't make sense, I keep trying to figure it out. I keep trying to do my schoolwork even if it is hard. I work really hard in school. I don't give up on my schoolwork even when I am frustrated. I try harder when I don't understand.	Before I start my schoolwork, I make sure I have all the things I need. I use my time in class to do my work and keep up with the rest of the class. I usually take part in what we do in class. When I'm in class, I think about what we are working on. I listen carefully in class. I am interested in the things we work on in class. I think most of my classes are fun.
Response Options: not at all like me, sort of like me, and a lot like me	

Table C-3. Survey Items Related to Sense of Competence as a Learner (2013–14)

Sense of Competence as a Learner **(5 Items)**	***Sense of Competence as a Reader*** **(5 Items)**
I like to give new things a try, even if they look hard. In school, I'm as good as other kids. I'm good at solving problems. I'm as good as other kids my age at learning new things. When I can't learn something right away, I keep trying until I get it.	I like to read at home during my free time. I enjoy reading when I'm at school. I enjoy reading when I'm at this afterschool program. I'm good at reading. I like to give new books a try, even if they look hard.
Sense of Competence in Mathematics **(6 Items)**	
I like to learn new things in math. I like to do math when I'm at school. I like to do math when I'm at this afterschool program. Math is something I'm good at. I'm interested in math. I like to give new math problems a try, even when they look hard.	
Response Options: not at all like me, sort of like me, and a lot like me	

Appendix D. Analysis Methods

Rasch Analysis on Survey Responses

At its most basic level, the use of Rasch modeling techniques yields estimates of an individual respondent's ability and the relative difficulty of a given item on the instrument in question (Bond & Fox, 2007). Working from the proposition that persons with greater ability will have a greater likelihood of successfully completing a given bank of test items (or find it easier to endorse survey items that demonstrate greater ability) than will less skilled persons, Rasch modeling techniques take person and item difficulty estimates yielded from an instrument, transform them by using a log function, and display them on a logit scale that allows person and item difficulties to be compared directly.[50]

One of the benefits of using Rasch approaches is that they result in true interval-level scores that can be used when conducting analyses. In order to create true interval measures that could be employed effectively in supporting the domain of analyses needed for the report, we employed Rasch analysis techniques to create scale scores for scales associated with several instruments used to support data collection efforts during the 2013–14 school year. Three different Rasch models were employed in this undertaking.

1. **Rasch Rating Scale Model (Linacre, 2005).** This model was used to calibrate scales appearing on the student surveys and took the following form:

   $$\text{Log}(P_{nix} / P_{ni(x-1)}) = B_n - (D_i + R_x)$$

   where

   P_{nix} = the probability of person n of ability B_n being observed in category x of item i with difficulty D_i

   $P_{ni(x-1)}$ = the probability of person n of ability B_n being observed in category $x - 1$ of item i with difficulty D_i

   B_n = the ability of respondent n

   D_i = the difficulty of item i

   R_x = rating scale structure parameter for category x (indicates how much of the latent construction is covered by a given response category of the rating scale)

[50] Item difficulty reflects how positively an item is endorsed. Items with low item difficulty will be frequently and positively endorsed (e.g., a high frequency of "Strongly Agree").

2. **Rasch Dichotomous Model (Wright & Masters, 1982).** This model was used to calibrate scales appearing on the APT-O and took the following form:

$$\text{Log}(P_{ni} / (1 - P_{ni})) = B_n - D_i$$

where

P_{ni} = the probability of activity *n* succeeding on item *i*

B_n = the ability of activity *n*

D_i = the difficulty of item *i*

3. **Many-Facet Rasch Measurement (Linacre & Wright, 2004).**– This model was used to calibrate scales appearing on the following observation instruments:

   a. PQA (both the Youth and School-Age versions of this instrument)
   b. OCES

   The many-facet Rasch measurement model employed in calibrating measures on the aforementioned instruments took the following form:

$$\text{Log}(P_{nijk} / P_{nij(k-1)}) = B_n - D_i - C_j - F_k$$

where

P_{nijk} = the probability of activity *n* being given a rating of *k* on item *i* by rater *j*

$P_{nij(k-1)}$ = the probability of activity *n* being given a rating of *k* - 1 on item *i* by rater *j*

B_n = the ability of activity *n*

D_i = the difficulty of item *i*

C_j = the severity of rater *j*

F_k = the difficulty of category *k* relative to category *k* -1

Analyses Used in RQ 1

PQA Subscale Development

For the analyses described in this report, the evaluation team created subscale scores to explore better whether specific types of PQA practices were found to be related especially to youth engagement. In undertaking analyses to create subscale scores, some subscales demonstrated poor reliability, given a lack of variation in scores across activities (almost all activities scored highly on such subscales) and were therefore

dropped from further analyses. Other subscales with poor reliability but good variability in scores were combined with other subscales to construct a usable scale. Ultimately, six subscale scores were calibrated from PQA data, as shown in Table D-1.

Table D-1. PQA Subscales

New Subscale Name	Old Subscales That Make Up New Subscale
Activity Characteristics That Support Active Engagement	Activities support active engagement (*Supportive Environment* subscale H).
Encouragement and Feedback	Staff support youth in building new skills (*Supportive Environment* subscale I). Staff support youth with encouragement (*Supportive Environment* subscale J).
Cooperative Learning and Belonging	Youth have opportunities to develop a sense of belonging (*Interaction* subscale L). Youth have opportunities to collaborate (*Interaction* subscale M). Youth have opportunities to act as group facilitators and mentors (*Interaction* subscale N).
Positive Adult Interactions	Youth have opportunities to partner with adults (*Interaction* subscale O).
Planning and Choice	Youth have opportunities to set goals and make plans (*Engagement* subscale P). Youth have opportunities to make choices based on their interests (E*ngagement* subscale Q).
Reflection	Youth have opportunities to reflect (*Engagement* subscale R).

As in analyses conducted previously on observation scores for this evaluation, scores were placed on a scale from 0 to 100. The mean of these scales was then taken to create a total PQA score. Average scale scores for activities observed in SPP and ACE-only centers are outlined in Table D-2. Although some differences in the average PQA scores between activities in SPP and ACE-only centers are noticeable, none were significant.

Table D-2. Average PQA Scale Scores by Subscale and Center Type (2013–14)

PQA Subscales	Activities in SPP Centers (*n* = 45)	Activities in ACE-Only Centers (*n* = 44)
Activity Characteristics That Support Active Engagement	65.069	64.030
Encouragement and Feedback	67.051	61.977
Cooperative Learning and Belonging	62.590	58.301
Positive Adult Interactions	59.504	56.707
Planning and Choice	40.678	48.634
Reflection	46.722	45.360
Total PQA Score	54.828	54.147

SOURCE: Texas 21st CCLC Evaluation, PQA Observation Scores, 2013–14.

Developing Higher and Lower Quality Program Designation

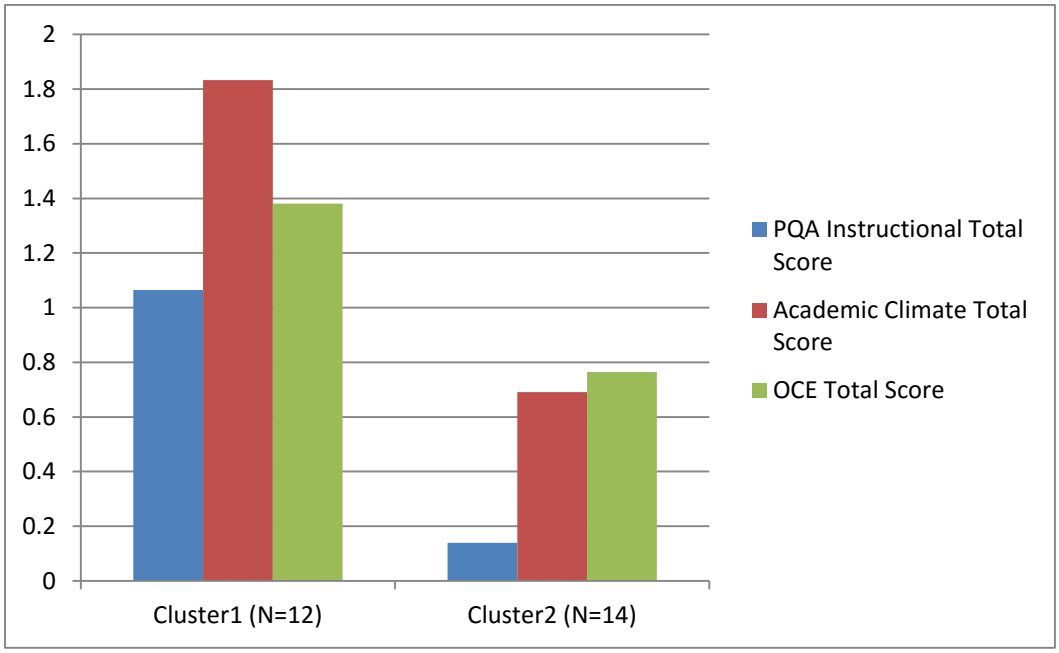
As noted earlier, during the spring semester of 2014, members of the evaluation team conducted two-day site visits at 26 centers—13 ACE centers with SPP funding and 13 ACE-only centers. During these visits, a total of 89 OST activities serving youth in Grades K–12 were observed by members of the evaluation team.

At the conclusion of each activity, the PQA (including the academic climate scale) was scored to obtain an estimate of activity-level quality. In addition, observers scored the OCES to obtain an estimate of student engagement as well as distributed a youth engagement survey for activities serving students in Grades 4–12. Nineteen of the 89 activities (approximately 21%) were observed by more than one member of the evaluation team. This method allowed the evaluation team to calibrate PQA scores in a way in which systematic observer bias (some raters are inherently more severe in scoring the instruments, and others are inherently more lenient) could be quantified and adjusted for when determining a final activity score.

Next, the evaluation team created higher and lower quality profiles of participating centers using a cluster analysis approach. Variables related to program instructional quality (PQA), academic climate (from the academic portion of the PQA), and youth engagement (OCES) were used in the cluster analysis. Data from the observation level were first analyzed using the Rasch analysis technique, which produced logit scores for the variables in question. These logit scores were used as the variables of interest in the cluster analysis. Two- and three-cluster solutions were produced using the k-means clustering (with Ward method, standardizing variables, and using squared Euclidian distance) first by taking a look at overall total scores from the PQA, academic climate scale, and the OCES. Second, cluster solutions were produced using the subscales on these measures, where applicable. Findings were inserted into Excel to examine the

graphics and output. Researchers determined that the two-cluster solution using the broader total scores on the PQA, academic climate scale, and the OCES produced the most telling picture in terms of low- and high-quality profiles of centers. Significant differences were found between each of the two groups on the mean scale score for each of the instruments included in the analysis. Figure D-1 shows the differences between the two clusters. As a result, 14 of the 26 observed centers fell within the low-quality profile, while the remaining 12 centers fell within the high-quality profile.

Figure D-1. Differences Between Higher and Lower Quality Clusters on Key Indicators, 2013–14



Conducting Mediation Analyses

In order to explore the relationships among all of the components of the theory of change (i.e., RQs 1.3 and 1.4), the evaluation team used mediation analysis techniques. A mediation analysis explores how an independent variable affects a dependent variable through one or more intermediate variables. The evaluation team employed the path analysis model for the mediation investigation. The path analysis is a special case of structural equation modeling (SEM) with multiple regressions employed in the same path model to investigate both direct effects of the independent variable(s) on dependent variable(s) and indirect effects between the two types of variables via the mediator(s). The diagrams below show the mediation analysis “path” and the variables that the evaluation team used to explore the mediating effects of engagement (RQ 1.3) and mindsets and behaviors (RQ 1.4).

Figure D-2. Path Analysis Variables for RQ 1.3

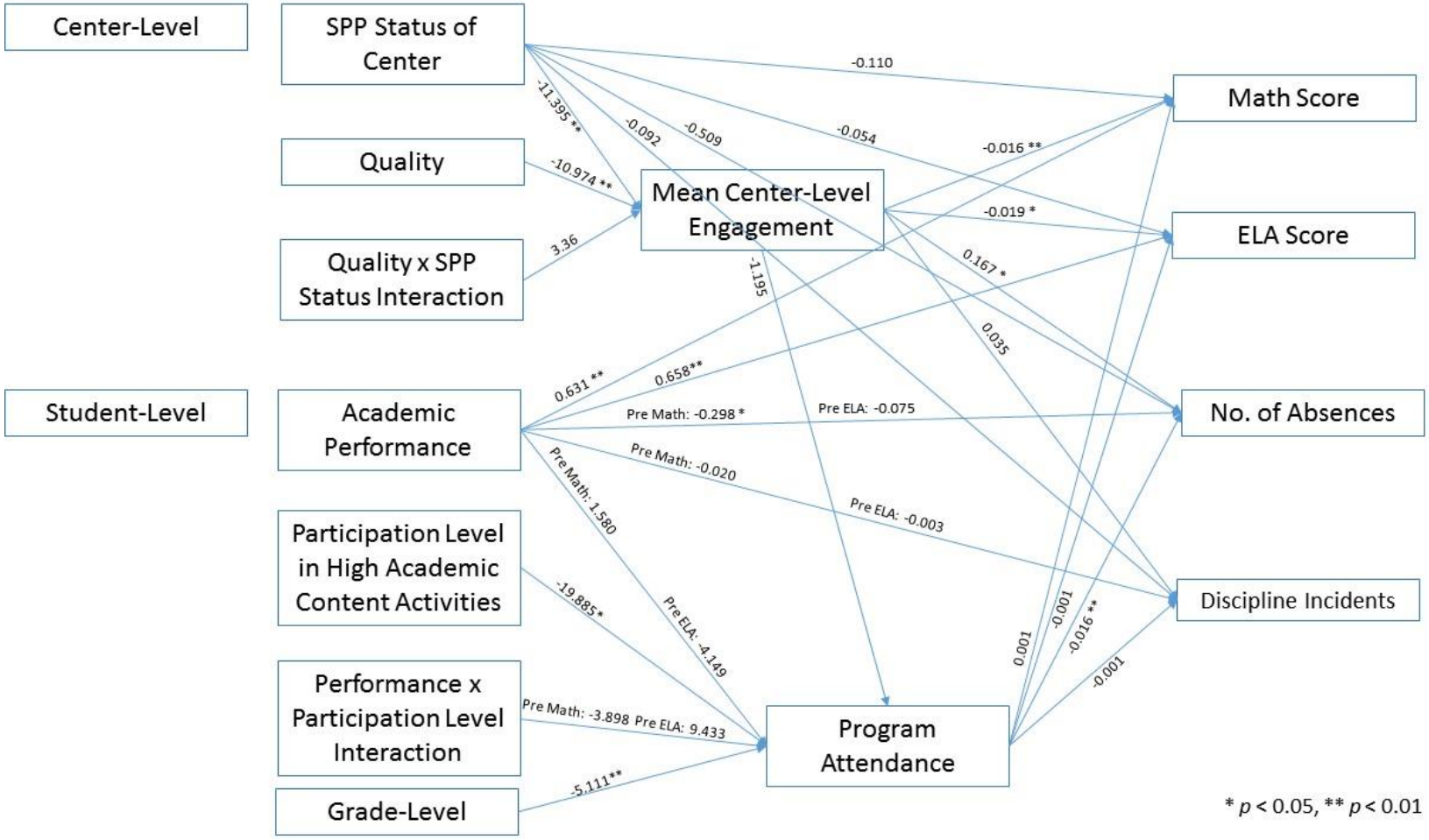
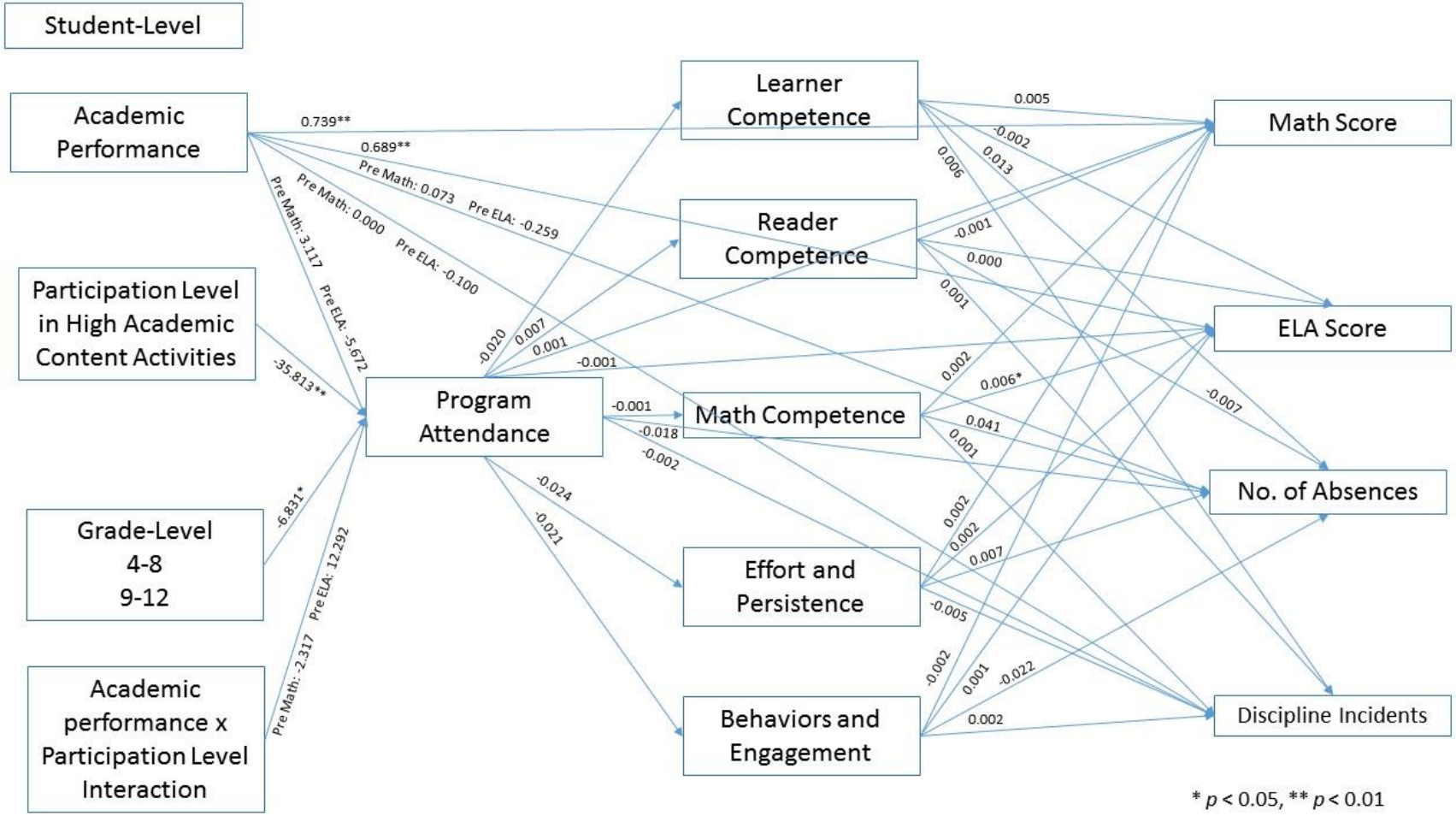


Figure D-3. Path Analysis Variables for RQ 1.4



Impact Analyses

Propensity Score Matching and Hierarchical Linear Modeling

The evaluation used the propensity score matching (PSM) and Hierarchical Linear Modeling (HLM) approaches for many of the research questions outlined in this report. PSM approach is a statistical technique designed to mitigate any selection bias that may occur because the programs and activities in question were not randomly assigned. HLM is a process used to account for the nested structure of data. Both methods are described here.

In any evaluation of a program where participants are not randomly assigned to participate, the problem of selection is paramount. We know that it is likely that students who participate in 21st CCLC programming are different from those who do not attend. Likewise, students who participate in SPP programs are likely different from those not participating, students participating in SPP+ACE are different from ACE-only students, and so on. These differences can bias estimates of program effectiveness because they make it difficult to disentangle preexisting differences between students who attended one type of programming and those who did not from the effect of attending the program. In general, we found that students who attended ACE programming tended to be lower achieving students than those who did not, prior to the start of the current academic year, and those attending SPP programming tended to be lower achieving than ACE-only students. The quasi-experimental approach outlined here is a method for mitigating that existing bias in program effect (i.e., if one were to simply compare the students who attended and those who did not).

PSM is a two-stage process designed to address this problem. In the first stage, the probability that each student participates was modeled on available observable characteristics. By modeling selection into the program, this approach allowed us to compare participating and nonparticipating students who would have had a similar propensity to select into the program based on observable characteristics that were available in the data received from TEA. In the second stage, the predicted probability of participation was used to model student outcomes while accounting for selection bias using a hierarchical linear model (HLM) approach. We balanced pretreatment group differences in observed covariates using a propensity score stratification and marginal mean weighting approach (Hong & Hong, 2009).

Stage 1: Creation of the Comparison Group. The outcome of interest in modeling propensity scores is treatment status (1 for students in the treatment group, 0 for the comparison group). To account for this binary outcome, logistic regression was used to model the logit (or log-odds) of student group assignment status. Examples of student-level variables used to fit the propensity score models included:

- Prior achievement in reading and mathematics

- Prior measures for other outcomes (grade-level promotion, behavior and attendance)
- Student demographic information, including:
  - Gender
  - Ethnicity
  - Socioeconomic status
  - At-risk status
  - English language proficiency
  - Special education status

In addition to the student-level variables, the propensity score model also included school-level variables shown as the following:[51]

- School type
- Total enrollment
- Student race/ethnicity composition
- School locale
- Campus rating
- Number of students identified as economically disadvantaged
- Number of English language learners
- Number of special education students

A total of 39 variables were considered for the propensity score model. Data were not available for each of these covariates for all students. To account for this, indicator variables were used to model the relationship between the pattern of missing data and propensity to participate in the program (Rosenbaum & Rubin, 1984). The propensity score model was fit separately for each grade (Grades 4–11) and separately for each definition of treatment (30+ days; 60+ days). The final propensity score models for each grade were checked to ensure that the analysis sample was balanced across relevant covariates. The propensity score models all produced comparison samples that were balanced with the treatment across 39 variables examined for balance.

Stage 2: Statistical Modeling of Student Outcomes. Outcomes of students in the treatment group were then were compared with the outcomes of comparison group

[51] For school-level variables, the evaluation team used the school that the majority of ACE participants at a given program attended. In most cases, a center that was based at a specific school drew the majority of its participants from that school and the evaluation team used the demographics and other characteristics of that school in the PSM model.

students. We balanced pretreatment group differences in observed covariates by using a propensity score stratification and marginal mean weighting approach (Hong & Hong, 2009). Various strata were used, based on the spread and overlap of the data. The propensity score logit along with the pretreatment measure of the outcome also were included in the outcome model to control for within-strata differences and residual bias (Schafer & Kang, 2008). Student outcomes were then modeled using two-level hierarchical linear models to account for the nested nature of the data (students within schools) as follows:

Level 1—Students

$$y_{ij} = \beta_{0j} + \beta_{1j}\ Participation_{ij} + \sum_{S=2}^{15} \beta_s L_{sij} + \beta_{16j} LP_{ij} + \beta_{17j} Pretest_{ij} + r_{ij}$$

where y_{ij} is a student-level outcome (e.g., student mathematics achievement), $Participation_{ij}$ is an indicator of whether the student participated in the SPP program, L_{sij} is an indicator variable for each of the logit propensity score strata, LP_{ij} is the logit propensity score, and $Pretest_{ij}$ is the pretreatment measure of the outcome. Subscripts *i*, *j*, and *s* correspond to student, school, and strata, respectively.

Level 2—Center

$$\beta 0j \quad = \gamma_{00} + u_{0j}$$

The Level 2 equation includes only β_{0j} because the chosen hierarchical linear model is a random intercept model; all other coefficients (i.e., participation indicator, logit propensity score stratum, logit propensity score, and pretreatment indicator) at Level 1 are fixed and, therefore, not listed at Level 2. Because the treatment and comparison groups were matched using all of the covariates described earlier, it is not necessary to include these variables in the final outcome model.

The two-level model of correlation between program participation and student performance (written in mixed model format) is as follows:

$$Y_{ij} = \gamma_{00} + \beta_{1j} Participation_{ij} + \sum_{S=2}^{15} \beta_s L_{sij} + \beta_{16j} LP_{ij} + \beta_{17j} Pretest_{ij} + u_{0j} + r_{ij}$$

where

- Y_{ij} is the performance of student i in school j;
- β_0 is a constant term showing average student performance in the comparison group;

- $StudentCovariate_{ij}$ is a vector of student-level covariates, including their prior performance on the outcome of interest, gender, ethnicity, economically disadvantaged status, limited English proficiency status, number of years enrolled in the program, and total days of SPP program participation;
- β_1 is a vector of coefficients associated with each of those covariates showing the association of each student-level characteristic and the outcome;
- $Treatment_{ij}$ is the treatment status for student i in center j;
- β_2 shows the average difference in performance between treatment group and comparison group;
- u_j is a school-level random error term, with an assumed normal distribution with mean zero and variance τ; and
- r_{ij} is a student-level error term, also assumed to have a normal distribution with mean zero and variance σ^2.

Appendix E. RQ 2—Cost Analysis Study Detail

The models used to estimate the association between student performance and program spending are two-level HLMs to account for the nested nature of the data (students within schools).[52] The full output for these models and the control variables used are provided below.

In order to examine the association between per-student program expenditures and student performance, the evaluation team used an HLM to predict student performance, accounting for program expenditure measured at the school level. The model controls for individual student characteristics, and the effect sizes for these controls are included in this appendix. The likelihood ratio test justifying the multilevel model shows strong evidence that the between-school variance was nonzero and the second-level variance explained by schools in the models was substantial for the Grades 4–8 model (.349 variance and .09 interclass correlation for mathematics; .247 and .06 interclass correlation for reading).[53] This means that 9% of the variance in the model for mathematics is explained by the school-level units.

Table E-1. Association Between SPP Program Total per Student Expenditures and Student Performance on STARR and EOC Mathematics and Reading (2013–14)

Variable	Mathematics Met Standard, Grades 4–8	Mathematics Met Standard, Grade 9	Reading Met Standard, Grades 4–8	Mathematics Met Standard, Grade 9
	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]
Per-student program expenditures	1.000	1.000	1.000	0.998
	0.000	0.001	0.000	0.001*
Ethnicity				
Asian	0.670		0.643	
	0.742		0.671	

[52] These are a form of generalized linear mixed models that adjust for nonindependence among students and allow for random effects across schools.

[53] The likelihood ratio test is used in lieu of the Wald test for testing for the significance of the Level 2 random effect (school effect) and helping justify the use of an HLM (Raudenbush & Bryk, 2002). It is computed by comparing the deviances of the Level 1 standard logistic regression model against the two-level HLM. The difference in the deviances is distributed chi-squared with a single degree of freedom under which the null hypothesis is that the variance of the Level 2 intercept components is zero, which we were able to reject here. Next, the ICCs, or total variance attributable to Level 2 average passing rates, were calculated to estimate the degree of nonindependence in the student performance across Level 1 units (students). Importantly, the magnitude of the ICCs calculated here are ample justification for the use of multilevel modeling (see Kreft & de Leeuw 1998; further, any nonzero (particularly >.05) ICC can invalidate the hypotheses and confidence intervals if multilevel models were not used), and the postulatory justification for these models is the fact that the students are nested into schools and the independence assumptions of OLS are violated.

Variable	Mathematics Met Standard, Grades 4–8	Mathematics Met Standard, Grade 9	Reading Met Standard, Grades 4–8	Mathematics Met Standard, Grade 9
	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]
Black	0.459	1.303	0.238***	0.486
	0.197	1.948	0.100	0.764
Hispanic	0.851	0.470	0.346**	0.680
	0.313	0.580	0.125	0.957
Other	1.001		0.301	
	0.781		0.235	
School-Day Attendance	1.032	1.104	1.036	1.102
	0.032	0.112	0.031	0.089
Female	0.773	0.799	1.760**	2.263
	0.134	0.536	0.305	1.847
Number of School-Day Discipline Incidents	0.838	0.723	0.916	0.950
	0.085	0.193	0.072	0.235
Number of Students	0.999	0.998	1.001	1.000
	0.001	0.002	0.001	0.002
Grade Level				
5	0.756		0.859	
	0.229		0.260	
6	0.614		0.309*	
	0.300		0.143	
7	1.016		1.430	
	0.485		0.612	
8	3.014*		1.267	
	1.484		0.556	
Passed Prior Year	10.989***	4.478*	11.813***	11.289**
	2.558	2.913	2.700	9.033
Constant	0.077	0.001	0.028	0.000
	0.229	0.008	0.084	0.001
Statistics				
N	760	69	768	74
Log Likelihood	-419.155	-32.818	-422.267	-25.972
McKelvey & Zavoina R^2	0.406	0.468	0.416	0.471
Level 2 Variance	0.349	0.000	0.247	0.000

Notes: *p<.05 **p<.01 ***p<.001

[1] SE stands for standard error. For each variable, the odds ratios are presented above the standard errors.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14.

Table E-2. Association Between SPP Program Computer-Based and Face-to-Face Instruction per-Student Expenditures and Student Performance on STARR and EOC Mathematics and Reading (2013–14)

Variable	Mathematics Met Standard, Grades 4–8	Mathematics Met Standard, Grade 9	Reading Met Standard, Grades 4–8	Reading Met Standard, Grade 9
	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]
Computer-Based instruction per-student expenditures	1.000	1.001	0.998*	0.991
	0.000	0.008	0.001	0.008
Face-to-Face instruction per-student expenditures	1.000	1.000	1.000	0.999
	0.000	0.003	0.000	0.003
Ethnicity				
Asian	0.674		0.634	
	0.746		0.659	
Black	0.453	1.235	0.243***	0.548
	0.195	1.910	0.101	0.894
Hispanic	0.870	0.477	0.358**	0.606
	0.317	0.589	0.128	0.894
Other	1.009		0.322	
	0.786		0.245	
Attendance	1.032	1.103	1.033	1.102
	0.032	0.112	0.031	0.091
Female	0.772	0.804	1.760**	2.272
	0.134	0.541	0.305	1.865
Number of School-Day Discipline Incidents	0.837	0.727	0.913	0.927
	0.085	0.196	0.072	0.235
Number of Students	0.999	0.998	1.001	1.002
	0.001	0.002	0.001	0.002
Grade Level				
5	0.724		0.798	
	0.221		0.237	
6	0.593		0.303**	
	0.284		0.136	
7	0.984		1.487	
	0.461		0.619	
8	2.917*		1.287	
	1.409		0.555	

Variable	Mathematics Met Standard, Grades 4–8	Mathematics Met Standard, Grade 9	Reading Met Standard, Grades 4–8	Reading Met Standard, Grade 9
	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]	Odds Ratio/SE[1]
Passed Prior Year	11.009***	4.640*	12.350***	8.898*
	2.560	3.254	2.821	7.648
Constant	0.077	0.001	0.048	0.000
	0.232	0.009	0.141	0.001
Statistics				
N	760	69	768	74
Log Likelihood	-418.590	-32.809	-419.756	-25.756
McKelvey & Zavoina R^2	0.408	0.480	0.455	0.488
Level 2 Variance	0.322	0.000	0.213	0.000

Notes: *p<.05 **p<.01 ***p<.001

[1] SE stands for standard error. For each variable, the odds ratios are presented above the standard errors.

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14; State of Texas Assessments of Academic Readiness, 2012–13 and 2013–14.

Appendix F. RQ 3—Impact Analyses Data Table Detail

Table F-1. Effect of ACE Program Participation on School-Day Discipline Incidents for 30+ and 60+ Day Attendees, Relative to Nonparticipants (2013–14)

Grade Level[1]	Number of School-Day Discipline Incidents—30+ Day Attendees				Number of School-Day Discipline Incidents—60+ Day Attendees			
	Mean Change	Effect Size	S.E.[2]	*p*	Mean Change	Effect Size	S.E.[2]	*p*
Grade 4	-2%	-0.023	0.029	0.430	-13%	-0.143	0.035	< 0.001***
Grade 5	-3%	-0.030	0.024	0.210	-10%	-0.106	0.029	< 0.001***
Grade 6	-7%	-0.067	0.013	< 0.001***	-21%	-0.235	0.018	< 0.001***
Grade 7	-2%	-0.021	0.012	0.088*	-17%	-0.192	0.018	< 0.001***
Grade 8	3%	0.032	0.014	0.020**	-3%	-0.035	0.020	0.075*
Grade 9	-11%	-0.121	0.020	< 0.001***	-27%	-0.314	0.036	< 0.001***
Grade 10	-12%	-0.123	0.024	< 0.001***	-15%	-0.164	0.040	< 0.001***
Grade 11	-21%	-0.230	0.033	< 0.001***	-27%	-0.321	0.057	< 0.001***

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Grade 12 was excluded from analysis because of the small sample size.

[2] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

Table F-2. Effect of ACE Program Participation on School-Day Absences for 30+ and 60+ Day Attendees, Relative to Nonparticipants (2013–14)

Grade Level[1]	Number of School-Day Absences—30+ Day Attendees				Number of School-Day Absences—60+ Day Attendees			
	Mean Change	Effect Size	S.E.[2]	*p*	Mean Change	Effect Size	S.E.[2]	*p*
Grade 4	-12%	-0.128	0.005	< 0.001***	-15%	-0.162	0.006	< 0.001***
Grade 5	-11%	-0.113	0.005	< 0.001***	-14%	-0.155	0.006	< 0.001***
Grade 6	-16%	-0.179	0.005	< 0.001***	-22%	-0.253	0.007	< 0.001***
Grade 7	-14%	-0.148	0.005	< 0.001***	-22%	-0.245	0.007	< 0.001***
Grade 8	-17%	-0.182	0.005	< 0.001***	-24%	-0.274	0.008	< 0.001***
Grade 9	-14%	-0.150	0.006	< 0.001***	-26%	-0.297	0.011	< 0.001***
Grade 10	-13%	-0.135	0.006	< 0.001***	-16%	-0.170	0.010	< 0.001***
Grade 11	-14%	-0.151	0.006	< 0.001***	-21%	-0.236	0.011	< 0.001***

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Grade 12 was excluded from analysis because of the small sample size.

[2] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

Table F-3. Effect of ACE Program Participation on Grade-Level Promotion for 30+ and 60+ Day Attendees, Relative to Nonparticipants (2013–14)

Grade Level[1]	Grade-Level Promotion—(30+ Day Attendees				Grade-Level Promotion—60+ Day Attendees			
	Odds Ratio	Effect Size	S.E.[2]	*p*	Odds Ratio	Effect Size	S.E.[2]	*p*
Grade 4	2%	0.016	0.104	0.878	12%	0.117	0.121	0.334
Grade 5	34%	0.291	0.100	0.003**	37%	0.317	0.119	0.008**
Grade 6	11%	0.100	0.133	0.453	6%	0.060	0.176	0.735
Grade 7	35%	0.300	0.136	0.027**	42%	0.348	0.202	0.085*
Grade 8	0%	-0.004	0.150	0.977	23%	0.211	0.226	0.350
Grade 9	48%	0.394	0.125	0.002**	36%	0.308	0.072	< 0.001***
Grade 10	20%	0.184	0.153	0.230	32%	0.278	0.095	0.003**
Grade 11	62%	0.485	0.208	0.020**	55%	0.437	0.001	< 0.001***

***Statistically significant at the .001 level; **Statistically significant at the .05 level; *Statistically significant at the .10 level

[1] Grade 12 was excluded from analysis because of the small sample size.

[2] Standard error

SOURCE: Public Education Information Management System data, 2013–14; TX21st CCLC Student Tracking System, 2013–14.

Appendix G. Program Typology Development Process

The SPP typologies were developed as part of the 2012–13 evaluation based on data collected from one-day site visits. During these site visits, evaluation team members observed 22 activities at 15 centers representing each of the 15 grantees that received SPP funding. The typologies were based on program activities—how they were designed—rather than on other features, such as the grade levels of students served or the subject area. The SPP theory of change indicates that it is essential for activities to be designed with core implementation factors in mind in order to provide students with a set of critical experiences. Thus, the SPP typologies were created based on the types of experiences SPP program activities offered students. Only SPP programs were included in the typology analysis in order to better understand what was happening in those specific activities. The evaluation team identified two categories of typologies that indicate how students might experience the activities (both with respect to what they do and what they learn). The first typology is associated with mode of delivery—the general structure of the sessions that directs what students do. The second typology is associated with the overall curricular approach of the activities—what students learn.

Mode of Delivery Typology

Two modes of delivery were identified: Computer-Based delivery and Face-to-Face delivery. The majority of SPP programs used a combination of Computer-Based and Face-to-Face delivery modes. The three categories can be defined as follows:

- **Combined Mode (Computer-Based Delivery and Face-to-Face Delivery).** Nineteen of the 38 SPP centers included in this analysis used both types of formats within the same activity. In these SPP centers, either some activities were Computer-Based and other activities were Face-to-Face, or an instructor actively facilitated Computer-Based sessions by directing student activities and providing consistent coaching and feedback.
- **Face-to-Face Delivery Only.** Ten of the 38 SPP centers included in the analysis used only Face-to-Face interventions and did not use Computer-Based learning programs.
- **Computer-Based Delivery Only.** Nine of the 38 centers provided interventions only through Computer-Based programs, with minimal or no activity leader facilitation.

Approach Typology

In addition to the mode of delivery for the activities, the SPP programs demonstrated two different approaches for improving students' academic performance. Those approaches were as follows:

- A ***Learning Strategies* approach**, in which specific curriculum guided the application of learning strategies that potentially could be applied to improve study skills broadly across various subject areas. This approach, although used in activities specifically focused on one subject area (e.g., reading), emphasized teaching students a process for learning, which students could apply to different subject areas and learning experiences. For example, in one center, the curricula focused on the importance of metacognition, in which students think about their thinking and how they solve problems. In another, the curriculum emphasized connecting learning to real-world activities. Twelve of the 38 SPP centers included in this analysis were identified as using a *Learning Strategies* approach.
- A ***Skills-Building* approach**, in which a curriculum, usually a Computer-Based program, was used to develop specific subject-area skills, and/or small-group instruction focused on developing specific skills. The programs usually had an assessment component through which both students and activity leaders could track student progress. Twenty-six of the 38 SPP programs included in this analysis were identified as using a *Skills-Building* approach.