RAND Summer Learning Series

RAND CORPORATION

Every Summer Counts

A Longitudinal Analysis of Outcomes from the National Summer Learning Project

Jennifer Sloan McCombs, Catherine H. Augustine,
John F. Pane, Jonathan Schweig

Every Summer Counts

A Longitudinal Analysis of Outcomes from the National Summer Learning Project

Jennifer Sloan McCombs, Catherine H. Augustine, John F. Pane,
Jonathan Schweig

RAND CORPORATION

Commissioned by

Wallace

For more information on this publication, visit www.rand.org/t/RR3201

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-1-9774-0451-0

Published by the RAND Corporation, Santa Monica, Calif.

Front cover image: SDI Productions/Getty Images

Back cover image: fstop123/GettyImages

Image credits—pp. iv: SDI Productions/Getty Images; p. vi: Creatas Images/Getty Images; p. xxii: Monkey Business/Adobe Stock; p. 8: Kalani/Adobe Stock; p. 32: fstop123/Getty Images; p. 50: highwaystarz/Adobe Stock.

PREFACE

The National Summer Learning Project (NSLP) consisted of five school districts—Boston, Massachusetts; Dallas, Texas; Duval County, Florida; Pittsburgh, Pennsylvania; and Rochester, New York—and their local community partners. The NSLP study was launched to determine whether—and, if so, how—voluntary summer programs with both academics and enrichment can benefit students. The study spanned three phases. The research team from the RAND Corporation (1) collected formative data for strengthening the five summer programs in 2011 and 2012; (2) examined student outcomes after one summer (2013) and after two summers of programming (2014 and 2015); and (3) examined student outcomes in spring 2017, at the end of three school years after the second summer of programming. This report summarizes the findings of this third phase in the context of earlier findings and offers implications for policy and practice.

This report is the seventh in a series of reports based on findings from the NSLP. The previous six are as follows:

1. Catherine H. Augustine et al., *Getting to Work on Summer Learning: Recommended Practices for Success*, 1st ed., RR-366-WF, 2013
2. Jennifer Sloan McCombs et al., *Ready for Fall? Near-Term Effects of Voluntary Summer Learning Programs on Low-Income Students' Learning Opportunities and Outcomes*, RR-815-WF, 2014
3. Catherine H. Augustine et al., *Learning from Summer: Effects of Voluntary Summer Learning Programs on Low-Income Urban Youth*, RR-1557-WF, 2016
4. Catherine H. Augustine and Lindsey E. Thompson, *Making Summer Last: Integrating Summer Programming into Core District Priorities and Operations*, RR-2038-WF, 2017
5. Heather L. Schwartz et al., *Getting to Work on Summer Learning: Recommended Practices for Success*, 2nd ed., RR-366-1-WF, 2018.
6. Catherine H. Augustine and Lindsey E. Thompson, *Getting Ready for Summer Learning: How Federal, State, City, and District Policies Affect Summer Learning Programs*, RR-2347-WF, 2020.

The first report, the first edition of *Getting to Work on Summer Learning* offers lessons learned from our detailed formative evaluations of the NSLP district programs in summer 2011. The second report, *Ready for Fall?*, describes how students in this study performed on mathematics, language arts, and social-emotional assessments in fall 2013, after one summer of programming. In the third report, *Learning from Summer*, we examined student outcomes at four different time points: in fall 2013, at the end of the 2013–2014 school year, in fall 2014 after the second summer of programming, and at the end of the 2014–2015 school year. The fourth report, *Making Summer Last*, describes how summer program leaders are integrating their programs into their districts' core priorities and operations as a quality improvement and sustainability strategy. The fifth report, the second edition of *Getting to Work on Summer Learning*, updates the first report and is based on lessons learned from our evaluation of the NSLP district programs in summers 2011–2014 and informed by our outcomes study. The sixth report, *Getting Support for Summer Learning*, examines the policies at the federal, state, and local levels that support or constrain the ability of districts to scale and sustain summer programs.

This research was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking.

The overarching study was commissioned by The Wallace Foundation, which seeks to support and share effective ideas and practices to foster improvements in learning and enrichment for disadvantaged children and the vitality of the arts for everyone. Its objectives are to improve the quality of schools, primarily by developing and placing effective principals in high-need schools, promoting social and emotional learning in elementary school and out-of-school-time settings, reimagining and expanding learning time both during the traditional school day and year and during the summer months, expanding access to arts learning, and developing audiences for the arts. For more information and research on these and other related topics, please visit the Foundation's Knowledge Center at www.wallacefoundation.org.

More information about RAND can be found at www.rand.org. Questions about this report should be directed to Jennifer McCombs (jennifer_mccombs@rand.org), and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

Contents

Figures

Tables

SUMMARY

Persistent achievement and opportunity gaps between students from low-income families and their peers from higher-income families widen during the summer months when school is out. Although a contributor to inequality, summer is also an opportune time to provide activities, interventions, and programs that promote positive student outcomes, such as academic achievement and access to enriching activities. A recent evidence review of rigorous evaluations of summer programs identified several programs that succeeded in benefiting children and youth academically, socially, emotionally, and in terms of summer employment and career knowledge, aspirations, and skills (McCombs, Augustine, Unlu, et al., 2019).

Although a contributor to inequality, summer is also an opportune time to provide activities, interventions, and programs that promote positive student outcomes, such as academic achievement and access to enriching activities.

One of these successful programs is the National Summer Learning Project (NSLP), which began in 2011 when The Wallace Foundation selected five school districts—Boston, Massachusetts; Dallas, Texas; Duval County, Florida; Pittsburgh, Pennsylvania; and Rochester, New York—to participate. The Foundation launched the NSLP to expand summer opportunities for low-income students in urban settings and to understand whether and how district-led, voluntary summer learning programs that include academic instruction and enrichment opportunities can improve student outcomes. Although districts made their own choices about some aspects of their programs, such as the specific academic curriculum and type of enrichment offered, they each implemented the following common elements:

1. voluntary, full-day programming that included academic instruction and enrichment activities (the latter mainly provided by community partners) for five days per week for no less than five weeks of the summer
2. at least three hours of language arts and mathematics instruction per day provided by a certified teacher
3. small class sizes of no more than 15 students per instructor
4. no fees to families for participation
5. free transportation and meals.

The Foundation commissioned the RAND Corporation to study implementation and student outcomes as part of the NSLP. Research was conducted in three phases:

- Phase I was a formative phase during which the selected programs received feedback and improved their programs in preparation for the evaluation phase.
- Phase II was a summative evaluation phase that consisted of a randomized controlled trial (RCT) and implementation evaluation administered over two summers (2013 and 2014) with outcomes measured through spring 2015.
- Phase III was a follow-up phase examining NSLP student outcomes in spring 2017, three school years after the second summer of programming.

For the RCT, we randomly assigned summer program applicants into two groups: a treatment group that had the opportunity to participate in two consecutive summers of programming and a control group that did not. This lotterylike process, which resulted in statistically equivalent groups, ensured that any differences between the groups at the end of the study (barring differential attrition between the two groups) were attributable to the program and not to external factors, such as motivation to apply for the summer program.

In spring 2013, 5,639 eligible third-grade students from the five participating districts applied to the programs. The number of applicants in each district exceeded recruitment goals, signaling strong demand. Across the districts, 47 percent of study students were African-American and 40 percent were Hispanic. The majority of students, 89 percent, were eligible for the national school lunch program (an indicator of low family income), 30 percent were English language learner (ELL) students, and 42 percent had scored at the lowest level of proficiency in one or both of their language arts and mathematics standardized state assessments in spring 2013.

Throughout Phases I and II, we also collected an extensive set of implementation data from each program through classroom observations, teacher surveys, teacher and administrator interviews, and administrative attendance records. We analyzed these implementation data to provide formative feedback to the districts to support their continuous improvement, develop lessons for the field (see Schwartz et al., 2018), and examine the links between implementation and student outcomes. Student outcomes tracked in Phases II and III were mathematics and language arts performance, social-emotional skills as measured by teachers, and school-year behaviors (e.g., school-year attendance, suspensions).

We addressed several research questions throughout this longitudinal study. Augustine, McCombs, Pane, and colleagues (2016) addressed the following research questions in Phases I and II:

1. How well are the programs implemented, including site management, quality of academic and enrichment instruction, time spent on academic instruction, site culture, and cost?

2. What is student participation in one summer and two summers of programming?

3. What is the effect of offering two consecutive years of voluntary summer programming on student achievement, behavior, and social-emotional outcomes, measured in the fall and spring after the first summer?

4. What is the effect of offering two consecutive years of voluntary summer programming on student achievement, behavior, and social-emotional outcomes, measured in the fall and spring after the second summer?

5. Do student characteristics, such as achievement level, family income, or ELL status, moderate outcomes?

6. What factors, including program implementation and student attendance, influence student outcomes?

This report summarizes findings from questions 2–6 and discusses them in relationship to new findings from the additional two research questions posed in Phase III, which examined the longer-term effects of summer programming:

1. What is the longer-term effect of offering two consecutive summers of voluntary summer programming on student achievement, behavior, and social-emotional outcomes, measured in spring 2017, at the end of the third school year after the second summer of programming?

2. What factors, including program implementation and student attendance, influence longer-term student outcomes?

Our study provides an opportunity to gain additional understanding of outcomes within three academic years after the intervention ended.

We examined these final two research questions to better understand the nature of summer learning program effects and whether they persisted, increased, or dissipated over time. Most evaluations of educational interventions only look at immediate outcomes; those that have followed students longitudinally tend to find that impacts dissipate over time. However, some studies of early childhood interventions that have observed a fade-out of effects find a subsequent reemergence of effects (in adulthood). Our study provides an opportunity to gain additional understanding of outcomes within three academic years after the intervention ended. Students entered the study at the end of the third grade. Students in the treatment group were offered the opportunity to attend the summer program as incoming fourth- and fifth-graders. Our final outcome measures for study students are from spring 2017, at the end of seventh grade. This study of summer learning programs is unique in its scope; it is the longest study of summer learning programs that we know of, beginning in 2011 and concluding in 2017.

Key Findings

When examining student outcomes, we conducted causal and correlational analyses, each of which provided important information. As described earlier, RAND researchers randomly assigned students to two groups: One group received the opportunity to attend voluntary summer programing (treatment) while the other group did not (control). Because of the voluntary nature of the programs, our causal analyses evaluated the effect of *offering* summer programming, regardless of whether students actually attended the program. Because 20 percent of treatment students did not attend in the first summer and 48 percent did not attend in the second summer, these analyses underestimate the effect of the program for students who did attend.

To build on the causal analysis, we used correlational methods to examine how implementation features of the summer program and student attendance related to student outcomes. Here we compared only subsets of treated students (e.g., students with high attendance) with students in the control group. Student attendance in the summer program was not randomly assigned, so we cannot rule out other factors that affected both summer attendance and student outcomes. To help mitigate the possibility of bias, we used a broad set of student characteristics (including prior academic performance) in our correlational analyses. Although we cannot rule out the possibility that unmeasured characteristics caused or contributed to the correlational results we describe here, the sum of evidence gives us confidence that the academic results are because of participation in the summer learning programs. We are less confident in the social-emotional results because we lacked a pretreatment measure of those outcomes for use as a statistical control.

Summary of Causal Findings

After the first summer, students offered the program (treatment group students) outperformed control group students on fall mathematics assessments.

In fall 2013, students in the treatment group outperformed students in the control group by an estimated effect size of 0.08 in mathematics. Using average annual gains on standardized assessments as a benchmark, students in our treatment group experienced about 15 percent of that annual gain, which is appropriately sized for a five-week program. However, we did not find a statistically significant effect of offering the program on the 2014 spring state assessment in mathematics. Neither did offering the program significantly improve other measured outcomes, notably language arts performance, social-emotional skills, and school-year attendance. Also, we found no discernible difference in program benefits among subgroups of students in the treatment group; students eligible for free or reduced-price lunch, students who had the lowest performance on prior achievement tests, and ELL students experienced approximately the same effects as other students in the treatment group.

After the second summer, offering the program did not significantly affect any of the measured outcomes among students in the treatment group.

Although estimates for mathematics and language arts were positive in fall 2014 and in spring 2015 and 2017, none of the estimates was statistically significant. This result was not entirely surprising because almost half (48 percent) of the treatment group students did not attend the second summer of programming. Some students had left the school district altogether, but others likely wanted different experiences in that second summer. Effects on social-emotional and behavioral outcomes were also nonsignificant and, unlike achievement outcomes, were not consistently positive.

Summary of Correlational Findings

After the first summer, high attenders outperformed control group students in mathematics in the fall and on the subsequent spring 2014 state assessment.

After the first summer, high attenders (students attending 20 or more days of the summer program) performed better than their control group peers in mathematics on the fall assessment (an effect size of 0.13 or 25 percent of an average annual gain) and on the spring 2014 state mathematics assessment (0.07 or 13 percent of the expected annual gain; see Lipsey, Puzio, et al. [2012]). After the first summer, high attenders did not significantly outperform control group students in language arts, social-emotional outcomes, or school-year behaviors.

After the second summer, high attenders performed better than control group students in mathematics and language arts through spring 2015.

After the second summer, in the fall of fifth grade, high attenders performed better than the control group on measures of mathematics, language arts, and social-emotional skills. The academic benefits were also observed in the spring 2015 state assessments (0.14 in mathematics and 0.09 in language arts). Students participating in two consecutive summers of programming also performed better than control group students in mathematics and language arts in the fall and spring. Because the vast majority of high attenders also attended both summers, we cannot be certain whether the benefits observed after the second summer were the result of two consecutive summers of attendance or improved program quality in the second summer; however, we hypothesize that both might have contributed to the results.

Greater amounts of time on task and higher quality of instruction were correlated with better outcomes through spring 2015.

Students with high academic time on task (who received a minimum of about 25 hours of summer mathematics instruction or 34 hours of summer language arts instruction) outperformed control group peers on mathematics and language arts assessments in fall 2013, fall 2014, and spring 2015. The amount of academic time on task that a student received in a summer program was dependent on the student's attendance and how teachers used time during the program. To achieve language arts benefits, our analyses suggest that both the quantity and quality of instruction are predictors of better outcomes. We found consistent positive correlations between the quality of language arts instruction and language arts achievement for each assessment through spring 2015, although this effect was statistically significant only on the fall 2013 assessment immediately after the first summer of programming.

After the second summer, high attenders received higher social-emotional skill ratings in the fall than control group students, but that advantage did not persist.

On return to school in fall 2014 after the second summer of programming, teachers rated high attenders as demonstrating stronger social-emotional skills than control group students. However, this advantage did not persist through seventh grade. As of spring 2017, teacher ratings of social-emotional skills did not differ discernably between high attenders and their control group peers. As discussed in *Learning from Summer* (Augustine, McCombs, Pane, et al., 2016), the correlational analyses of social-emotional skills lacked a baseline measure; resulting estimates were also less precise than those for achievement and did not exhibit clear patterns over the course of the study.

Summary of Follow-Up Findings

Three school years after the second summer of programming, academic benefits for attenders decreased in magnitude and were not statistically significant, but they might be important in practical terms.

The magnitude of the benefits observed for high and consecutive attenders in language arts and mathematics in spring 2015 (one school year after the second summer) declined by spring 2017 when the students were finishing seventh grade. Other education studies tracking impact after an intervention ended have also found impacts that dissipate over time (e.g., Puma et al., 2012; Bailey, Fuchs, et al., 2018; Lipsey, Farran, and Kelly, 2018).

The program effects in spring 2017 for these groups were not large enough to reach statistical significance; however, when benchmarked against typical achievement gains at the same grade level, they remained large enough to be educationally meaningful. Typical annual achievement growth decreases as students progress from kindergarten through 12th grade, when measured in the same standardized effect units. For students with high attendance both summers, the 2017 estimated effects (0.04 in language arts and 0.07 in mathematics) represent 19 percent of typical annual growth in language arts and 23 percent in mathematics.

Implications for Policy and Practice

These findings have implications for policy and practice.

Urban districts should consider offering voluntary summer programs as part of their overall efforts to improve outcomes for students from low-income families and with low academic achievement, particularly if they can offer these programs over multiple summers.

[D]istricts should actively work to promote high rates of student attendance within and across summers and encourage students to attend for multiple, consecutive summers.

Offering a five-week voluntary summer program with both academics and enrichment can produce short-term benefits in mathematics among late elementary students. High attenders and students who participated for consecutive summers benefited from these programs not just in mathematics, but in language arts as well. These benefits were observed in the fall and in the spring, using both study-administered and state assessments.

Because benefits of the program were greatest for students who attended consecutive summers and those who had strong attendance, districts should actively work to promote high rates of student attendance within and across summers and encourage students to attend for multiple, consecutive summers.

Although the magnitude of the academic benefits we observed in spring 2015 did not persist at the same level or grow years after the program, those benefits also did not fade away completely. Districts willing to develop quality programming that promotes strong attendance can consider this type of summer program a good option to help improve academic achievement.

Districts offering voluntary summer programs that seek to provide academic benefits should offer at least five weeks of

programming, and preferably six, with at least three hours of academic instruction per day.

Districts with academic programs should offer programs for at least five weeks to boost the number of students who attend more than 20 days as a method to maximize program effectiveness. Offering six or more weeks of programming could increase the proportion of students meeting this threshold of attendance.

To increase program effectiveness and maximize return on investment, districts should focus on ensuring strong student attendance, productive use of instructional time, and high-quality instruction.

Our analyses identified strong attendance, productive use of instructional time, and instructional quality as key mechanisms that promoted positive academic benefits. This is not surprising, given the importance of these factors in learning during the school year. Districts recognize these as priorities; however, effectively executing them can be even more challenging in the summer than during the school year and requires intentional planning (Schwartz et al., 2018). Districts and partners interested in learning how to plan and implement effective programs that provide positive experiences for students can find detailed guidance in another report in this series—*Getting to Work on Summer Learning* (Schwartz et al., 2018)—that is freely available on the RAND and The Wallace Foundation websites. In addition, The Wallace Foundation's Knowledge Center includes a set of accompanying tools and resources that provide concrete examples and templates for districts and their partners developing voluntary summer learning programs.

ACKNOWLEDGMENTS

We would like to thank Ann Stone, Edward Pauly, Lucas Held, and Dan Browne at The Wallace Foundation for their valuable guidance on this work and for other substantive and financial support.

We thank our RAND colleagues who contributed to this report. Stephanie Lonsinger assisted with editing for our drafts. During the quality assurance and production process, Fatih Unlu, James Kim, and Andrew McEachin provided a careful review and valuable feedback. Arwen Bicknell edited the final document and Katherine Wu designed the document and cover.

Finally, we thank the district and community leaders in Boston, Massachusetts; Dallas, Texas; Duval County, Florida; Pittsburgh, Pennsylvania; and Rochester, New York for their participation in the study and dedication to the children in their communities.

ABBREVIATIONS

DESSA-RRE	Devereux Student Strengths Assessment–RAND Research Edition
ELL	English language learner
GMADE	Group Mathematics Assessment and Diagnostic Evaluation
GRADE	Group Reading Assessment and Diagnostic Evaluation
ITT	intention-to-treat
NSLP	National Summer Learning Project
RCT	randomized controlled trial
TOT	treatment-on-the-treated
WWC	What Works Clearinghouse

CHAPTER ONE

Introduction

Across the United States, most students are out of school over the summer months. Although many children and youth engage in a set of enriching activities that promote their development or allow them to explore skills and interests, families with low incomes are less likely to be able to afford high-quality summer experiences. Moreover, students from lower-income families are frequently concentrated in communities that lack the resources to support access to high-quality summertime opportunities and are also at increased risk of exposure to adverse neighborhood conditions, such as crime, overpolicing, and environmental hazards, that undermine their development and learning (National Academies of Sciences, Engineering, and Medicine, 2019). Given the inequity of access to enriching summer experiences, research has found, perhaps unsurprisingly, that summer might contribute to achievement gaps between students from low-income families and their peers from higher-income families (Downey, Von Hippel, and Broh, 2004; Kim, 2004; McCoach et al., 2006; Benson and Borman, 2010; Ready, 2010; White et al., 2013; Von Hippel, Hamrock, and Kumar, 2016).

Despite efforts to improve academic achievement of students from low-income families, there is a persistent achievement gap related to family income in the United States, one that has arguably worsened over time for those from the lowest-income families (Reardon, 2011). On the National Assessment of Educational

Progress, 52 percent of fourth-grade students ineligible for the national school lunch program (an indicator of low family income) scored at or above the proficient level in reading compared with 22 percent of students eligible for the lunch program. Similar proficiency gaps exist in mathematics and for other grade levels. Sizable achievement gaps also exist among racial-ethnic groups and between native English speakers and English language learners (ELLs). These achievement gaps are also found in state assessments. The gaps are troubling because they translate into attainment gaps; students from low-income families graduate from high school at lower rates than peers from higher-income families (70 percent versus 85 percent) and college (10 percent versus 60 percent) (National Center for Education Statistics, 2015; Pell Institute, 2015).

Studies of summer achievement find that students from low-income families experience setbacks over the summer relative to their more economically advantaged peers.

Summer may contribute to this achievement gap. Studies of summer achievement find that students from low-income families experience setbacks over the summer relative to their more economically advantaged peers. A seminal meta-analysis of summer learning (Cooper, Nye, et al., 1996) found that all students lost mathematics and reading knowledge over the summer, although the loss in mathematics knowledge was generally greater than in reading. This meta-analysis also indicated that losses were larger for low-income students, particularly in reading. Recent studies are inconclusive on the absolute loss of achievement over the summer or even whether loss takes place (e.g., Burkham et al., 2004; McCoach et al., 2006; Benson and Borman, 2010; Ready, 2010; Fitzpatrick, Grissmer and Hastedt, 2011; Zvoch and Stevens, 2013; Von Hippel, and Hamrock, 2019); however, research consistently finds evidence of differential outcomes for students related to family income. Many studies find that students from lower-income families learn less than their peers from wealthier families over the summer, even if they do not experience knowledge losses during that time (Downey, Von Hippel, and Broh, 2004; McCoach et al., 2006; Benson and Borman, 2010; Ready, 2010; Von Hippel, Hamrock, and Kumar, 2016). Studies have also found that students living in low-income neighborhoods (Benson and Borman, 2010) and attending poorer schools (White et al., 2014; Atteberry and McEachin, 2016) experience larger losses over the summer relative to peers in wealthier neighborhoods or schools.

Throughout their lives, students from low-income families have different opportunities and experiences outside school than do students from higher-income families. Approximately 59 percent of school-age children from low-income families participate in sports, compared with 84 percent of children from wealthier families (i.e., those with annual incomes of $75,000 or more). These types of opportunity gaps also exist for private lessons (e.g., piano lessons) and engagement in clubs (Pew Research Center, 2015).

Family income affects students' summer experiences similarly. In 2014, for example, 61 percent of families living in concentrated poverty reported that they wanted to enroll their children in a summer program, but only 41 percent were able to do so, given costs and availability (Afterschool Alliance, 2014). A more recent study found that 38 percent of incoming first-graders from households above the federal poverty level attended a day camp in the summer compared with 13 percent of children from near-poor families and 7 percent of children from poor families. This study also found that children from low-income families are less likely to engage in such experiences as visits to the beach, a state or national park, the zoo or aquarium, or an amusement park during the summer (Redford, Burns, and Hall, 2018). Another analysis examining children's time use during the summer months found that children from low-income households watched more television and spent less time talking with parents than children from higher-income households (Gershenson, 2013).

Although summer is a contributor to inequitable outcomes, it is also an opportune time to provide activities, interventions, and programs that promote positive student outcomes, including student academic achievement and access to enriching activities. Indeed, research provides evidence that summer programs can achieve some of these goals. A recent evidence review identified 43 programs with evidence meeting the top three tiers specified in the Every Student Succeeds Act legislation and subsequent federal guidance (McCombs, Augustine, Unlu, et al., 2019). The majority of evaluations that met review criteria identified at least one positive and statistically significant finding. Authors found evidence of effectiveness in summer programs designed for in-person academic learning, learning at home, social and emotional well-being, and employment- and career-focused issues. Although much is known about the effectiveness of summer programming, it is unclear what we should expect in terms of the size of benefits for these relatively short programs, how long we should expect statistically significant benefits to persist over time, or the effects

Authors found evidence of effectiveness in summer programs designed for in-person academic learning, learning at home, social and emotional well-being, and employment- and career-focused issues.

of participating in a sequence of activities and programs over the course of childhood and youth (National Academies of Sciences, Engineering, and Medicine, 2019).

The National Summer Learning Project

In 2011, The Wallace Foundation selected five school districts—Boston, Massachusetts; Dallas, Texas; Duval County, Florida; Pittsburgh, Pennsylvania; and Rochester, New York—to participate in the National Summer Learning Project (NSLP), which ran through 2017. The Foundation launched the NSLP to expand summer opportunities for low-income students in urban settings and to understand whether and how district-led, voluntary summer learning programs that include academic instruction and enrichment opportunities can improve student outcomes. When this project began, there was evidence that voluntary academic programs could, but would not necessarily, produce positive effects on achievement outcomes. However, there was no evidence regarding the effectiveness of large-scale, voluntary, district-run summer learning programs serving large numbers of low-income elementary students. Nor were there any studies that tracked student outcomes years after the summer intervention.

The NSLP programs all had five common elements that were anchored in research and expert guidance:

1. voluntary, full-day programming that included academic instruction and enrichment activities (the latter mainly provided by community partners) for five days per week for no less than five weeks of the summer
2. at least three hours of language arts and mathematics instruction per day provided by a certified teacher
3. small class sizes of no more than 15 students per instructor
4. no fees to families for participation
5. free transportation and meals.

The programs were also designed to reduce barriers to participation, such as cost and lack of transportation. Districts and their partners made other programmatic design choices, such as the curriculum used and the type of enrichment activities provided. Table 1.1 shows some of the variation in programmatic choices by district from summer 2014, which also was representative of summer 2013.

TABLE 1.1

Characteristics of NSLP Summer 2014 Programs for Elementary Students

Program Characteristic	Boston	Dallas	Duval County	Pittsburgh	Rochester
Name of summer program	Summer Learning Project	Thriving Minds Summer Camp	Super Summer Academy	Summer Dreamers Academy	Rochester Summer Scholars
Program leader(s)	Boston After School and Beyond with Boston Public Schools	Dallas Independent School District with Big Thought	Duval County Public Schools	Pittsburgh Public Schools	Rochester City School District
Number of summer sites serving study students	10	8	8	3	1 organized into 3 “houses”
Duration (days)	25–30	24	29	25	25
Daily hours	Varied: typically seven-hour days	8:00 a.m.–4:00 p.m.	8:15 a.m.–3:45 p.m.	8:30 a.m.–4:00 p.m.	7:30 a.m.–3:30 p.m.
Program structure	Varied by site: Typically academics in the morning and enrichment in the afternoon	Academics in the morning, enrichment in the afternoon	Sections of academics and enrichment offered throughout the day	Academics in the morning, enrichment in the afternoon	Academics in the morning, enrichment and writing in the afternoons
Enrichment activities	Varied by site: Tennis Sailing Nature walks Ropes course Archery Arts and crafts Swimming Boat building	Dance Music Physical education Theater Visual arts	Varied by site: Dance Music Physical education Theater Visual arts Arts and crafts	Varied by site: Fencing Music Science Visual arts Water polo	Cooking Dance Rock Climbing Sand sports Swimming

The NSLP Study

The Wallace Foundation selected our RAND Corporation research team to conduct the NSLP study on whether and how these programs benefited students. We addressed several research questions throughout this three-phased project. Augustine, McCombs, Pane, and colleagues (2016) addressed the following research questions in the first two phases:

1. How well are the programs implemented, including site management, quality of academic and enrichment instruction, time spent on academic instruction, site culture, and cost?

2. What is student participation in one summer and two summers of programming?

3. What is the effect of offering two consecutive years of voluntary summer programming on student achievement, behavior, and social-emotional outcomes, measured in the fall and spring after the first summer?

4. What is the effect of offering two consecutive years of voluntary summer programming on student achievement, behavior, and social-emotional outcomes, measured in the fall and spring after the second summer?

5. Do student characteristics, such as achievement level, family income, or ELL status, moderate outcomes?

6. What factors, including program implementation and student attendance, influence student outcomes?

This report summarizes findings from questions 2–6 and discusses them in relation to new findings from the additional two research questions that, in this third phase, examined the longer-term effects of summer programming:

1. What is the longer-term effect of two consecutive summers of voluntary summer programming on student achievement, behavior, and social-emotional outcomes, measured in spring 2017, at the end of the third school year after the second summer of programming?

2. What factors, including program implementation and student attendance, influence longer-term student outcomes?

We examine the effect of the summer learning program three school years after the end of summer programming to better understand the longer-term nature of summer learning program effects and whether effects persisted at the same level, increased, or dissipated over time. This study of summer learning programs is unique in its scope; it is the longest study of summer learning programs that we know of, beginning in 2011 and concluding in 2017.

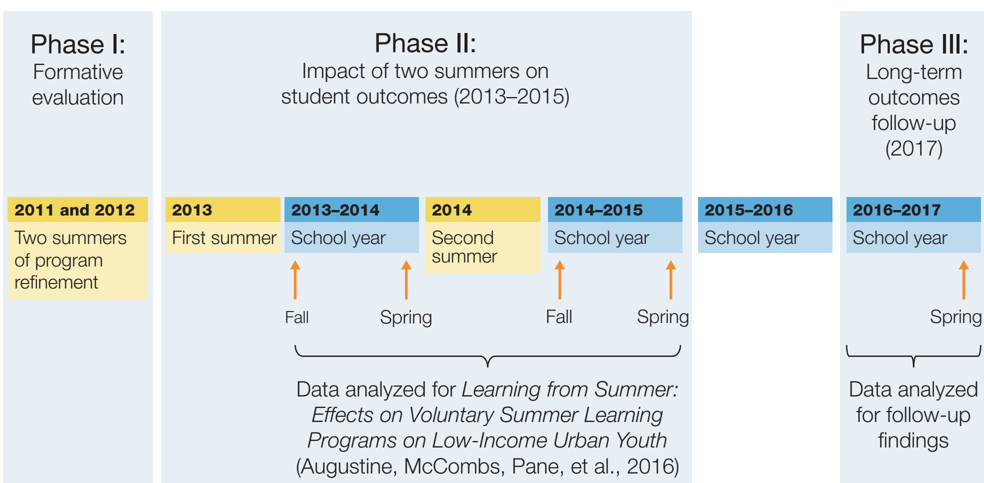
Study Phases

As depicted in Figure 1.1, this study was conducted in three phases:

- Phase I was a formative phase during which the selected programs received feedback and improved their programs in preparation for the evaluation phase.
- Phase II was a summative evaluation phase consisting of a randomized controlled trial (RCT) and implementation evaluation over two summers (2013 and 2014) with outcomes measured through spring 2015.
- Phase III was a follow-up phase examining NSLP student outcomes in spring 2017, three school years after the second summer of programming.

The orange arrows in the figure depict the points at which we measured student outcomes.

FIGURE 1.1
Phases of the NSLP Study



Phase I: Formative Evaluation (2011–2012)

In anticipation of the launch of the RCT in spring 2013, The Wallace Foundation funded two preparatory years in each of the five school districts. Specifically, for summers 2011 and 2012, the Foundation partially funded the summer programs and provided additional funding for curricular consultants, peer collaboration, and external formative evaluation. We conducted the formative evaluation in each summer, providing the districts and their partners each fall with feedback and recommendations, which they used to strengthen their programs. Lessons learned from these early years of programming were published as a guide for practitioners (Augustine, McCombs, Schwartz, et al., 2013).

Phase II: Impact of Two Summers—Student Outcomes Analyses (2013–2015)

The second research phase started in spring 2013. During this phase, the activities of Phase I continued (Wallace financial support, peer learning, curricular support, formative evaluation) and the RCT began. The trial participants were a cohort of third-grade students in spring 2013 who applied to the summer program. We randomly assigned students who applied to the summer program into two groups: a treatment group that had the opportunity to participate in two consecutive summers of programming and a control group that did not. (For details regarding recruitment and randomization, see McCombs, Pane, et al., 2014.) This lotterylike process, which resulted in statistically equivalent groups, ensured that any differences between the groups at the end of the study, barring differential attrition between the two groups, were attributable to the program and not to external factors, such as motivation to apply for the summer program.

Throughout Phase II, we collected implementation, demographic, and outcomes data. For example, we gathered detailed summer attendance data and observed each classroom of students for an entire day in both summers. (For details on this part of the study, see Augustine, McCombs, Pane, et al., 2016, Appendix B.) In an effort to aid districts and their partners in their continuous improvement of their summer programs, we used these implementation data to provide the districts and their partners with feedback. We also used implementation data in our descriptive and correlational analyses to determine the factors correlated with positive outcomes and developed lessons for the field on how to implement summer learning programs. (For this guidance, see Schwartz et al., 2018.)

We received administrative data from each school district, including such background data as prior achievement, race, and eligibility for the national school lunch program (an indicator of low family income), which served as control variables in statistical models. Additional administrative data, such as state assessment scores, course grades, attendance, and suspensions, served as outcome measures.

Phase III Long-Term Outcomes Follow-Up (2017)

The third research phase examined whether the summer program affected student outcomes in the longer term: three school years after the second summer of programming. When students entered the study, they were finishing third grade. At the point of the longer-term follow-up, students were in seventh grade.

For the most part, we adhered to the same methods for the analyses of spring 2017 student outcomes that we used for examining student outcomes in Phase II. (For details regarding data and methods used, see Augustine, McCombs, Pane, et al., 2016.) In limited circumstances, we determined that methodological adjustments were necessary. The appendix of this report describes those changes and their rationale. We believe the changes enable coherent interpretation of the longitudinal series of results produced by this entire study.

Table 1.2 provides an overview of the student outcomes we examined at different time points in Phases II and III: study-administered broad, generalized assessments of language arts and mathematics in fall 2013 and 2014; a validated teacher-report instrument measuring student social-emotional competencies (Devereux Student Strengths Assessment–RAND Research Edition or DESSA-RRE); and state assessments in language arts and mathematics. We also examined measures related to achievement and social-emotional competencies, such as grades, student attendance, and suspensions. Our focus on academic outcomes is self-explanatory. Although we also examined social-emotional outcomes, the programs did not have an explicit social-emotional learning curriculum. Nonetheless, program leaders had hypothesized that their program would affect students' self-motivation and self-regulation skills during the school year, partly by maintaining a school-like routine during the summer.

TABLE 1.2
Student Outcomes Examined in the NSLP

Outcome Measure	Fall 2013	Spring 2014	Fall 2014	Spring 2015	Spring 2017
Group Reading Assessment and Diagnostic Evaluation (GRADE)	X		X		
Group Mathematics Assessment and Diagnostic Evaluation (GMADE)	X		X		
DESSA-RRE	X		X		X
State assessment in language arts and mathematics		X		X	X
Course grades in language arts and mathematics		X		X	X
Suspensions		X		X	X
School-year attendance		X		X	X

Students in the Study

In spring 2013, 5,639 eligible third-grade students from the five districts applied to the program, exceeding recruitment goals. Students were recruited through fliers sent home to parents and through personal outreach efforts, such as teachers who wrote handwritten notes to parents and school counselors who talked with parents during drop-off and pick-up times.

The students who participated in the study were largely non-white and came from low-income families (Table 1.3). Across the districts, 47 percent of study students were African American and 40 percent were Hispanic. The majority of students, 89 percent, were eligible for the national school lunch program, an indicator of low family income. Overall, 31 percent of study students were ELL students; Dallas had the highest ELL proportion at 59 percent. Approximately 42 percent of study students scored at the lowest level of proficiency in language arts, mathematics, or both on their statewide standardized spring 2013 assessments. However, there was wide variation at the district level, ranging from a low of 12 percent of students in Duval County to a high of 81 percent of students in Rochester. This variation might stem from the varying difficulty of state assessments or overall achievement levels in the district, but it also reflects district policies that affected student eligibility for the program. In Duval County, for example, students scoring at the lowest level on the state language

TABLE 1.3
Demographic Profile of All Study Students, by District

District	Number of Students in the Study	African American (%)	Hispanic (%)	Asian (%)	White (%)	Low Income (%)	ELL (%)	Lowest Achieving[a] (%)	IEP (%)
Boston	957	42	41	6	8	NA	30	24	15
Dallas	2,056	19	77	1	1	95	59	43	5
Duval County	888	79	5	1	12	87	3	12	8
Pittsburgh	656	70	3	3	17	83	7	39	17
Rochester	1,080	65	22	4	8	82	16	81	15
Total	5,637[b]	47	40	3	7	89	31	42	10

SOURCE: District student-level data from the 2012–2013 school year.
NOTES: Racial and ethnic categories may not add to 100 percent because "other" is not shown. Low-income students are eligible for the national school lunch program. IEP = students with individualized education plans (special education).
[a] *Lowest-achieving* is defined as students scoring at the lowest proficiency level on either the spring 2013 mathematics or language arts state tests, prior to the start of the study.
[b] Two students initially randomized are not represented in this table because of withdrawal of parental consent to use the students' data for this study.

arts assessment were mandated to attend a separate summer program and were thus ineligible to participate in our research.

Across the districts, more than 3,000 students were assigned to the treatment group (57 percent) and 2,445 (43 percent) were assigned to the control group. We assigned the larger percentage of students to the treatment group to admit as many students as possible while maintaining sufficient statistical power. As expected from a random selection process, characteristics between the treatment and control groups were very similar.

We were unable to track all the students in the study through spring 2017. About 11 percent of our total study sample had left their districts by summer 2014. By the time we measured outcomes in spring 2017, almost one-third of the study sample had left their districts. We found no difference in attrition rates between the treatment and control groups.[1]

Despite the low participation rates in the second summer, offering these programs helped close the summer opportunity gap in the study districts.

Student Participation and Attendance

Despite having applied to the summer program, not all of the treatment students attended. Twenty-one percent of treatment students did not attend the summer program in 2013. In summer 2014, the no-show rate increased to 48 percent. The no-show rate in summer 2014 included students who were invited 14 months earlier to attend the program in both summer 2013 and summer 2014 but chose not to in summer 2014 and students who had left their districts and were unable to attend. We did not find differences based on observable characteristics (e.g., achievement, race/ethnicity, family income) between students who did not show up (in either summer) and students who attended.

Of students who showed up for the summer program, the average daily attendance rate in each summer was 75 percent. This average masked differences among districts, where average daily attendance ranged from a low of 60 percent to a high of 80 percent. In each summer, approximately 60 percent of students who showed up for the program were high attenders, attending 20 or more days of the program.

Despite the low participation rates in the second summer, offering these programs helped close the summer opportunity gap

[1] Additional detail regarding attrition is presented in the appendix.

in the study districts. We surveyed all of the students at the end of the first summer and learned that a far larger percentage of treatment students reported attending a camp or summer program (81 percent) compared with the control group (42 percent). Treatment students were also far more likely to report language arts and writing at camp or summer school than were control group students.

Reporting Outcomes

In examining the effects of the summer programs on student outcomes, we report standardized effect sizes to quantify the difference between the treatment and control groups. By using standardized effect sizes, we can compare the magnitude of program effects across the various outcome measures. For example, we use effect sizes to examine whether the programs have a larger impact on language arts or mathematics outcomes. Standardization also allows us to compare program effects with those of other programs. Despite the standardization, we caution that the magnitude of an effect size is influenced by a variety of factors—including the type of assessment used, grade level and subject, and type of study conducted. It might be useful to consider the following data—all shown in standardized effect size units—to help set realistic benchmarks for what effect sizes to expect in this case.

- Measured in effect size units, typical annual spring-to-spring gains on broad standardized assessments vary by subject and grade level from as large as 1.52 in language arts between spring of kindergarten and spring of first grade to as small as 0.01 in mathematics from spring of 11th grade to spring of 12th grade (Lipsey, Puzio, et al., 2012). In general, typical gains are larger in mathematics than in language arts and decline as students age.

- For the grade span covered by Phase II of the study (spring of third grade to spring of fifth grade), typical annual gains are 0.38 in language arts and 0.54 in mathematics. For the year including seventh grade, in which we obtained outcome measures for Phase III, the typical gains are 0.23 in language arts and 0.30 in mathematics. A five- to six-week summer program represents 10 percent of a calendar year and 15 percent of a school year, so the effects of those programs would likely be correspondingly smaller.

- Among RCT studies of elementary-grade interventions, mean effect sizes have been largest (0.40) when the outcome was measured by specialized tests, such as researcher-developed or curriculum-based assessments, and smallest (0.08) when measured by broadly focused standardized tests, such as those used in this research (Lipsey, Puzio, et al., 2012).

Based in part on these observations, we designed the research to have sufficient statistical power to detect effects as small as about 0.10 on academic outcomes for attenders, or about 0.08 for all students offered admission to the program assuming a 25-percent no-show rate.

Report Organization

In the next chapter, we present our outcomes findings from Phase II and Phase III of the study. At the end of the chapter, we offer an interpretation of the findings, drawing partly on the broader evidence base for summer programs and other interventions aimed at improving outcomes for disadvantaged youth. In the last chapter, we summarize our findings and present implications for policy and practice. The appendix provides details regarding our data and modeling.

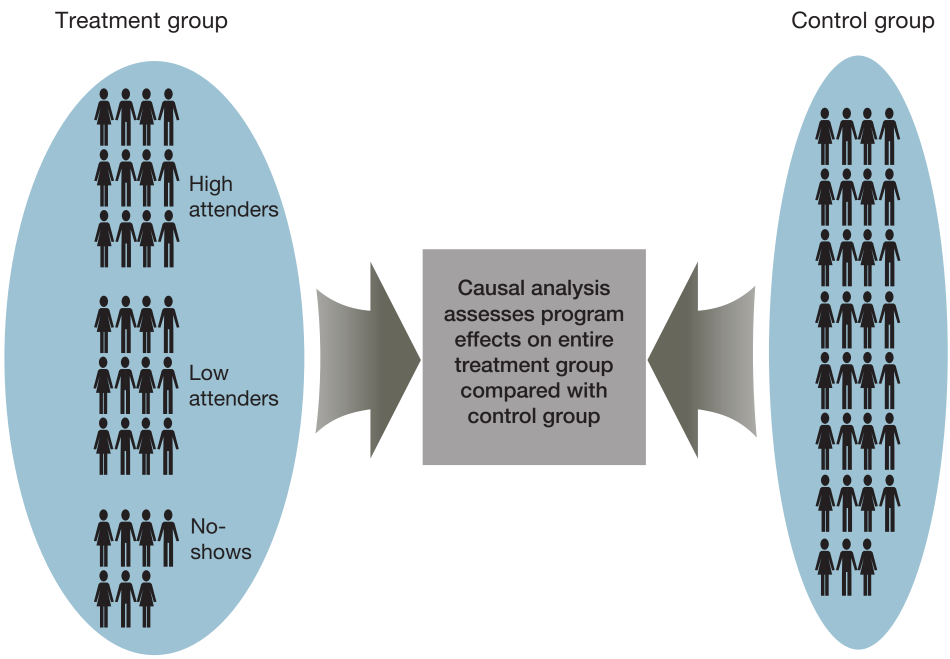
CHAPTER TWO

Student Outcomes: Findings and Interpretation

In this chapter, we summarize the causal and correlational findings from Phase II of the NSLP (2013–2015 student outcomes) and present new findings from Phase III (2017 student outcomes). Our causal (or confirmatory) estimates compare the outcomes of all students who were randomly offered admission to two summers of programming (2013 and 2014) with the outcomes of all students who were randomly assigned to the control group, regardless of whether the students actually attended the summer program (Figure 2.1). As such, these estimates represent the impact of *offering admission* into the summer learning program. Because many students who received the offer did not show up or had poor attendance, if there was a program effect, we would expect the absolute value of the estimates for all invited students to be smaller than the effects experienced by students who did attend regularly. For every causal result, statistical significance has been adjusted to account for all of the causal statistical tests we performed in this study, both past and present, regardless of whether we reported them in the main texts of the reports.

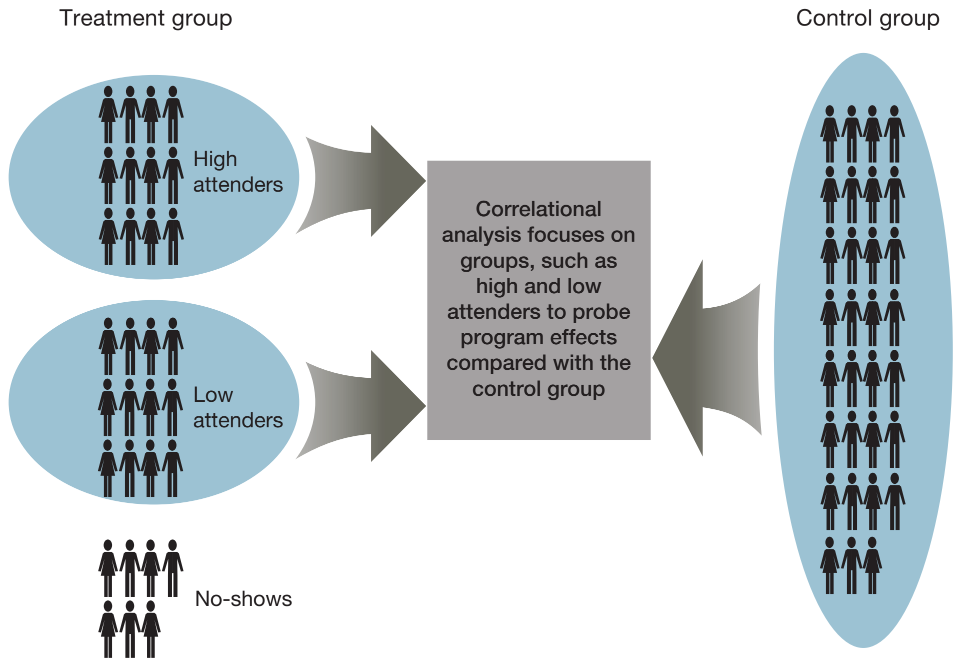
FIGURE 2.1

Causal Analyses Compare Outcomes for All Treatment and Control Group Students



Our correlational (or exploratory) analyses examine the relationship between implementation factors of interest (e.g., attendance, amount of academic time on task) and student outcomes. These analyses compare all of the control group students with subsets of the treatment group that were not randomly determined (Figure 2.2). For example, whether students attended or how much they attended were not experimentally controlled but rather determined by the students themselves or their circumstances. For that reason, selection bias is a possibility, meaning that a subset of interest within the treatment group, such as high attenders, may have differed from the control group even before the summer program began; thus, the summer program is not the only possible explanation for any subsequent differences we measure. To help mitigate the effects of potential selection bias, the models for correlational analyses controlled for a broad set of student characteristics and prior academic performance, but it is important to note that these results may still be biased.

FIGURE 2.2
Correlational Analyses Estimate Program Effects for Subsets of the Treatment Group



Summary of Previous Findings (Fall 2013 to Spring 2015)

In this section, we review findings on student outcomes measured through spring 2015, starting with the causal findings and then turning to correlational findings. These results were previously reported in *Learning from Summer* (Augustine, McCombs, Pane, et al., 2016).

Causal Findings Through Spring 2015 Showed a Modest Near-Term Benefit in Mathematics That Did Not Persist at a Statistically Significant Level

In the fall after the first summer of programming, students in the treatment group outperformed the control group in mathematics on a study-administered standardized assessment. The standardized average effect (i.e., effect size[2]) of offering the program was 0.08 for mathematics and was statistically significant (see Figure 2.3). Lipsey, Puzio, and colleagues (2012) estimated that

[2] All effect estimates are reported in effect sizes that represent the magnitude of the effect on an outcome divided by the standard deviation of that outcome.

students typically experience growth in mathematics of about 0.52 standardized units from the spring of third grade to the spring of fourth grade. By that benchmark, students in our treatment group experienced about 15 percent of that annual gain. A five-week summer program is about 10 percent of a calendar year and 15 percent of a school year. The effect of 0.08 is also comparable to effects reported by Lipsey, Puzio, et al. (2012) for RCTs studying elementary grade–level interventions and measuring outcomes with broad-scope assessments like those used in our research.

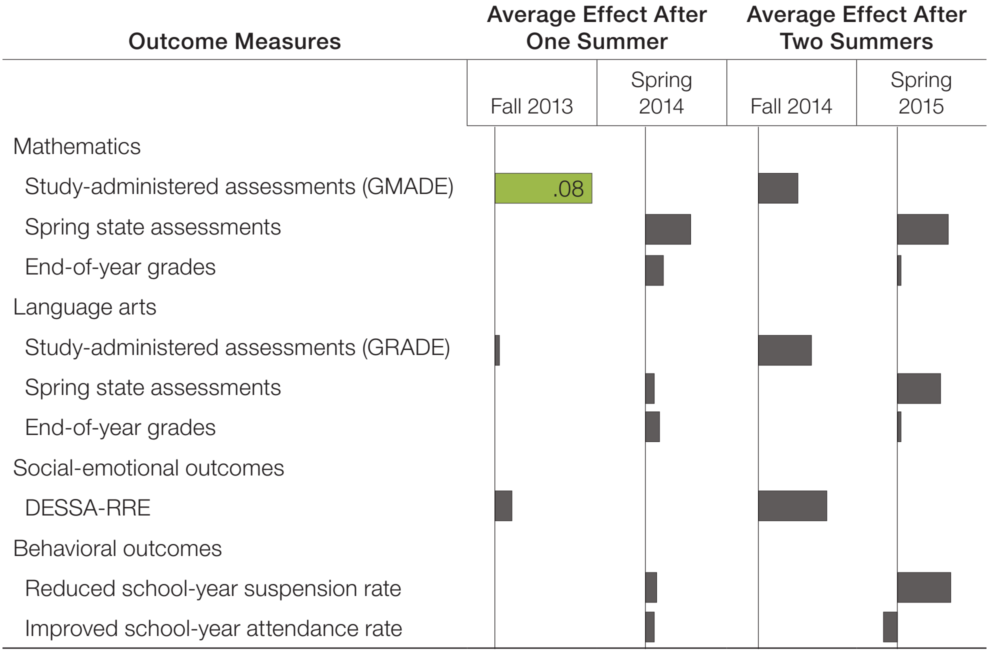
We did not find statistically significant benefits for any other outcomes we measured after that first summer in the near term (fall) or longer term (spring); these consisted of language arts achievement, mathematics and language arts grades, social-emotional competencies, attendance, and suspensions. Nor did we find that certain subgroups of students in the treatment group benefited more or less than others; ELL students, students eligible for free or reduced-price lunches, and students who had the lowest performance on prior achievement tests experienced approximately the same effects as other students in the treatment group.

When analyzing the effects of offering two summers of programming on all treatment students, we found no statistically significant effects in mathematics, language arts, social-emotional or school-year behavioral outcomes. This result was not entirely surprising because nearly half of the treatment students did not attend the program at all during the second summer. These students were still considered part of the treatment group in the causal analyses. If these low attendance rates are typical of summer programs, the causal estimates set realistic expectations for the effects of offering a program for two years on all students who receive the offer; however, the effects for attenders were underestimated in this analysis, diluted by the high proportion of nonattenders. The higher the no-show rate, the larger the effect of the program would have to have been on those who did attend in order to be detected. For the same reason, if effects accumulated over consecutive summers, they would have had to accumulate by a substantial amount for us to have been able to detect this trend statistically. The correlational analyses that we discuss in the next section provide additional insight into the effects on attenders.

The higher the no-show rate, the larger the effect of the program would have to have been on those who did attend in order to be detected.

FIGURE 2.3

Causal Effects of Summer Learning Programs on Outcomes Measured in Phase II for All Treatment Group Students Relative to All Control Group Students



NOTES: The horizontal length of the bar represents the magnitude of the standardized program effect estimate, with the vertical line representing zero. Bars are green where results are statistically significant after correction for multiple hypothesis tests; otherwise, the bars are gray. All models controlled for student baseline characteristics, including prior mathematics and language arts achievements, prior attendance and suspensions, poverty, race, gender, and classification as an ELL student or a special education student. Blanks indicate data were not available for the particular outcome and time point.

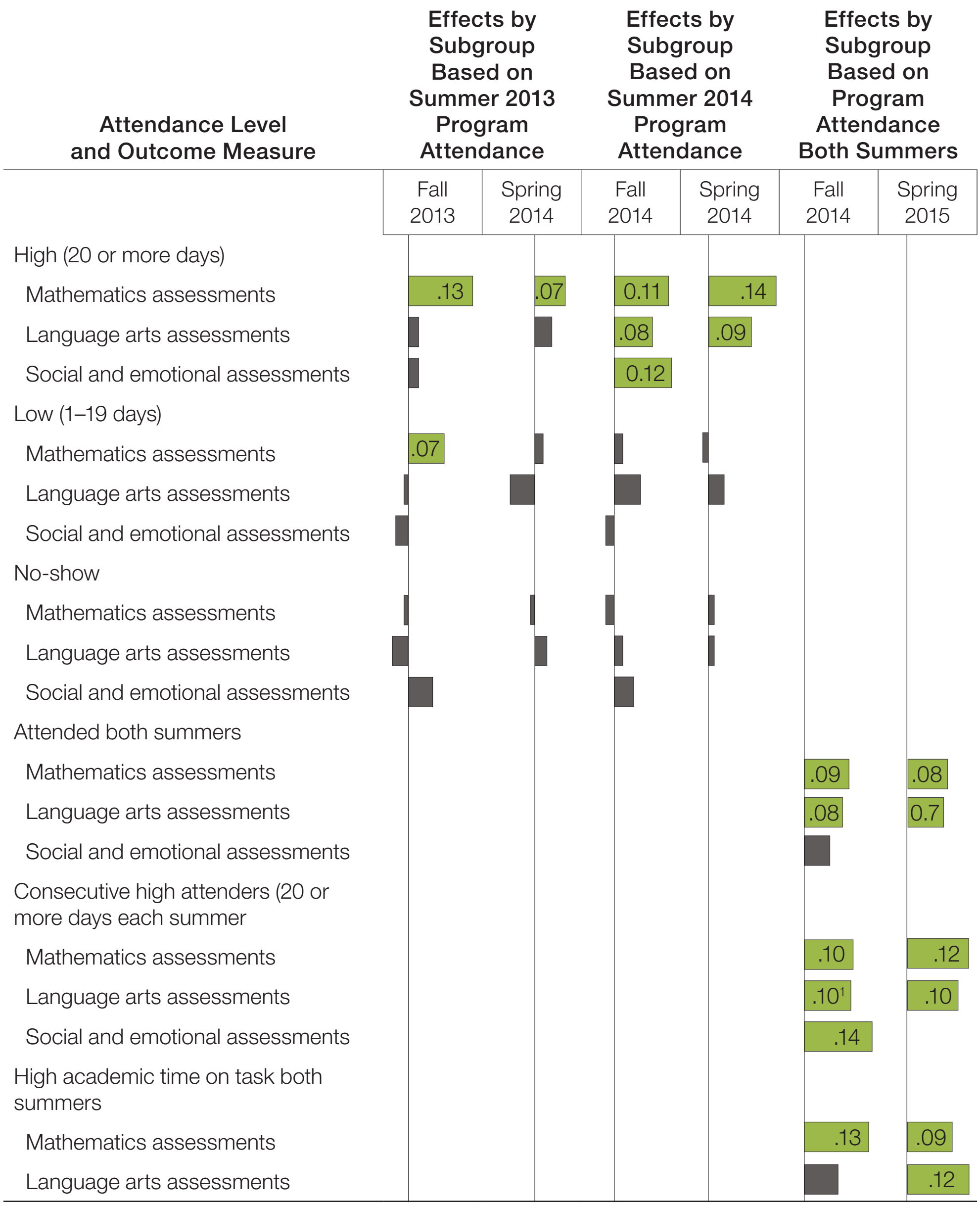
Correlational Findings Through Spring 2015 Showed Benefits in Mathematics and Language Arts for High Attenders

Correlational results through spring 2015 are summarized in Figure 2.4. For the outcomes measured after the first summer, students who attended at least 20 days of the program (high attenders) demonstrated higher achievement in mathematics than control group students in the fall (with an effect size of 0.13 or 25 percent of the typical annual mathematics gain) and on the spring 2014 state assessment (0.07 or 13 percent of the typical annual gain). However, high attenders did not perform better than their control group peers on other outcomes measured after the first summer. High attenders accounted for about 60 percent of all students who attended at least one day in summer 2013.

For the outcomes measured after the second summer, students who attended at least 20 days of the second summer of

FIGURE 2.4

Correlational Effects of Program Attendance on Outcomes Measured in Phase II for Subsets of Treatment Group Students Relative to All Control Group Students



NOTES: The horizontal length of the bar represents the magnitude of the standardized program effect estimate, with the vertical line representing zero. Bars are green where results are statistically significant after correction for multiple hypothesis tests; otherwise, the bars are gray. Blanks indicate data were not available for the particular outcome and time point. All models controlled for student baseline characteristics, including prior mathematics and language arts achievement, prior attendance and suspensions, poverty, race, gender, and classification as an ELL or a special education student. High academic time on task is defined as 25.5 or more hours of instruction for mathematics, and 34 or more hours of instruction for language arts.

[1] Because of an error in figure production, this result was shown as 0.09 in Augustine, McCombs, Pane, et al., 2016.

programming (2014) demonstrated higher achievement than control group students. High attenders outperformed the control group in mathematics and language arts in the fall (0.11 and 0.08, respectively) and in the subsequent spring on the 2015 state assessments (0.14 and 0.09, respectively). These differences represented 14 percent to 21 percent of typical annual gains in mathematics, and 17 percent to 25 percent of the typical annual gains in language arts for students at this age (Lipsey, Puzio, et al., 2012).[3]

We also conducted correlational analyses on the 2014–2015 outcomes for students who attended both summers and those who had consecutive high attendance both summers. Of the students in the treatment group, 46 percent attended both summers and 29 percent were high attenders both summers. Students who attended both summers performed better than control group students in mathematics (0.09) and language arts (0.08) in fall 2014 and again on spring 2015 assessments (0.08 in mathematics and 0.07 in language arts). Students who were consecutive high attenders also performed better than control group students in fall 2014 (0.10 in both mathematics and language arts) and again in spring 2015 (0.12 in mathematics and 0.10 in language arts).

Although we wanted to discern whether benefits found during the 2014–2015 school year come from cumulative program exposure or improved programming in the second summer, we were unable to do so because the vast majority of students who were high attenders in summer 2014 were also high attenders in summer 2013. Using the pattern of results and our knowledge of program implementation, we hypothesize that a combination of cumulative program benefits and improved programming during the second summer might have contributed to the positive correlational findings for academic outcomes observed during the 2014–2015 school year.

Statistically significant correlational effects on social-emotional outcomes also emerged for high attenders after the second summer. Students who had high attendance the second summer scored higher on DESSA-RRE than their control group peers (0.12), as did students who had high attendance in both

[W]e hypothesize that a combination of cumulative program benefits and improved programming during the second summer might have contributed to the positive correlational findings for academic outcomes observed during the 2014–2015 school year.

[3] Students at this grade have an average effect size gain from spring of one year to the following spring of 0.40 in language arts and 0.56 in mathematics (Lipsey, Puzio, et al., 2012).

summers (0.14). Unlike the mathematics and language arts analyses, where we were able to use measures of prior achievement to help control for selection bias, there were no available baseline (pretreatment) measures for social-emotional skills. Thus, we cannot rule out the possibility that these results were driven by selection; for example, it is possible that students who had high attendance systematically exhibited more positive social-emotional behaviors prior to program participation.

Correlational Findings Through Spring 2015 Showed That Instructional Time and Quality Were Positively Related to Student Outcomes

The amount of academic time on task that a student received in a summer program was dependent on the student's attendance and how teachers used time during the program. Using our extensive observations (for details, see Augustine, McCombs, Pane, et al., 2016) and collection of attendance records, we were able to estimate the amount of academic time on task that each student received for language arts and mathematics during the summer program. We defined high academic time on task as a minimum of about 25 hours of summer mathematics instruction or 34 hours of summer language arts instruction each year. Figure 2.4 shows that students who had high academic time on task both summers outperformed control group peers on mathematics assessments (0.13 in fall 2014 and 0.09 in spring 2015) and on language arts assessments (0.12 in spring 2015).

We found consistent positive correlations between the quality of language arts instruction and language arts achievement for each assessment through spring 2015.

Our analyses suggest that both the quantity and quality of instruction were correlated with the language arts achievement benefits. Our instructional quality measure considered clarity of instruction, on-task behavior, and teachers' assessment of student understanding. We found consistent positive correlations between the quality of language arts instruction and language arts achievement for each assessment through spring 2015, although this effect was statistically significant only on the fall 2013 assessment immediately after the first summer of programming (for further details, see Augustine, McCombs, Pane, et al., 2016, p. 68).

Examining Longer-Term Effects (Spring 2017)

During Phase II, we were seeing signals of relatively important positive effects for attenders. In this section, we examine whether effects of the summer program were discernable in spring 2017, three school years after the second summer of programming concluded. After the second summer, students had finished fourth grade and were entering fifth grade. In spring 2017, students were finishing seventh grade.

Causal Findings Through Spring 2017

When we analyzed the effects of offering the 2013 and 2014 summer program on spring 2017 outcomes, we found no statistically significant effects. Figure 2.5 summarizes the results over the course of the study (fall 2013 to spring 2017).

On spring 2017 state assessments, we estimated a standardized effect of 0.01 in mathematics and 0.02 in language arts. On the DESSA-RRE measure of social-emotional skills, we estimated a standardized effect of –0.03. We estimated that the program induced a reduction in suspensions of 0.5 percentage points and a decrease in school-year attendance of 0.3 percentage points (these are displayed in Figure 2.5 as standardized effect sizes of 0.02 and –0.03, respectively). None of these results was statistically significant.

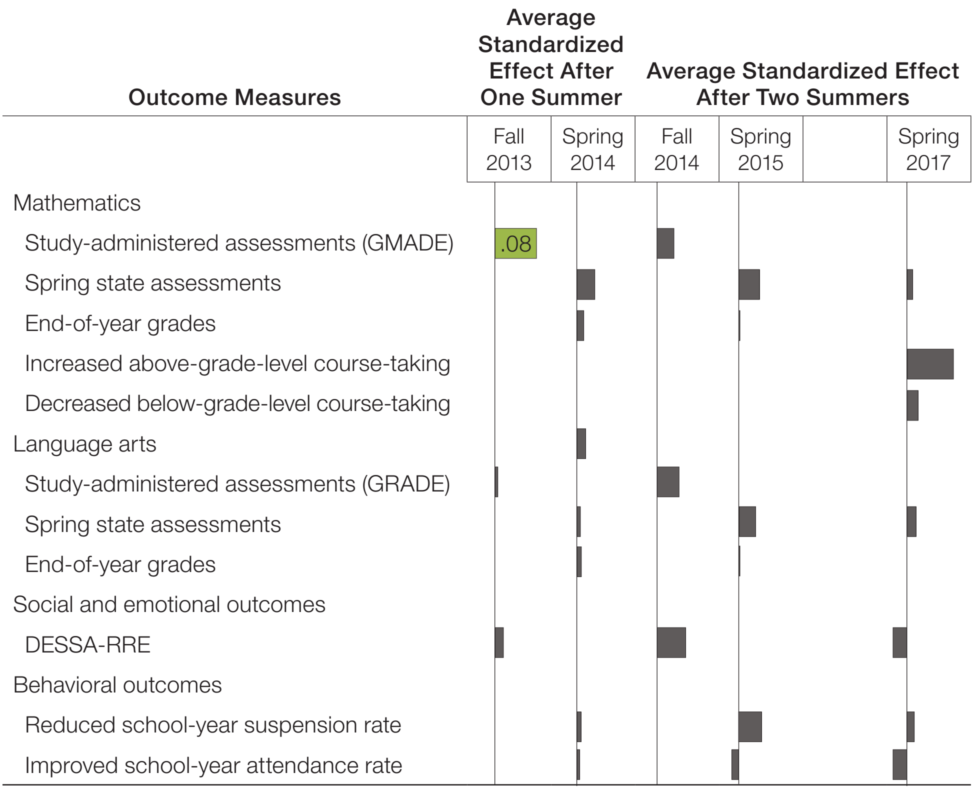
Because the students were in middle school, which offers different levels of mathematics courses depending on student readiness, we also examined whether the program resulted in differential course-taking for treatment and control students. District officials helped us classify courses as being above, at, or below grade level. For mathematics course-taking, we estimated that students in the treatment group were 3 percent more likely to enroll in an mathematics course above grade level and 1 percent less likely to enroll in a mathematics course below grade level (presented in Figure 2.5 as average effect estimates of 0.09 and 0.02, respectively). Neither of these was statistically significant.

The appendix to this report contains tabulations of all spring 2017 causal results, including treatment-effect-on-the treated analyses and other secondary analysis that we do not report here in the main text.

[W]e estimated that students in the treatment group were 3 percent more likely to enroll in an mathematics course above grade level and 1 percent less likely to enroll in a mathematics course below grade level.

FIGURE 2.5

Causal Effects of Summer Learning Programs on Outcomes for All Treatment Group Students Relative to All Control Group Students



NOTES: The horizontal length of the bar represents the magnitude of the standardized program effect estimate, with the vertical line representing zero. Bars are green where results are statistically significant after correction for multiple hypothesis tests; otherwise, the bars are gray. All models controlled for student baseline characteristics, including prior mathematics and language arts achievement, prior attendance and suspensions, poverty, race, gender, and classification as an ELL or a special education student. Blanks indicate data were not available for the particular outcome and time point.

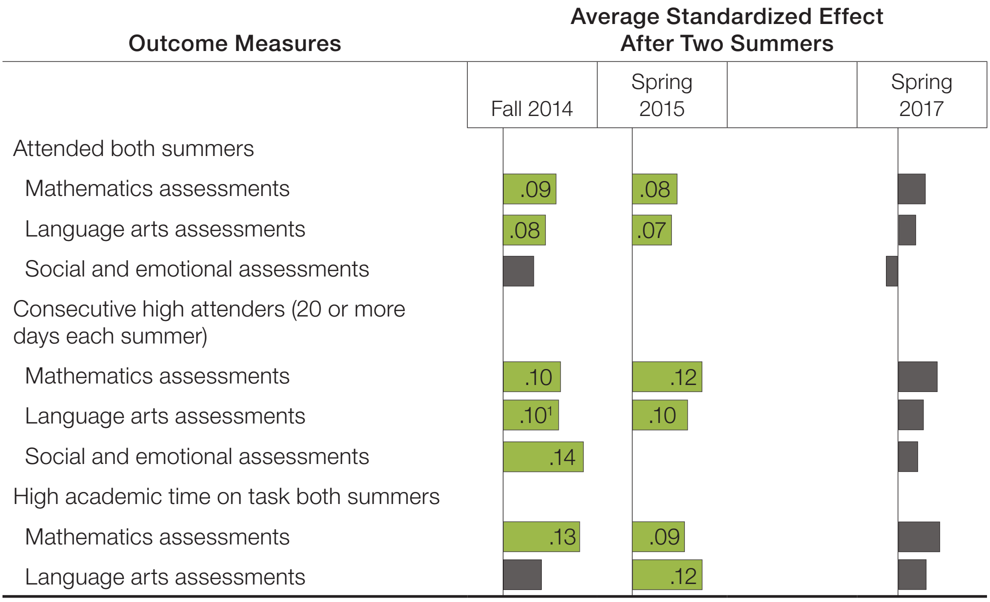
Correlational Findings Through Spring 2017

We now turn to correlational estimates of the effects of the program on three groups of students: those who attended both summers, those who were high attenders in both summers, and those who received high academic time on task in both summers. Figure 2.6 summarizes the average effects in fifth grade (2014–2015, just after the second summer of programming) and in seventh grade (spring 2017) for these subsets of the treatment group.

Consistent with the patterns seen in prior rounds of analysis, we estimate positive effects for students who attended both summers, and the estimates are larger for those who had high attendance or high levels of instructional time on task both summers. However, unlike results measured in the fifth grade, the longer-term estimates are not statistically significant.

FIGURE 2.6

Correlational Effects of Attending Two Years of Summer Programing on Assessment Outcomes for Subsets of Treatment Group Students Relative to All Control Group Students



NOTES: The horizontal length of the bar represents the magnitude of the standardized program effect estimate, with the vertical line representing zero. Bars are green where results are statistically significant after correction for multiple hypothesis tests; otherwise, the bars are gray. Blanks indicate data were not available for the particular outcome and time point. All models controlled for student baseline characteristics, including prior mathematics and language arts achievement, prior attendance and suspensions, poverty, race, gender, and classification as an ELL student or a special education student. High academic time on task is defined as 25.5 or more hours of instruction for mathematics, and 34 or more hours of instruction for language arts.

[1] Because of an error in figure production, this result was shown as 0.09 in Augustine, McCombs, Pane, et al., 2016.

For students who were high attenders both summers, the estimates are 0.07 in mathematics, 0.04 in language arts, and 0.03 on the DESSA-RRE. For students who attended both summers, we estimate effects on 2017 outcomes of 0.05 in mathematics, 0.03 in language arts, and –0.02 on the DESSA-RRE social-emotional assessment. For students with high academic time on task, the estimates are 0.07 in mathematics and 0.05 in language arts.

The appendix contains tabulations of all spring 2017 correlational results, including some that we view as secondary and do not report here in the main text.

Interpretation of Results

Our interpretation of the effects of the summer learning program on student outcomes measured over the course of this study takes a holistic approach that synthesizes the causal and correlational findings. Although the correlational findings are vulnerable to selection bias, their consistency with the causal experimental results helps to reduce bias concerns. Benefits evident in the pattern of positive causal estimates could only have accrued to the students who attended the program, and the correlational estimates for attendees do not appear to be overestimated because of bias—they are in numeric ranges consistent with the causal estimates (discussed in greater detail in Augustine, McCombs, Pane, et al., 2016). Moreover, the correlational findings echo a pattern of smaller causal effect estimates in spring 2017 than were estimated two years earlier, which would not necessarily occur if the correlational findings were influenced by factors other than program effects.

Briefly, we interpret the synthesis of results as indicating that the summer programs conferred benefits to attenders on outcomes closely linked to the instructional content offered (in mathematics and language arts). The advantage the program bestowed on attenders did not persist at the same magnitude of effect size over the three years since the programs ended. In the following sections, we discuss this interpretation in more detail.

A Holistic Interpretation of Causal and Correlational Evidence Suggests That the Program Conferred Academic Benefits Beyond Fall 2013 to Students Who Attended

The causal results show positive effects in mathematics and language arts at all measured time points, although statistically significant only for mathematics achievement in fall 2013. The consistency of these results, which are assumed to be unbiased coming from a randomized experiment, suggest that the programs may have conferred some lasting benefits, though ones that are not large enough to be statistically confirmed by our analyses.[4]

The causal estimates were determined by analyzing data from all students in the study, even students who were admitted to the summer program but did not attend because of alternative plans or exit from the participating districts. Twenty percent of the treatment group students did not participate at all during the first summer, and this increased to nearly 50 percent the second summer. If the causal estimates reflect real program benefits, they would be expected to accrue to the students who attended. This aligns with the correlational results showing stronger positive

[4] The study had sufficient statistical power to detect achievement causal effects of approximately 0.06 or larger.

effects in mathematics and language arts—relative to the estimates for the whole sample—for students who attended both summers, had high attendance both summers, or had high academic time on task both summers. Thus, although we lack strong causal evidence of impacts except for the near-term mathematics estimate after the first summer, the whole set of causal and correlational results is consistent with academic benefits in both mathematics and language arts for students who attended.

Thus, although we lack strong causal evidence of impacts except for the near-term mathematics estimate after the first summer, the whole set of causal and correlational results is consistent with academic benefits in both mathematics and language arts for students who attended.

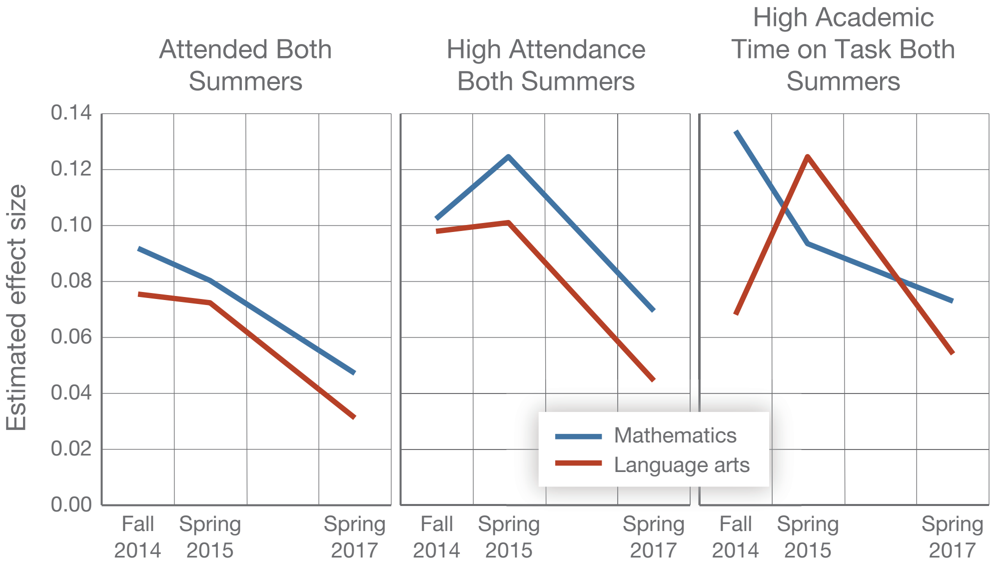
Although we consistently found correlational evidence that students who had high attendance and high academic time on task experienced academic benefits through spring 2015, there were many other outcomes, such as school-year attendance and school year suspensions, for which we did not find statistically significant effects. The effects on outcomes that were most closely aligned to program content (instruction in mathematics and language arts) were most likely to be significant while effects on less proximal outcomes, such as school-year attendance, were generally not statistically discernable. An exception is that teachers rated high attenders as having higher social-emotional skills than control group students in fall 2014 after the second summer. This advantage did not persist through seventh grade, when there was no discernable effect on social-emotional skills. As discussed in *Learning from Summer* (Augustine, McCombs, Pane, et al., 2016), the correlational analyses of social-emotional skills lacked a baseline measure; they were also less precise than achievement effect estimates and did not exhibit clear patterns over the course of the study. For these reasons, we do not have high confidence that the positive fall 2014 estimate for high attenders represents a meaningful positive effect in social-emotional learning.

Three School Years After the Summer Program Concluded, Academic Benefits Decreased in Magnitude, Yet May Remain Educationally Meaningful

The combination of causal and correlational results suggest that the summer program likely had positive effects that mainly accrued to high and consecutive attenders. Figure 2.7 focuses on students who attended both summers and achievement outcomes measured at several time points after the second summer of programming. Longer-term effects measured in spring 2017 were uniformly smaller in magnitude than those measured in 2014–2015, the school year after the second summer of programming ended.

FIGURE 2.7

Trends in Achievement Effect Estimates After Two Summers of Programming for Students Who Attended Both Summers, Had High Attendance Both Summers, or Had High Academic Time on Task Both Summers



NOTES: Plots display estimated effect sizes from correlational analyses. High academic time on task is defined as 25.5 or more hours of instruction for mathematics, and 34 or more hours of instruction for language arts.

(Fluctuations in effect sizes from fall to spring of 2014–2015 are difficult to interpret because of variations in content covered by the assessments. The fall assessments were study-administered and the spring assessments were state tests.)

To reiterate, between spring 2015 and spring 2017, we estimated decreasing program effects in standardized effect units. This trend should be considered alongside a well-established empirical observation that typical annual achievement growth, when measured in the same standardized effect units, also decreases as students progress from kindergarten through 12th grade.[5] Table 2.1 compares 2015 and 2017 effect estimates with typical achievement growth for the corresponding grade levels, as reported in Lipsey, Puzio, et al. (2012). For students with high attendance both summers, the 2017 estimated effect in mathematics represents 23 percent of typical growth in seventh grade, comparable with 22 percent for the 2015 estimated effect benchmarked against typical growth in fifth grade. For language arts, the 2017 benchmark

[5] Lipsey, Puzio, et al. (2012) reports typical effect sizes in mathematics of 1.14 for first grade, and 0.01 for 12th grade. For language arts, these values are 1.52 and 0.06, respectively. Gains are measured from spring of the prior year.

TABLE 2.1

Achievement Effect Estimates After Two Summers of Programming Benchmarked Against Typical Grade-Level Academic Growth

		High Attendance Both Summers		High Academic Time on Task Both Summers	
	Benchmark of Typical Annual Achievement Growth	Estimated Effect	Estimated Effect as a Percentage of Benchmark Growth	Estimated Effect	Estimated Effect as a Percentage of Benchmark Growth
Mathematics					
(2015) Grade 5	0.56	0.124	22%	0.093	17%
(2017) Grade 7	0.30	0.069	23%	0.073	24%
Language arts					
(2015) Grade 5	0.40	0.101	25%	0.125	31%
(2017) Grade 7	0.23	0.044	19%	0.054	23%

NOTE: Benchmarks of typical growth are annual spring-to-spring achievement gains reported by Lipsey, Puzio, et al. (2012, Table 5). High academic time on task is defined as 25.5 or more hours of instruction for mathematics, and 34 or more hours of instruction for language arts.

of 19 percent compares with 25 percent for 2015. Similar patterns are seen for high academic time on task both summers.

From this perspective, although the estimated effects for these high attenders decreased in absolute magnitude between 2015 and 2017, they appear more stable and large enough to remain important in practical terms when viewed relative to typical grade-level achievement growth.

Results from This Study Are Consistent with Other Studies of Educational Interventions That Document Impacts Dissipating Over Time

Although longitudinal designs are rare in program evaluation, and it is not very common for researchers to continue examining outcomes of participants beyond the end of the initial period of the study, an accumulating body of evidence suggests that even when interventions show initially positive impacts on student achievement and other cognitive outcomes, the effects tend to decline or completely fade out over time.

One of the few summer studies to take repeated measures of academic outcomes is Schacter and Jo (2005), although on a shorter time frame than our study. These authors studied the effect of a first-grade summer learning program for decoding and reading comprehension at three, six, and nine months after the intervention. They found that effect sizes for decoding were strongest at the first post-test (0.96), positive but reduced at the second post-test (0.59), and insignificant at the last post-test. Effects on reading comprehension also declined over time.

A recent meta-analysis by Bailey, Duncan, and colleagues (2017) of studies of 67 high-quality early childhood interventions published between 1960 and 2007 showed a pattern of declining effect sizes over time. In fact, the meta-analysis found that the average impact had diminished by more than 50 percent only 12 months after treatment had concluded. In a study of the Head Start program, for example, Puma and colleagues (2012) found evidence of fade-out in early elementary grades. An RCT of an early mathematics intervention targeting the conceptual and procedural bases that support arithmetic found significant impacts on mathematics achievement after one year of intervention, but none of the effects measured in subsequent years was statistically significant (Bailey, Fuchs, et al., 2018). In a study of the Tennessee Prekindergarten Program, Lipsey, Farran, and Durkin (2018) found that treatment students outperformed the control group students on achievement tests after one year of program exposure. However, the control children subsequently closed this gap and generally surpassed treatment students. The summary of studies provided in the meta-analysis suggests that most of the existing studies that have examined longer-term impacts focus on early childhood and prekindergarten interventions, and that studies focused on elementary school or middle school grades are rare.

It is important to consider whether fade-out of program effects necessarily means that the initially observed benefits dissipated. Program effects capture the difference in the outcomes of a treatment and a comparison group. That treatment-comparison difference could decrease over time because the program benefit for the treatment group dissipated, the comparison group later received a boost, or both. Many schools target interventions to their lowest-performing students. For example, in a study of program A, a group of treatment students received benefits that moved them out of the lowest-performing group of students. If the new group of lowest-performing students participate in program B, any benefits of program B could disproportionately raise the performance of program A's comparison group, causing a decrease in future estimates of program A's effects. Thus, the benefits of a program could theoretically persist on an absolute basis even if they appear to fade relative to the comparison group. Studies are typically not designed to shed light on this theory. The measurable treatment effect of program A (the difference between the treatment and comparison groups) will have faded, yet the benefits may have persisted on an absolute basis.

Less is known about fadeout for social-emotional and other non-academic domains. Several existing meta-analyses and research syntheses suggest that positive youth development programs and programs focused on social-emotional learning can have positive impacts initially, but that relatively little is known about the longer-term effects of such programs and interventions (Durlak et al., 2011; Weare and Nind, 2011). A recent meta-analysis suggests that impacts on these domains might be more persistent (Taylor et al., 2017), though the extent to which these impacts persisted over time was moderated by participant age, with older students having less-persistent impacts.

Because only a handful of studies have investigated impacts beyond the high school grades and into adulthood, little is known about program impacts in the extended long term. However, the limited evidence that does exist suggests the possibility of sleeper effects, where initial fadeout is followed by later-stage impacts (Barnett, 2011; Bailey, Duncan, et al., 2017). Several studies have found evidence of these sleeper effects on behavioral and academic outcomes (Deming, 2009; Puma et al., 2012; Chetty, Friedman, and Rockoff, 2014; Dodge et al., 2015).

In summary, although there is some evidence that well-implemented, high-quality early childhood programs can have shorter-term impacts on cognitive, behavioral, and social-emotional outcomes, most evidence suggests that these impacts dissipate over time into early adolescence, though a small handful of studies suggest that these impacts could reemerge in adulthood.

Results from this study are consistent with these prior findings, with effects diminished in magnitude three years after the program ended. Nonetheless, benefits for high attenders appear to remain substantively important and might persist into adulthood.

Results from this study are consistent with these prior findings, with effects diminished in magnitude three years after the program ended. Nonetheless, benefits for high attenders appear to remain substantively important and might persist into adulthood.

CHAPTER THREE

Conclusions and Implications for Policy and Practice

Our research on summer learning programs is unique in its scope and analytic rigor. It is the longest study of summer learning programs, beginning in 2011 and concluding in 2017, tracking outcomes for three years after students entered the second (and final) summer of programming. It featured five large school districts across the country; it examined many program outcomes. It employed an RCT design and the collection of extensive implementation data, which allowed us to conduct a rigorous set of causal and correlational analyses. By following students for years after the second summer of programming ended, we were able to understand more about the persistence of program effects over time. Through our analyses linking our implementation and outcomes data, we were able to identify key factors that mattered the most for program effectiveness and to provide detailed guidance on how to design and implement effective programs (documented in Schwartz et al., 2018). Here, we summarize what we have learned about the effectiveness of voluntary summer learning programs and discuss implications for policy and practice.

Key Findings Regarding the Effectiveness of Summer Learning Programs

First, a caveat. Although we learned a great deal about the effectiveness of summer learning programs, we acknowledge that the programs might have affected students in positive ways that our research was not designed to detect. For example, the programs provided many students with opportunities that they might not have had otherwise, such as to swim, rock climb, cook, and experience new environments. In one district, we heard students comment that they had never before left the city to visit a nature preserve, hike in the woods, go on a boat, or visit an island—all things students experienced in the summer program. Programs also provided students with daily supervision and meals—both breakfast and lunch—and some even provided snacks or dinner to take home for the evening. In another district, we observed program leaders providing needed clothing to students. We do not have measures of how these aspects of the programs influenced students and their families. These aspects of the programs and the benefits outlined below bolster the concept that summer is an opportune time to provide experiences, services, and interventions to children and youth (McCombs, Augustine, Unlu, et al., 2019).

[W]e acknowledge that the programs might have affected students in positive ways that our research was not designed to detect.

Here we highlight the key findings and insights we obtained from this longitudinal research, with an indication of the strength of evidence on which these findings are based.

- **Offering the summer learning program increased access to opportunities.** Far more treatment students participated in any summer program or camp or in a summer program or camp that included language arts and mathematics than students in the control group.
- **Offering the summer learning program provided short-term benefits to students in mathematics after one summer.** This finding is based on our causal analysis, which provided strong evidence of effects, without risk of selection bias. The causal analyses estimate the effect of *offering* the program, which underestimates the effect of the program for those who actually attended.

- **High rates of attendance and consecutive summers of attendance generated academic benefits.** After the first summer of programming, high attenders outperformed control group students in mathematics in the fall and on state assessments the following spring. After two summers of programming, high attenders and students attending both summers outperformed control group students in mathematics and language arts tests in the fall and the spring. These findings emerge from our correlational analyses, which carry a risk of the estimates being influenced by selection bias. However, because these findings are consistent with the causal analyses and we were able to control for prior achievement, we have confidence in them.

- **The amount and quality of instruction influenced the amount of academic benefit that attenders received from the program.** These findings are based on correlational analyses, which carry a risk of the estimates being influenced by selection bias. However, because these findings are consistent with the causal analyses and we were able to control for prior achievement, we have confidence in them.

- **After the second summer of programming, students with high attendance were rated higher on social-emotional skills than comparable control group students; however, this advantage did not persist into spring 2017**. This finding is based on correlational analyses and lacks a baseline measure. The estimates are also less precise than the achievement effect estimates and do not exhibit clear patterns over the course of the study. For these reasons, we do not have high confidence that the positive fall 2014 estimate for high attenders represents a meaningful positive effect.

- **The summer program did not affect outcomes that were not directly addressed in the program design and content at a level detectible in our study.** Across our analyses, we do not see evidence of program effects for outcomes that were not directly targeted by programming, such as suspension and attendance rates during the school year. These findings are consistent with a systematic evidence review of summer programs, which find that although the majority of programs were effective, most studied programs were not effective in generating statistically significant outcomes in all measured potential benefits, particularly those that were not directly addressed in program content (McCombs, Augustine, Unlu, et al., 2019).

- **Three school years after the summer program, academic benefits for high attenders decreased in magnitude and were not statistically significant, yet may be important in practical terms.** The magnitude of the benefits in language arts and mathematics observed in spring 2015 for high attenders over both summers declined by spring 2017 when the students were finishing seventh grade. Although no longer statistically significant, when benchmarked against typical achievement gains at the same grade level, they remained large enough to be important in practical terms. This finding is based on correlational analyses and the interpretation of the finding is informed by the causal analysis and by research on typical annual achievement growth measured in standardized effect sizes.

Implications for Policy and Practice

Urban districts should consider offering voluntary summer programs as part of their overall efforts to improve outcomes among students from low-income families and with low academic achievement, particularly if they can offer these programs over multiple consecutive summers.

Offering a five-week voluntary summer program with both academics and enrichment can produce short-term benefits in mathematics among late elementary students. Our review of the evidence on summer programs (McCombs, Augustine, Unlu, et al., 2019) found this effect to be "strong" (Tier I) under the standards set forth in the Every Student Succeeds Act (Public Law 114-95, Section 8101 [21] [A]). Therefore, a summer learning program following the NSLP model might be eligible for federal funding under this law if the program targets mathematics skills.

As defined by *Every Student Succeeds*, we find "promising" evidence (Tier III) that high-attending students reap benefits in mathematics and language arts, as do students who attend for two consecutive summers (in general or at high rates). These results can be used to demonstrate eligibility for federal funding if districts can establish a track record of high attendance in their summer programs.

Our findings also shape the expectations that we should have for five-week summer programs. The studied programs were found to be effective in terms of short-term mathematics achievement when looking at all students offered admission. After the second summer, high attenders and consecutive attenders benefited on fall and spring assessments in mathematics and language arts. The magnitudes of those estimates are commensurate with the short duration of the program.

The magnitude of the benefits do not persist or grow over time, but neither do they fade away completely. Offering one or two summers of programming in elementary school appears to be insufficient to significantly alter the learning trajectory of participants as they move through later schooling. Growing evidence regarding the longer-term effects of education interventions suggests that multiple ongoing efforts are needed to improve student achievement over time. High-quality summer programs are a viable option to consider as one of those efforts.

Benefits of the program were greatest for students who attended consecutive summers and those who had strong attendance. Districts should encourage students to attend regularly and for consecutive summers to maximize the academic impact of the programs. Furthermore, prior research shows that districts that offer summer programs sporadically or begin planning late in the year struggle to develop programs that are well implemented and well attended (McCombs, Augustine, Schwartz, et al, 2011; Schwartz et al., 2018). Offering summer programs consistently enables districts and their partners to work on improving the quality of programming each summer. The programs we studied targeted late elementary school students. Districts extending programming up through middle school may need to modify the NSLP model to ensure that it is developmentally appropriate for those students.

Offering summer programs consistently enables districts and their partners to work on improving the quality of programming each summer.

Districts offering voluntary summer programs that seek to provide academic benefits should offer at least five weeks of programming—and preferably six—with at least three hours of academic instruction per day.

Given the correlational evidence suggesting benefits for high attenders and the average daily attendance rates for these programs, districts should offer programs for at least five weeks to boost the number of students who attend more than 20 days.

Offering six or more weeks of programming could increase the proportion of students meeting this threshold of attendance.

To increase program effectiveness and maximize their return on investment, districts should focus on ensuring strong student attendance, productive use of instructional time, and high-quality instruction.

Our research identified key factors that were correlated with program effectiveness: attendance, productive use of instructional time, and instructional quality. This is not surprising, given the importance of these factors in education generally. Although all districts strive for this in their programs, our evaluation of program implementation found that execution of these priorities requires intentional planning (Schwartz et al., 2018). Districts and partners interested in learning how to plan and implement effective programs that provide positive experiences for students can find detailed implementation guidance in another report in this series—*Getting to Work on Summer Learning* (Schwartz et al., 2018)—that is freely available on the RAND and The Wallace Foundation websites. In addition, The Wallace Foundation's Knowledge Center includes a set of accompanying tools and resources that provide concrete examples and templates for districts and their partners developing voluntary summer learning programs.

APPENDIX

In this appendix, we discuss the details of the outcomes data used for analyses presented in this report and the attrition rates for each outcome, and we briefly summarize the statistical models used for estimating both causal and exploratory (nonexperimental) effects. We focus this discussion on modifications to these models that we made for the analyses presented in this report. Full details on the statistical models is reported in *Learning from Summer* (Augustine, McCombs, Pane, et al., 2016). The appendix concludes with tabulations of all spring 2017 causal results.

Data Used for Spring 2017 Analyses

We collected data related to five outcomes:

1. state assessments in mathematics and language arts administered in spring 2017
2. suspensions
3. end-of-year course grades in mathematics and language arts
4. school-year attendance
5. social-emotional competencies.

As in previously conducted analyses, state assessment scores in mathematics and language arts are standardized within each district, labeled A-E, using our study sample to have a mean of 0 and standard deviation of 1. In 2016–2017, there were students in all five districts who took tests that were not at grade level, particularly in mathematics. For example, some students in seventh grade took either the sixth-grade or eighth-grade state mathematics assessment. Often, these deviations in test-taking were associated with course enrollment. For example, in District B, students who took the eighth-grade mathematics assessment were enrolled in advanced seventh-grade mathematics courses.

We received end-of-course grades in mathematics and language arts for all districts. However, course-taking was examined as a substitute analysis for course grades analyses conducted in previous years. We explain this decision in more detail.

For suspensions, we created a variable (ever suspended) to indicate whether a student had been suspended (either in school or out of school) at least once during the 2016–2017 academic year. For any students who were missing all other spring outcomes (i.e., course grades, spring standardized test scores and school year attendance), suspension data were assumed missing. School-year attendance indicates the percentage of total school days in school year 2016–2017 that the student was marked as being in attendance.

Social-emotional competencies were measured using the DESSA–RRE, which was administered to school-year teachers who reported on the behaviors of individual study students.

Attrition Rates

Table A.1 lists the percentage missing (or attrition) for each of the main outcomes data categories that are color coded, aligned to the attrition rate boundaries set by the What Works Clearinghouse (WWC). For example, highlighted cells are above the WWC established threshold for acceptable attrition (i.e., attrition rates that meet WWC's attrition standards).

Generally, rates of total attrition (treatment and control groups combined) for the full sample are in the range of 30 percent to 40 percent for all outcomes except DESSA-RRE, which had nearly 50-percent attrition. However, these values vary widely across districts, with Duval County presenting the best overall attrition rates and Rochester presenting the worst. Boston has particularly bad attrition for DESSA-RRE.

TABLE A.1
Overall Attrition of the Experimental Sample for Spring 2017 Outcomes

		Percentage Missing (Spring 2017 Data)					
		Mathematics		Language Arts	Social-Emotional/Behavioral		
District	Original Sample	State Test	Course-Taking	State Test	DESSA-RRE	Attendance	Suspension
A	888	21%	20%	18%	39%	15%	15%
B	2,056	35%	35%	36%	51%	33%	33%
C	957	38%	48%	39%	65%	33%	36%
D	1,080	56%	39%	52%	47%	40%	40%
E	656	34%	30%	34%	39%	26%	26%
Full sample	5,637	38%	35%	36%	49%	31%	31%

NOTE: Red shading indicates an overall attrition rate that is too high to meet WWC attrition standards. Yellow shading indicates an overall attrition rate that is too high to meet WWC standards under conservative assumptions. WWC standards also consider differential attrition, which is not shown in the table but is discussed in the text.

For the full sample, there were no appreciable attrition differences between the treatment and control groups (not shown in the table). The treatment group displayed slightly higher attrition rates between 0.1 and 0.6 percentage points across these variables (approximately 0 for suspension data). Thus, the combination of overall and differential attrition rates for the full-sample attrition are within WWC standards for low attrition even under conservative attrition standards, and analyses are eligible to meet WWC standards without reservations provided the study uses an acceptable approach to address missing data (WWC, 2017).

Although attrition does not threaten bias for the full-sample analyses reported in the main text, this is not necessarily the case for the district-specific estimates reported in this appendix. The high overall attrition in District C for the DESSA-RRE and District D for the mathematics state test are highlighted in Table A.1. Not shown in the table, differential attrition was considerable in District E, where there was higher attrition in the treatment group across all variables, by 6.7 to 10 percentage points (all significant except DESSA-RRE). Although it may be interesting to explore possible explanations for these differences, we believe this differential attrition in District E jeopardizes the validity or eligibility to meet WWC standards only for estimates specific to District E. Differential attrition in the other districts was not appreciable.

Analytic Approach

As detailed in previous reports (Augustine, McCombs, Pane, et al., 2016), our preferred approach for estimating causal effects uses an intention-to-treat (ITT) approach, comparing the outcomes of all students who were randomly admitted to two summers of programming (2013 and 2014) with the outcomes of all students who were randomly assigned to the control group, regardless of whether the students actually attended the summer program. These analyses produce the causal effect estimates reported in the main text using the following model:

where:

$$Y_{qispc} = \alpha T_{ispc} + \beta X_{ispc} + \delta PreTestMean_{qc} + \gamma_s + \pi_p * Z_{ispc} + \mu_c * Z_{ispc} + \varepsilon_{qispc}$$

- Y_{qispc} is the standardized post-test score in subject q for student i in strata s in summer site p in summer classroom c, where p and c are defined to be 0 for students who did not attend the summer program.

- T_{ispc} is an indicator of assignment to the treatment group
- X_{ispc} is a vector of baseline covariates
- $PreTestMean_{qc}$ is a vector of mean pretest values of all students who were assigned to the same summer classroom in subject *q*. This is 0 for all students who did not attend the summer program. There are four classroom means, one for each of the four pretests (spring 2013 mathematics and language arts, and the earlier assessments in mathematics and language arts that were used for stratification).
- γ_s are strata fixed-effects (dummy variables)
- Z_{ispc} is an indicator variable taking a value of 1 if a student is a member of class *c* in site *p*, used to define random effects. Every student in summer site *p* and classroom *c* is associated with a random effect, including those students assigned to treatment who take up the program and those students assigned to control who take up the program (i.e., "crossovers"). This is 0 for the rest of the control group students and all treatment group students who do not take up treatment ("no shows").
- $\pi_p * Z_{ispc}$ is a random-effect common to all students in summer site *p*.
- $\mu_c * Z_{ispc}$ is a random-effect common to all students in summer classroom *c*.
- $\breve{\ }_{qispc}$ is a residual, the variance of which is allowed to vary by pattern of available pretests.

Details about the complete set of baseline covariates and software implementations of these models are reported in *Learning from Summer* (Augustine, McCombs, Pane, et al., 2016). Several outcomes, such as student suspension and course-taking, involved a binary outcome (i.e., 1 if suspended once or more during school year, 0 otherwise). Consistent with past reports (Augustine, McCombs, Pane, et al., 2016), these outcomes are analyzed using linear probability models, which use binary outcomes in the random effects modeling framework.

In addition to the ITT estimates, we also estimated the causal effect of attending the summer program (i.e., the treatment-on-the-treated or TOT effects) using two-stage least squares models.

Although TOT results are not discussed in Chapter Two of this report, these results are available in Table A.9. By using randomization status as an instrumental variable for program attendance (defined as the student appearing at least one time, regardless of which summer), these models control for endogenous selection into program attendance (additional details are available in Augustine, McCombs, Pane, et al., 2016).

Correlational analyses use simple extensions to this model. In the case of models for attendance and academic time on task, the treatment assignment indicator was replaced with continuous or categorical variables for these mediators. (For details, see Augustine, McCombs, Pane, et al., 2016.) In these models, high attendance is defined as 20 or more days of attendance in a particular summer and high academic time on task is defined as receiving 25.5 hours of mathematics instruction or 34 hours of language arts instruction in a particular summer.

Except where the project administered a common measure in all districts (DESSA-RRE, as well as GMADE and GRADE in earlier rounds of analysis), our approach, as detailed in previous reports, has been to conduct analyses using this model within each district and use fixed-effects meta-analysis techniques to produce overall results. The meta-analysis weights each district-level result by its precision, which is very similar to weighting by sample size.

Multiple Imputation Remedy for Missing Data

In the spring 2017 data, we observed large differences in attrition by district (see Table A.1). For example, District D, which had the second-largest sample in the study, now ranks fourth in the number of student observations available. District A, which was fourth originally, now ranks second. These changes affect the relative precisions of the district-level estimates, causing them to become more or less influential in the overall result produced by meta-analysis. Under this approach, the overall estimates look more like those of District A and less like those of District D than they did in prior years. This can complicate or mislead interpretation of changes in estimates of overall program impact over time.

Our remedy was to adopt a different WWC-approved approach for addressing missing data—specifically, multiple imputation. In deciding to implement multiple imputation, we adhered closely to guidance provided in the WWC Version 4.0 standards (2017).

In multiple imputation, each missing value in the data set is replaced with a plausible value. This process is repeated multiple times to incorporate the uncertainty that is involved in identifying and selecting those plausible values. We implemented this procedure using an R package called mice, which automatically generates plausible values for each missing data point (Van Buuren and Groothuis-Oudhoorn, 2011) and also enables subsequent analysis to incorporate the uncertainty in these plausible values. Multiple imputation enables us to retain all students in the original sample in the district-level analyses, making the precision of the district-level estimates, and thus their weight in the meta-analytic overall estimate, more similar to prior rounds of analysis. This makes it easier to interpret changes in impact estimates over time.

The use of multiple imputation also enabled us to address off-grade-level testing of some students. Rather than excluding these students from main analysis of state test scores because their on-grade-level test was missing, multiple imputation estimated plausible scores that represent the scores these students would have received had they been administered the on-grade-level state test.[6] For all analyses, we used multiple imputation with both missing covariate and missing outcome values.

Our overall approach to estimating ITT and TOT estimates was similar to the methods described in the technical appendix of Augustine, McCombs, Pane, et al. (2016). However, several modifications were made to facilitate the use of multiple imputation. In our original models, we used mean imputation for the spring 2013 (baseline) assessment scores for all missing values, and missing value indicators were incorporated into the covariate set. Additionally, the random effects model allowed for separate variance structures to be estimated for nonmissing and missing scores. Because we switched to a multiple imputation framework, it was no longer necessary to use mean imputation for missing baseline assessment scores, or to use missing value indicators or separate variance structures.

[6] Another approach to handling the off-grade-level testing is to put it on the same standardized scale. However, this approach potentially conceals important differences in test difficulty. For example, below-average scores on the eighth-grade assessment are not equivalent to below-average scores on the seventh-grade assessment, and it is plausible that a student who took the eighth-grade assessment would have had higher scores on the seventh-grade assessment had that student taken that test.

Our first sensitivity test was to ensure that these modifications to estimation did not affect our results. We reran 2015 analyses for mathematics and language arts state test scores in the multiple imputation framework and confirmed that they produced numerically and substantively similar results to those from the original analytic framework. This provided us with evidence that using multiple imputation did not affect findings.

Second, we confirmed that the substantive conclusions about spring 2017 outcomes—that there were no significant treatment effects—are the same under the new and old analytic frameworks, with some notable exceptions. When we used the old analytic framework, we obtained a significant positive ITT effect of 0.05 on language arts achievement, and a significant positive TOT effect of 0.07. There were also positive effects for students who had lower or higher academic time on task in summer 2014 (0.17 and 0.12, respectively). When we used the new multiple imputation–based framework, these estimates were smaller and no longer significant. The estimates obtained under the old framework seemed less trustworthy because the estimated effects were larger for lower academic time on task in summer 2014 than for higher academic time on task in both summers, and larger than the 2015 result for higher academic time on task both summers. This, coupled with the fact that there was differential attrition across districts that influenced the relative contributions of districts to the meta-analysis, led us to believe that these significant results based on the old analytic framework were spurious.

As a final check on the imputation methods, we confirmed that students who were administered tests that were above grade level received relatively higher imputed scores on the on-grade-level test, and students who were administered tests that were below grade level received relatively lower imputed scores.

Data Issues With Course Grades and Substitute Analysis

As students get older and move into middle school, they substantially diversify their course-taking. In the 2017 grades data, we found that students in our sample were enrolled in many different mathematics and language arts courses within each district. For example, in District D, we had grades from 18 different courses, with sample sizes ranging from one student to 226 students. At another extreme, District B gave us course grades with no identifying course information. In all of the districts, we suspect that the courses that students took varied in difficulty (meaning a grade of “B” in one course was not the same as a “B” grade in another). We had no way of accounting for or understanding differences in course difficulty, and small sample sizes per course also posed problems for estimation. As a result, we concluded that it was impossible to conduct a valid analysis of the treatment effects on grades.

However, through a clarification process with district personnel and assumptions related to the tested grade level for the state mathematics assessments, we were able to classify student course-taking in mathematics into three rough categories: below, at, and above grade level. We were not able to obtain similar data for language arts courses. Thus, as a substitute for analysis of effects on grades, we analyze effects on the likelihood of a student taking a mathematics course that is below grade level and, similarly, the likelihood of a student taking a mathematics course that is above grade level.

Tables of the Results of All 2017 Analyses

In Tables A.2 through A.10, we present tabulations of all spring 2017 causal and correlational results. These causal results represent the effects of being admitted to the summer programs (i.e., ITT effects). We begin by presenting causal and correlational results for mathematics outcomes. We then present results for language arts outcomes. Third, we present social-emotional and behavioral results. We conclude this section with tables of causal effects of attending in the summer program (i.e., the TOT effects).

TABLE A.2

Overall Causal Effects of Summer Learning Programs on Mathematics Outcomes for All Treatment Group Students Relative to All Control Group Students

2017 Spring State Assessment	2017 Increased Above-Grade-Level Course-Taking	2017 Decreased Below-Grade-Level Course-Taking
0.01 (0.03)	0.03 (0.02)	0.01 (0.01)

NOTE: Standard error is shown in parentheses.

TABLE A.3

Correlational Effects of Attending Two Years of Summer Program Attendance, Mathematics Achievement

Analyses	2017 Spring State Assessment	2017 Increased Above-Grade-Level Course-Taking	2017 Decreased Below-Grade-Level Course-Taking
Attendance category			
Attended summer 2013 only	–0.05 (0.04)	0.03 (0.03)	0.01 (0.01)
Attended summer 2014 only	0.01 (0.07)	0.02 (0.06)	0.01 (0.02)
Attended both summers	0.05 (0.04)	0.05 (0.03)	0.01 (0.01)
Consecutive high attendance	0.07 (0.04)	0.06 (0.04)	0.01 (0.01)

NOTE: Standard error is shown in parentheses. Consecutive high attendance is defined as attending 20 or more days each summer.

TABLE A.4

Correlational Effects of Academic Time on Task, Mathematics Achievement

Analyses	2017 Spring State Assessment	2017 Increased Above-Grade-Level Course-Taking	2017 Decreased Below-Grade-Level Course-Taking
Academic time on task			
No-show	–0.02 (0.03)	0.03 (0.02)	0.00 (0.01)
Low	0.02 (0.04)	0.01 (0.03)	0.02 (0.02)
High	0.07 (0.04)	0.01 (0.03)	0.01 (0.01)
High both summers	0.07 (0.06)	0.03 (0.03)	0.01 (0.01)

NOTE: Standard error is shown in parentheses. High academic time on task is defined as 25.5 or more hours of instruction for mathematics.

TABLE A.5

Causal Effects of Summer Learning Programs on Language Arts Achievement for All Treatment Group Students Relative to All Control Group Students

2017 Spring State Assessment
0.02 (0.03)

NOTE: Standard error is shown in parentheses.

TABLE A.6

Correlational Effects of Attending Two Years of Summer Program Attendance, Language Arts Achievement

Analyses	2017 Spring State Assessment
Attendance category	
Attended summer 2013 only	–0.01 (0.04)
Attended summer 2014 only	0.02 (0.07)
Attended both summers	0.03 (0.03)
Consecutive high attendance	0.04 (0.04)

NOTE: Standard error is shown in parentheses. Consecutive high attendance is defined as attending 20 or more days each summer.

TABLE A.7

Correlational Effects of Academic Time on Task, Language Arts Achievement

Analyses	2017 Spring State Assessment
Academic time on task	
No-show	0.02 (0.03)
Low	0.03 (0.04)
High	0.03 (0.05)
High both summers	0.05 (0.06)

NOTE: Standard error is shown in parentheses. High academic time on task is defined as 34 or more hours of instruction for language arts.

TABLE A.8

Overall Causal Effects of Summer Learning Programs on Social-Emotional and Behavioral Outcomes for All Treatment Group Students Relative to All Control Group Students

2017 DESSA-RRE	2017 Reduced School-Year Suspension Rate	2017 Improved School-Year Attendance Rate
–0.03 (0.04)	0.01 (0.01)	0.00 (0.01)

NOTE: Standard error is shown in parentheses.

TABLE A.9

Correlational Effects of Attending Two Years of Summer Program Attendance, Social-Emotional and Behavioral Outcomes

Analyses	DESSA-RRE	Reduced School-Year Suspension Rate
Attendance category		
Attended summer 2013 only	–0.02 (0.05)	0.00 (0.02)
Attended summer 2014 only	0.00 (0.08)	0.01 (0.02)
Attended both summers	–0.02 (0.05)	0.00 (0.01)
Consecutive high attendance	0.03 (0.05)	0.01 (0.01)

NOTE: Standard error is shown in parentheses. Consecutive high attendance is defined as attending 20 or more days each summer.

TABLE A.10

Causal Effects of Summer Learning Programs on Outcomes for Treatment Group Students Who Attended the Summer Program (Treatment Effect on the Treated)

Analyses	Estimate (SE)
Mathematics	
Spring 2017 state assessment	0.02 (0.03)
Increased above-grade-level course-taking	0.00 (0.01)
Decreased below-grade-level course-taking	0.01 (0.01)
Language arts	
Spring 2017 state assessment	0.03 (0.03)
Social-emotional outcomes	
DESSA-RRE	–0.01 (0.03)
Behavioral outcomes	
Reduced school-year suspension rate	0.01 (0.01)
Improved school-year attendance rate	0.00 (0.00)

NOTE: Standard error is shown in parentheses.


Oxidation

REFERENCES

Afterschool Alliance, *America After 3PM: Afterschool Programs in Demand*, Washington, D.C., October 2014. As of September 28, 2019: http://www.afterschoolalliance.org/documents/AA3PM-2014/AA3PM_National_Report.pdf

Atteberry, Allison, and Andrew McEachin, "School's Out: Summer Learning Loss Across Grade Levels and School Contexts in the U.S. Today," in K. Alexander, S. Pitcock, and M. Boulay, eds., *The Summer Slide: What We Know and Can Do About Summer Learning Loss*, New York: Teachers College Press, 2016, pp. 35–54.

Augustine, Catherine H., Jennifer Sloan McCombs, John F. Pane, Heather L. Schwartz, Jonathan Schweig, Andrew McEachin, and Kyle Siler-Evans, *Learning from Summer: Effects of Voluntary Summer Learning Programs on Low-Income Urban Youth*, Santa Monica, Calif.: RAND Corporation, RR-1557-WF, 2016. As of October 28, 2018: https://www.rand.org/pubs/research_reports/RR1557.html

Augustine, Catherine H., Jennifer Sloan McCombs, Heather L. Schwartz, and Laura Zakaras, *Getting to Work on Summer Learning: Recommended Practices for Success*, Santa Monica, Calif.: RAND Corporation, RR-366-WF, 2013. As of October 31, 2018: https://www.rand.org/pubs/research_reports/RR366.html

Bailey, Drew H., Greg J. Duncan, Candice L. Odgers, and Winnie Yu, "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions," *Journal of Research on Educational Effectiveness,* Vol. 10, No. 1, 2017, pp. 7–39.

Bailey, Drew H., Lynn S. Fuchs, Jennifer K. Gilbert, David C. Geary, and Douglas Fuchs, "Prevention: Necessary but Insufficient? A 2-Year Follow-Up of an Effective First-Grade Mathematics Intervention," *Child Development*, 2018.

Barnett, W. Steven, "Effectiveness of Early Educational Intervention," *Science*, Vol. 333, No. 6045, 2011, pp. 975–978.

Benson, James, and Geoffrey D. Borman, "Family, Neighborhood, and School Settings Across Seasons: When Do Socioeconomic Context and Racial Composition Matter for the Reading Achievement Growth of Young Children?" *Teacher's College Record*, Vol. 112, No. 5, 2010, pp. 1338–1390.

Burkham, David T., Douglas D. Ready, Valerie E. Lee, and Laura E. LoGerfo, "Socio-Class Differences in Summer Learning Between Kindergarten and First Grade: Model Specification and Estimation," *Sociology of Education*, Vol. 77, No. 1, January 2004, pp. 1–31.

Chetty, Raj, John N. Friedman, and Jonah Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, Vol. 104, 2014, pp. 2633–2679.

Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsay, and Scott Greathouse, "The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review," *Review of Educational Research*, Vol. 66, No. 3, 1996, pp. 227–268.

Deming, David, "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start," *American Economic Journal: Applied Economics*, Vol. 1, No. 3, 2009, pp. 111–134.

Dodge, Kenneth A., Karen L. Bierman, John D. Coie, Mark T. Greenberg, John E. Lochman, Robert J. McMahon, and Ellen E. Pinderhughes, "Impact of Early Intervention on Psychopathology, Crime, and Well-Being at Age 25," *The American Journal of Psychiatry*, Vol. 172, No. 1, 2015, pp. 59–70.

Downey, Douglas B., Paul T. Von Hippel, and Beckett A. Broh, "Are Schools the Great Equalizer? Cognitive Inequality During the Summer Months and the School Year," *American Sociological Review*, Vol. 69, No. 5, 2004, pp. 613–635.

Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger, "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions," *Child Development*, Vol. 82, No. 1, 2011, pp. 405–432.

Fitzpatrick, Maria D., David Grissmer, and Sarah Hastedt, "What a Difference a Day Makes: Estimating Daily Learning Gains During Kindergarten and First Grade Using a Natural Experiment," *Economics of Education Review*, Vol. 30, No. 2, 2011, pp 269–279.

Gershenson, Seth, "Do Summer Time-Use Gaps Vary by Socioeconomic Status?" *American Educational Research Journal*, Vol. 50, No. 6, 2013, pp. 1219–1248.

Kim, James S., "Summer Reading and the Ethnic Achievement Gap," *Journal of Education for Students Placed at Risk*, Vol. 9, No. 2, 2004, pp. 169–188.

Lipsey, Mark W., Dale C. Farran, and Kelley Durkin, "Effects of the Tennessee Prekindergarten Program on Children's Achievement and Behavior Through Third Grade," *Early Childhood Research Quarterly*, Vol. 45, 4th Quarter 2018, pp. 155–176.

Lipsey, Mark W., Kelly Puzio, Cathy Yun, Michael A. Herbert, Kasia Steinka-Fry, Mikel W. Cole, Megan Roberts, Karen S. Anthony, and Matthew D. Busick, *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms*, Washington, D.C.: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education, NCSER 2013-3000, 2012.

McCoach, D. Betsy, Ann A. O'Connell, Sally M. Reis, and Heather A. Levitt, "Growing Readers: A Hierarchical Linear Model of Children's Reading Growth During the First Two Years of School," *Journal of Educational Psychology*, Vol. 98, No. 1, 2006, pp. 14–28.

McCombs, Jennifer Sloan, Catherine H. Augustine, Heather L. Schwartz, Susan J. Bodilly, Brian McInnis, Dahlia S. Lichter, and Amanda Brown Cross, *Making Summer Count: How Summer Programs Can Boost Children's Learning*, MG-1120-WF, Santa Monica, Calif.: RAND Corporation, 2011. As of March 16, 2020: https://www.rand.org/pubs/monographs/MG1120.html

McCombs, Jennifer Sloan, Catherine H. Augustine, Fatih Unlu, Kathleen M. Ziol-Guest, Scott Naftel, Celia J. Gomez, Terry Marsh, Goke Akinniranye, and Ivy Todd, *Investing in Successful Summer Programs: A Review of Evidence Under the Every Student Succeeds Act*, Santa Monica, Calif.: RAND Corporation, RR-2836-WF, 2019. As of February 27, 2020: https://www.rand.org/pubs/research_reports/RR2836.html

McCombs, Jennifer Sloan, John F. Pane, Catherine H. Augustine, Heather L. Schwartz, Paco Martorell, and Laura Zakaras, *Ready for Fall? Near-Term Effects of Voluntary Summer Learning Programs on Low-Income Students' Learning Opportunities and Outcomes*, Santa Monica, Calif.: RAND Corporation, RR-815-WF, 2014. As of February 27, 2020: https://www.rand.org/pubs/research_reports/RR815.html

National Academies of Sciences, Engineering, and Medicine, *Shaping Summertime Experiences: Opportunities to Promote Healthy Development and Well-Being for Children and Youth*. Washington, D.C.: National Academies Press, 2019.

National Center for Education Statistics, "Table 2. Public High School 4-Year Adjusted Cohort Graduation Rate (ACGR), by Race/Ethnicity and Selected Demographics for the United States, the 50 States, and the District of Columbia: School Year 2012-13," *EDFacts/Consolidated State Performance Report, SY 2012-13*, Washington, D.C.: U.S. Department of Education, February 2015. As of July 6, 2018: https://nces.ed.gov/ccd/tables/ACGR_RE_and_characteristics_2012-13.asp

Pell Institute, *Indicators of Higher Education Equity in the United States—45 Year Trend Report*, Washington, D.C., December 17, 2015. As of July 6, 2018: http://www.pellinstitute.org/publications-Indicators_of_Higher_Education_Equity_in_the_United_States_45_Year_Report.shtml

Pew Research Center, *Parenting in America: Outlook, Worries, Aspirations Are Strongly Linked to Financial Situation*, Washington, D.C., December 17, 2015.

Public Law 114-95, Every Student Succeeds Act, December 10, 2015.

Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, Pam Broene, Frank Jenkins, Andrew Mashburn, and Jason Downer, *Third-Grade Follow-Up to the Head Start Impact Study: Final Report*, OPRE Report 2012-45, Washington, D.C.: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, October 2012.

Ready, Douglas D., "Socioeconomic Disadvantage, School Attendance, and Early Cognitive Development: The Differential Effects of School Exposure," *Sociology of Education*, Vol. 83, No. 4, 2010, pp. 271–286.

Reardon, Sean F., "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations," in Greg J. Duncan and Richard J. Murnane, eds., *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, New York: Russell Sage Foundation, 2011, pp. 1033–1681.

Redford, Jeremy, Stephanie Burns, and L. Jane Hall, *The Summer After Kindergarten: Children's Experiences by Socioeconomic Characteristics*, Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2018.

Schacter, John, and Booil Jo, "Learning When School Is Not in Session: A Reading Summer Day-Camp Intervention to Improve the Achievement of Exiting First-Grade Students Who Are Economically Disadvantaged," *Journal of Research in Reading*, Vol. 28, No. 2, 2005, pp. 158–169.

Schwartz, Heather L., Jennifer Sloan McCombs, Catherine H. Augustine, and Jennifer Tamargo, *Getting to Work on Summer Learning: Recommended Practices for Success*, 2nd ed., Santa Monica, Calif.: RAND Corporation, RR-366-1-WF, 2018. As of October 28, 2018:
https://www.rand.org/pubs/research_reports/RR366-1.html

Taylor, Rebecca D., Eva Oberle, Joseph A. Durlak, and Roger P. Weissberg, "Promoting Positive Youth Development Through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-Up Effects," *Child Development*, Vol. 88, No. 4, 2017, pp. 1156–1171.

Van Buuren, Stef, and Karin Groothuis-Oudshoorn, "mice: Multivariate Imputation by Chained Equations in R," *Journal of Statistical Software*, Vol. 45, No. 3, 2011, pp. 1–67.

Von Hippel, Paul T., and Caitlin Hamrock, "Do Test Score Gaps Grow Before, During, or Between the School Years? Measurement Artifacts and What We Can Know in Spite of Them," *Sociological Science*, Vol. 6, No. 3, 2019, pp. 43–80.

Von Hippel, Paul T., Caitlin Hamrock, and Mina Kumar, "Do Test Score Gaps Grow Before, During or Between the School Years? Measurement Artifacts and What We Can Know in Spite of Them," working paper, Social Science Research Network, 2016.

Weare, Katherine, and Melanie Nind, "Mental Health Promotion and Problem Prevention in Schools: What Does the Evidence Say?" *Health Promotion International* , Vol. 26, No. S1, 2011, pp. i29–i69.

What Works Clearinghouse, *What Works Clearinghouse Standards Handbook*, Version 4.0, Washington, D.C.: Institute of Education Sciences, U.S. Department of Education, 2017.

White, Thomas G., James S. Kim, Helen C. Kingston, and Lisa Foster, "Replicating the Effects of a Teacher-Scaffolded Voluntary Summer Reading Program: The Role of Poverty," *Reading Research Quarterly*, Vol. 49, No. 1, 2013, pp. 5–30.

WWC—*See* What Works Clearinghouse

Zvoch, Keith, and Joseph J. Stevens, "Summer School Effects in a Randomized Field Trial," *Early Childhood Research Quarterly*, Vol. 28, No. 1, 2013, pp. 24–32.

The National Summer Learning Project (NSLP) examined the implementation and effectiveness of voluntary summer learning programs developed by five school districts—Boston, Massachusetts; Dallas, Texas; Duval County, Florida; Pittsburgh, Pennsylvania; and Rochester, New York—and their local community partners. The study spanned three phases. The RAND research team (1) collected formative data for strengthening the five summer programs in 2011 and 2012; (2) examined student outcomes after one summer (2013) and after two summers of programming (2014 and 2015); and (3) examined student outcomes in spring 2017, at the end of three school years after the second summer of programming. This seventh report in a series summarizes the findings of this third phase in the context of earlier findings and offers implications for policy and practice. Overall long-term findings show that, by spring 2017, the academic benefits for high attenders decreased in magnitude and were not statistically significant—although when benchmarked against typical achievement gains at the same grade level, they remained large enough to be educationally meaningful.

Commissioned by The Wallace Foundation

$33.00

RAND EDUCATION AND LABOR

www.rand.org

ISBN-10 1-9774-0451-0
ISBN-13 978-1-9774-0451-0

53300

9 781977 404510

RR-3201-WF