ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 21-379

# The Impact of Summer Programs on Student Mathematics Achievement: A Meta-Analysis

Kathleen Lynch
University of Connecticut

Lily An
Harvard University

Zid Mancenido
Harvard University

We present results from a meta-analysis of 37 contemporary experimental and quasi-experimental studies of summer programs in mathematics for children in Grades pre-K-12, examining what resources and characteristics predict stronger student achievement. Children who participated in summer programs that included mathematics activities experienced significantly better mathematics achievement outcomes, compared to their control group counterparts. We find an average weighted impact estimate of +0.10 standard deviations on mathematics achievement outcomes. We find similar effects for programs conducted in higher- and lower-poverty settings. We undertook a secondary analysis exploring the effect of summer programs on non-cognitive outcomes and found positive mean impacts. The results indicate that summer programs are a promising tool to strengthen children's mathematical proficiency outside of school time.

VERSION: July 2022

Suggested citation: Lynch, Kathleen, Lily An, and Zid Mancenido. (2022). The Impact of Summer Programs on Student Mathematics Achievement: A Meta-Analysis. (EdWorkingPaper: 21-379). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/da7r-4z83

# The Impact of Summer Programs on Student Mathematics Achievement: A Meta-Analysis

Kathleen Lynch, Lily An, & Zid Mancenido

**Abstract**

We present results from a meta-analysis of 37 contemporary experimental and quasi-experimental studies of summer programs in mathematics for children in Grades pre-K-12, examining what resources and characteristics predict stronger student achievement. Children who participated in summer programs that included mathematics activities experienced significantly better mathematics achievement outcomes, compared to their control group counterparts. We find an average weighted impact estimate of +0.10 standard deviations on mathematics achievement outcomes. We find similar effects for programs conducted in higher- and lower-poverty settings. We undertook a secondary analysis exploring the effect of summer programs on non-cognitive outcomes and found positive mean impacts. The results indicate that summer programs are a promising tool to strengthen children's mathematical proficiency outside of school time.

*Keywords*: Summer learning, summer programs, summer school, mathematics, learning loss

# The Impact of Summer Programs on Student Mathematics Achievement: A Meta-Analysis

The critical need to improve children's mathematics performance is a widely documented problem. Three out of four low-income children in the U.S. fail to meet standards for mathematical proficiency in the fourth grade, as do 43% of middle-income children (McFarland et al., 2017), and sizeable income-related gaps in mathematics achievement are also evident in cross-national research (Chmielewski & Reardon, 2016). Due to the cumulative nature of mathematical knowledge (Hiebert & Wearne, 1996), early difficulties in mathematical understanding can diminish children's likelihood of later success in advanced mathematics coursework—a key gatekeeper to science, technology, engineering, and mathematics (STEM) careers (National Research Council [NRC], 2011). Given the significant wage premium of STEM employment (Deming & Noray, 2020), unequal access for children from economically disadvantaged backgrounds can effectively inhibit opportunities for socioeconomic mobility and reinforce social inequality (Carter, 2006).

To address these inequities, high-poverty school districts frequently operate summer programs to help struggling students recover academic ground and avoid grade repetition (Jacob & Lefgren, 2004; Mariano & Martorell, 2013; Matsudaira, 2008). These programs mostly focus on reading and mathematics, domains considered foundational for student learning.

Despite the ubiquity of summer school and the critical need to strengthen students' mathematics ability, we lack contemporary evidence on the impacts of summer programs on mathematics learning, and an understanding of what features predict stronger student impacts. In the current study, we present results of a meta-analysis of the effects of summer mathematics programs. Specifically, we address the following research questions:

1. What are the main effects of summer programs on mathematics achievement?
2. What program activities, resources, and study characteristics moderate the effectiveness of summer programs in mathematics?

We also present one of the first efforts to synthesize the impacts of academic summer programs on outcomes beyond achievement through the following question: What is the relationship between summer mathematics learning programs and children's non-cognitive outcomes, such as attendance and motivation? To address these questions, we use data from 37 contemporary studies. Both outcomes we examine, mathematics achievement and non-cognitive outcomes, are of strong relevance to research and policy (e.g., McKown, 2017).

This work is especially timely given the impacts of COVID-19. While estimates of the educational ramifications of the COVID-19 pandemic to date have varied (e.g., Kuhfeld et al., 2020; Pier et al., 2021), it is generally acknowledged that inequity has been exacerbated and that substantial efforts are needed to help low-income students recover (Darling-Hammond et al., 2020). Summer school is a key policy mechanism for addressing these learning disruptions. A notable example is the American Rescue Plan Act of 2021, which allocated $29 billion for "planning and implementing activities related to summer learning and supplemental afterschool programs, including providing classroom instruction or online learning during the summer months."

**The Importance of Summer Programs**

A robust history of research has investigated the potential for seasonal school closures to exacerbate inequalities in children's learning. Early research studies comparing children's learning trajectories across seasons often indicated that low-income children were disproportionately affected, particularly in reading (e.g., Cooper et al., 1996; Downey et al.,

2004; Heyns, 1978), and that summer learning disparities may contribute to long-run achievement gaps (Alexander et al., 2007). These early studies drew attention to the influence of summer learning and were generative to the field; however, their focus was often limited to specific school districts or grade levels, and many used test scores that were not vertically linked (von Hippel & Hamrock, 2019). More recent research has posited the sensitivity of conclusions about summer gap widening to, for example, choice of whether pretest scores are included in models of summer learning gains (Dumont & Ready, 2020; Quinn, 2015). The issue of measuring summer learning and parsing its potential contribution to inequality remains an active source of scholarly inquiry (Atteberry & McEachin, 2021) and debate (e.g., von Hippel, 2019a, and Alexander, 2019). However, there is general agreement that children learn reading and mathematics more slowly during the summer than the school year, and that summer therefore affords children opportunities to catch up and to enrich their learning (e.g., von Hippel, 2019b).

The challenge for policymakers and families is that during summer vacation, the school resources 'faucet' is turned off (Borman et al., 2005). As a result, children's summer time use is often determined by family resources. Children from more advantaged families are more likely to participate in summer camps and enrichment activities, whereas low-income children are disproportionately exposed to TV (Burkam et al., 2004). Many attribute these patterns to cost: typical weekly summer program tuition in the U.S. in 2013 was $288, which is over 60% of the household income for a family of four at the federal poverty threshold (Afterschool Alliance, 2015). In response to these issues, many school districts have adopted summer learning programs to advance remediation and equity goals, supported in part by research indicating that extending school time can support student learning for those at risk of school failure (Patall et al., 2010).

**Previous Reviews of Research on Summer Programs**

The first systematic review of the impact of summer programs was undertaken by Cooper et al. (2000) who conducted a meta-analysis of summer school programs focused on remediation, primarily in reading and mathematics. Pooling mathematics and reading outcomes, the review found that pretest-posttest only studies (with no control group) had an average effect size of 0.30 SD ($k$ = 81) using a random-effects model, while studies employing a comparison group had an average effect size of 0.09 SD ($k$ = 44). The authors labeled results from randomized experiments as most trustworthy ($d$ = 0.14, $k$ = 11). Cooper et al. reported that the benefits of summer school were larger for middle-class than low-income children, but did not conduct moderator analyses separately for mathematics versus reading. By contrast, Kim and Quinn (2013) meta-analyzed summer reading programs and concluded that summer reading had larger impacts on lower-income children compared to mixed-income samples.

Lauer et al. (2006) undertook a meta-analysis of the impacts of out-of-school time programs targeting students at risk for school failure. The authors reviewed 35 studies evaluating after-school and summer programs and reported a pooled mean mathematics effect size for summer programs of 0.09 SD (fixed-effects model). The authors did not find a consistent relationship between program duration and effect size magnitude, but did find that effect sizes were significantly greater than zero only for programs that lasted more than 45 hours. The review included 12 studies of summer school programs that reported mathematics achievement outcomes, only one of which was judged of high research quality. The most recent included study was published in 2002. In addition, because the study's moderator analyses did not disaggregate after-school versus summer school programs, it could not disentangle specific factors that predict positive impacts of summer school.

More recently, there have been two narrative reviews of the research literature. McCombs et al. (2019) collected information on summer programs that met criteria for 'evidence-based interventions' required under the Every Student Succeeds Act (ESSA). The authors provided descriptive summaries of 43 programs that showed evidence of effectiveness, targeting domains including academics, social support, and employment/career readiness. The authors concluded that they were unable to determine why some summer programs were effective while others were not, due in part to limited available implementation data. Meanwhile, the National Academies of Sciences, Engineering, and Medicine (NASEM, 2019) conducted a narrative synthesis of the evidence of the impacts of summer youth programs that targeted physical and mental health, safety, social skills, and academic learning. Based on themes they gleaned from the literature along with expert opinions, they concluded that summer programs appeared to be more successful when content was aligned with both desired outcomes and student needs, when student attendance was high, and when programs were of sufficient duration.

**The Present Study**

The current study differs from previous reviews in several key respects. First, the most recent research studies included in prior meta-analyses of summer mathematics programs are nearly two decades old, and use samples and methodologies that are now dated. As a point of reference, the What Works Clearinghouse generally does not review studies that are more than 20 years old due to considerable changes in educational environments and interventions over time (WWC, n.d.). As noted above, recent research has synthesized the updated literature on summer reading (Kim & Quinn, 2013); our study synthesizes the contemporary evidence in summer mathematics.

We conjectured that using contemporary data, the overall estimated impacts of summer mathematics programs, as well as conclusions about relative effects in lower- versus higher-income settings, may differ from Cooper et al. for two primary reasons. First, the mean effect size from the current meta-analysis may be expected to be smaller due to the stronger research designs generally employed in contemporary studies. The older literature synthesized in the Cooper et al. review tended to include designs whose results may have been upwardly biased, such as studies that only compared student learning pre- and post-intervention (no control group) and thus conflated program effects with maturation effects. Since the early 2000s, research agencies such as the Institute of Education Sciences (IES) have increasingly emphasized randomized controlled trials and other designs that support causal inference (Angrist, 2004), and more researchers have taken them up. Effect sizes from studies using such designs tend to be smaller in magnitude than those from non-causal designs commonly used in prior decades (e.g., Lortie & Inglis, 2019).

Summer learning programs are often designed as compensatory programs to support children who are in need of additional learning time, including children from low-income backgrounds. Cooper et al. (2000) concluded that middle-class children benefited more academically from summer programs than did low-income children. However, the conclusions from the Cooper et al. review merit re-evaluation (Kim & Quinn, 2013). Income inequality has widened in the period since the Cooper et al. studies were conducted (Dabla-Norris et al., 2015), changing the contexts in which summer programs operate. High-income parents have increased their spending on their children in recent decades (Kornich & Furstenberg, 2013), and children spend their summers in neighborhoods that are increasingly segregated by family income (Owens, 2016). These trends suggest that the relative benefits of summer programming for

higher-income children may be smaller than decades ago, for instance if higher-income parents provide greater educational opportunities for their children in the control group, thus attenuating the treatment-control contrast of summer programs for higher-income children (Kim & Quinn, 2013). We re-examine the robustness of Cooper et al.'s (2000) finding using contemporary research in the summer mathematics context.

Second, the more recent syntheses of the literature are narrative reviews rather than formal meta-analyses, and thus do not present quantitative estimates of mean pooled impact nor of characteristics that predict outcomes.

Third, to our knowledge, no prior meta-analysis has examined the impacts of academic summer programs on children's non-cognitive[1] outcomes—a topic that the NASEM (2019) panel report highlighted as a 'priority research need.' A sizable body of research finds that non-cognitive outcomes, such as academic motivation, school attendance, and social skills, predict both academic achievement and long-run educational attainment and career results (Heckman & Kautz, 2014; Steinmayr & Spinath, 2009). Students with stronger school attendance (Gottfried, 2017) and more positive academic beliefs (Yeager & Walton, 2011) tend to demonstrate better academic performance. In recent decades, employment and earnings growth have been especially strong in careers that require both mathematics and social skills (Deming, 2017). Of urgent concern, the COVID-19 pandemic precipitated stark declines in students' social-emotional well-being and mental health, leading to a pressing need for policy options to help students to rebuild non-cognitive skills (Hamilton & Gross, 2021).

[1] Because personal attributes and skills beyond those measured by achievement tests also involve cognition, 'non-cognitive skills' is a misnomer (e.g., West et al., 2016). We retain the term here as it is in widespread use in the research literature. According to Duckworth and Yeager (2015), "debate over the optimal name for this broad category of personal qualities obscures substantial agreement about the specific attributes worth measuring" (p. 237).

Socioeconomic disparities in children's social and emotional skills may be exacerbated during the summer, when children have reduced access to school-based supports for social-emotional learning and enrichment (NASEM, 2019). In theory, common summer program elements, such as a focus on hands-on inquiry and small class sizes, may improve children's motivation, which could carry over into the school year. Gaining skills during the summer may also bolster students' confidence in learning mathematics, begetting more skills (e.g., Ceci & Papierno, 2005). On the other hand, it is possible that academic summer programs could diminish students' attitudes and other non-cognitive outcomes, for example if students lose out on recreation opportunities. Notwithstanding the relatively small number of studies reporting impacts on domains beyond achievement, compiling the emerging evidence is important given well-documented income gaps in these outcomes (e.g., Downey et al., 2019), and the importance of non-cognitive skills for overall educational and career success.

Lastly, unlike prior meta-analyses, which pooled programs across subject areas or across summer school and afterschool when examining moderators, we explicitly test for summer program characteristics that predict stronger mathematics learning. This is important given variability in findings of recent evaluations. For example, a randomized evaluation of the BELL summer program reported mostly non-significant findings on reading, mathematics, and social-emotional outcomes (Somers et al., 2015). However, a randomized study of summer programs in five school districts found positive impacts on mathematics scores after the first year, but null results for reading, social-emotional skills, and effects at longitudinal follow-up (McCombs et al., 2020).

Specifically, summer programs vary on several malleable programmatic features that may explain disparities in their results. In recent decades, scholars have produced new studies

that include components that did not exist in summer learning programs from previous decades. Whether we would expect the inclusion of these elements to strengthen or attenuate the observed mean effects of summer programs is not obvious. As we discuss below, newer programs often incorporate elements that may be expected to lead to stronger learning impacts, including novel curricula that reflect reform-oriented mathematics standards and social-emotional learning goals. Newer studies often provide more information about program implementation, allowing us to examine moderators of program impact in greater detail. However, novel formats such as online-only summer programs may be expected to show smaller effects. In the following section, we describe these potential moderators, which as we note below were adapted from prior literature in summer reading (Kim & Quinn, 2013) and updated to reflect a focus on summer mathematics.

**Programmatic Features of Summer Programs that May Moderate Effects**

Summer programs may focus on either remediation, including review of material from the previous year (Cooper et al. 2000), or on non-remedial goals, including enrichment and preview of future coursework. Programs may be broad or subject-specific in their programmatic focus. Inclusion of instruction in other academic subjects (e.g., reading, science, social studies) may hypothetically improve mathematics learning, for example, if there is cross-domain transfer such that reading instruction strengthens mathematics outcomes (e.g., Glenberg et al., 2012). On the other hand, it is conceivable that focusing on mathematics alone may have resulted in stronger mathematics learning, for example if this meant that more program resources, such as teacher professional learning and curriculum support, were directed to mathematics instruction. In light of research pointing toward positive relationships between time on task and student learning gains (e.g., Stronge et al., 2011), we also sought to examine whether program impacts varied by overall program duration as well as daily time allocated specifically to mathematics.

Summer programs may be conducted in-person or fully online. With the rapid proliferation of educational media for children, school districts have increasingly implemented online summer programs with the goal of lessening summer learning loss at low cost (Lynch & Kim, 2017), and this trend may be growing in the wake of schools' widespread use of virtual learning during the COVID-19 pandemic. However, a growing body of evidence documents that online instruction is less effective than in-person school for children across grade levels (Woodworth et al., 2015); further, earlier research has documented that learning risks may occur when children are left to learn mathematics independently, without support from a teacher (Erlwanger, 1973).

Lastly, given widespread calls in mathematics education for increased attention to students' engagement with core disciplinary concepts and practices (NRC, 2011), we examined whether each program's content as described included student activities aligned with the National Council of Teachers of Mathematics (NCTM, 2000) process standards (e.g., problem solving, communication) and/or Common Core State Standards (CCSS) for mathematical practice (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010). Children may engage with deeper mathematical thinking, including applications and problem solving, via hands-on project-based learning, in which students are involved in investigations of authentic problems (e.g., Meyer et al., 1997). Another feature that may enhance students' mathematical learning is engagement in group work (Zakaria et al., 2010), via mechanisms that encourage students to discuss, challenge and defend points of view (Gilbert-Macmillan & Leitz, 1986). The use of textbook exercises in summer program curricula may strengthen student learning by, for example, providing guidance to teachers in sequencing lessons and structuring lesson plans (Mesa, 2004); it is also conceivable that textbook use in a

summer program could result in lower levels of learning, if students find textbook activities unengaging and lose motivation. The commercial availability of curriculum materials is of policy interest because it aids replication of a key summer program element.

The resources available to summer programs may also predict the strength of students' learning benefits (Cooper et al., 2000; Kim & Quin, 2013). Interventions that provide mathematics teachers with high-quality professional development have positive effects on student achievement, on average (Lynch et al., 2019). The provision of lesson plans to teachers may reduce their preparation burden during the summer, thus enhancing lesson quality and subsequent student learning (Cai, 2005). Bus transportation and the provision of free school meals have also been linked to improved school attendance (Gottfried, 2017) and learning (Schwartz & Rothbart, 2020). Small class sizes are another resource that may strengthen learning (e.g., Krueger, 2003).

In summary, the current review synthesizes the recent empirical literature on summer learning programs in order to understand what characteristics and contextual factors are associated with stronger student outcomes in mathematics. It explores the impacts of summer learning programs on non-cognitive outcomes, and highlights directions for future research.

## Method

We conduct a meta-analysis of the experimental and quasi-experimental literature on summer learning programs in mathematics. Meta-analysis allows us to pool information across multiple studies, and to examine multiple hypothesized moderators of program impact.

### Search Procedures

For this review, we define summer mathematics programs as summer programs that aim to improve children's academic achievement in mathematics, including both mandatory

programs, such as district-required summer school programs for students who have failed the previous grade, as well as optional programs, such as those parents may elect for enrichment or child care purposes. Summer programs may be either classroom-based, with children attending in person at local schools or other community sites, such as college campuses; or home-based, with mathematics activities given to the child to complete at home, either alone or with family members. Although our definition did not exclude ex ante alternative types of home-based programs (such as mathematics books or packets mailed to children), the only qualifying studies of home-based summer mathematics programs that we found in the literature were of virtual (online) interventions. We include interventions that focused exclusively on mathematics, as well as more broad-based programs that also provided instruction in additional content areas.

We developed a database of studies via a four-phase search process similar to that used in Kim and Quinn (2013) for reading. We searched these channels from August 1998, as this was the last date for which searches were conducted in the previous comprehensive meta-analysis of the literature on summer school in mathematics (Cooper et al., 2000). Our review period is similar to that of the What Works Clearinghouse's (WWC, n.d.) 20-year review time limit. Searches were completed through April 2020. In the first search phase, we conducted an electronic search using the databases Academic Search Premier, Education Abstracts, ERIC, PsycINFO, EconLit, and ProQuest Dissertations and Theses, for the period August 1998 through April 2020. Searches were conducted using subject-related keywords relating to summer programs and methodology-related keywords designed to capture experimental and quasi-experimental designs, adapted from Kim and Quinn (2013).[2] Second, we searched targeted

[2] The specific search strings applied were as follows: ("summer program*" OR "summer school*" OR "summer math" OR "summer science" OR "summer STEM" OR "summer engineering" OR "summer enrichment" OR "summer remedia*" OR "summer instruction*" OR "summer education*" OR "summer learning") AND

internet sites including the What Works Clearinghouse, MDRC, NBER, RAND, AIR, Mathematica, Wallace Foundation, and the National Summer Learning Association. We also searched the abstracts of the Society for Research on Educational Effectiveness (SREE) conference. Third, we scanned the reference lists of previous review articles (Alexander et al., 2016; Bodilly & Beckett, 2005; Lauer et al., 2006; McCombs et al., 2011, 2019; NASEM, 2019; Terzian et al., 2009). Lastly, via a RAND report (Marsh et al., 2009), we identified U.S. states and districts that may have had required summer school, and contacted government agencies in these localities requesting any relevant research reports.

The search procedures described above yielded 2,544 records identified via database screening, and an additional 17 records identified through other sources (see Figure 1 for screening flowchart). After removing duplicates, we were left with 1,960 records.

**Study Inclusion Criteria**

In addition to being published after August 1998, we required that studies meet the following criteria to be included in the meta-analysis: (1) Evaluate the impacts of a summer mathematics intervention; (2) Present mathematics learning outcomes for treatment and control groups of students; (3) Include students who were in Grades pre-K-12 following their enrollment in a summer mathematics intervention; (4) Compare the performance of students in a treatment group to the performance of students in a control group who did not participate in the treatment or an alternative treatment; and (5) Present sufficient information to calculate one or more effect sizes (Hedges's *g*). Included studies could be conducted in any country, and also needed to provide evidence that the achievement levels of treatment and control groups were comparable at

("*experiment*" OR "control*" OR "regression discontinuity" OR "compared" OR "comparison" OR "field trial*" OR "effect size*" OR "evaluation").

baseline, as discussed below. We admitted studies that used randomized experimental and regression discontinuity designs, as well as quasi-experiments that met standards for group equivalence at baseline. Following guidance from the What Works Clearinghouse, if studies presented information on student achievement outcomes for which pretest differences were between 0.05 and 0.25 SD, we required that the authors had performed statistical adjustments for pretest differences (e.g., ANCOVAs); in cases where these were not presented in study reports, we manually calculated a difference-in-differences adjustment by subtracting the standardized pre-intervention difference from the standardized difference in outcomes, per What Works Clearinghouse guidelines (WWC, 2020).

**Study Screening**

We conducted screening in two phases. First, two raters screened each of the studies' titles and abstracts to identify potentially relevant studies, advancing studies to the second phase when they met criteria #1-4. All studies flagged as potentially relevant by either rater were reviewed by one of the authors, who made a final decision about advancing the study forward. This screening round resulted in the exclusion of many reports that were off-topic, such as articles on summer institutes for college faculty or summer research for college students, and descriptive articles about the phenomenon of summer learning loss. A total of 103 studies met the initial relevance criteria and proceeded to full-text screening.

In the second screening phase, two raters working independently, including at least one study author, examined the full text of each study and applied a more detailed set of methodological inclusion criteria. We required that studies present sufficient information to calculate an effect size (criterion #5), along with evidence that the treatment and control groups'

achievement levels were comparable at baseline, as discussed above.[3] We excluded summer programs with no mathematics component, such as programs focused exclusively on social skills or book reading. We required that participating students were entering Grades pre-K-12; because of our conceptual interest in summer learning during seasonal school closures, we excluded studies that examined only preschool children who had no formal schooling prior to participating in a summer program. The most frequent exclusion reasons were for characteristics of the intervention (e.g., off-topic, did not evaluate the effects of a classroom- or home-based summer mathematics intervention; $n = 17$), methodological issues [e.g., no control group, $n = 8$; no pretest data or pretest data not equivalent at baseline; $n = 22$ (e.g., Kendall, 2009)], and lack of outcome data (i.e., did not present mathematics learning outcomes for treatment and control; $n = 14$) (e.g., Hart et al., 2016). Note that some studies had multiple exclusion reasons (see Figure 1).

These search procedures netted a total of 37 studies that met the full review inclusion criteria and advanced to study coding. Of these, only two were included in Lauer et al.'s (2006) synthesis, and none were included in Cooper et al.'s (2000) synthesis. The number of studies included in the final dataset is in the same range as that included in Kim and Quinn's (2013) meta-analysis of summer reading programs ($k = 35$), Lauer et al.'s (2006) meta-analysis of out-of-school time programs for at-risk students ($k = 35$), and Cooper et al.'s (2000) meta-analysis of remedial summer programs ($k = 41$). In situations where study authors produced multiple reports on the same study, we used all available study documents to locate information about the intervention and study impacts, and used the most recent version (often the peer-reviewed version) for final impact estimates. Many studies contributed multiple effect sizes because they

[3] We did not require demonstration of outcome-specific baseline equivalence for non-cognitive outcomes. Pretest data for these outcomes was often not reported, and for certain outcomes (e.g., dropout) there is no directly corresponding baseline variable.

reported information for multiple outcome measures, multiple samples, multiple versions of the same program with a common control group, and/or multiple programs.

The final meta-analytic sample includes 149 effect sizes nested within these 37 studies. The sample includes a separate effect size for each treatment contrast, each measure of mathematics achievement and non-cognitive outcomes, and each sample of students that the study reported. (See Online Appendix B for references for the included studies.)

**Study Coding**

Study authors and trained graduate research assistants conducted full-text coding using the following procedures. Before beginning double-coding, we established inter-rater reliability. We began by having each member of the team code studies separately, then we held meetings to reconcile disagreements and refine codebook descriptions. We repeated this procedure until we reached a stable set of codes and an 80% agreement threshold. Each study was then coded by two researchers, including at least one study author. Each researcher coded the studies independently, then the coding pair met and reconciled discrepant codes via discussion.[4]

**Outcome Variables**

We examine two categories of dependent variables, *mathematics achievement outcomes* and *non-cognitive outcomes*. The first, *mathematics achievement outcomes* (112 effect sizes

[4] Inter-rater agreement from independent coding prior to reconciliation was computed as follows: study design (randomized experiment or regression discontinuity design vs. other quasi-experiment): 0.95; publication type, 0.92; sample grade level, 1.00; poverty level (percentage of students low-income or FRPL eligible), 0.74; high-poverty sample flag, 0.92; program hours per day, 0.79; program timespan, 0.82; total program hours, 0.82; hours per day spent on mathematics, 0.88; broad academic focus, 0.92; remediation or preview focus, 0.85; fully online versus in person, 1.00; alignment with NCTM or CCSS standards, 0.72; activity participation in hands-on projects, 0.97; textbook exercises, 0.85; group work, 0.95; computer-based skills practice, 0.92; teacher professional development, 0.87; explicit direction for summer instruction, 0.90; provision of transportation, 0.87; on-site meals, 0.85; average class size, 0.85.

contributed from 37 studies), comprises outcomes from both standardized mathematics achievement tests (96 effect sizes extracted from 34 studies), such as tests administered by U.S. states and those available through commercial vendors (e.g., NWEA, ITBS), as well as broader school mathematics attainment outcomes (16 effect sizes pooled from six studies), which we define to include mathematics course grades, mathematics course-taking, and completing a STEM degree. Both test scores and course grades and attainment outcomes are important outcomes for policy (Kautz & Heckman, 2014). As such, we pool both types of outcomes in our primary analyses. However, as a sensitivity check, we also re-fit all models using test score outcomes only.

We defined three categories of *non-cognitive outcomes* aligned with Farrington et al.'s (2012) conceptual framework of non-cognitive factors related to academic performance. The first category assessed *academic mindsets, attitudes, and effort,* including students' tendency to persevere in schoolwork, psychosocial attitudes, and mindsets about academics. The second category included types of *social skills and behavioral adjustment*, such as interpersonal skills and school discipline. The third category, *academic behaviors*, indexed "the visible, outward signs that a student is engaged and putting forth effort to learn" (Farrington et al., 2012, p. 8), including attendance and absenteeism. We identified 37 relevant effect sizes from eight studies encompassing outcomes such as absenteeism, self-efficacy, self-regulation, and social skills. When outcomes were from scales that included items from multiple categories (e.g., Devereux Student Strengths Assessment) the outcome was classified into the category that matched most closely. See Online Appendix Table A5 for a list of the included non-cognitive outcomes.

**Effect Sizes Calculation**

Standardized mean difference effect sizes were calculated using Hedges's *g*:

$$g = J \times \frac{(\overline{Y_E} - \overline{Y_C})}{S^*}$$

Here, $\overline{Y_E}$ represents the average treatment group outcome, $\overline{Y_C}$ represents the average control group outcome, and $S^*$ represents the pooled within-group standard deviation. $J$ represents a correction factor that adjusts the standardized mean difference to avoid bias in small samples:

$$J = 1 - \frac{3}{4 \times (N_E + N_c - 2) - 1},$$

In this equation, $N_E$ represents the number of students in the treatment group and $N_C$ represents the number of students in the control group (Borenstein et al., 2009). Effect sizes were calculated using the software package Comprehensive Meta-Analysis (CMA).[5] In three cases, study authors presented information about an outcome that was insufficient to calculate an effect size. All came from studies that did report an effect size for at least one additional outcome, which meant that no studies were dropped from the analysis due to missing outcomes. We exclude these missing outcomes in the primary analyses, but then conduct a sensitivity check in which we impute a range of plausible values for these outcomes, then re-estimate our models.

## Empirical Strategy

### *Estimating Effects of Summer Programs*

[5] We used the following decision rules to calculate effect sizes: If the authors reported Hedges's *g*, we used this effect size and calculated its standard error when necessary (12% of effect sizes). If the authors reported a standardized mean difference effect size, such as Cohen's *d* or Glass's delta, we converted author-reported effect sizes to Hedges's *g* (72% of effect sizes). If authors did not report a standardized mean difference effect size but did report a covariate-adjusted unstandardized mean difference (e.g., a coefficient from a multilevel model) and raw standard deviations, we calculated a standardized mean difference and converted to Hedges's *g* (4%). If covariate-adjusted mean differences were not reported, we calculated effect sizes based on raw posttest means and standard deviations (5%). In the remaining cases, effect sizes were calculated from other results (e.g., studies that reported the results of analyses of variance [ANOVAs]; 7%).

Study authors often measure interventions' impacts on several different outcomes, raising a frequent issue in meta-analysis: single studies that present multiple effect sizes. Effect sizes nested within a single study are likely to be correlated, which violates the assumption of statistical independence. Previous meta-analyses of the impacts of summer programs have approached this problem either by averaging effect sizes, or by selecting a single effect size per study to 'represent' that study in analyses. Here we use a robust variance estimation (RVE) approach (Tanner-Smith & Tipton, 2014) so that we can include multiple effect sizes per study while accounting for the nested nature of our data. This method adjusts standard errors to account for the dependencies among effect sizes within the individual studies, in a comparable manner to adjusting standard errors in ordinary least squares (OLS) regression models for heteroscedasticity (e.g., using Huber–White standard errors) or to account for the nesting of data within clusters (e.g., clustered standard errors). This approach permits us to include multiple effect sizes from a single study in our analysis (see Tanner-Smith & Tipton, 2014), and we have used a similar method as described below in our prior research (Lynch et al., 2019).

We compute the weight for effect size *i* in study *j* using the following formula:

$$w_{ij} = \frac{1}{\{(v_{*j}+\tau^2)[1+(k_j-1)\rho]\}},$$

where $v_{*j}$ is the mean of within-study sampling variances $\left(SE_{ij}^2\right)$ within each study, $\tau^2$ is the estimate of the between-studies variance component, $k_j$ is the number of effect sizes within each study, and $\rho$ is the assumed correlation between all pairs of effect sizes within each study. The formula assigns lower weight to effect sizes from studies contributing more effect sizes and with higher sampling variances. We use the recommended default value of $\rho$ = .80, with $\rho$ assumed constant across studies (Tanner-Smith & Tipton, 2014), and also conduct a series of sensitivity checks to gauge the robustness of our findings to alternative values of $\rho$. All analyses except the

computation of *F* tests were conducted in Stata, with the *robumeta* package in Stata 15 (Tanner-Smith & Tipton, 2014) used to estimate our RVE models, including the recommended small-sample correction (Tipton & Pustejovsky, 2015). We conducted *F* tests using the *robumeta* and *clubSandwich* packages in R, to test the joint significance of the program features included in the RVE models (Fisher & Tipton, 2015).

Effect size heterogeneity is addressed somewhat differently in RVE compared with traditional meta-analysis methods. The RVE developers explain that the core objective of this method is to estimate fixed effects, specifically meta-regression coefficients, rather than to model effect size variation; thus, tests for heterogeneity presented in traditional meta-analysis are unavailable within RVE (Tanner-Smith & Tipton, 2014; Tanner-Smith et al., 2016). For each of our primary models, however, we report the method-of-moments estimate of $\tau^2$ as measures of between-study heterogeneity in effect sizes. To estimate average impacts of summer programs, we fit separate RVE models for the two categories of dependent variables: mathematics achievement outcomes and non-cognitive outcomes.

### *Examining Predictors of Summer Programs' Effectiveness*

To identify potential moderators of summer program impact for coding and analysis, we began by adapting codes from a prior meta-analysis of summer reading programs (Kim & Quinn, 2013), revising items as appropriate to reflect a focus on mathematics. We then identified other potential moderators of program impact by examining prior meta-analyses and reviews of the literature on summer learning, out of school time, and instructional effectiveness in mathematics. Based on this review, we labeled overarching categories of potential moderators (e.g., activities, foci, resources), as well as specific codes (e.g., computer-based skills practice, group work, textbook exercises) that frequently emerged in the literature. After compiling a draft codebook,

we jointly coded a sample of studies, iteratively refining the codes as needed until we reached a stable set of codes.

We grouped potential moderators of program impact into five categories: (1) *study design and sample characteristics*; (2) *duration/intensity*; (3) *program foci*; (4) *program activities*; and (5) *program resources* (see below for descriptions). To examine whether specific features in each category moderated program impact, we then fit five sets of conditional meta-regression models with RVE, including the coded features as moderators and treating these moderators as fixed. Within each category, following recent meta-analyses (e.g., Garrett et al., 2019; Lynch et al., 2019), we first modeled the effect of each code separately, then probed their joint relationships by fitting a model with all codes in the category together. In cases for which there is within-study variability in program features (e.g., among studies with multiple treatment arms), we included the study-level mean value of each covariate and moderator (Tanner-Smith & Tipton, 2014). For covariates with within-study variability in at least 10% of studies, we also included a within-study version of the covariate by subtracting the study-level mean values from the original covariate values. All models controlled for whether the study used a randomized controlled trial (RCT) or regression discontinuity (RD) design, and an indicator for whether the study was conducted with elementary students. We do not fit these models examining moderators on non-cognitive outcomes due to data limitations; we provide a descriptive summary of the impacts.

Below, we describe the five categories of coded moderators of program impact.

**Study Design and Sample Characteristics.** We coded each study on a set of methodological criteria, categorizing whether the study design was a randomized experiment or regression discontinuity design versus another type of quasi-experiment. We captured

publication type, indexing whether the study was a peer-reviewed journal publication, dissertation, or technical report including contract researchers' reports, conference reports, and district, state, or federal government reports. To identify moderators related to study sample, we coded for whether the summer program included elementary students (pre-K-Grade 5) or was focused on middle/high school students. To operationalize poverty level, we coded the percentage of students in the sample reported as low-income or as eligible for free or reduced-price school lunch (FRPL). If the study did not report FRPL information for the sample but did report FRPL information for the school or district from which the sample originated, following Kim and Quinn (2013), we used that information to code the study's sample. We also created a dichotomous indicator indexing whether the study was conducted with a high-poverty sample, which we operationalized as studies in which low-income children comprised greater than 75% of program participants.

**Duration/Intensity.** We captured information about the *duration* of the program, using codes for program hours per day, timespan in weeks over which the program was conducted, the total program hours offered (summing across years for multiyear programs), and the number of program hours per day spent specifically on mathematics.

The remaining three categories of codes captured summer program characteristics. We coded program characteristics as 'present' if the study report indicated the feature was present, and 'not present' either if the report indicated that the feature was not a part of the intervention, or if the report was silent on the feature (following e.g., Garrett et al., 2019).

**Program Foci.** The first set of codes examined the summer program's *focus.* We classified each program as focused on either mathematics only, or as possessing a broad academic focus, including other academic subjects (e.g., reading, science) in addition to

mathematics. We classified the goals of each program as primarily focused on either remediation or on preview of future coursework. We captured whether the summer program was conducted fully online versus in person. We used a dichotomous indicator to index whether each program's content as described included student activities aligned with NCTM or CCSS standards.

**Program Activities.** A second set of codes examined the *activities* in which children participated during the summer program. We coded each study for evidence that children participated in hands-on projects, completed textbook exercises, engaged in group work, and/or completed computer-based skills practice, over the course of the summer program. We also computed a composite index of the total number of these activities that were reported per study, and coded whether the curriculum materials used were commercially available.

**Program Resources.** A third set of codes indexed the *resources* available at each summer program. We coded each study for information about summer program *staffing*, including evidence that the summer program instructors received professional development, as well as whether teachers received explicit direction in preparing for summer instruction, such as pre-made lesson plans. We examined *district and community supports*, including whether programs provided transportation for students (i.e., bus rides) and whether the program provided free meals on site (breakfast and/or lunch); however, this information was unreported in many studies. Lastly, we captured information about the *average class size* in the summer program.

## Publication Bias

A common concern in research syntheses is the possibility that estimates of average effects may be influenced by publication bias. We used three strategies to examine this issue (Kim et al., 2021). We first examined whether peer reviewed status was a significant predictor of effect size magnitude. We then used a trim-and-fill analysis (Duval & Tweedie, 2000), and

plotted a cumulative meta-analysis forest plot (Borenstein et al., 2009). We further conducted leave-one-out meta-analysis as an additional sensitivity check (StataCorp, 2021).

## Results

The results section is organized as follows. First, we present descriptive information on the included studies and samples. Next, we present estimates of the pooled mean effects of summer programs on mathematics outcomes. We then analyze moderators of program impacts. Lastly, we explore the relationship between summer programs and non-cognitive outcomes.

### Descriptive Information for the Included Studies and Programs

Table 1 presents descriptive statistics regarding the studies and summer programs included in our dataset. Thirty percent of included studies were randomized experiments or regression discontinuity designs, including several large-scale studies conducted in large, high-poverty urban school districts (e.g., Jacob & Lefgren, 2004; Mariano & Martorell, 2013; McCombs et al., 2020). The remaining 70% of studies employed propensity score matching or other quasi-experimental designs that demonstrated satisfactory group equivalence at baseline, as described above. These studies included evaluations of large programs such as Upward Bound Mathematics and Science (Olsen et al., 2007) and Building Educated Leaders for Life (BELL) (Somers et al., 2015), along with smaller programs conducted at the school and district levels. The studies comprised peer-reviewed journal publications (19% of reports), dissertations (43%), and technical reports including contract researchers' reports, conference reports, and district, state, or federal government reports (38%). All but two studies were conducted in the United States: Davies et al. (2019) examined the Summer Numeracy Program in Ontario, Canada; and Gorard et al. (2015) investigated a summer school model established by the Future Foundations in England. Of the included mathematics achievement effect sizes, 86% were standardized test

outcomes, and 14% were school attainment outcomes, such as course grades; 22% of studies ($k$ = 8) presented one or more non-cognitive outcomes.

The programs examined in our dataset primarily served low-income students. Among studies with available data ($k$ = 32), the mean percentage of low-income participants was 65%. The National Center for Education Statistics has characterized high-poverty schools as those where more than 75% of students are FRPL eligible (Irwin et al., 2021). 41% of included studies had greater than 75% of program participants classified as FRPL eligible or low-income. Only 6% of studies had 25% or fewer low-income program participants, the NCES benchmark for low-poverty schools. Among the 18 studies that reported information about students' English language learner (ELL) status, 29% of students were ELLs. Among the 24 studies that reported full sample student race, on average 72% of students were non-White. Programs served a mix of elementary students (46% of studies) and middle/high school students (54% of studies).

Table 1 also presents study-level frequencies for summer programs' characteristics, including duration, foci, activities, and resources. Most summer programs evaluated were conducted in person (89%), while 11% were fully online. Among studies that reported on program time, mean program duration was 158.2 hours (reported in $k$ = 31 studies), and the average timespan over which the programs were spread in a summer was 5.2 weeks, with five studies examining multiyear programs. Mean reported length of the program day was 4.6 hours ($k$ = 28 studies), and mean hours per day spent on mathematics was 2.1 hours ($k$ = 22 studies).

Most programs (78%) were focused on remediation of previous years' academic content. Most were also broad-based in academic focus, with 78% offering instruction on a range of other academic subjects in addition to mathematics. Approximately a third of programs (32%) reportedly used curriculum materials or activities aligned to CCSS and/or NCTM standards.

Nearly all in-person programs that provided information on teachers' qualifications were taught by either certified teachers or a mix of teachers and aides. Most studies (54%) reported that instructors received professional development; 27% of studies reported that specific lesson plans or structures were provided. Among in-person programs, 47% reported providing transportation, and 46% reported that meals were provided; however, this information was unreported in many studies. Among studies reporting class size data, mean class size was 17 students.

**Did Summer Programs Impact Students' Mathematics Learning Outcomes?**

Compared to control group students, students who participated in summer programs that included mathematics activities experienced significantly greater improvements on average in mathematics learning. We found an average weighted impact estimate of +0.10 standard deviations on mathematics outcomes (Table 2, Column 1). Examining specifically outcomes on standardized mathematics achievement tests (Table 2, Column 2), we found an average weighted impact estimate of +0.10 standard deviations. To contextualize the magnitude of this effect, a typical treatment group student who participated in a summer program would be expected to rank approximately 4 percentile points higher than a typical control group student (Lipsey et al., 2012). The pattern of results for broader attainment outcomes (e.g., subsequent mathematics course grades and course-taking) is similar (Table 2, Column 3), albeit less precisely estimated given the smaller number of studies reporting such effects. Pooled across both types of mathematics achievement outcomes (standardized tests and school mathematics attainment), of the 112 effect sizes included in the meta-analysis, 72 were positive in sign (64%), and 29 of these were statistically significant. Thirty-seven effect sizes were negative in sign (33%), and only 2 of these were statistically significant. Three effect sizes had point estimates of zero (3%). (See Table S1 [online only] for a summary of included outcomes and effect sizes.)

Table S2 (online only) shows that there were not statistically significant differences in effect sizes for mathematics learning outcomes based on whether the study employed a randomized experimental or regression discontinuity design versus other quasi-experimental designs; on whether or not the study was a dissertation; or on student grade level (elementary versus middle/high school). (For a breakdown of estimated mean effect sizes based on unconditional RVE meta-regression models by grade level, see Table S3 [online only].)

## Features That Moderate Program Impacts

We next examine factors that may moderate impacts on mathematics learning outcomes.

### *Poverty Level of Sample*

As discussed above, the extant research was conducted mostly in low- and mixed-income settings, consistent with the populations many summer learning programs primarily aim to support. We did not find a significant relationship between the poverty level of the student sample and program impacts (Table 3). For this analysis, poverty level was operationalized using a continuous indicator for the proportion of program participants classified as low-income or eligible for free or reduced-price school lunch. These results indicate that studies of summer programs tended to show similar, positive impacts on children's learning when conducted with both higher poverty and relatively lower poverty samples. We also explored whether the impacts on mathematics learning were different for higher versus lower income children attending the same summer program. Following Kim and Quinn (2013), we conducted within-study analyses that compared the magnitude of effect sizes for children from low-income versus mixed-income backgrounds using the subset of six studies that reported outcomes broken out by student poverty level. This analysis employed random-effects meta-analysis to summarize pooled mean effect sizes for the low- and higher-income samples within studies, then compared these magnitudes.

We did not find significant within-study differences in impacts by student poverty level. The results are consistent with a conclusion that children in both lower- and higher-income settings garner similar, positive mathematics learning impacts from summer program participation.

***Duration/Intensity***

We turn next to summer programs' duration and intensity (Table 4). We did not observe significant relationships between total program hours or program hours per day and students' mathematics outcomes. The results are consistent with a pattern of larger mean effect size magnitudes among programs that spent more hours per day on mathematics (+0.10 SD); however, as only 22 of the 37 studies provided data on this indicator, caution is warranted in interpreting this finding (Tipton & Pustejovsky, 2015).

***Program Foci***

Next, we examined the associations between the focus of the summer program and effect sizes via a series of models (Table 5). Average effect sizes were larger among programs focused specifically on mathematics, as compared with those having a broader focus on multiple academic subjects (+0.18 SD, $p < .05$). This result remained significant in the final model which controlled for other program foci indicators. As another robustness check, we sought to examine whether the advantage of mathematics-only programs versus combined programs was retained if the duration of mathematics content was controlled, by fitting a model containing both variables jointly (Table 5). In the joint model, the magnitudes of the indicators are positive but not statistically significant; however, as in the prior model including mathematics hours, caution is warranted in interpreting this model due to missing data (e.g., only 5 studies of mathematics-specific programs also provided information on content hours) (Tipton & Pustejovsky, 2015). Neither a focus on remediation, as compared with preparation for future coursework, nor the

inclusion of content judged to be aligned with NCTM standards and/or CCSS was a significant predictor of effect size magnitude. Descriptively, programs that were fully online had smaller impacts on average than did in-person programs, although as noted above, the number of fully online programs was relatively small, and this relationship was not statistically significant.

### *Program Activities*

We then turned to the relationships between effect sizes and summer program activities (Table 6). We found that the use of textbook exercises was significantly associated with effect size magnitude. This association was negative (-0.11 SD, $p < .05$), indicating that summer programs that reportedly assigned mathematics textbook work had smaller impacts than those that did not, on average. We did not find significant relationships between any of the other program activities for which we coded—use of a commercially available curriculum, hands-on projects, group work, or computer-based skills practice—and the magnitude of effect sizes.

### *Program Resources*

Table 7 displays the results from models investigating the relationships between summer program resources and effect sizes. None the activities for which we coded, including the provision of teacher professional development, teacher direction in lesson planning, student transportation, and average class size, were significantly associated with effect size magnitude, either individually or in the combined model.

We note that programs that lack features associated with larger-than-typical effect sizes may still have positive impacts on student outcomes, on average. Therefore, in Table S3 [online only], we display the results of these moderator analyses summarized using regression-adjusted mean effect sizes. We first present average effect sizes based on subgroup analyses without controls for additional program features, which are derived from unconditional meta-regression

models estimated using RVE to account for the nesting of effect sizes within studies. We next display mean effect sizes based on conditional meta-regression models, corresponding to the primary moderation analyses, with each predictor included separately and controlling for the same program features as discussed above. Lastly, we display average effect sizes corresponding to our final moderator analyses with all predictors within each category included simultaneously. Below for parsimony we discuss only those program features that were statistically significant moderators of effect size magnitude in the primary models.

As shown in Table S4 [online only], the mean effects of summer programs that did and did not have the moderators analyzed were typically positive. Even among programs that did not have the features previously identified as predictors of larger effect sizes, summer programs typically had positive impacts on mathematics learning. For example, programs that did not focus specifically on mathematics had positive effects, on average ($\overline{g_{c+}}$ = 0.07, $\overline{g_c}$ = 0.07, $\overline{g_{uc}}$ = 0.06, $p_{uc} < .01$), as did programs that incorporated the use of mathematics textbook exercises ($\overline{g_{c+}}$ = 0.02, $\overline{g_c}$ = 0.03, $\overline{g_{uc}}$ = 0.04, $p_{uc} < 0.05$), and programs that dedicated relatively fewer hours per day specifically to mathematics ($\overline{g_c}$ = 0.21, $\overline{g_{uc}}$ = 0.07, $p_{uc} > .10$). The differences in mean effect sizes based on estimating unconditional and conditional models are generally similar in direction and magnitude. Lastly, in Table S5 [online only], we show the results of fitting an omnibus model including all predictors simultaneously except those correlated above 0.6; however, fitting this model results in significant loss of information, retaining only 38% of the study pool ($k$ = 14), and given the sample size restrictions across predictors, the model findings must be interpreted with caution (Tipton & Pustejovsky, 2015). Similar to other recent meta-analyses (e.g., Garrett et al., 2019), the non-omnibus models are our preferred models given these data restrictions. The limitations of RVE with larger numbers of moderator variables and

study pools typical in education and the social sciences have been previously discussed in the literature and are a subject of ongoing research (e.g., Tipton & Pustejovsky, 2015).

### *Additional Study Design and Sample Moderators*

We report results from examining the associations between additional study design features and the magnitude of effect sizes in Table S6 (online only). No significant differences in effect size magnitudes were observed related to whether the study was conducted in an urban or nonurban setting, whether the study setting was one district, multiple districts, and/or states, nor the gender composition of the sample. Lastly, we examined other study design features, including whether the study design was a randomized trial; whether students with low attendance were dropped from the analysis; whether the study reported sizeable student attrition (20% or more of participants); amount of time elapsed between the summer program and the assessment; and the treated sample size; however, attendance and attrition information were unreported in many studies. None of these features were significantly related to effect size magnitudes.

## Did Summer Programs Impact Students' Non-Cognitive Outcomes?

With our data, we had a unique opportunity to explore the impacts of summer programs on outcomes beyond achievement. A total of eight studies presented information on the impacts of summer programs on 37 non-cognitive outcomes aligned with Farrington et al.'s (2012) conceptual framework. The relatively small number of studies reporting non-cognitive outcomes is consistent with other domains of educational interventions (e.g., teacher professional development; Yoon et al., 2007) where synthesists previously found few rigorous impact studies. We urge future primary researchers to measure and report noncognitive outcomes to permit moderator testing.

Compared with control group students, students who participated in summer programs that included mathematics had significantly better average non-cognitive outcomes. We found an average weighted impact estimate of +0.11 standard deviations (Table 2, Column 4). To put the magnitude of this effect into context, a typical treatment group student who participated in a summer program would be expected to rank approximately 5 percentile points higher on non-cognitive skills than a typical control group student (Lipsey et al., 2012). We summarize the effect sizes and outcomes in Table S7 (online only). Of the 37 effect sizes included in the meta-analysis, 27 were positive in sign (73.0%), with 10 being statistically significant. Seven effect sizes were negative in sign with one being statistically significant. Three effect sizes had point estimates of zero.

The 37 outcomes were grouped into three categories, including academic mindsets, attitudes, and effort; social skills and behavioral adjustment indicators; and academic behaviors. Due to missing data in study reports, the number of studies and effect sizes represented in each category is small; as such, and given that power in the context of dependent effect sizes and with the inclusion of moderators is an area of ongoing research (Pigott, 2012), we interpret the estimated mean effect sizes by category from unconditional RVE meta-regression models depicted in Table S8 (online only) with caution. While not statistically significant, the magnitude of the pooled effect for academic behaviors (i.e., attendance and chronic absenteeism) is larger than those for the other categories, which are close to zero. The pattern of findings suggests that summer programs' average positive non-cognitive impacts may be driven by improvements to students' subsequent academic year attendance, a hypothesis that warrants follow-up.

The summer programs we examined used different approaches to support students' non-cognitive outcomes. Although we do not conduct formal moderator tests due to data limitations

in this category, in this section we discuss program elements that the study authors emphasized as relevant to their reported non-cognitive impacts.

A common theme identified among several of the programs that demonstrated positive impacts on non-cognitive outcomes was an *explicit program focus* on improving social-emotional and/or behavioral skills and well-being. The Horizons National Student Enrichment Program (Scher, 2018) was among the most intensive interventions studied, with an explicit goal of having students enroll for multiple summers, and participants in the impact evaluation having attended for four or more summers. A key feature of the program was that Horizons teachers "create positive relationships with students that are sustained across many years, and students develop friendships that also encourage multi-year attendance" (Scher, 2018, pp. 1-2). In addition to academics, the program also provided "access to cultural and recreational opportunities like those enjoyed by their peers in middle-income households" (Scher, 2018, p. 1). Study participants were found to have better subsequent school attendance and fewer high school disciplinary referrals than their nonattending matched peers. On the other hand, Mac Iver and Mac Iver's (2015) evaluation of a 5-week STEM robotics program for middle school students in a high-poverty urban district suggests that a less intensive program may also be beneficial. According to the authors, "participation in the robotics enrichment was expected to increase student engagement in general (measured by attendance the following year)" (p. 5). The authors found that participating students had better attendance the following school year; attendance impacts in the follow-up year were positive in sign but not statistically significant.

Meanwhile, McCombs et al.'s (2014, 2020) randomized evaluation of five voluntary summer programs is instructive both because of its rigorous methodological design and because the programs it examines were district-run and likely similar to those offered in many urban

school districts, albeit adhering to study-specific implementation standards. Although program details differed across each district (Boston; Dallas; Duval County, FL; Rochester, NY; Pittsburgh), participating districts committed to providing 5 weeks of full-day programming for two summers, with at least 3 hours per day of language arts and mathematics instruction taught by certified teachers. Experimental impacts of the programs on self-regulation and self-motivation skills, attendance, and suspensions after the first and second summers of programming, and three years later, were mostly positive in sign, but small in magnitude and not statistically significant. In hypothesizing why null results were found after the first summer, McCombs et al. (2014) noted that only one out of the five programs took specific actions to focus on social-emotional skills by providing teachers with professional development on the topic. They stated that "the effect estimate in this district is positive and larger than the other districts, although not statistically significant" (p. xiii). Summarizing the analyses across years, the authors concluded that "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" (McCombs et al., 2020, p. 20).

It is also worth noting that even among programs that did not highlight a specific emphasis on non-cognitive skills, most impact estimates were positive in sign. The one impact estimate that was statistically significant and negative was for school suspensions, reported in Harlow and Benson's (2001) study of Wake Summerbridge, a middle school summer enrichment program focused on preparing students to succeed in high school, attend college, and become leaders. The authors reported that summer participants were more likely to be suspended during the school year; however, they were also significantly less likely to drop out of school. One possibility is that the summer program may have helped some students at risk for school

suspension avoid dropout, leading to an observed uptick in the school suspension rate for summer participants. This finding suggests the importance of collecting evidence on dropout along with discipline records, particularly in high school. Overall, the pattern of mostly positive findings is consistent with the conclusion that there is unlikely to be a tradeoff or harm to non-cognitive skills from participating in academic summer programs. Rather, the evidence, albeit suggestive, points in the direction of positive non-cognitive benefits from summer programs.

**Publication Bias**

Finally, we examined the potential role of publication bias using three sensitivity analyses that used the aggregate mean effect size per study as the unit of analysis (e.g., Kim et al., 2021). First, we employed trim-and-fill analysis (Duval & Tweedie, 2000). This analysis indicated no studies missing from the funnel plot representing potentially unpublished studies with smaller mean effects, a scenario consistent with a lack of influence of publication bias on estimates of mean effects. We next plotted a cumulative meta-analysis forest plot (Borenstein et al., 2009), which depicts how the average effect size varies with the inclusion of smaller studies by adding one study at a time to each subsequent analysis (Figure S1 [online only]). The results suggest that while the mean effect size shifted upward as small-sample studies were added to the meta-analysis, we retain the overall conclusion of average positive impacts of summer mathematics programs. Lastly, we conducted leave-one-out meta-analysis (StataCorp, 2021), which performs a series of meta-analyses that exclude one study from each analysis to investigate the influence of each study on the overall effect size estimate (see Figure S2 [online only]). The overall mean effects remained generally similar and positive when individual studies were omitted. The combined checks are consistent with the conclusion that the results are robust

to publication bias. A series of additional sensitivity checks are reported in Tables S9-S16 [online only].

## Discussion

In summary, we found that studies of summer programs in mathematics had positive effects on mathematics achievement outcomes, on average, with a mean pooled effect size across studies of +0.10 standard deviations. Summer programs had similar positive impacts on standardized mathematics tests (+0.10 SD) and broader school mathematics attainment outcomes, such as course grades (+0.11 SD).

To contextualize the magnitude of these achievement impacts, prior research has estimated that a typical teacher who raises student achievement on standardized tests by +0.14 SD produces marginal gains of approximately US$7,000 per child in present value future earnings (Chetty et al., 2014). Extrapolating from this, the estimated average test score impact of summer programs of +0.10 SD would be expected to net approximately US$5,000 in present value future earnings per child. Summer programs have larger mean achievement effects than do several other categories of school-based interventions summarized in Fryer (2017), such as teacher merit pay, teacher professional development, data-driven instruction, and school choice, and the typical impact of summer programs is similar to the pooled estimate of the causal impact of charter schools. Considered a different way, if children were to accrue the pooled average benefit every summer in grades K-12 and these results were to accumulate linearly, the cumulative benefit would be greater than the size of the Black-White test score gap in fourth-grade mathematics (i.e., a potential +1.30 SD gain; McFarland et al., 2017). The current overall estimate of the mean impact of summer programs on mathematics achievement is in the same range as Cooper et al.’s (2000) estimate of the impacts of summer mathematics and reading

programs derived from studies with comparison groups (0.09 SD) as well as Lauer et al.'s (2006) estimate from a fixed-effects model (0.09 SD). Our findings thus confirm prior syntheses' substantive conclusion that summer mathematics programs tend to produce positive learning impacts.

Another relevant benchmark is the potential cost-benefit ratio of summer school as an investment. The following example adapted from Matsudaira (2008) provides one point of comparison. Examining the results of the Tennessee STAR experiment, Krueger (2003) estimated that reducing class size in the early grades by one third improved student achievement by 0.20 SD, at an estimated cost per student of roughly $13,000 (in current dollars). By contrast, Matsudaira (2008) and Augustine et al. (2016) reported summer school costs per student in major urban districts of approximately $1,500-3,300 in current dollars. If summer programs improve student achievement on standardized mathematics tests by approximately 0.10 SD, as suggested by the meta-analytic findings, extrapolating from the above would imply that the cost-benefit payoff of summer school may be more than twice as large as a class-size reduction with respect to boosting student achievement, consistent with Matsudaira's (2008) conclusion. Cost-benefit estimates should be considered suggestive in nature. Detailed intervention cost data would allow us to estimate comparative payoffs more precisely, and we encourage future studies to report this information. Moreover, summer programs often provide benefits beyond improved academic achievement, such as extracurricular experiences and child care coverage (Augustine et al. 2016; Cooper et al., 2000); these affordances are not captured in standard cost-benefit analyses. As a different yardstick, Augustine et al. (2016) found that three district-run summer programs had hourly costs per student that were lower than the school year per-pupil hourly

costs, both within the district and compared to the national average.[6] Together, the combined evidence is broadly consistent with a conclusion that summer programs provide a positive return on districts' investments.

**Features Associated With Summer Mathematics Program Effectiveness**

Via a suite of analyses, we examined the extent to which summer program characteristics and contextual factors predicted the magnitude of impacts on student mathematics achievement. One characteristic significantly associated with stronger than typical student mathematics learning outcomes was focusing program content specifically on mathematics. This finding is consistent with a sizeable body of research linking time on task to student achievement (e.g., Stronge et al., 2011), and indicating that programs tend to improve outcomes in the specific domains that they target (Kraft, 2020). However, this predictor did not retain its significance when modeled jointly with mathematics content hours, although joint analysis of these variables was limited because many studies lacked content hours data. Meanwhile, programs that targeted a broader variety of academic subjects may well have produced academic benefits in other subject areas that we did not capture in the present synthesis, given our specific interest in mathematics. Since summer programs with mathematics and reading content (Kim & Quinn, 2013) tend to improve each of these outcomes, respectively, policymakers may wish to match summer programs' foci to students' areas of perceived need. Indeed, some U.S. states appear to have adopted such a targeted approach in their COVID disaster recovery spending, by directing summer programs to focus on literacy for early elementary children, and mathematics for older students (e.g., Massachusetts Department of Elementary and Secondary Education, 2021).

[6] Estimated average summer program hourly cost per student was $6.70 in 2014; school-year costs in the same districts ranged from $7.65 to $20.06, and the 2013 per-student national average school-year costs were $10.52 per hour (Augustine et al., 2016).

The moderator analysis also suggested a negative link between textbook use and effect size magnitude. This finding may seem counterintuitive given that textbooks are generally considered an important contributor to students' potential opportunity to learn (Tornroos, 2005) and a key support for teachers (Mesa, 2004). However, it is possible that student engagement may have suffered if summer programs too closely mirrored typical school year offerings through the use of textbooks (McCombs et al., 2011).

**Mathematics Learning Impacts for Children of Different Family Income Backgrounds**

We found that summer programs serving lower-income children and those serving children from a mix of higher income backgrounds were similarly beneficial to children's mathematics learning. We analyzed impacts by income level within studies as a sensitivity check and found the pattern of findings was consistent with the study-level analysis.

By contrast, Kim and Quinn (2013) found that the impacts of summer reading programs were larger for low-income children than their higher income counterparts. One possible explanation for these differences in mathematics versus reading is that income-based patterns in the home activities that children do over the summer may differ by subject area. Children in higher income families tend to read more at home during the summer than their lower income peers (Heyns, 1978); thus, summer programs may induce a greater differential boost in summer literacy habits for low-income students than for higher income students (Kim & Quinn, 2013). It is also conceivable that few students of any income level do a significant amount of mathematics at home during the summer (Cooper et al., 1996), which may make the treatment-control contrast of summer mathematics programs similar for children across income groups. Overall, the finding that summer programs improve mathematics learning for children across income levels is important for policy given both the broad need to strengthen students' STEM opportunities, and

the current pressing demand for malleable policy factors that can aid in COVID learning recovery.

**The Relationship Between Summer Learning Programs and Non-Cognitive Skills**

The current findings support the notion that summer programs can improve students' non-cognitive outcomes. The potential for positive non-cognitive impacts is noteworthy because prior ethnographic research has suggested that "summer inequities in *nonacademic* learning may be even more egregious than the academic disparities that past research has emphasized" (Chin & Phillips, 2004, p. 206). In their study of social-class differences in children's summer experiences, Chin and Phillips (2004) found that middle-class children, via their opportunities to attend camps and participate in structured enrichment, received exposure to new environments with the potential to catalyze their future interest in areas such as science, history, arts, and culture. These opportunities promoted middle-class children's pride in their skills grown and satisfaction in their accomplishments over the summer. Meanwhile, poor and working-class children were more likely to spend their summers in circumscribed environments. The authors concluded that social-class differences in children's opportunities to develop their talents during summer likely contribute to "both a 'talent development gap' and a 'cultural exposure gap,' which, if exacerbated each summer, contribute to disparities in children's future life chances" (p. 206). The excess time that low-income children spend watching TV during the summer as compared with their higher income peers amounts to the equivalent of approximately a full month of school days (Gershenson, 2013), a concerning level given that television viewing has been linked to aggressive behavior (e.g., Manganello & Taylor, 2009) and obesity (Rey-López et al., 2008). The current findings provide supportive evidence that summer learning programs targeted toward low-income children have the potential not only to aid students academically,

but also to counteract inequities in nonacademic skills that may grow during summer vacation (for broader discussion of the effects of universal and targeted interventions on learning and equity, see Ceci & Papierno, 2005; also, e.g. Fuchs et al., 2001).

Study reports suggested that targeting social-emotional skills specifically in programming, such as via providing teacher professional development on the topic or including relationship-building and positive engagement as program goals, may have been linked to stronger impacts; however, not all studies that reported positive impacts had such an emphasis. We were unable to empirically test moderators of summer programs' impacts on non-cognitive outcomes given data limitations; however, this analysis would be a fruitful avenue for future research after more original studies presenting impacts on non-cognitive outcomes are conducted. Meanwhile, evidence-based approaches to helping children build personal and social skills in other out-of-school time settings, such as active learning and a focus on such skills, are likely also beneficial in the summer program context (Durlak et al., 2010).

**Limitations and Future Research Directions**

The limitations of this study point toward several potentially productive avenues for future research. Missing data presented the first challenge. A common issue in research syntheses is that programs subjected to rigorous evaluations may not fully represent the kinds of programs that children are typically offered (Institute of Education Sciences & National Science Foundation, 2013). Many of the study reports identified for the current review were evaluations of district-run, classroom-based summer school programs. While these evaluations reflect often-typical programming, more data on alternative types of programs would aid our understanding of how a broader range of program modalities may influence outcomes. For example, a large body of research indicates the effectiveness of tutoring in mathematics (e.g., Ritter et al., 2009).

However, most research on mathematics tutoring has been conducted in schools during term-time. We identified no studies of summer mathematics tutoring, yet such research could shed light on how to design effective tutoring programs for children when they are away from daily school mathematics practice. Similarly, a growing number of school districts are turning to online programs as a low-cost strategy to encourage summer mathematics practice (e.g., Lynch & Kim, 2017). Yet despite the popularity of such programs, the evidence base on their efficacy is thin as we were able to identify only four studies that examined online-only interventions. Online offerings also lack many of the affordances of in-person programs, such as meals, socialization, physical fitness, and child care for working parents. In addition, although we did not exclude studies based on country setting, all but two of the studies that met our methodological and substantive inclusion criteria were conducted in the United States. Future design and efficacy research on new and understudied kinds of summer programs, including studies conducted in multiple country settings, would move the field forward, and could help schools and districts to structure their summer programs more effectively.

The observation that summer programs improved children's mathematics learning, on average, spurs inquiry into what makes some summer programs more effective than others. While a handful of studies presented informative portraits of children's classroom activities (e.g., McCombs et al., 2014; Roderick et al., 2003), in many study reports detailed information about children's and teachers' activities during the intervention was lacking. Missing data thus precluded us from examining some moderators that prior researchers have hypothesized to influence program impact. For example, Cooper et al. (1996) hypothesized that procedural knowledge in mathematics may be more subject to forgetting over the summer than conceptual knowledge. The study reports often lacked detail about the contents of the assessments

administered, and as such we were unable to examine this issue, nor potential differential impacts of summer programs by mathematical domain. In addition, although we originally hoped to analyze attendance as a moderator of program impact, given evidence on a link identified in some primary studies (e.g., Augustine et al., 2016), too few reports provided sufficient attendance details to make this analysis feasible. Instead, following Lauer et al. (2006), we examined program duration, which captures potential for program exposure. Comparative data documenting how far behind summer participants were academically compared to their average-achieving peers was also often missing from reports, precluding us from probing the extent to which summer participants were on par with grade level performance by the fall. Many studies also presented outcomes from only one posttest; more consistent reporting of follow-up data would permit a more detailed examination of potential fade-out effects.

A longstanding focus in the design and funding of summer programs has been on supporting low-income children with high levels of need (Borman & D'Agostino, 1996). Consistent with this emphasis, the samples of students included in the research studies identified for the current synthesis were primarily low-income. We have limited data on the impacts of summer learning programs in higher-income contexts. Such research could illuminate disparities in program offerings for low- versus high-income children, and point toward programmatic steps that could be taken to reduce those inequities. In addition, although we had initially hoped to examine impacts for student subgroups by race/ethnicity and ELL status, too few studies reported this information to permit this analysis. This information would allow us to examine impacts by these characteristics, and we urge future research to report this data.

Although we examined a sizable pool of studies to estimate mean impacts on mathematics achievement, we could examine non-cognitive impacts using only a smaller

subsample of studies that reported these outcomes. As such, we consider these analyses exploratory. More consistent reporting of non-cognitive outcomes in study reports would facilitate more precise estimation of these impacts. Especially in the aftermath of the COVID pandemic, the potential for summer programs to support social-emotional development is noteworthy. School closures and social isolation have harmed students' mental health and well-being (Hamilton & Gross, 2021). Many students suffered lower attendance, or dropped out of school (Korman et al., 2020). The data present a form of existence proof for the notion that summer programs can also support non-cognitive outcomes.

Lastly, ethnographic research studying the summer experiences of children across sociodemographic lines could illuminate other means by which summer programs can better support children and families (Cooper et al., 2000). We could identify only one ethnographic study of children's summer experiences (Chin & Phillips, 2004); although this study is quite informative, it was conducted in a single elementary school over two decades ago. Parents often choose to send their children to summer programs for reasons beyond improving academics, such as for socialization, physical activity, and child care (Chin & Phillips, 2004). Qualitative studies, including interviews and observations of children that vary by geographic regions, family resources, and local availability of summer programming, could shed light on summer program structures and features that may provide holistic benefits beyond improved academic achievement.

However, despite the noted limitations, we compiled the evidence from dozens of studies synthesizing over two decades of the most rigorous extant evidence on the impacts of summer programs on children's mathematics learning. In summary, contemporary summer programs are a malleable factor to improve children's mathematics learning, including in high-poverty settings

where children possess a persistent need for support. By bolstering children's mathematics learning, summer programs have the potential to advance long-run STEM educational opportunities and outcomes.

## References

Afterschool Alliance. (2015). *America after 3PM survey*.

Alexander, K. L. (2019). Summer learning loss sure is real. *Education Next*.

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review, 72*(2), 167-180.

Alexander, K., Pitcock, S., & Boulay, M. C. (Eds.). (2016). The summer slide: What we know and can do about summer learning loss. *Teachers College Press*.

Angrist, J. D. (2004). American education research changes tack. *Oxford Review of Economic Policy, 20*(2), 198–212.

Atteberry, A., & McEachin, A. (2021). School's out: The role of summers in understanding achievement disparities. *American Educational Research Journal, 58*(2), 239–282.

Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J., McEachin, A., & Siler-Evans, K. (2016). *Learning from Summer: Effects of Voluntary Summer Learning Programs on Low-Income Urban Youth. RR-1557-WF*. RAND Corporation.

Bodilly, S. J., & Beckett, M. K. (2005). *Making out-of-school-time matter: Evidence for an action agenda.* RAND. https://www.rand.org/pubs/monographs/MG242.html

Borenstein, M., Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis.* Russell Sage Foundation.

Borman, G. D., Benson, J., & Overman, L. T. (2005). Families, schools, and summer learning. *The Elementary School Journal, 106*(2), 131–150.

Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, *18*(4), 309-326.

Burkam, D. T., Ready, D. D., Lee, V. E., & LoGerfo, L. F. (2004). Social-class differences in

summer learning between kindergarten and first grade: Model specification and estimation. *Sociology of Education*, *77*(1), 1-31.

Cai, J. (2005). US and Chinese teachers' constructing, knowing, and evaluating representations to teach mathematics. *Mathematical Thinking and Learning*, *7*(2), 135-169.

Carter, D. F. (2006). Key issues in the persistence of underrepresented minority students. *New Directions for Institutional Research*, 130, 33–46.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the" have-nots" gain but the "haves" gain even more. *American Psychologist*, *60*(2), 149.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-79.

Chin, T., & Phillips, M. (2004). Social reproduction and child-rearing practices: Social class, children's agency, and the summer activity gap. *Sociology of Education*, *77*(3), 185-210.

Chmielewski, A. K., & Reardon, S. F. (2016). Patterns of cross-national variation in the association between income and academic achievement. *Aera Open*, *2*(3), 2332858416649593.

Cooper, H., Charlton, K., Valentine, J. C., Muhlenbruck, L., & Borman, G. D. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, i-127.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*(3), 227-268.

Dabla-Norris, M. E., Kochhar, M. K., Suphaphiphat, M. N., Ricka, M. F., & Tsounta, M. E.

(2015). *Causes and consequences of income inequality: A global perspective*. International Monetary Fund.

Darling-Hammond, L., Schachner, A., & Edgerton, A. K. (2020). *Restarting and Reinventing School: Learning in the Time of COVID and Beyond.* Learning Policy Institute.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), 1593-1640.

Deming, D. J., & Noray, K. (2020). Earnings dynamics, changing job skills, and STEM careers. *The Quarterly Journal of Economics*, *135*(4), 1965-2005.

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*(5), 613-635.

Downey, D. B., Workman, J., & von Hippel, P. T. (2019). Socioeconomic, ethnic, racial, and gender gaps in children's social/behavioral skills: Do they grow faster in school or out? *Sociological Science*, *6*, 446–466.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251.

Dumont, H., & Ready, D. D. (2020). Do schools reduce or exacerbate inequality? How the associations between student achievement and achievement growth influence our understanding of the role of schooling. *American Educational Research Journal*, *57*(2), 728-774.

Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American*

*Journal of Community Psychology*, *45*(3), 294-309.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI Mathematics. *Journal Of Children's Mathematical Behavior, 1*(2), 7–26.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review*. Consortium on Chicago School Research. Chicago, IL.

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. arXiv preprint arXiv:1503.02220.

Fryer, R. G., Jr. (2017). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In E. Duflo & A. Banerjee (Eds.), *Handbook of field experiments* (Vol. 2, pp. 95-322). Amsterdam: North-Holland.

Fuchs, D., Fuchs, L., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., Braun, M., & O'Connor, R. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*, 251–267.

Garrett, R., Citkowicz, M., & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education, 43*(1), 106-137.

Gershenson, S. (2013). Do summer time-use gaps vary by socioeconomic status? *American Educational Research Journal*, *50*(6), 1219–1248.

Glenberg, A., Willford, J., Gibson, B., Goldberg, A., & Zhu, X. (2012). Improving reading to

improve math. *Scientific Studies of Reading*, *16*(4), 316-340.

Gilbert-Macmillan, K., & Leitz, S. J. (1986). Cooperative small groups: A method for teaching problem solving. *The Arithmetic Teacher*, *33*(7), 9-11.

Gottfried, M. A. (2017). Linking getting to school with going to school. *Educational Evaluation and Policy Analysis*, *39*(4), 571-592.

Hamilton, L., & Gross, B. (2021). How has the pandemic affected students' social-emotional well-being? *Center on Reinventing Public Education*.

Hart, K. C., Graziano, P. A., Kent, K. M., Kuriyan, A., Garcia, A., Rodriguez, M., & Pelham Jr, W. E. (2016). Early intervention for children with behavior problems in summer settings. *Journal of Early Intervention*, *38*(2), 92-117.

Heckman, J. J., & Kautz, T. (2014). Fostering and measuring skills: Interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, & T. Kautz (Eds.), *The myth of achievement tests: The GED and the role of character in American life* (pp. 341-430). University of Chicago Press.

Heyns, B. (1978). *Summer learning and the effects of schooling.* Academic Press.

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction*, *14*(3), 251-283.

Institute of Education Sciences & National Science Foundation. (2013). *Common guidelines for education research and development*.

Irwin, V., Zhang, J., Wang, X., Hein, S., Wang, K., Roberts, A., York, C., Barmer, A., Bullock Mann, F., Dilig, R., & Parker, S. (2021). *Report on the condition of education 2021*. U.S. Department of Education. National Center for Education Statistics.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-

discontinuity analysis. *Review of Economics and Statistics*, *86*(1), 226-244.

Kendall, W. (2009). The effect of a summer academy on math achievement. [Unpublished doctoral dissertation]. Aurora University.

Kim, J., Gilbert, J., Yu, Q., & Gale, C. (2021). Measures matter: A meta-analysis of the effects of educational apps on preschool to grade 3 children's literacy and math skills. *AERA Open*, *7*.

Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, *83*(3), 386–431.

Korman, H., O'Keefe, B., & Repka, M. (2020, October). *Missing in the margins: Estimating the scale of the COVID-19 attendance crisis.*

Kornrich, S., & Furstenberg, F. (2013). Investing in children: Changes in parental spending on children, 1972–2007. *Demography*, *50*(1), 1-23.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, *113*(485), F34-F63.

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher*, *49*(8), 549–565.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*(2), 275-313.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K., & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. *National Center for Special Education Research.* https://ies.ed.gov/ncser/pubs/20133000/

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166.

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, *41*(3), 260-293.

Lynch, K., & Kim, J. S. (2017). Effects of a summer mathematics intervention for low-income children: A randomized experiment. *Educational Evaluation and Policy Analysis*, *39*(1), 31-53.

Mac Iver, M. A., & Mac Iver, D. J. (2015). The Baltimore City Schools Middle School STEM Summer Program with VEX Robotics. *Baltimore Education Research Consortium*.

Manganello, J. A., & Taylor, C. A. (2009). Television exposure as a risk factor for aggressive behavior among 3-year-old children. *Archives of Pediatrics & Adolescent Medicine*, *163*(11), 1037-1045.

Mariano, L. T., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis*, *35*(1), 96-117.

Marsh, J. A., Gershwin, D., Kirby, S. N., & Xia, N. (2009). *Retaining students in grade: Lessons learned regarding policy design and implementation.* RAND Corporation.

Massachusetts Department of Elementary and Secondary Education. (2021). *2021 summer learning opportunities*. https://www.doe.mass.edu/asost/summer-learning.html

Matsudaira, J. D. (2008). Mandatory summer school and student achievement. *Journal of Econometrics, 142*(2), 829-850.

McCombs, J. S., Augustine, C. H., Pane, J. F., & Schweig, J. (2020). *Every summer counts: A longitudinal analysis of outcomes from the National Summer Learning Project.* RAND Corporation. https://www.rand.org/pubs/research_reports/RR3201.html

McCombs, J. S., Augustine, C. H., Schwartz, H. L., Bodilly, S. J., McInnis, B. I., Lichter, D. S., & Cross, A. B. (2011). *Making summer count: How summer programs can boost children's learning.* RAND Corporation.

McCombs, J. S., Augustine, C. H., Unlu, F., Ziol-Guest, K. M., Naftel, S., Gomez, C. J., Marsh T., Akinniranye, G., & Todd, I. (2019). *Investing in successful summer programs: A review of evidence under the Every Student Succeeds Act.* RAND Corporation.

McCombs, J. S., Pane, J. F., Augustine, C. H., Schwartz, H. L., Martorell, P., & Zakaras, L. (2014). *Ready for fall? Near-term effects of voluntary summer learning programs on low-income students' learning opportunities and outcomes*. RAND Corporation.

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., & Hinz, S. (2017). *The condition of education 2017.* National Center for Education Statistics.

McKown, C. (2017). *Social and emotional learning: A policy vision for the future* [Policy brief]. The Future of Children. Princeton University.

Mesa, V. (2004). Characterizing practices associated with functions in middle school textbooks: An empirical approach. *Educational Studies in Mathematics, 56,* 255–286.

Meyer, D. K., Turner, J. C., & Spencer, C. A. (1997). Challenge in a mathematics classroom: Students' motivation and strategies in project-based learning. *The Elementary School*

*Journal*, *97*(5), 501-521.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

National Academies of Sciences, Engineering, and Medicine. (2019). *Shaping summertime experiences.* The National Academies Press. https://doi.org/10.17226/25546

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards*. Author.

National Research Council. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics.* The National Academies Press.

Olsen, R., Seftor, N., Silva, T., Myers, D., DesRoches, D., & Young, J. (2007). *Upward Bound Math-Science.* U.S. Department of Education.

Owens, A. (2016). Inequality in children's contexts: Income segregation of households with and without children. *American Sociological Review*, *81*(3), 549-574.

Patall, E. A., Cooper, H., & Allen, A. B. (2010). Extending the school day or school year: A systematic review of research. *Review of Educational Research*, *80*(3), 401-436.

Pier, L., Hough, H. J., Christian, M., Bookman, N., Wilkenfeld, B., & Miller, R., (2021). *COVID-19 and the educational equity crisis.* Policy Analysis for California Education.

Pigott, T. (2012). *Advances in meta-analysis*. Springer Science & Business Media.

Quinn, D. (2015). Black–White summer learning gaps: Interpreting the variability of estimates

across representations. *Educational Evaluation and Policy Analysis*, *37*(1), 50-69.

Rey-López, J. P., Vicente-Rodríguez, G., Biosca, M., & Moreno, L. A. (2008). Sedentary behaviour and obesity development in children and adolescents. *Nutrition, Metabolism and Cardiovascular Diseases, 18*(3), 242-251.

Ritter, G., Barnett, J., Denny, G., & Albin, G. (2009). The effectiveness of volunteer tutoring programs for elementary and middle school students: A meta-analysis. *Review of Educational Research*, *79*, 3–38.

Roderick, M., Engel, M., & Nagaoka, J. (2003). *Ending Social Promotion: Results from Summer Bridge. Charting Reform in Chicago Series*. Consortium on Chicago School Research.

Schwartz, A., & Rothbart, M. (2020). Let them eat lunch: The impact of universal free meals on student performance. *Journal of Policy Analysis & Management, 39,* 376-410.

StataCorp. 2021. *Stata 17 Base Reference Manual*. College Station, TX: Stata Press.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*(1), 80-90.

Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education*, *62*(4), 339-355.

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, *5*(1), 13-30.

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology,* 2, 85–112.

Terzian, M., Moore, K. A., & Hamilton, K. (2009). Approaches for economically disadvantaged children and youth: A white paper for the Wallace Foundation. *Child Trends*, 1-42.

Tipton, E., & Pustejovsky, J. E. (2015) Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604-634.

Tornroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, *31*(4), 315–327.

von Hippel, P. T. (2019a). Is summer learning loss real? How I lost faith in one of education research's classic results. *Education Next*, *19*(4), 8-15.

von Hippel, P. T. (2019b). Summer learning: Key findings fail to replicate, but programs still have promise. *Education Next.*

von Hippel, P. T., & Hamrock, C. (2019). 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*, *6*, 43.

West, M., Kraft, M., Finn, A., Martin, R., Duckworth, A., Gabrieli, C., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, *38*(1), 148-170.

What Works Clearinghouse. (n.d.). *Questions and answers from demystifying the What Works Clearinghouse: A webinar for developers and researchers.*

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, version 4.1.* U.S. Department of Education, Institute of Education Sciences.

Woodworth, J. L., Raymond, M. E., Chirbas, K., Gonzalez, M., Negassi, Y., Snow, W., & Van Donge, C. (2015). Online charter school study 2015. Stanford University CREO.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267-301.

Yoon, K. S., Duncan, T., Lee, W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement. Washington, DC: U.S. Department of Education.

Zakaria, E., Chin, L. C., & Daud, M. Y. (2010). The effects of cooperative learning on students' mathematics achievement and attitude towards mathematics. *Journal of Social Sciences*, *6*(2), 272-275.

## Tables and Figures

### Table 1

*Categories and Descriptions of Codes*

| Code | Code description | Code present [a] |
|---|---|---|
| Effect size type | | |
| Standardized mathematics test outcome | Percentage of mathematics achievement outcomes that are standardized test scores. | 86% |
| School mathematics attainment outcome | Percentage of mathematics achievement outcomes that are school attainment measures (e.g., mathematics course grades). | 14% |
| Non-cognitive outcomes | Percentage of studies contributing one or more non-cognitive outcomes. | 22% |
| Adjusted for covariates | Effect size is adjusted for covariates (e.g., pretest score). | 92% |
| Study design and sample characteristics | | |
| RCT or RD | Study used a randomized controlled trial or regression discontinuity design. | 30% |
| Publication type | Study is a dissertation. | 43% |
| | Study is a peer-reviewed journal publication. | 19% |
| | Study is a technical report including contract researchers' reports, conference reports, and district, state, or federal government reports. | 38% |
| Grade level – Elementary | Study sample included elementary (pre-K-5) students (versus middle/high school). | 46% |
| Poverty level | Percentage of students reported eligible for free or reduced-price school lunch. | 65% |
| Duration/intensity | | |
| Duration in weeks | Average timespan in weeks over which the summer program occurred. | 5.2 |
| Total program hours | Average number of total summer program hours. | 158.2 |
| Program hours per day | Average hours per day that the summer program met. | 4.6 |
| Hours per day on mathematics | Average hours per day dedicated to mathematics. | 2.1 |
| Summer program focus | | |
| Mathematics-specific focus | The summer program focused specifically on mathematics, in contrast to broad-based programs that also included other academic subjects (e.g., reading, science, social studies). | 22% |
| Program goals | The summer program focused on remediation, learning loss, or 'catch up.' | 78% |
| | The summer program focused on future coursework or the next grade level via preparation and/or preview of future content. | 22% |

| | | |
|---|---|---|
| Standards alignment | The summer program content was aligned with NCTM standards and/or CCSS. | 32% |
| Online only | Study examined a summer program conducted exclusively online. | 11% |
| Summer program activities | | |
| Children's activities | Variables indexing children's participation in hands-on projects, textbook exercises, group work, and computer-based skills practice, as well as total number of activities reported (range 0–3). | 0.92 |
| Curriculum: Commercial program | The study reported the summer program's use of a commercially available curriculum. | 27% |
| Summer program resources | | |
| Staffing | Program instructors received PD, either prior to or during the summer. | 54% |
| | Teacher direction (lesson plans or structure) was provided. | 27% |
| District/community support [b] | Transportation was provided. | 47% |
| | Meals (breakfast and/or lunch) were provided. | 46% |
| Average class size | Average class size | 16.6 |

*Note*. *N* = 37 studies. NCTM = National Council of Teachers of Mathematics; CCSS = Common Core State Standards; PD = professional development.

[a] Figures in the third column include the percent of studies which feature the row code for binary variables, or the sample average calculated at the study level for continuous variables. For studies that had the feature present in one treatment arm but not another treatment arm, the code is counted as present if it is present in any treatment arm.

[b] Conditional on the summer program being offered in person.

**Table 2**

*Results of Estimating Unconditional Meta-Regression Models With Robust Variance Estimation (RVE)*

| | Dependent variable: All mathematics outcomes effect size (Hedges's *g*) | Dependent variable: Standardized mathematics achievement tests effect size (Hedges's *g*) | Dependent variable: School mathematics attainment outcomes effect size (Hedges's *g*) | Dependent variable: Non-cognitive skills effect size (Hedges's *g*) |
|---|---|---|---|---|
| Constant | 0.096** | 0.101*** | 0.111 | 0.114* |
| | (0.024) | (0.025) | (0.074) | (0.049) |
| N effect sizes | 112 | 96 | 16 | 37 |
| N studies | 37 | 34 | 6 | 8 |
| $\tau^2$ [a] | 0.008 | 0.008 | 0.015 | 0.032 |
| 95% prediction interval [b] | (-0.076, 0.268) | (-0.071, 0.273) | (-0.127, 0.348) | (-0.234, 0.463) |

*Note.* We assume the average correlation between all pairs of effect sizes within studies is 0.80.
[a] $\tau^2$ is the method of moments estimate of the between-study variance in the underlying effects provided by the *robumeta* package in Stata 15 (Tanner-Smith & Tipton, 2014).
[b] The 95% prediction interval is calculated as the estimated average effect size +/- 1.96* $\tau$.
*$p$ < .10. **$p$ < .05. ***$p$ < .01.

**Table 3**

*Results of Estimating Meta-Regression Models With Robust Variance Estimation (RVE) for Mathematics Achievement Outcomes Including Sample Characteristics (Poverty Level of Sample) as Moderators*

| | Dependent variable: Effect size (Hedges's *g*) |
|---|---|
| *Between-study effects* | |
| % of sample low-income (standardized) | -0.039 (0.037) |
| N effect sizes | 105 |
| N studies | 32 |
| $\tau^2$ [a] | 0.009 |
| | Weighted mean: Effect size (Hedges's g) |
| High-poverty sample (% low income >0.75) | 0.083*** |
| Mid-low poverty sample (% low income<=0.75) | 0.122* |

*Note.* We assume the average correlation between all pairs of effect sizes within studies is 0.80. Models include controls for randomized controlled trial or regression discontinuity study design and elementary school sample at the between-study and within-study levels. RVE = robust variance estimation.

[a] $\tau^2$ is the method of moments estimate of the between-study variance in the underlying effects provided by the *robumeta* package in Stata 15 (Tanner-Smith & Tipton, 2014).

*$p$ < .10. ***$p$ < .01.

**Table 4**

*Results of Estimating Meta-Regression Models With Robust Variance Estimation (RVE) for Mathematics Achievement Outcomes Including Program Duration/Intensity Indicators as Moderators*

| | Dependent variable: Effect size (Hedges's *g*) | | |
|---|---|---|---|
| *Between-study effects* | | | |
| Total program hours | 0.000<br>(0.000) | | |
| Program hours per day | | -0.004<br>(0.016) | |
| Hours per day on mathematics | | | 0.095*<br>(0.043) |
| N effect sizes | 100 | 73 | 54 |
| N studies | 31 | 28 | 22 |
| $\tau^2$ [a] | 0.009 | 0.010 | 0.006 |

*Note.* We assume the average correlation between all pairs of effect sizes within studies is 0.80. Models include controls for randomized controlled trial or regression discontinuity study design and elementary school sample at the between-study and within-study levels. RVE = robust variance estimation.

[a] $\tau^2$ is the method of moments estimate of the between-study variance in the underlying effects provided by the *robumeta* package in Stata 15 (Tanner-Smith & Tipton, 2014).

*$p < .10$.

**Table 5**

*Results of Estimating Meta-Regression Models With Robust Variance Estimation (RVE) for Mathematics Achievement Outcomes Including Summer Program Foci as Moderators*

| | Dependent variable: Effect size (Hedges's *g*) | | | | | |
|---|---|---|---|---|---|---|
| *Between-study effects* | | | | | | |
| Mathematics-specific focus | 0.176*<br>(0.078) | | | | 0.170**<br>(0.061) | 0.124<br>(0.204) |
| Program goal: Remediation | | -0.128<br>(0.086) | | | -0.104<br>(0.065) | |
| Standards-aligned | | | 0.075<br>(0.067) | | 0.025<br>(0.052) | |
| Online-only program | | | | -0.056<br>(0.089) | -0.094<br>(0.061) | |
| Hours per day on mathematics | | | | | | 0.066<br>(0.062) |
| N effect sizes | 112 | 112 | 112 | 112 | 112 | 54 |
| N studies | 37 | 37 | 37 | 37 | 37 | 22 |
| $\tau^2$ [a] | 0.010 | 0.011 | 0.012 | 0.012 | 0.010 | 0.006 |
| Results of joint *F* test | | | | | *F* = 2.47, *df* = 4, *p* = 0.152 | |

*Note.* We assume the average correlation between all pairs of effect sizes within studies is 0.80. Models include controls for randomized controlled trial or regression discontinuity study design and elementary school sample at the between-study and within-study levels. RVE = robust variance estimation.

[a] $\tau^2$ is the method of moments estimate of the between-study variance in the underlying effects provided by the *robumeta* package in Stata 15 (Tanner-Smith & Tipton, 2014).

*$p$ < .10. **$p$ < .05.

**Table 6**

*Results of Estimating Meta-Regression Models With Robust Variance Estimation (RVE) for Mathematics Achievement Outcomes Including Summer Program Activities as Moderators*

| | Dependent variable: Effect size (Hedges's *g*) | | | | | | |
|---|---|---|---|---|---|---|---|
| *Between-study effects* | | | | | | | |
| Commercially available curriculum | -0.016 (0.060) | | | | | | |
| Hands-on projects | | 0.062 (0.089) | | | | 0.077 (0.101) | |
| Textbook exercises | | | -0.112** (0.042) | | | -0.115** (0.044) | |
| Group work | | | | -0.016 (0.089) | | -0.044 (0.087) | |
| Computer-based skills practice | | | | | -0.003 (0.074) | -0.026 (0.079) | |
| Number of summer program activities | | | | | | | -0.024 (0.039) |
| N effect sizes | 112 | 112 | 112 | 112 | 112 | 112 | 112 |
| N studies | 37 | 37 | 37 | 37 | 37 | 37 | 37 |
| $\tau^2$ [a] | 0.013 | 0.012 | 0.015 | 0.012 | 0.012 | 0.017 | 0.014 |
| Results of joint *F* test | | | | | | *F* = 0.283, *df* = 4, *p* = 0.881 | |

*Note.* We assume the average correlation between all pairs of effect sizes within studies is 0.80. Models include controls for randomized controlled trial or regression discontinuity study design and elementary school sample at the between-study and within-study levels. RVE = robust variance estimation.

[a] $\tau^2$ is the method of moments estimate of the between-study variance in the underlying effects provided by the *robumeta* package in Stata 15 (Tanner-Smith & Tipton, 2014).

**$p < .05$.

**Table 7**

*Results of Estimating Meta-Regression Models With Robust Variance Estimation (RVE) for Mathematics Achievement Outcomes Including Summer Program Resources as Moderators*

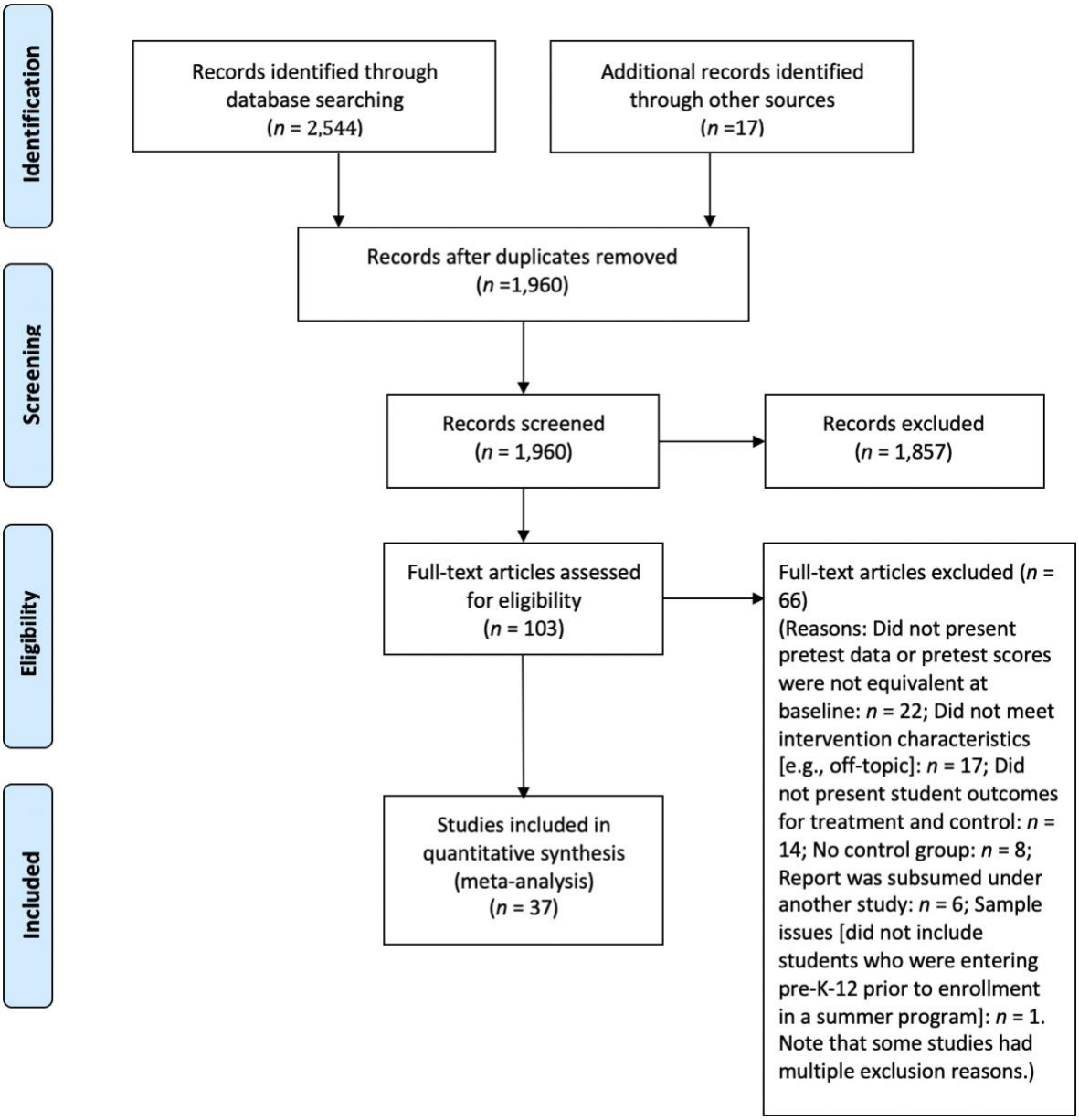
| | Dependent variable: Effect size (Hedges's *g*) | | | | |
|---|---|---|---|---|---|
| *Between-study effects* | | | | | |
| Teacher PD | 0.007<br>(0.053) | | | -0.040<br>(0.068) | |
| Lesson plans | | 0.069<br>(0.075) | | 0.081<br>(0.080) | |
| Transportation | | | -0.024<br>(0.053) | -0.024<br>(0.065) | |
| Class size | | | | | 0.013<br>(0.009) |
| N effect sizes | 112 | 112 | 105 | 105 | 51 |
| N studies | 37 | 37 | 34 | 34 | 17 |
| $\tau^2$ [a] | 0.009 | 0.013 | 0.013 | 0.018 | 0.006 |
| Results of joint *F* test | | | | *F* = 0.436, *df* = 3, *p* = 0.731 | |

*Note.* We assume the average correlation between all pairs of effect sizes within studies is 0.80. Models include controls for randomized controlled trial or regression discontinuity study design and elementary school sample at the between-study and within-study levels. Studies of online-only programs are excluded from the analysis of transportation. Average class size information was available for a subset of studies. There are no statistically significant effects at the $p < .10$ level. RVE = robust variance estimation.

[a] $\tau^2$ is the method of moments estimate of the between-study variance in the underlying effects provided by the *robumeta* package in Stata 15 (Tanner-Smith & Tipton, 2014).

**Figure 1**

*PRISMA Flow Diagram*




Source. Moher, Liberati, Tetzlaff, Altman, and The PRISMA Group (2009).
*Note.* PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.