

Fadeout and Persistence of Intervention Impacts on Social–Emotional and Cognitive Skills in Children and Adolescents: A Meta-Analytic Review of Randomized Controlled Trials

Emma R. Hart¹, Drew H. Bailey², Sha Luo³, Pritha Sengupta⁴, and Tyler W. Watts¹

¹ Department of Human Development, Teachers College of Columbia University

² School of Education, University of California, Irvine

³ Department of Life Sciences Communication, University of Wisconsin–Madison

⁴ Department of Psychology, Wesleyan College

Researchers and policymakers aspire for educational interventions to change children's long-run developmental trajectories. However, intervention impacts on cognitive and achievement measures commonly fade over time. Less is known, although much is theorized, about social–emotional skill persistence. The current meta-analysis investigated whether intervention impacts on social–emotional skills demonstrated greater persistence than impacts on cognitive skills. We drew studies from eight preexisting meta-analyses, generating a sample of 86 educational randomized controlled trials targeting children from infancy through adolescence, together involving 56,662 participants and 450 outcomes measured at posttest and at least one follow-up. Relying on a metaregression approach for modeling persistence rates, we tested the extent to which posttest impact magnitudes predicted follow-up impact magnitudes. We found that posttest impacts were equally predictive of follow-up impacts for cognitive and social–emotional skills at 6- to 12-month follow-up, indicating similar conditional persistence rates across skill types. At 1- to 2-year follow-up, rates were lower, and, if anything, cognitive skills showed greater conditional persistence than social–emotional skills. A small positive follow-up effect was observed, on average, beyond what was directly predicted by the posttest impact, indicating that interventions may have long-term effects that are not fully mediated by posttest effects. This pattern of results implied that smaller posttest impacts produced more persistent effects than larger posttest impacts, and social–emotional skill impacts were smaller, on average, than cognitive skill impacts. Considered as a whole, intervention impacts on both social–emotional and cognitive skills demonstrated fadeout, especially for interventions that produced larger initial effects. Implications for theory and future directions are discussed.

Public Significance Statement

This study found that a broad array of educational programs improved child social–emotional and cognitive skills at the program end. However, these improvements faded significantly in the following years. Although boosts to social–emotional skills are expected to persist and lay the foundation for children's future success, this study suggests that initial improvements to cognitive and social–emotional functioning were often short-lived. Although some educational programs produce life-altering benefits in the long-term, our review suggests that these long-term impacts are unlikely to be explained by fully persistent effects on targeted cognitive or social–emotional skills during childhood.

Keywords: meta-analysis, educational randomized controlled trials, fadeout, social–emotional skills, cognitive skills

Supplemental materials: <https://doi.org/10.1037/bul0000450.supp>

Blair T. Johnson served as action editor.

Emma R. Hart  <https://orcid.org/0000-0003-3808-0838>

Tyler W. Watts  <https://orcid.org/0000-0002-2741-0873>

The codebooks, analytic syntax, and data necessary to replicate these findings can be found on the LDbase at <https://doi.org/10.33009/ldbbase.1719529626.152e>. This study was not preregistered. Preliminary findings from this study have been presented at a convening of the Consortium of Early Childhood Intervention Impact, the Society for Research on Child Development's biennial conference, the European Association for Developmental Psychology conference, and a seminar series for the American Educational Research Association. Additionally, an earlier

version of this article has been posted as a preprint on EdWorkingPapers and is accessible at <https://www.edworkingpapers.com/ai23-782>.

Time spent on this project was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant 1R01HD095930-01A1 awarded to Tyler W. Watts and the National Science Foundation Grant DGE-2036197 awarded to Emma R. Hart. We are grateful to Mark Lipsey, Greg Duncan, Jens Dietrichson, Benjamin Lovett, Dana McCoy, Susan Kruglinski, and members of the Consortium of Early Childhood Intervention Impact (1R01HD095930-01A1) for their helpful comments on this work. The authors thank the following research assistants whose efforts made this work possible (in alphabetical order): Helen Ding, Precious Elam, Simran Juneja, Gabby Lammano, Siyu Liang, Sha Luo, Opal

continued

Creating and implementing educational interventions that support children to thrive in the moment and achieve desired outcomes in the long run are core priorities in the fields of psychology and education. Drawing on developmental theory, researchers and policymakers often anticipate that researcher-created educational interventions will improve child outcomes both initially and through cascading long-term impacts. Critically, however, an accumulating body of evidence suggests that initial intervention impacts commonly fade across subsequent follow-up assessments (Bailey, Duncan, et al., 2020). This phenomenon, termed *fadeout*, has been met with considerable concern. Fadeout raises questions about the translation of developmental theory to practice and casts doubt on the ability of interventions to change children's short- and long-run development trajectories in meaningful ways.

Despite growing concern around fadeout, the extent to which fadeout is, in fact, a pervasive phenomenon across broad classes of educational interventions and child outcomes is poorly understood. Most empirical work to date has focused on fadeout for cognitive outcomes (e.g., performance-based measures of math, language and literacy, intelligence), making it unclear whether this pattern of diminishing effects exists for social-emotional outcomes (Abenavoli, 2019). Given arguments that children's social-emotional skills are critical to adult success based on longitudinal analyses (Duckworth et al., 2018; Heckman & Kautz, 2012; Nagaoka et al., 2015; Soto et al., 2024), it is commonly hypothesized that intervention-driven boosts in social-emotional skills will persist more than boosts in cognitive skills.

This hypothesis aligns with the prevailing theory that unmeasured social-emotional skills drive long-term adult impacts of educational programs. Indeed, among a handful of evaluations that have collected adult follow-up data, several highly cited educational interventions have found emerging longer run impacts on measures of attainment despite observing fadeout, or consistently null effects, on cognitive test scores (e.g., Chetty et al., 2011; Deming, 2009; Gray-Lobe et al., 2022; for review, see Bailey, Duncan, et al., 2020). In such cases, the persistence of impacts on social-emotional skills, which are commonly unmeasured, has often been inferred to explain how long-term impacts on important life outcomes could be observed despite cognitive skill fadeout (Chetty et al., 2011; Heckman et al., 2013; Heckman & Kautz, 2012).

The hypothesis that impacts on social-emotional skills are more likely to persist than impacts on cognitive skills raises questions that can be empirically tested. Namely, are fadeout effects limited to measures of cognitive functioning and academic achievement? Or do fadeout effects also occur for measures of social-emotional skill development? The present study aimed to address this question. Using meta-analytic methods, the present study systematically investigated patterns of fadeout and persistence across a comprehensive collection of educational interventions and child outcomes.

To our knowledge, this is the first study to test whether fadeout is observed across both cognitive and social-emotional skills.

The Case for Social-Emotional Skill Persistence

Two explanations support the possibility that social-emotional skill development systematically differs from cognitive skill development, driving greater impact persistence for social-emotional skills. The first comes from skill-building models, which assert that more rudimentary skills lay the foundation for advanced skills. Intervention-driven skill boosts may increase the productivity of subsequent skill investments by enabling a child to leverage opportunities for growth (i.e., "skills beget skills"; Cunha & Heckman, 2007). Intervention-driven boosts in social-emotional skills could trigger cascades that reinforce further social-emotional skill development (i.e., *self-productivity*) and development in other domains (i.e., *cross-productivity*). Although many would predict similar dynamics for cognitive skills, interventions that effectively boost social-emotional skills could initiate uniquely effective and socially driven feedback loops between the child and their context, resulting in a developmental trajectory reflective of sustained treatment impacts. For example, an adolescent intervention targeting risky behaviors (e.g., Botvin & Griffin, 2004) may diminish the likelihood that teens will engage in social drinking. This could, in turn, improve relations with their parents and teachers, leading to further positive reinforcement of prosocial behavior from their environment (e.g., see also social information processing theory; Dodge et al., 1986). Targeting skills at key developmental moments may also produce crucial changes that lead to further skill advancement. For example, reductions in risky behavior could curtail school expulsion, altering long-run trajectories and ultimately propagating low-risk behaviors long term.

Second, impacts on social-emotional skills may be less prone to fadeout via control group *catch-up*. Catch-up occurs when posttreatment contexts and experiences provide children in the control group with opportunities to develop the skills that children in the treatment group acquired from the intervention (Bailey, Duncan, et al., 2020). Control group catch-up has been demonstrated to help explain cognitive skill fadeout (Elango et al., 2016; Watts et al., 2022). Bailey et al. (2017) argued that skill-building interventions are more likely to persist if they target skills unlikely to develop in counterfactual conditions. Whereas many cognitive skills (e.g., math, reading) are explicitly targeted in traditional school settings, social-emotional skills may receive less explicit focus in schools and other learning contexts.

A Skill-Type Null Hypothesis

A skill-based null hypothesis whereby fadeout for social-emotional skills is similar to that for cognitive skills is also possible, as social-emotional skill building may suffer from the same

Ofstedal, Fatmanur Ozay, Xinyu Pan, Spruha Reddy, Mindy Rosengarten, John Schupbach, Maddie Scricco, Pritha Sengupta, Jessica Sperber, Devon Turner, Leo Weaver, and Josefa Westerman.

Emma R. Hart played a lead role in data curation, formal analysis, investigation, project administration, writing—original draft, and conceptualization and methodology. Drew H. Bailey played a lead role in conceptualization and methodology, a supporting role in data curation, and an equal role in formal analysis and writing—review and editing. Sha Luo played a supporting role in writing—review and editing and an equal role in

data curation and investigation. Pritha Sengupta played a supporting role in writing—review and editing and an equal role in data curation and investigation. Tyler W. Watts played a lead role in conceptualization, methodology, resources, and supervision, a supporting role in data curation, and an equal role in formal analysis and writing—review and editing.

Correspondence concerning this article should be addressed to Tyler W. Watts, Department of Human Development, Teachers College of Columbia University, 462 Grace Dodge Hall, 525 West 120th, New York, NY 10027, United States. Email: tw2108@tc.columbia.edu

challenges as cognitive skill building. For example, control group catch-up may occur if interventions spur social-emotional development that would have naturally occurred through subsequent experiences. Alternately, educational interventions may struggle to initiate skill-building cascades for social-emotional outcomes because treatment impacts are unlikely to overcome the power of the individual-level (e.g., genetics, family environment) and contextual-level (e.g., socioeconomic resources, neighborhood) factors that contribute to the stability of individual differences after interventions end. Indeed, social-emotional capacities demonstrate traitlike stability (Rieger et al., 2017), albeit somewhat less than cognitive skills (Soland et al., 2019).¹

This raises the question: Could any process *other* than the persistence of social-emotional skill impacts explain long-run emergent impacts? A more nuanced alternative hypothesis could be that self- and cross-productivity of social-emotional skills, like cognitive skills, are real but limited. Longer term effects could emerge as a product of small carryover and transfer effects among various skill domains over time. In such cases, impacts on a given skill may diminish, but the intervention could affect long-run outcomes via initially declining treatment impact ripples that spread through a complex interconnected network of skills, contexts, and opportunities. The latter includes institutional gateways—environmental opportunities influenced by time-specific advantages in social-emotional or cognitive skills that produce more positive long-term outcomes (Bailey et al., 2017; Bailey, Duncan, et al., 2020). In the adolescent intervention example, imagine the boost in social skills prevented expulsions. Here, the short-term gain in social functioning itself may fade, but longer term effects could emerge as a result of staying in school.

The Present Study

The present meta-analysis set out to investigate whether the impacts of educational interventions are more persistent for social-emotional skills than for cognitive skills. For this study, we compiled a new meta-analytic data set: the Meta-analysis of Educational Randomized Controlled Trials (RCTs) with Follow-up (MERF). To our knowledge, MERF is the first meta-analysis of educational RCTs that has systematically investigated whether longitudinal impacts unfold distinctly by skill type. The analytic sample for this study was comprised of 86 interventions from the larger MERF sample with follow-ups on cognitive and/or social-emotional outcomes, sampled from eight preexisting meta-analyses. MERF allowed us to make several methodological and analytic innovations to address issues that have clouded previous work in this area. First, to limit internal validity issues, we only included RCTs. Second, to reduce bias due to selective outcome measurement and reporting across follow-ups, we only included constructs consistently measured at posttest and follow-up. For our main examination of the differences in persistence for social-emotional and cognitive skills, we made a priori analytic decisions but did not have strong a priori hypotheses.

Method

Process

Inclusion Determinations

Interventions were drawn for consideration from eight recent and influential meta-analyses on social-emotional and cognitive

intervention effects. These included: Bailey, Jenkins, and Alvarez-Vargas (2020); Burns et al. (2016); Kraft et al. (2018); Li et al. (2020); Protzko (2015, 2017); Suggette (2016); and Taylor et al. (2017). Our intention in starting with these eight meta-analyses was to create a broad and diverse sample of educational interventions to investigate patterns of persistence and fadeout. We intended to create a sample comprised of educational interventions that covered a range of common intervention foci and tactics. Further, we purposely selected meta-analyses that reported follow-up impacts (e.g., Bailey, Jenkins, & Alvarez-Vargas, 2020; Li et al., 2020; Protzko, 2015; Suggette, 2016; Taylor et al., 2017) and meta-analyses that have been widely cited and influential in the field (e.g., Kraft et al., 2018: 1,250 citations; Suggette, 2016: 469 citations; Taylor et al., 2017: 2,362 citations). Together, these eight meta-analyses captured some of the most important educational interventions considered by researchers and policymakers today, including public preschool and early childhood education, phonics-based reading programs, school-wide social-emotional interventions, and teacher professional development.

It should be noted that we did not attempt to capture the entire population of educational interventions reporting impacts on cognitive or social-emotional skills. Rather, we chose to sample from a broad set of preexisting meta-analyses with the goal of generating a varied set of studies with the same underlying causal expectation: Intervention boosts in skills at posttest will change children's developmental trajectories, as evidenced by later intervention impacts on follow-up measures. We intended to select a broad set of interventions that would allow us to examine whether patterns of treatment impact fadeout are observed across interventions that vary in key functions but share this underlying causal expectation.

As Table 1 illustrates, the sample includes reports published over the past 50 years, with the earliest article published in 1969 and the latest published in 2022 ($M = 2005$; $Mdn = 2008$). Among the interventions coded for inclusion in our study, approximately 52% aimed to improve child social-emotional skills broadly, 52% targeted child language/literacy skills, 13% targeted substance use, 10% targeted psychological well-being, and approximately 6%–7% targeted math and/or general cognitive functioning. Less than 2% of interventions targeted science, executive functioning skills, and/or general learning skills. The majority of these interventions (86%) involved a change in context (i.e., curricular intervention, enhanced pre-K) rather than the provision of an entirely new environment (i.e., after-school program, pre-K; 14%).

Together, these eight meta-analyses provided 426 unique reports, 400 of which had accessible PDFs in the English language and were reviewed. These 400 reports included impacts on 305 unique educational studies and were reviewed for inclusion in our sample. Figure 1 presents inclusion decisions. Inclusion criteria and decisions are briefly described below and are further described in the Supplemental Materials. First, we reviewed each study to determine if it utilized an RCT design. Only RCTs were included to limit the need to evaluate the internal validity of quasi-experimental studies, which can be subjective and difficult to determine. Of the original 305 studies, 196 utilized an RCT design (109 were excluded). Second, RCTs had to report at least one effect size (or data that could

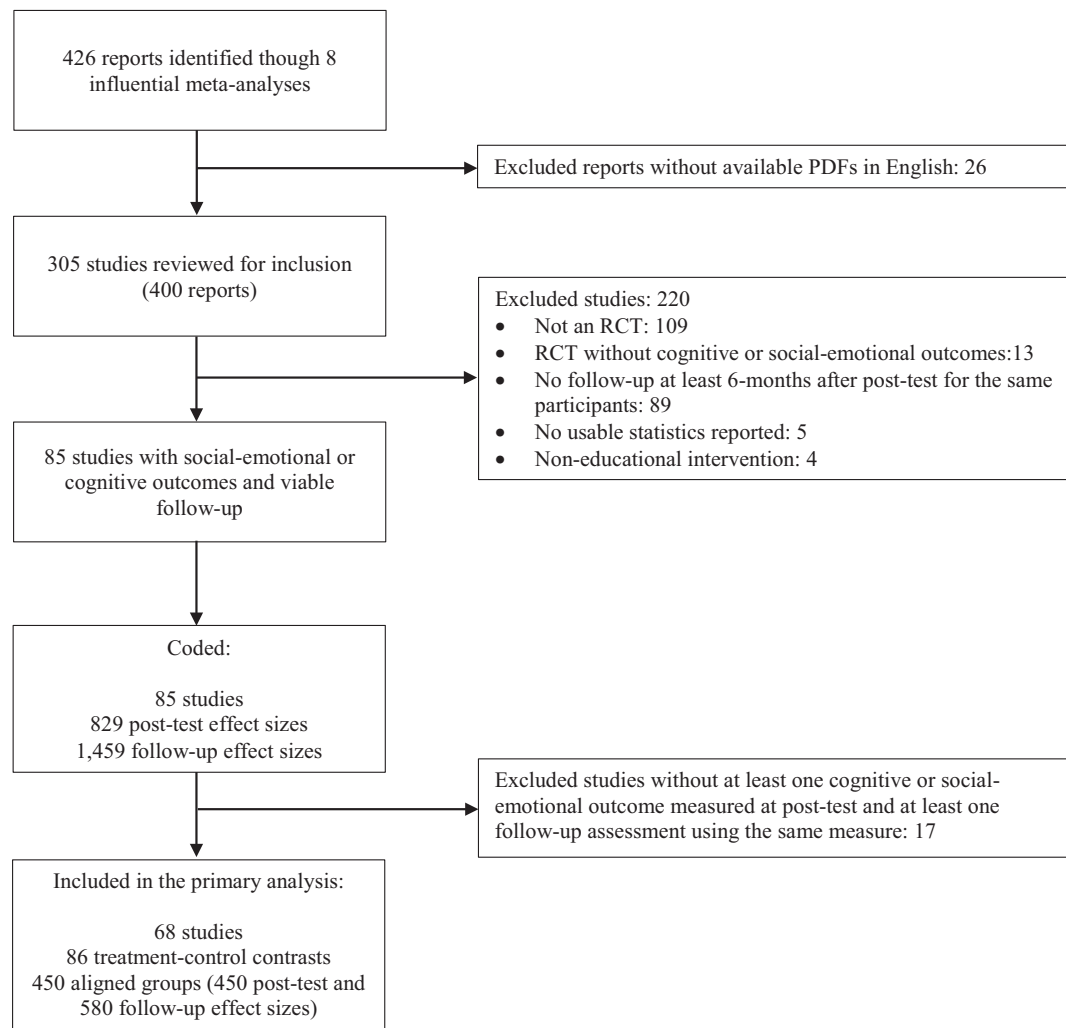
¹ Factors contributing to stability may also vary by skill type; meta-analytic work has suggested that environmental factors contribute more to stability in personality, whereas genetic factors contribute more to cognitive stability (Briley & Tucker-Drob, 2017).

Table 1
Intervention and Participant Characteristics, Mean (Minimum–Maximum)

Intervention/participant characteristic	All outcome		Social–emotional outcome		Cognitive outcome	
	Average (1)	% Nonmissing (2)	Average (3)	% Nonmissing (4)	Average (5)	% Nonmissing (6)
Report publication year	2005 (1969–2022)	100%	2006 (1987–2015)	100%	2005 (1969–2022)	100%
Baseline age (months)	95.56 (0–170)	99%	125.67 (42–170)	98%	67.80 (0–122)	100%
Intervention length (months)	7.29 (1–36)	81%	6.40 (1–18)	75%	8.12 (1–36)	85%
Intervention time (hours)	80.98 (4–1,075)	65%	21.73 (6–163)	50%	115.22 (4–1,075)	69%
Intervention type (%)						
Change in environment	85.88	99%	89.74	98%	83.33	100%
New environment	14.12	99%	10.26	98%	16.67	85%
Intervention targets (%)						
Math	6.98	100%	7.50	100%	11.11	100%
Language/literacy	52.33	100%	10.00	100%	83.33	100%
Science	1.16	100%	0.00	100%	1.85	100%
General cognition	5.81	100%	2.50	100%	9.26	100%
Executive functioning	1.16	100%	0.00	100%	1.85	100%
Learning skills	1.16	100%	0.00	100%	1.85	100%
Social–emotional skills	52.33	100%	100.00	100%	24.07	100%
Substance use	12.79	100%	27.50	100%	0.00	100%
Psychological well-being	10.47	100%	22.50	100%	0.00	100%
Intervention targets, categorized (%)						
Social–emotional only	41.86	100%	90.00	100%	7.41	100%
Cognitive only	45.35	100%	0.00	100%	72.22	100%
Broad	12.79	100%	10.00	100%	20.37	100%
Adult involvement (%)						
Teachers	54.65	100%	60.00	100%	55.56	100%
Parents	22.09	100%	25.00	100%	22.22	100%
Participant race/ethnicity (%)						
Asian	13.99	17%	10.46	15%	19.08	19%
Black	42.07	48%	35.18	48%	46.99	56%
White	54.77	42%	60.81	50%	42.85	41%
Hispanic	25.23	33%	17.34	35%	33.08	37%
Intervention location (%)						
Europe	23.26	100%	15.00	100%	25.93	100%
Middle East	1.16	100%	2.50	100%	0.00	100%
Multiple countries	3.49	100%	2.50	100%	3.70	100%
North America	61.63	100%	65.00	100%	64.81	100%
United States	60.47	100%	62.50	100%	64.81	100%
Oceania	10.47	100%	15.00	100%	5.56	100%
Female participants (%)	47.39	84%	50.92	90%	44.85	81%
Aligned groups <i>n</i>	5.23 (1–33)	100%	5.95 (1–19)	100%	4.93 (1–33)	100%
Participant <i>n</i> (at posttest)	727.71 (24–10,170)	100%	1332.33 (42–10,170)	100%	375.17 (24–3,928)	100%
Number of interventions		86		40		54

Note. This table presents intervention and participant characteristics for interventions that contributed “aligned groups” to the analytic sample. The level of observation in this table is the intervention (i.e., treatment–control contrast). In some cases, there was more than one intervention per study. Column 1 presents these characteristics for all outcomes contributing aligned groupings to the analytic sample, whereas Columns 3 and 5 present characteristics contributing at least one social–emotional or cognitive aligned group, respectively. Columns 2, 4, and 6 present the percentage of interventions that contributed to the averages (i.e., how representative the averages are of the full analytic sample). Note that the intervention target and adult involvement sections contain items that are not mutually exclusive (i.e., interventions could target both math and language/literacy skills or parents and teachers). Means are presented with accompanying rounded ranges (minimum to maximum) when appropriate.

Figure 1
Flow of Reports and Studies Into the Meta-Analysis



Note. RCT = randomized controlled trial.

be used to calculate this) for a cognitive or social-emotional outcome to be considered for inclusion; 183 studies included either cognitive or social-emotional outcomes (13 studies were excluded). Third, studies had to report follow-up treatment impacts for the same sample of children at least 6 months after the posttest, the identification of which required an extensive search process to gather all available reports of follow-up impacts. For this step, 94 studies met the criterion (89 studies were excluded for inadequate follow-up). Fourth, each study had to provide usable statistics (i.e., at least one follow-up effect size or data that could be used to calculate an effect size). Five studies were removed due to insufficient data, leaving 89 studies. Finally, although we had not initially excluded studies based on intervention focus, we revisited this decision for four studies that exclusively focused on nutrition supplementation. These were subsequently excluded because they were not educational in focus, making them qualitatively dissimilar to the other studies. Thus, the final sample contained 85 studies with impacts documented across 139 reports (see [Supplemental Materials](#)

for a complete list). The sample of studies included in the current analysis was further limited by additional criteria (see the Analytic Sample section).

Coding

Reports that met our inclusion criteria were double coded for extensive details and results. The coding team comprised a master coder (the first author, a doctoral student) and two additional coders (master's-level students). A doctoral-level study principal investigator (the last author) supervised the coding process. Before coding, the master coder led 6 months of coding training to ensure coders understood each data element (see [Supplemental Materials](#) for full coding protocol). Coders documented a variety of information about the intervention, including basic study information (e.g., level of randomization, publication year), intervention and control group details (e.g., intervention duration and intensity), treatment targets (e.g., parents, teachers) and inputs

(e.g., math skills, self-regulation), internal validity (e.g., whether baseline equivalence was addressed), and participant demographics (e.g., race, sex).

Treatment impacts for cognitive and social-emotional outcomes were also coded (e.g., means, standard deviations, effect sizes, p values), as were details on each reported treatment impact (e.g., author-reported construct name, measure used, the timing of assessment). Importantly, we coded pretest, posttest, and follow-up results. A follow-up wave was included if it reported impacts at least 6 months after the intervention ended. Broad definitions of what constituted cognitive (e.g., IQ, working memory, math, reading) and social-emotional (e.g., behavioral problems, prosocial behaviors, substance use, depressive symptoms) outcomes were used to guide coding. For the purposes of the current analyses, outcomes were further classified as “cognitive” if they were coded into these categories: achievement composites, general cognition, language and literacy, math, and other academic abilities. Outcomes were considered “social-emotional” if they were coded into these categories: crime, externalizing behaviors, internalizing symptoms, general social-emotional skills, or substance use. (Supplemental Table S2 provides examples of constructs and measures for each category.) Additional information on this construct categorization process, the coding process more generally, and a link to the coding protocol are provided in the Supplemental Materials.

After the training period, we tested reliability by checking discrepancies in coding on a random selection of 10 reports in the sample. Across the three coders, agreement ranged from 82% to 89%. All reports were subsequently double coded. A master’s-level research assistant identified discrepancies in coding, and the coding team frequently met to reach a consensus on all discrepancies. Study principal investigators were consulted when the coding team could not reach a consensus.

Effect Size and Standard Error Calculations

Following coding, effect sizes were calculated for each outcome on a case-by-case basis. In the most straightforward cases, author-reported effect sizes were used, or effect sizes were calculated using treatment and control group means and standard deviations. Details related to these cases are reported below. The supplement details other crucial aspects of this process, including deviations from typical calculation approaches and calculation formulas (see Supplemental Figure S1).

When treatment and control group means and standard deviations were reported, effect sizes were calculated using the formula for Glass’s Delta:²

$$ES = \frac{M_{tx} - M_{cntrl}}{Sd_{cntrl}}. \quad (1)$$

We used the author-reported effect sizes when they were reported in standardized units, so long as the estimation method produced a viable main effect (e.g., longitudinal effects could not be modeled using parametric assumptions in a growth curve model; the treatment indicator could not be used in interaction terms; mediators could not be included). There were some cases when viable effect sizes were reported, and it was also possible to calculate effects using means and standard deviations. In these cases, we relied on decision criteria to determine which effect size to use in our

analyses. The supplement provides details regarding these criteria, but the overarching approach in making effect size determinations was to arrive at the best estimate of the average treatment effect. In all cases, effect sizes were rescaled so that positive effect sizes reflected more desirable or “better” outcomes for the treatment group (e.g., a reduction in behavioral problems for the treatment group was rescaled as a positive treatment impact).

Standard errors and p values were also assigned to each effect size. When using author-reported effect sizes, we took the corresponding author-reported standard errors and p values so long as they were precisely reported. When these statistics were not available or when effect sizes were calculated using descriptive statistics, standard errors were calculated using the following formula (see Borenstein et al., 2009, p. 27):

$$SE_{ES} = \sqrt{\frac{n_{tx} + n_{cntrl}}{n_{tx}n_{cntrl}} + \frac{ES^2}{2(n_{tx} + n_{cntrl})}}. \quad (2)$$

In these cases, p values were estimated by calculating a t statistic from the effect size (i.e., effect size divided by standard error) and determining the associated p value. Degrees of freedom were set to the total sample size minus 2.

Clustered randomization was employed in several of the studies included in our sample. For author-reported effect sizes from studies that used cluster randomization, we assumed that the standard errors were generated from a model that properly adjusted for clustering. To adjust standard errors associated with effects that we calculated using reported descriptive information, we scaled standard errors by a variance inflation factor that assumed 20 clusters and an intraclass correlation coefficient of 0.10 (see Supplemental Materials for additional discussion and further sensitivity models).

Analysis

Previous research has identified several methodological factors that may bias meta-analytic estimates of short- and long-run impacts. These factors include selective reporting of short-run impacts, selective reporting of longer run impacts, and selection into follow-up data collection in the absence of selective reporting. (The first two are addressed in detail in Bailey, Duncan, et al., 2020; Bailey & Weiss, 2022; Watts et al., 2019 addresses the third.) With limited resources and pressure to publish, researchers may use a variety of selective reporting techniques to increase the likelihood of observing positive follow-up effects. Such decisions may be based on justifiable theoretical rationale, or they could be seen as efforts to curate the data to support a narrative of intervention effectiveness. For example, researchers may only collect additional follow-up measures of constructs that showed promising posttest impacts because they reason that it would be a waste of time and resources to follow up on measures that did not “yield” interesting results. Further, grant funders may only extend funding to studies that show “promising” short-term impacts (Watts et al., 2019), biasing the set

² Standard deviations from the control group were used to calculate effect sizes because a given treatment could have affected the variance of scores. Control group standard deviations reported in concordance with the treatment and control group means were used (i.e., concurrent standard deviations; see Supplemental Materials for a model in which only control group posttest standard deviations were used).

of studies that report follow-up effects at all. If follow-up carries on over multiple waves, researchers may shift their focus toward skills that they believe will show positive effects based on previous observations, and they may altogether stop measuring skills that show no effects over time. Alternatively, researchers may change the measures of a construct across follow-up assessments if an earlier assessment did not produce a positive effect.

When considered as a whole, it is impossible to predict exactly how these decisions will shape estimates of effect persistence. However, it is clear that such decisions could bias a meta-analytic examination of long-run follow-ups if one were to simply aggregate whatever measures were reported at each wave for a given study. To address these issues, we have devised an alternative approach that relies on careful alignment of measures within a study over time, and we have adopted an analytic technique that moves beyond simply observing average effects for each study at each follow-up wave. We address both of these issues, in turn, below.

Analytic Sample

The analytic sample for the current analysis was created by specifying groupings of constructs measured at posttest and follow-up assessments. We employed a strict operationalization of these groupings for which effect sizes had to meet several criteria. The effect sizes had to come from the same study and treatment-control contrast, they had to capture the same author-reported construct, they had to be measured using the same measure and subscale from the same reporter (e.g., self, parent, teacher), and they had to be collected at posttest and at least one follow-up assessment. Each sequence of effect sizes that met these criteria constituted one unit of analysis in our models and is referred to as an “aligned group.”

Of note, as stated above, aligned groups were comprised of effects from the same study and treatment-control contrast. Some studies had more than one randomly assigned treatment group (i.e., participants were randomized to more than one treatment condition). We refer to each unique treatment group created through randomization as a “treatment-control contrast” or “intervention” (both terms are used interchangeably hereinafter), for which some studies had multiple. For any one intervention, there could be none, one, or many sets of effect sizes depending on how many measures were collected consistently at posttest and at least one follow-up.

This grouping approach limits selection bias concerns, especially related to the collection and reporting of different measures across follow-up assessment waves. However, our approach does not perfectly address these measurement selection bias concerns, as we are ultimately limited by what has been reported in the literature. Although our approach is likely to mitigate bias due to selective reporting of newly added measures at follow-up, it may still contain bias if researchers only follow up on measures for which large posttest effects are observed. We attempt to probe these issues through sensitivity analyses (see Publication Bias and Selection into Follow-Up in the Results section) described in more detail below. We also recognize that researchers may introduce new measures at subsequent waves for legitimate scientific reasons (e.g., maturation necessitates a different measure to capture the same construct). Thus, we also ran supplemental analyses that relaxed our most restrictive sample inclusion criteria to allow for measures to vary for the same construct across follow-up assessments (see Results section for further discussion).

In addition to limiting the impact of researcher measurement selection issues described above, our exclusive focus on constructs measured consistently over time also increases the likelihood that our analysis contains measures that researchers determined were the key outcomes for a given study. Although we did not code for whether a given outcome was considered “confirmatory” or “exploratory” by the report authors (this is rarely mentioned), measures administered consistently at posttest and follow-up likely constituted the outcomes that the researchers most heavily prioritized. This should assuage concerns that report authors may have never expected persistent impacts for the outcomes included in our study.

Analytic Approach

Our empirical strategy for modeling the persistence of intervention impacts relies on a theoretical model that predicts a causal relation between the posttest effect and follow-up effect on a given skill measured consistently over time. Here, we assume that a given intervention impacts skill *X* at follow-up via two pathways: (a) changes in skill *X* at posttest and (b) changes in other skills or environments, *Y*, at posttest. The exact functional form of the decay or persistence of impacts on skill *X* will depend on several unknown parameters that we cannot estimate in the present study, such as the number of skills or contextual factors impacted by the intervention (measured or unmeasured) in the vector *Y* and the effects of skill *X* and *Y* at posttest on subsequent levels of skill *X* and *Y* (i.e., the intertemporal stability of a given skill and the cross-productivity of related skills over time). The causal links between skill *X* and *Y* over time can also vary across the type of skill, intervention, and population of interest. However, under any realistic model predicting intervention impact persistence or fadeout on a given skill over time, follow-up effect sizes on skill *X* will depend on the posttest effect size on skill *X*. Indeed, a priori, the magnitude of the observed difference between the treatment and control group on skill *X* at posttest is expected to drive the magnitude of the difference between groups on skill *X* at follow-up.

This assumption of an underlying causal pathway, by which interventions should generate posttest impacts that should predict subsequent follow-up treatment impacts, is well supported in the educational intervention literature. Indeed, educational interventions that target specific skill development in both cognitive and social-emotional domains are typically motivated by theories of skill building that conceptualize skill development on a continuum. This is reflected in the measures that are typically used to capture skill acquisition, which allow for operationalizations that reflect hierarchical levels of skill development (e.g., item response theory based measures of cognitive achievement, Likert-based measures of behavior). It is also reflected in regression-based approaches that relate skill development at one time point to skill development at a later time point (e.g., Duncan et al., 2007). Our modeling approach follows this conceptualization, as we hypothesize that larger posttest skill impacts should lead to larger intervention-driven skill impacts at follow-up.

Thus, we regressed impacts for skill *X* at follow-up on impacts for skill *X* at posttest. This model generates two important and interesting parameters that provide inferences regarding skill *X* and *Y*. First, the slope term from this model represents the effect that an impact on skill *X* at posttest has on the subsequent impact on skill *X* at follow-up, plus the effects of any additional factors that an

intervention might change in accordance with skill X at posttest (i.e., typical omitted variables bias). The slope term is inherently conditioned on posttest impact magnitude and, as such, we refer to this term as the *conditional persistence rate*.³ This rate can be understood as capturing the component of a follow-up intervention effect on skill X that is driven by posttest intervention effects on skill X . The second term of interest is the intercept, which represents the effects of Y at posttest on skill X at follow-up, for all parts of Y that do not reliably change with skill X in response to intervention (i.e., changes that are uncorrelated with posttest impacts on skill X). To the extent that measures of skill X at posttest do not fully capture the underlying skill X construct at posttest, the intercept will also capture these effects.

While we discuss the slope and intercept as separate parameters, given that each provides unique and theoretically valuable information, both terms must be considered together when forming inferences about expected follow-up effects. Indeed, considered additively, the intercept and conditional persistence rate (i.e., slope term) forecast follow-up impacts conditional on posttest impact magnitude. Given our interest in the extent to which *posttest boosts* to child skills remain constant or diminish, we place particular emphasis on the conditional persistence rate as a key parameter of interest in estimating skill-specific persistence and fadeout. In other words, it would be surprising if we found that posttest impacts were totally unpredictable of follow-up impacts, and such a result would be at odds with any reasonable theory of skill building. However, because of the possibility that factors other than posttest impact could influence longer term impacts, both the intercept and slope must be considered together when generating overall conclusions regarding persistence or fadeout.

Our regression-based approach to measuring fadeout and persistence overcomes several limitations imposed by simply observing absolute changes in intervention impacts (i.e., absolute changes in follow-up effects from posttest effects). First, whereas an absolute approach does not distinguish the contribution of intervention impacts on skill X versus vector Y in estimating persistence, the relative regression-based approach parses these components into the slope and intercept terms, respectively. Second, the absolute approach fails to account for the fact that small changes in effects are more or less meaningful at different magnitudes of posttest impact. For example, a reduction in treatment impacts of 0.05 SD is evidence of substantially more fadeout for an intervention with a posttest impact of 0.08 SD than a posttest impact of 0.50 SD . To further highlight the advantages of this approach, consider an absolute approach that would simply take the difference in standard deviation units between follow-up effect sizes and the posttest impact. A given intervention that produced a posttest effect size of 1 SD with a 1-year follow-up effect of 0.25 SD would be described as producing an effect that faded by 0.75 SD . However, if another iteration of this intervention produced a posttest effect of 0.50 SD , would we then predict that the 1-year follow-up would be $-0.25 SD$ based on our previous observation of 0.75 SD unit fadeout in 1 year? Using the same hypothetical example, a relative approach to modeling fadeout would find a rate of 25% conditional persistence from posttest to 1 year-follow-up (i.e., $0.25/1$) and would predict a follow-up effect of 0.13 SD for the second intervention that produced a posttest impact of 0.50 SD (i.e., $0.50 SD \text{ Posttest Impact} \times 25\% \text{ Persistence Rate}$).

This analytic approach is also useful in that it allows for an examination of differences in conditional persistence by skill type across a diverse array of outcomes from a diverse array of interventions. Insofar as varying intervention and outcome features are salient enough to drive differences in posttest impacts, such features are “controlled for” in determining rates of posttest skill persistence using this approach, assuming that the subsequent skill-building dynamics are not influenced by intervention features present before posttest impacts are observed. Thus, by determining the conditional persistence rate, we can make an apples-to-apples comparison of the extent to which an intervention posttest impact (of any magnitude) proportionally persists at follow-up, holding constant absolute differences in posttest effects due to intervention and outcome features that could vary systematically across social-emotional and cognitive skills.

One final and important feature of this modeling approach is that it embraces the assumption of nonlinearity of treatment impact fadeout across follow-ups. Given the expectation of nonlinearity in persistence based on previous fadeout work in early childhood education interventions (see Bailey et al., 2017), we modeled the intercept and conditional persistence rate at three distinct periods: at least 6 months to 1 year (e.g., 6-month follow-up, 12-month follow-up), greater than 1 year and up to 2 years (e.g., 14-month follow-up, 24-month follow-up), and greater than 2 years (e.g., 25-month follow-up, 60-month follow-up).⁴ We fit separate linear regression models in which posttest impacts predicted follow-up impacts for each of the aforementioned periods. We binned follow-up waves and treated impacts binned at various time points as independent outcomes because we expected that fadeout effects would be nonlinear over time, but we doubt that this is an exactly exponential function. Some previous studies have attempted to impose structured functional forms to estimate fadeout, such as linearity. Such models impose unrealistic constraints on the data: They predict effects will eventually flip to negative, which rarely happens, and the most distal time points (which are most likely to suffer from selection artifacts) are heavily weighted because they are farthest from the centroid. It should also be noted that our binning of the data followed common conceptualizations of follow-up time points in intervention research, as many educational studies follow students on an annual basis during follow-up. Indeed, in our full data set, a plurality of follow-ups were collected at 12 months after posttest (31%), followed by 10% collected at 24 months after posttest.

Analytic Models

Using a sample of the total available cognitive and social-emotional posttest impacts carefully linked with follow-up impacts (i.e., same intervention, construct, and measure at both times), we

³ We are not the first to conceptualize of a linear-regression-based slope term as “persistence” (e.g., see Blanden, 2019; Marques & Dias, 2005).

⁴ Occasionally, there were multiple assessments of the same measure and construct within these categorical time bins. In these cases, estimates were averaged within the following “bins” so that there was one estimate per time bin and measure: 6- to 12-month follow-up, greater than 1 year and up to 2 years follow-up, greater than 2 years and up to 3 years follow-up, greater than 3 years and up to 4 years follow-up, and greater than 4 years follow-up. For the purposes of these regression models, the “greater than 2 years” bin was then reconstructed to capture the longest term follow-up effect in the case that more than one was available.

assessed the extent to which posttest effect sizes were predictive of follow-up effect sizes and the portion of follow-up impacts driven by factors other than skill-specific posttest effects. We then tested whether these parameters varied by skill type (i.e., social-emotional vs. cognitive). Because many studies contributed multiple outcomes in our analysis, we addressed the nonindependence of effect sizes from the same study using a random-effects model. Given the expectation of nonlinearity in persistence, we assessed persistence in posttest impacts at three distinct periods: at least 6 months to 1 year, greater than 1 year and up to 2 years, and greater than 2 years. We ran separate linear regression models in which posttest impacts predicted follow-up impacts for each of these follow-up periods.

Multilevel random-effects metaregressions were executed in R using the metafor package with aligned groups as the unit of observation, random effects at the study level, and study-level clustered standard errors. In effect, these models included effect sizes for a given measure and construct grouping (i.e., aligned group) at Level 1, with studies at Level 2. On average, our data included 6.62 aligned groups per study. Note that we also ran the correlated and hierarchical model (Pustejovsky & Tipton, 2022) as a supplemental analysis in which treatment-control contrast group was considered as a third level. On average, our data included 5.23 aligned groups per treatment-control contrast. Results were aligned with the findings from our more parsimonious two-level approach. Thus, we used the following model to estimate patterns of persistence in cognitive versus social-emotional treatment impacts:

Level 1 measure/construct groupings:

$$ES_{fsi} = \beta_{0s} + \beta_{1s}ES_{psi} + \varepsilon_{fsi}, \quad (3)$$

Level 2 study:

$$\beta_{0s} = \gamma_{00} + \tau_{0s}, \quad (4)$$

$$\beta_{1s} = \gamma_{10} + \tau_{1s}, \quad (5)$$

where i indicated aligned group (i.e., the same construct reported at multiple assessment waves collected using the same measure for the same intervention), s indicated study, f indicated the follow-up assessment wave, and p indicated the posttest assessment wave. Thus, ES_{psi} represents the corresponding effect size for aligned group i from study s at posttest, and ES_{fsi} represents the corresponding follow-up effect; ε_{fsi} captures error in the estimation of ES_{fsi} , which is assumed to be normally distributed and independent across observations. At Level 2, γ_{00} captures the grand mean intercept across all studies, and τ_{0s} is the study-level random effect that captures study-specific variance in the intercept term (i.e., study-level variance in predicted follow-up effects when posttest impacts are 0). The inclusion of this random effect adjusts for the nonindependence of effect sizes estimated within the same study (i.e., a random constant term is modeled for each respective study). We also included a random effect for the slope term, τ_{1s} , which assessed the extent to which the conditional persistence rate varied across studies. We specified an unconstrained covariance matrix, which allowed the two random effects to correlate.

Random-effects weighting was used to place greater weight on effect sizes estimated with greater precision. This approach incorporates the inverse variance-covariance matrix of the full metaregression model (i.e., both the error associated with each

effect and between-study variability given the study random effect). In other words, weights were determined by both estimation precision at the individual effect size level (i.e., the magnitude of each standard error associated with an effect) and between-study variability (i.e., the magnitude of variation across studies). Cluster-robust standard errors, clustered at the study level, were used in all models.

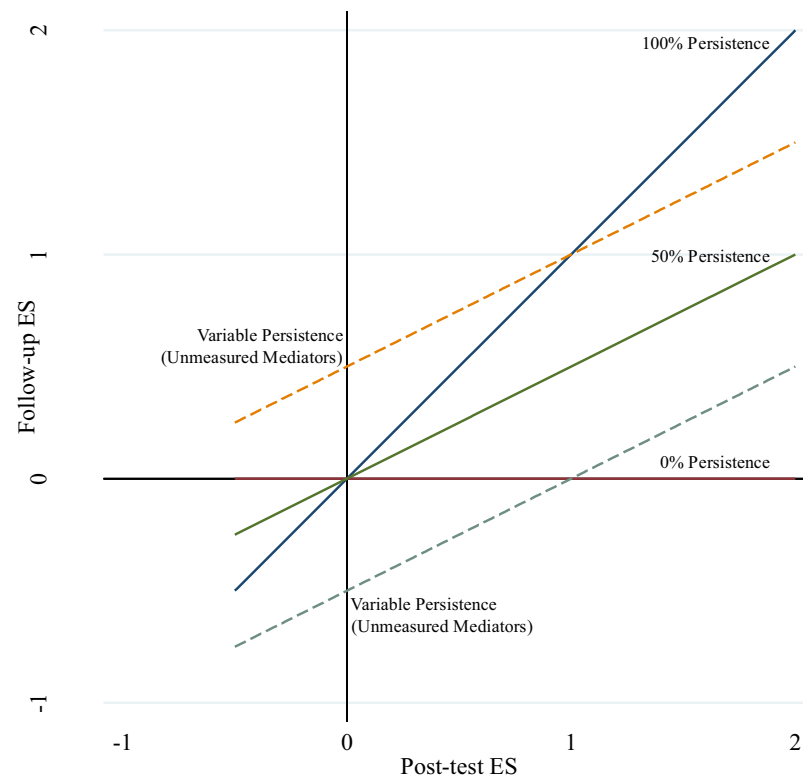
As we noted above, this specification produces estimates of two important parameters that must be considered together to estimate follow-up effects, each contributing novel insight about fadeout dynamics. Figure 2 demonstrates possible patterns for these terms. The first parameter is the slope term (β_{1s}), or conditional persistence rate. To understand the predictions made by β_{1s} , consider several scenarios in which β_{0s} is equal to 0. In Figure 2, the blue “100% Persistence” line demonstrates a case where posttest effects do not fade, and posttest magnitude perfectly predicts follow-up magnitude (i.e., $\beta_{1s} = 1$). If this were true, a posttest effect size of 0.50 SD would predict a follow-up impact of 0.50, with no fadeout observed. In contrast, the red “0% Persistence” line demonstrates the opposite case: Regardless of posttest magnitude, the follow-up effect size is 0. Under this scenario, all intervention impacts fade out regardless of the magnitude of the posttest effect. The green “50% Persistence” line represents a slope of $\beta_{1s} = 0.50$, in which the follow-up effect is expected to be 50% of the posttest impact. Here, a posttest impact of 0.50 SD would lead to a follow-up effect of 0.25 SD .

The second key term, β_{0s} , is the y-intercept. Because our data are comprised of effect sizes and we did not further normalize the data for analyses (i.e., regression parameters can be interpreted in raw “effect-size units”), β_{0s} is defined as the predicted follow-up impact when the posttest impact is zero. As such, in this model, the intercept indicates the extent to which factors *uncorrelated* with posttest effects contribute to average follow-up effects. In other words, a nonzero intercept would suggest that follow-up effects are still observed, on average, even when posttest effects are zero. This is demonstrated by the orange line, where a slope of .50 is graphed with a positive y-intercept. Such a pattern of effects is consistent with the “dark matter” hypothesis that early interventions may affect other latent skills not captured by posttest impacts (e.g., Elango et al., 2016; Pages et al., 2022). One could imagine that a positive intercept effect could be observed if an intervention had impacts on an array of skills that drove follow-up effects on the outcome of focus. For example, a broad early childhood intervention could have a zero posttest impact on mathematics achievement, but one could observe a positive follow-up impact on mathematics ability if the intervention produced impacts on other skills that support mathematical development in later periods (e.g., language). This pattern could instead, or additionally, be evidence of measurement error, if meaningful treatment-driven variance in a particular skill is captured at follow-up but not at posttest.

Likewise, an intercept below 0 would indicate that the follow-up effect is smaller than expected based on the posttest measure, again assuming a linear relation between the posttest and follow-up (see gray dashed line). A negative y-intercept could be expected if the intervention produced long-term adverse effects, even when short-term impacts are positive (e.g., results from the Tennessee voluntary pre-K study; Durkin et al., 2022).

With the current data, we tested Equation 3 on our full set of eligible follow-up impacts across the cognitive and social-emotional domains. We then fit the following model with an interaction term to

Figure 2
Hypothetical Patterns of Fadeout/Persistence



Note. This figure portrays various hypothetical patterns of fadeout and persistence with different modeled intercept and slope terms. The blue “100% Persistence” line portrays $y = x$, where the intercept is 0 and the slope is 1. The red “0% Persistence” line portrays $y = 0$, where the intercept and slope are 0. The “50% Persistence” line represents $y = .5x$ where the intercept is 0 and the slope is .50. The dashed “Variable Persistence (Unmeasured Mediators)” lines represent two hypothetical patterns: (a) $y = .5 + .5x$ where the intercept and slope are .50 (orange) and (b) $y = -.5 + .5x$ where the intercept is $-.50$ and the slope is .50 (gray). ES = effect size. See the online article for the color version of this figure.

test our primary question regarding differences between cognitive and social-emotional impacts:

Level 1 measure/construct groupings:

$$ES_{fsi} = \beta_{0s} + \beta_{1s}ES_{psi} + \beta_{2}SOC_{si} + \beta_{3}ES_{psi} \times SOC_{si} + \varepsilon_{fsi}, \quad (6)$$

Level 2 study:

$$\beta_{0s} = \gamma_{00} + \tau_{0s}, \quad (7)$$

$$\beta_{1s} = \gamma_{10} + \tau_{1s}, \quad (8)$$

where ES_{fsi} and ES_{psi} are defined as before. Here, we add a dummy indicator, SOC_{si} , capturing whether a construct/measure/intervention i in study s falls within the cognitive or social-emotional category (1 = social-emotional outcome; 0 = cognitive outcome). If β_2 were positive in the full interaction model, this would indicate that additional unmeasured factors lead to stronger follow-up effects for social-emotional outcomes than cognitive outcomes. We then include the interaction between the posttest effect and the social-emotional indicator, denoted by β_3 . If β_3 were positive, this would indicate that

the conditional persistence rate is greater for social-emotional skills than cognitive skills. Of note, we additionally ran a host of analyses to test the robustness of our primary model specifications to issues regarding underlying study quality. For these analyses, we probed whether our findings hinged on the inclusion of potentially biased studies or effect sizes and whether our findings were robust to alternative analytic decisions.

Transparency and Openness

We largely adhered to the Meta-analysis Reporting Standards (Appelbaum et al., 2018). The codebooks, analytic syntax, and data necessary to replicate our analyses can be found on the LDbase at <https://doi.org/10.33009/ldbase.1719529626.152e> (Hart et al., 2024). This study was not preregistered.

Results

Descriptive Information

Before limiting our sample to posttest and follow-up assessments collected using the same construct and measure over time, the

inclusion process—beginning with studies from eight preexisting meta-analyses—yielded 85 studies with 829 posttest effect sizes and 1,459 follow-up effect sizes. After imposing our aligned measures criterion, the sample contained 68 studies with 86 treatment-control contrast groups (i.e., interventions), 450 posttest impacts for 56,662 participants, and 580 follow-up impacts. The supplemental file includes forest plots showing the average cognitive and social-emotional posttest effects for these 86 treatment groups (see Supplemental Figures S2 and S3).

Table 1 details intervention and participant characteristics for these treatment groups and for treatments reporting social-emotional and/or cognitive outcomes specifically. The sample was comprised of reports published from 1969 to 2022 ($M = 2005$; $Mdn = 2008$). The majority of these interventions (86%) involved a change in context (i.e., curricular intervention, enhanced pre-K) rather than the provision of an entirely new environment (i.e., after-school program, pre-K; 14%). The interventions targeted a range of child skills and capacities. About 52% aimed to improve child social-emotional skills broadly, 52% targeted child language/literacy skills, 13% targeted substance use, 10% targeted psychological well-being, and around 6%–7% targeted math and/or general cognitive functioning. Fewer than 2% of interventions targeted science, executive functioning skills, and/or general learning skills. Of note, many interventions targeted more than one skill (e.g., social-emotional skills and language/literacy skills). When considered together, about 42% of interventions targeted only social-emotional skills, 45% targeted only cognitive skills, and 13% targeted both.⁵ For the social-emotional outcomes in our sample, 90% came from interventions targeting social-emotional skills only, whereas 72% of cognitive outcomes came from interventions targeting cognitive skills only.

In addition to targeting children's development, about 55% of the interventions also targeted teachers, and about 22% targeted parents. On average, participants were about 8 years old at baseline, though interventions reporting social-emotional outcomes involved older children ($M = 10$ years old) than those reporting cognitive outcomes ($M = 6$ years old). Interventions varied considerably in intended treatment length ($M = 7$ months, range = 1–36 months) and, among interventions for which it was possible to compute intended treatment time (only 65% of interventions), treatment time was higher for cognitive ($M = 115$ hr) than social-emotional skills ($M = 22$ hr). This intensity discrepancy was largely driven by a few outlier interventions, and our supplemental analyses address the role of age and intervention intensity as potential confounds. Fewer than half of the reports indicated characteristics on sample race and ethnicity, making these estimates less instructive. Sixty-two percent of interventions took place in North America (60% in the United States, 1% in Canada) followed by 23% in Europe, 10% in Oceania, 4% across multiple countries, and 1% in the Middle East. On average, each intervention contributed five aligned construct-measure groupings to our sample (range = 1–33).

Table 2 presents the descriptive information for outcomes included in our analytic sample and demonstrates the variety of skills measured at posttest and follow-up assessment waves. As Table 2 demonstrates, a total of 54 interventions (from 40 studies) contributed 236 cognitive constructs measured using the same assessment at posttest and at least one follow-up. On average, we observed between one and two follow-up assessments for each aligned group ($M = 1.41$; range = 1–3).⁶ The average time elapsed

between the posttest and follow-up was 10.52 months. The majority of the cognitive outcomes were language and literacy related (83%), followed by math (9%) and general cognitive outcomes (5%; e.g., IQ). On average, samples were comprised of 304 participants.

For social-emotional outcomes, 40 interventions (from 33 studies) contributed 214 effect sizes measured at posttest and follow-up using the same measure. In alignment with cognitive outcomes, we observed between one and two follow-up assessments for each grouping ($M = 1.57$; range = 1–3), and follow-up measures were collected approximately 9 months after the posttest. The most common (42%) social-emotional outcomes were broad assessments of social-emotional skills (e.g., composites of externalizing and internalizing behaviors, prosocial behaviors) followed by substance use (29%), and internalizing (16%) outcomes. Sample sizes for social-emotional outcomes were much larger than for cognitive outcomes (average $n = 1,392$).

Trajectories of Fadeout

Average Effect Sizes Across Assessments

First, we descriptively charted treatment impact trajectories across all follow-up assessments provided to gain a broad understanding of patterns of persistence and fadeout. Table 3 presents the average weighted effect sizes (and associated 95% prediction intervals; see Int'Hout et al., 2015) for all outcomes and for cognitive and social-emotional outcomes considered separately. Figures 3 and 4 display the longitudinal treatment impact trajectories for cognitive and social-emotional outcomes. For cognitive and social-emotional outcomes, effects for interventions with larger samples hovered closer to null than those with smaller samples. Across the board, the RCTs had positive impacts at posttest, though there was considerable range in posttest effect size magnitude across both types of skills. Overall, we observed a fading average intervention impact for both cognitive and social-emotional skills.

This pattern was particularly clear for cognitive outcomes, where we observed a 0.40 SD weighted impact ($p < .001$) that faded to 0.22 SD ($p < .001$) by the 6- to 12-month follow-up, 0.16 SD ($p = .03$) by the 1- to 2-year follow-up, and 0.05–0.14 SD ($p = .14$ –.65) at subsequent follow-ups conducted at least 2 years after posttest. Social-emotional impacts were smaller at posttest (0.14 SD , $p < .01$) and minimally different at the 6- to 12-month follow-up (0.16 SD , $p < .01$). Then, at all subsequent follow-ups, effects were more imprecisely estimated and hovered closer to 0: 0.05–0.11 SD ($p = .02$ –.12). Critically, patterns in these descriptives are susceptible to various biases, including selective (non)reporting of (non)significant follow-up estimates and selective collection of follow-up data.

⁵ To be categorized as a cognitive intervention, an intervention had to target at least one of the following cognitive skills: math, language and literacy, executive function, general cognition, or science. To be categorized as a social-emotional intervention, the intervention had to target at least one of the following social-emotional skills: general social-emotional skills, learning skills, substance use, or psychological well-being. To be categorized as an intervention with broad targets, the intervention had to target at least one cognitive skill and one social-emotional skill.

⁶ One follow-up assessment was constituted as having at least one impact estimate measured in one of the following assessment periods: 6 months to 1 year; greater than 1 year, up to 2 years; greater than 2 years, up to 3 years; greater than 3 years, up to 4 years; and greater than 4 years.

Table 2
Cognitive and Social-Emotional Outcome Descriptives

Outcome/treatment focus	ES from group (%)	Treatment group (no.)	Aligned grouping (no.)	M no. of follow-ups	M months since posttest	M posttest ES, weighted	M posttest ES (SE), unweighted	Posttest ES $p < .05$	M_n
Cognitive outcomes		54	236	1.41	10.52	0.40 (0.06)***	0.39 (0.06)***	40%	304
Language and literacy	83%	46	196	1.29	9.48	0.44 (0.07)***	0.43 (0.07)***	42%	256
Math	9%	17	21	1.30	11.46	0.07 (0.10)	0.02 (0.09)	14%	400
Cognitive-general	5%	7	12	2.54	20.86	0.40 (0.13)*	0.42 (0.13)*	58%	224
Other academic ability	2%	4	5	1.00	5.20	0.12 (0.02)†	0.09 (0.04)	20%	794
Achievement composite	1%	2	2	3.25	19.50	0.25 (0.06)	0.26 (0.06)	100%	2068
Social-emotional outcomes		40	214	1.57	8.72	0.14 (0.04)**	0.17 (0.05)**	23%	1,392
Social-Emotional-general	42%	28	89	1.57	8.38	0.17 (0.06)*	0.21 (0.09)*	25%	1,474
Substance use	29%	11	63	1.05	6.90	0.13 (0.06)†	0.18 (0.06)†	30%	2,262
Internalizing	16%	17	34	1.69	7.81	0.10 (0.06)	0.12 (0.07)	21%	565
Externalizing	11%	8	24	2.09	12.52	0.08 (0.03)†	0.08 (0.04)†	8%	608
Criminality	2%	2	4	3.00	16.95	0.09 (0.03)	0.08 (0.03)	100%	883
Treatment type									
Cognitive only TXs	44%	39	196	1.21	8.90	0.45 (0.07)***	0.40 (0.07)***	38%	118
Social-emotional only TXs	46%	36	207	1.42	7.80	0.15 (0.04)**	0.17 (0.05)**	24%	1,413
Broad TXs	10%	11	47	2.51	17.98	0.35 (0.10)**	0.29 (0.11)*	45%	1,047

Note. Effect sizes are in standard deviation units. A positive effect size indicates that children in the treatment group displayed a more desirable outcome than children in the control group. Weighted effect sizes were estimated using a random-effects meta-analytic model that included a random effect for study, weights, and robust standard errors. The average number of assessments reflects the average number of follow-up assessments that were collected (at least 6 months after posttest). Timing of assessments refers to the average number of months that elapsed between posttest at follow-up assessment(s). The number of aligned groups refers to the number of groupings in which a posttest and at least one follow-up assessment of the same construct measured using the same measure, subscales, and reporter within an intervention was reported. ES = effect size; M_n = average participant sample size; TX = treatment.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3
Average Effect Sizes Across Posttest and Follow-Up Assessment Waves

Assessment wave	All outcome		Social-emotional outcome		Cognitive outcome	
	<i>M</i> (<i>SE</i>) (1)	<i>n</i>	<i>M</i> (<i>SE</i>) (2)	<i>n</i>	<i>M</i> (<i>SE</i>) (3)	<i>n</i>
Posttest	0.29 (0.04)*** [−0.48, 1.06]	450	0.14 (0.04)** [−0.39, 0.67]	214	0.40 (0.06)*** [−0.42, 1.22]	236
6 months to 1 year	0.21 (0.03)*** [−0.37, 0.79]	420	0.16 (0.05)** [−0.52, 0.84]	197	0.22 (0.04)*** [−0.21, 0.65]	223
>1 year, up to 2 years	0.07 (0.02)** [−0.12, 0.26]	87	0.05 (0.01)* [0.03, 0.07]	59	0.16 (0.06)* [−0.26, 0.58]	28
>2 years, up to 3 years	0.08 (0.05) [−0.25, 0.41]	31	0.11 (0.05) [†] [−0.12, 0.34]	22	0.05 (0.10) [−0.42, 0.52]	9
>3 years, up to 4 years	0.10 (0.05) [†] [−0.19, 0.39]	29	0.09 (0.03) [0.02, 0.16]	9	0.12 (0.08) [−0.33, 0.57]	20
>4 years	0.14 (0.08) [−0.24, 0.52]	13			0.14 (0.08) [−0.24, 0.52]	13

Note. Effect sizes are in standard deviation units. A positive effect size indicates that children in the treatment group displayed a more desirable outcome than children in the control group. The values in brackets indicate the lower and upper bounds of 95% prediction intervals (see [Int'Hout et al., 2015](#)). The analytic sample was constituted by “aligned groupings” which included a posttest and at least one follow-up effect size for the same construct measured using the same measure, subscales, and reporter within an intervention. Average effect sizes were estimated using a random-effects meta-analytic model that included a random effect for study, weights, and cluster-robust standard errors (clustered at the study level). *SE* = standard error; *n* = number of aligned groups.

* $p < .05$. ** $p < .01$. *** $p < .001$.

While interesting, such descriptive patterns are insufficient for understanding whether persistence rates differ by skill type. Simply observing the meta-analytic impact at both posttest and follow-up does not directly address the extent to which posttest impacts on a specific skill are persisting or fading over time. Indeed, the skills that produce larger follow-up effects at 6- to 12-month follow-up could be different than the skills that produce larger posttest impacts, making the simple average a flawed indicator of skill-based persistence of posttest impacts.

Modeled Persistence

To examine patterns of fadeout more rigorously and with the same outcomes matched over time, we fit metaregression models to test the extent to which follow-up effects were a function of posttest impacts on the same skill (i.e., slope) and other uncorrelated factors (i.e., intercept). We then assessed whether these patterns differed for cognitive and social-emotional outcomes. Results for 6- to 12-month follow-up and 1- to 2-year follow-up can be found in [Table 4](#) and in [Figures 5](#) and [6](#). Results for estimates greater than 2 years after posttest exhibited low estimation precision and, as such, are detailed in the supplement (see [Supplemental Table S3](#)).

Overall, we found that social-emotional skills and cognitive skills displayed similar conditional persistence rates at 6- to 12-month follow-up and that social-emotional skills produced a smaller conditional persistence rate than cognitive skills at 1- to 2-year follow-up. The slope for 6- to 12-month follow-up effects was 0.45 ($p < .001$) and did not differ by outcome type (i.e., see interaction in [Table 4](#), Column 3; $p = .91$). This indicates that the skill-specific contribution to a 6- to 12-month follow-up effect was expected to be 45% of the magnitude of the observed posttest impact. For 1- to 2-year follow-up, we observed a statistically nonsignificant overall

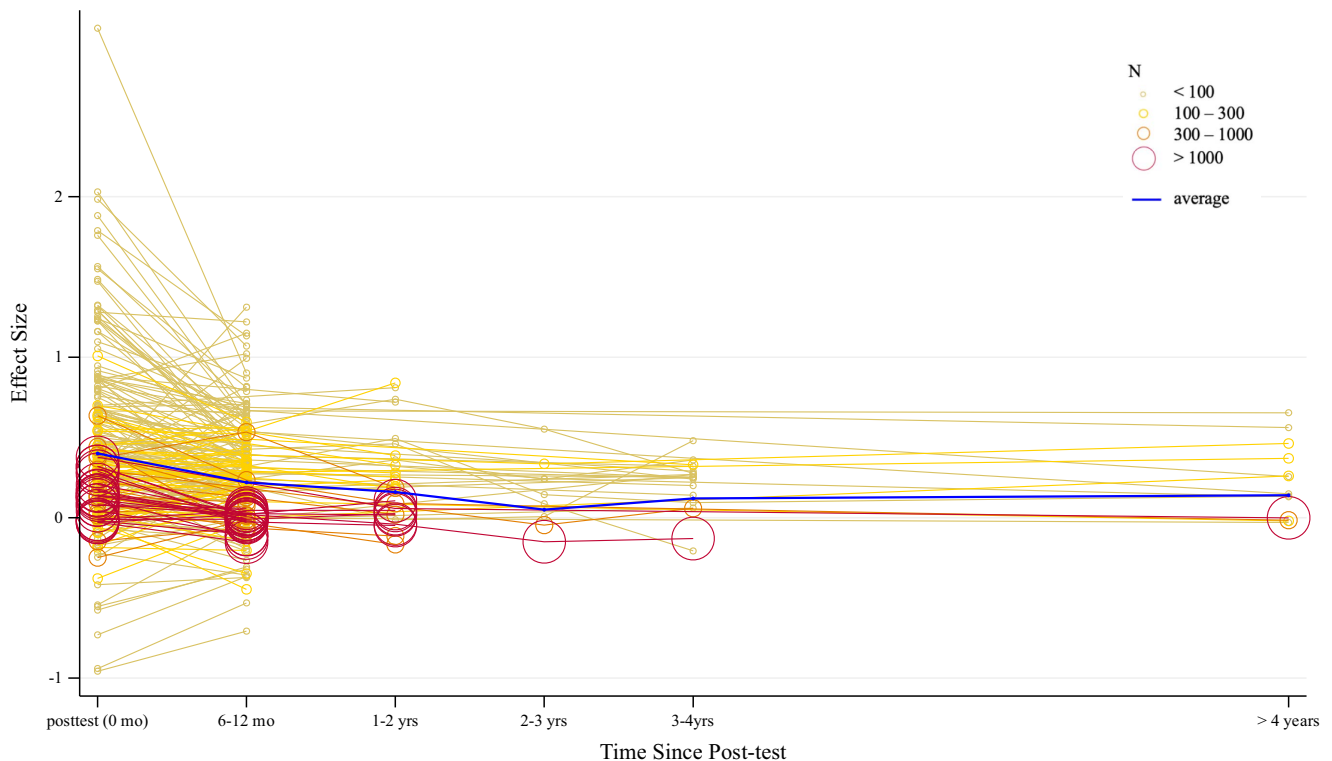
effect of posttest on follow-up, suggesting little predictive association between posttest and follow-up impacts across both skill types ($\beta_1 = 0.18$, $p = .15$). However, the posttest impacts on cognitive skills appeared to be more predictive of follow-up effects than were posttest impacts on social-emotional skills, though this difference was not statically significant ($p = .13$). Point estimates from this model should be interpreted with caution given that they were estimated with low precision on a small sample of effects (87 aligned groups).

Turning to the second parameter of interest, we observed evidence of small intercept effects across all of the models. In [Table 4](#) Column 1, we observed an intercept across skill types of 0.07 and 0.02 at the follow-up time points that was statistically significant at the first follow-up but not the second. There were some indications of slightly larger, or statistically stronger, intercept terms for social-emotional skills, especially in the skill-specific models at 1- to 2-year follow-up. However, the full interaction models suggested that these differences were small (β_2 ranged from 0.03 to 0.07) and not statistically significant ($p = .14$ – $.20$).

Heterogeneity in Persistence

At the 6- to 12-month follow-up, we observed substantial heterogeneity in follow-up effects between studies ($I^2 = 61\%$; $\tau_{\text{intercept}} = 0.23$). Consistent with our results, introducing posttest effect size and outcome type into the model considerably reduced the model heterogeneity ($\tau_{\text{intercept}} = 0.11$ and $I^2 = 36\%$), but the inclusion of the interaction between posttest and outcome type did not ($\tau_{\text{intercept}} = 0.11$ and $I^2 = 33\%$). The inclusion of random slopes in the model indicated substantial heterogeneity in the association between posttest and follow-up effect size that was similarly not

Figure 3
Effect Size Trajectories—Cognitive Outcomes



Note. Each line represents a cognitive construct that was measured at posttest and at least one follow-up assessment using the same measure for the same intervention. The average effect size trajectory is displayed in blue and was calculated using random-effects meta-analysis with weights. As detailed in the key, coordinates were weighted by the posttest sample size (larger coordinates represent estimates from larger samples) with aligned color coding (darker colors represent larger sample sizes). For display purposes, effect sizes within the -1 to 3 SD range are presented. As such, effect size trajectories less than -1 SD s at posttest are not displayed ($n = 1$ group). See the online article for the color version of this figure.

explained by the inclusion of outcome type in the model ($\tau_{\text{slope}} = 0.23$ – 0.25 across models).

We observed less between-study heterogeneity in 1- to 2-year follow-up effects (null model: $\tau_{\text{intercept}} = 0.07$; $I^2 = 22\%$). This variability was minimally explained by posttest effect size ($\tau_{\text{intercept}} = 0.04$; $I^2 = 20\%$) and skill type ($\tau_{\text{intercept}} = 0.04$; $I^2 = 20\%$), and marginally explained by the interaction between posttest and skill type ($\tau_{\text{intercept}} = 0.03$; $I^2 = 5\%$). Notably, there was substantial heterogeneity in the association between posttest and follow-up effects ($\tau_{\text{slope}} = 0.36$), partially explained by the interaction between posttest and skill type ($\tau_{\text{slope}} = 0.27$).

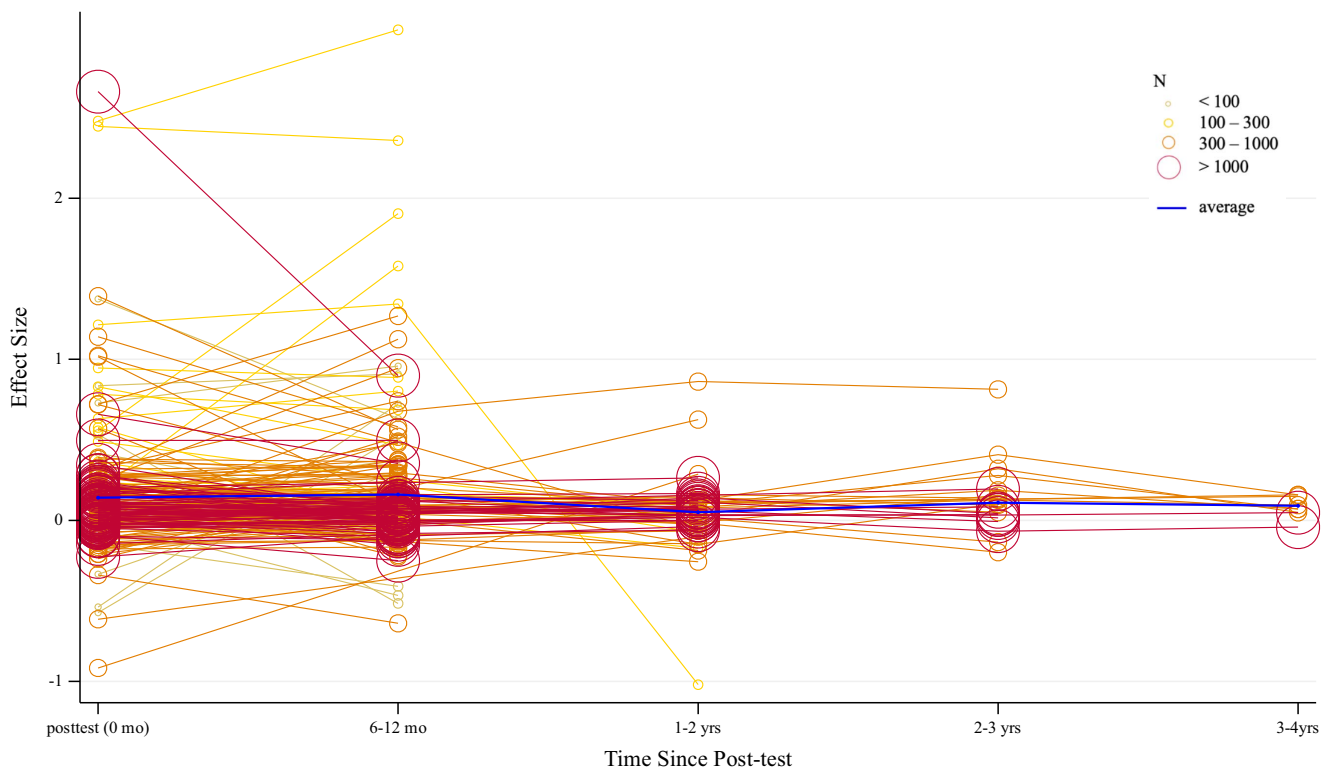
Forecasted Follow-Up Effects and Total Persistence

To concretize these results, we applied our findings to several example data points in our sample to estimate forecasted follow-up effects and overall persistence rates, considering both the slope and intercept from our predictive models. Using our regression-based parameters, we estimated the follow-up effect, absolute fadeout in standard deviation units, and total persistence using four plausible posttest intervention impacts: 0.14 SD , 0.25 SD , 0.40 SD , and 0.50 SD . Of note, 0.14 SD was the meta-analytic average posttest for

social-emotional skills in the sample, and 0.40 SD was the average for cognitive skills.

We note a few important takeaways from these forecasted effects. First, at 6- to 12-month follow-up, social-emotional skills generally showed greater total persistence than cognitive skills regardless of posttest impact magnitude. This pattern was due to the fact that the intercept effect was slightly larger for social-emotional skills than for cognitive skills. Second, for both cognitive and social-emotional skills, smaller posttest impacts showed greater persistence, driven by the intercept effect; when a posttest effect is small, the intercept is a larger proportion of the follow-up effect. This intercept-driven difference is particularly apparent when comparing the overall persistence for the social-emotional skill posttest average (0.14 SD) with that for the cognitive average (0.40 SD). Although we might observe more persistence for social-emotional skills than cognitive skills at 6- to 12-month follow-up, it is important to note that our regression results suggested that there were no statistically significant skill type differences in the slope term nor the intercept term, so any differences inferred from Table 5 should be made with caution (see the Discussion section). At 1- to 2-year follow-up, overall persistence was greater for cognitive skills than social-emotional skills across almost all posttest intervention impact magnitudes due to the conditional persistence rate. Indeed, social-emotional

Figure 4
Effect Size Trajectories—Social-Emotional Outcomes



Note. Each line represents a social-emotional construct that was measured at posttest and at least one follow-up assessment using the same measure for the same intervention. The average effect size trajectory is displayed in blue and was calculated using random-effects meta-analysis with weights. As detailed in the key, coordinates were weighted by the posttest sample size (larger coordinates represent estimates from larger samples) with aligned color coding (darker colors represent larger sample sizes). See the online article for the color version of this figure.

skill follow-up impacts were nearly entirely predicted by the intercept at the 1- to 2-year follow-up.

Exploratory Analyses

Intervention Type

The current analysis focused on charting patterns of persistence across social-emotional and cognitive skill *outcomes*. Although we operated from the assumption that outcomes measured at posttest and at least one follow-up assessment represented target outcomes that researchers anticipated their intervention would impact, it is certainly possible that persistence patterns could vary by the stated intervention focus (i.e., whether the intervention directly targeted social-emotional and/or cognitive skills). Thus, we conducted additional exploratory analyses to determine whether patterns of fadeout varied according to the skills targeted by the intervention: social-emotional skills only, cognitive skills only, or both (i.e., “broad” interventions; see Table 6). In brief, we found no strong or consistent evidence to suggest that the type of intervention heavily influenced persistence across follow-up waves. Results should be interpreted cautiously given the small sample sizes for outcome by treatment target subsamples and large standard errors on almost all estimates.

Sensitivity Analyses

Probing Intervention-Related Differences

We conducted several sensitivity analyses to test the robustness of the primary findings and found that they were generally robust to various model specifications. First, we ran a series of analyses to probe the extent to which intervention-related differences may have biased our estimates. To address the possibility that variation in study-level characteristics could bias estimated patterns of persistence, we tested a model in which we dropped the random effect and, instead, introduced a study-level fixed effect (i.e., the inclusion of a dummy variable control for each study). The inclusion of the study-level fixed effect controlled for any unobserved study-level intervention characteristics, thus constraining our key parameters to studies that reported effects on *both* cognitive and social-emotional outcomes. Estimates from this model should reflect differences in persistence for cognitive and social-emotional outcomes within the same study, holding constant study-related features. Of note, only eight interventions (from five studies) contained both social-emotional and cognitive skills measured consistently at posttest and follow-up. Overall, this model produced estimates that were substantively aligned with those from the primary model, though less precise (see Supplemental Table S4, Column 1).

Table 4*Regression-based Persistence Estimates for Social–Emotional and Cognitive Outcomes*

Parameter label	All outcome			Split outcome	
	Baseline β (SE) (1)	Main effect β (SE) (2)	Interaction β (SE) (3)	Cognitive β (SE) (4)	Social–emotional β (SE) (5)
Panel A: 6- to 12-month follow-up					
Posttest effect	0.45*** (0.05)	0.46*** (0.05)	0.46*** (0.06)	0.45*** (0.06)	0.41*** (0.09)
Soc (vs. Cog) outcome		0.03 (0.01)	0.03 (0.01)		
Posttest $\times$ Soc (vs. Cog) Outcome			0.01 (0.09)		
Constant	0.07** (0.02)	0.05* (0.02)	0.05 [†] (0.02)	0.06 [†] (0.03)	0.06* (0.02)
Sample size (study/int/outcome)	60/77/420	60/77/420	60/77/420	35/48/223	30/37/197
$\tau_{\text{intercept}}$ (null = .23, .17, .27)	.11	.11	.11	.10	.10
τ_{slope}	.25	.23	.24	.20	.31
I^2 (null = 60.50%, 53.60%, 66.24%)	35.79%	32.80%	32.47%	8.61%	47.89%
Panel B: >1 year, up to 2 years follow-up					
Posttest effect	0.18 (0.12)	0.21 (0.13)	0.35* (0.14)	0.35* (0.12)	−0.04 (0.18)
Soc (vs. Cog) outcome		0.03 (0.02)	0.07 (0.04)		
Posttest $\times$ Soc (vs. Cog) Outcome			−0.40 (0.24)		
Constant	0.02 (0.01)	0.00 (0.02)	−0.02 (0.03)	−0.03 (0.03)	0.05* (0.01)
Sample size (study/int/outcome)	23/24/87	23/24/87	23/24/87	11/11/28	13/14/59
$\tau_{\text{intercept}}$ (null = .07, .16, .00)	.04	.04	.03	.04	.02
τ_{slope}	.36	.37	.27	.23	.35
I^2 (null = 21.51%, 50.92%, 0.00%)	19.60%	20.12%	5.21%	28.44%	0.00%

Note. “Soc (vs. cog) outcome” is a dummy variable for outcome type (0 = cognitive outcome, 1 = social–emotional outcome). The unit of analysis is “aligned groupings” of posttest and follow-up impacts collected for the same construct using the same measure, subscales, and reporter at posttest and at least one follow-up assessment within an intervention. As such, each regression tests the association between posttest and follow-up effect sizes for the same measure. The conditional persistence rate is captured in the estimates of the posttest effect, and the interaction between posttest and the social–emotional skill indicator. Panel A presents associations between posttest and follow-up effect sizes collected 6 to 12 months after posttest. Panel B presents associations between posttest and follow-up collected more than 1 year and up to 2 years after posttest. Model parameters were estimated using a random-effects meta-analytic model that included a random effect for study, random slopes for posttest effect size, weights for effect size precision, and cluster-robust standard errors (clustered at the study level). Sample sizes are reported for studies, interventions, and outcomes. Heterogeneity statistics are presented for three null models estimating follow-up effect size magnitude using random-effects metaregression. First, they are presented for the model that includes all outcomes, followed by the model including only cognitive outcomes, and finally the model including only social–emotional outcomes. Negative I^2 values were rounded to zero. SE = standard error; Soc = social–emotional; Cog = cognitive.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

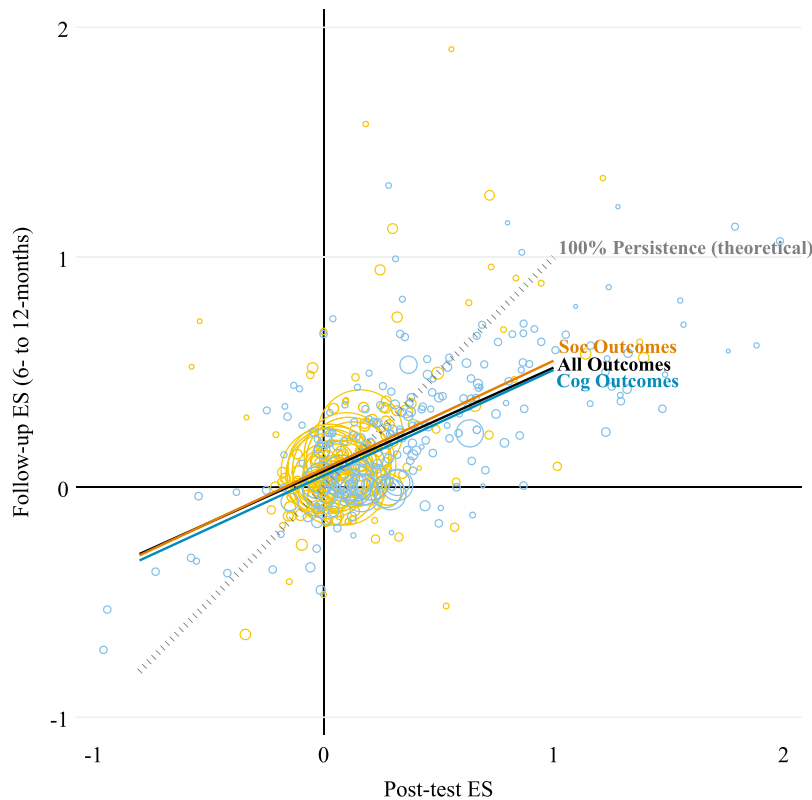
To further examine if key intervention differences could have affected our results, we ran additional models to test whether persistence was biased by the two intervention features that markedly varied across interventions reporting cognitive and social–emotional impacts: participant age and intervention duration. On average, social–emotional outcomes came from interventions with older participants and shorter timeframes than cognitive outcomes. Thus, we fit a model controlling for average participant age at baseline and intervention duration in months, as well as the interaction between these variables and posttest effect size. To address missingness on these covariates, we imputed missing age and duration values using the average for social–emotional and cognitive outcomes, respectively, and included dummy variables that indicated whether imputation was used. This model yielded less precise estimates with large standard errors but demonstrated the same patterns of conditional persistence as the primary model (Supplemental Table S4, Column 2).

Next, given that intervention intensity as measured by hours was only available for 65% of outcomes and controlling for this variable would significantly reduce our sample size, we ran models that compared persistence for cognitive and social–emotional outcomes from interventions that were more similar in intervention intensity than was the case in the sample at large. To do so, we limited the sample to interventions involving fewer than 200 hr of content and

then, further, into a subset involving less than 100 hr. This reduced the average hours of intervention for cognitive outcomes to 49.42 and 32.08 hr, respectively (social–emotional average = 14.30 hr). Interventions that did not report intensity were still included in these two models. Again, these checks produced largely consistent results (Supplemental Table S4, Columns 3 and 4).

Finally, we ran an expanded version of these control models in which we controlled for dichotomous indicators of additional, plausibly salient intervention features including: intervention duration, child age at baseline, parent involvement, whether the intervention added school time or involved a curricular and/or environmental change, whether the intervention targeted struggling students, whether the intervention targeted social–emotional and cognitive skills simultaneously, and publication year. We also controlled for the interaction of each of these variables and posttest effect size, as well as a dummy variable to indicate whether imputation was used to address missingness, as appropriate (as described above for the other control model). The 6- to 12-month estimates were in line with our primary results. The 1- to 2-year estimates were quite imprecise, with large standard errors, and suggested near-zero conditional persistence rate for both social–emotional and cognitive skills (likely due to the inclusion of highly correlated covariates). At 6- to 12-month follow-up, these intervention features did not appear to confound modeled

Figure 5
Persistence Patterns: Posttest to 6- to 12-Month Follow-Up



Note. Each coordinate represents an impact on a cognitive or social-emotional construct that was measured at posttest and at 6 to 12 months after posttest (i.e., one “aligned group”). Yellow coordinates represent social-emotional impacts, and blue coordinates represent cognitive impacts. Model parameters were estimated using a random-effects meta-analytic model with weights (see Table 4, Panel A). Coordinates were weighted by $1/SE^2$ (larger coordinates represent estimates with smaller standard errors at 6- to 12-month follow-up). The gray dashed “100% Persistence” line was included for reference. For display purposes, posttest and follow-up effects within the -1 to 2 SD range were presented. The plotted estimates are from Table 4, Panel A, Column 3. ES = effect size; Soc = social-emotional; Cog = cognitive. See the online article for the color version of this figure.

persistence and differences across social-emotional and cognitive skills (Supplemental Table S4, Column 5).

Probing Alternate Analytic Approaches and Concerns

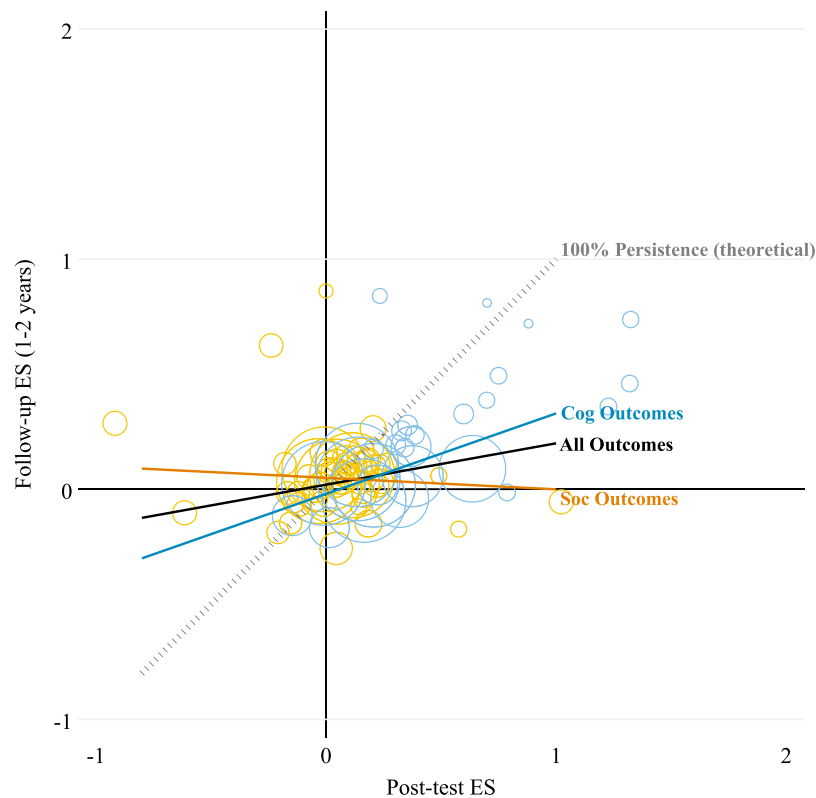
We then turned our attention to robustness tests in which we employed reasonable alternate analytic approaches and addressed data-related concerns.

First, we ran a series of analyses to address various dependencies in the data. Among these was the correlated and hierarchical model (Pustejovsky & Tipton, 2022) in which we used nested random effects; aligned groups were nested within interventions that were then nested within studies. This model accounts for the dependency of effects at the study and intervention levels and within-group heterogeneity in effects. These results were in line with our primary findings (see Supplemental Table S5, Column 1). Interestingly, the model indicated that at 6- to 12-month follow-up, intercept-level variation was largely explained by within-study/within-intervention

group variance ($\tau = .09$) rather than between-group differences ($\tau = .00$ for study and intervention). At 1- to 2-year follow-up, variance in intercept effects was estimated at both the between-study/ between-intervention levels ($\tau = .05$ for study, $\tau = .00$ for intervention) and within-study/within-intervention levels ($\tau = .06$).

Second, to account for the possibility that effects from studies gathered from the same meta-analysis could be correlated, we ran our primary model with study-level random effects (both intercept and slope) nested within random effects (both intercept and slope) for source meta-analysis (see Supplemental Table S5, Column 2). At 6- to 12-month follow-up, we observed little variation in follow-up effects explained by meta-analysis-level variation ($\tau = .01$). Similarly, there was little heterogeneity explained by meta-analysis level differences in follow-up effects at the 1- to 2-year follow-up assessment ($\tau = .01$). We view this as somewhat unsurprising, given the likelihood that the impacts of intervention features on follow-up effect sizes, which may be correlated within origin meta-analysis, are substantially captured by our key predictor: posttest

Figure 6
Persistence Patterns: Posttest to 1- to 2-Year Follow-Up



Note. Each coordinate represents an impact on a cognitive or social-emotional construct that was measured at posttest and at 1 to 2 years after posttest (i.e., one “aligned group”). Yellow coordinates represent social-emotional impacts, and blue coordinates represent cognitive impacts. Model parameters were estimated using a random-effects meta-analytic model with weights (see Table 4, Panel B). Coordinates were weighted by $1/SE^2$ (larger coordinates represent estimates with smaller standard errors at 1- to 2-year follow-up). The gray dashed “100% Persistence” line was included for reference. For display purposes, posttest and follow-up effects within the -1 to 2 SD range were presented. The plotted estimates are from Table 4, Panel B, Column 3. ES = effect size; Soc = social-emotional; Cog = cognitive. See the online article for the color version of this figure.

effect size. However, there was more heterogeneity in the meta-analysis-level random slopes at 6- to 12-month follow-up ($\tau = .23$) and 1- to 2-year follow-up ($\tau = .15$). Given that there were only eight meta-analyses, with several contributing very few aligned groups to our sample and two contributing the majority of aligned groups (Suggate, 2016; Taylor et al., 2017), these slope estimates may be heavily influenced by between-study differences. The slope coefficient is likely to be particularly unstable across small subgroups of studies, because small variation in follow-up effect sizes can affect it dramatically. For example, for two subgroups of studies with a posttest impact of $.18$ SD, if one subgroup has a follow-up impact of $.06$ SD and the other has a follow-up impact of $.12$ SD, the estimated persistence rates from these subsets would vary substantially (33% vs. 67%). However, in such cases, the absolute heterogeneity of the residuals would appear far smaller (follow-up effects would have a mean of around $.09$ SD, with one subset showing a positive residual of $.03$ SD and another showing a negative residual of the same magnitude). We think the very small residual variances estimated

from these models better convey the nature of heterogeneity across meta-analyses than the large slope variances.

Next, we attempted to address concerns related to large-sample interventions and the related issue of clustered randomization. In our primary model, interventions with larger samples were weighted more heavily. Although this is desirable for determining the best meta-analytic estimates, it is worth acknowledging that the interventions with larger samples may substantively differ from interventions with smaller samples. For example, interventions at scale may be implemented with weaker fidelity and program developer control. These larger sample interventions are also more likely to be those that employed clustered randomization (i.e., randomization at the level of classrooms, schools, districts). Insofar as these larger cluster-based RCTs generate smaller posttest impacts, we expect that these differences should not affect our key estimated parameters as the slope is conditioned on posttest impact magnitude. However, there may exist concerns that the dynamics of fadeout (i.e., association between posttest and follow-up impacts)

Table 5
Forecasted Follow-Up Effects and Persistence for Various Plausible Posttest Effects

Parameter label	ES/%							
	Social-emotional outcome				Cognitive outcome			
Hypothetical posttest effect	0.14 ^a	0.25	0.40	0.50	0.14	0.25	0.40 ^a	0.50
6- to 12-month follow-up	$ES_{fsi} = 0.08 + (.47) ES_{psi}$				$ES_{fsi} = 0.05 + (.46) ES_{psi}$			
Estimated follow-up effect	0.15	0.20	0.27	0.32	0.11	0.17	0.23	0.28
Absolute fadeout in <i>SD</i>	-0.01	0.05	0.13	0.19	0.03	0.09	0.17	0.22
% Persistence	104%	79%	67%	63%	82%	66%	59%	56%
1- to 2-year follow-up	$ES_{fsi} = 0.05 - (.05) ES_{psi}$				$ES_{fsi} = -0.02 + (.35) ES_{psi}$			
Estimated follow-up effect	0.04	0.04	0.03	0.03	0.03	0.07	0.12	0.16
Absolute fadeout in <i>SD</i>	0.10	0.21	0.37	0.48	0.11	0.18	0.28	0.35
% Persistence	31%	15%	8%	5%	21%	27%	30%	31%

Note. Effect sizes (ES) are in standard deviation units. A positive effect size indicates that children in the treatment group displayed a more desirable outcome than children in the control group. Each column presents the estimated follow-up effects and absolute proportional persistence for different posttest impacts, and by skill type. The estimated follow-up effect sizes were derived using the regression-based estimates indicated by skill type and follow-up wave. (See Table 4, Column 3). f = follow-up assessment wave; s = study; i = aligned group; p = posttest assessment wave.

^aIndicates that the posttest effect matches the meta-analytic average for the respective skill type (0.14 *SD* for social-emotional skills, 0.40 *SD* for cognitive skills).

varies for cluster randomized designs and that these differences are overrepresented in our models.

Thus, we tested three additional model specifications. First, we tested a model that controlled for whether cluster randomization was employed. Primary model estimates did not change with the inclusion of this covariate (Supplemental Table S5, Column 3). Second, we tested a model that did not use the standard error clustering adjustment; the results were consistent with our primary estimates (Supplemental Table S5, Column 4). Third, we also tested a model that used unweighted average treatment impacts, which essentially allowed each impact estimate to contribute equally to the model (Supplemental Table S5, Column 5; see Supplemental Table S6 for meta-analytic averages of unweighted effects at each time point). When the weights were dropped, the persistence rate for social-emotional skills was larger but not statistically significant due to the decrease in precision. This pattern was reversed at the 1- to 2-year follow-up (Supplemental Table S6).

Next, we set out to examine the validity of the observed negative posttest impacts. First, we returned to the original reports to check that all negative and statistically significant posttest impacts were correctly coded (see Supplemental Material for more discussion). Second, given that theory concerning fadeout generally operates from the assumption that interventions have *positive* posttest impacts, we ran models in which we dropped all outcomes that had negative posttest treatment impacts. (Approximately 20% of follow-up outcomes were dropped at both follow-up waves.) We again observed results that aligned with our primary estimates (Supplemental Table S5, Column 6).

We then explored the possibility that effect sizes that required extensive calculation by our team may have affected our results (i.e., effect sizes that had to be determined using calculation assumptions described in the Supplemental Materials). To address this concern, we fit a model in which we either dropped effect sizes that were calculated for dichotomous outcomes, estimated based on imprecise *p* values, or calculated using estimated standard deviations. These models produced estimates largely in line with our primary estimates (Supplemental Table S5, Column 7).

Finally, many have expressed concern that fadeout may be the result of changes in variance across assessment waves (see Cascio & Staiger, 2012; Wan et al., 2021, for discussion). Indeed, one could imagine cases in which raw differences in treatment and control group performance persist, but impact estimates shrink in magnitude because the variance increases as participants get older. Although it is debatable whether one should discount these real changes in variation over time when estimating follow-up effects, it is still valuable to identify the extent to which this scaling concern shapes fadeout trajectories in our sample. To probe this concern, we rescaled follow-up effect sizes using posttest control group standard deviations, as opposed to concurrent control group standard deviations (which are used in our primary models). In cases when this information was not reported and, therefore, not coded, we scaled follow-up effect sizes by the meta-analytic average of the ratio of follow-up standard deviations to posttest standard deviations, respectively, for cognitive and social-emotional skills and by follow-up assessment wave. (The ratio of follow-up standard deviation to posttest standard deviation for cognitive skills was: 6- to 12-month follow-up: 1.17, 1- to 2-year follow-up: .97; for social-emotional skills: 6- to 12-month follow-up: 1.05, 1- to 2-year follow-up: 1.00.) Then, we tested our primary model using these rescaled follow-up effects. As one would expect based on the meta-analytic ratios, we observed stronger conditional persistence for cognitive skills than social-emotional skills at the 6- to 12-month follow-up, though this difference was not statistically significant. At 1- to 2-year follow-up, the adjustment produced estimates that were aligned with the primary model (see Supplemental Table S5, Column 8).

Summary. Across these robustness checks, our primary findings were largely corroborated. For the 6- to 12-month follow-up, the sensitivity checks provided estimates that were consistent with estimates from our preferred model. For the 1- to 2-year follow-up, the robustness check estimates were generally aligned with those from the preferred model, though the magnitude and statistical significance of the difference between cognitive and social-emotional skill effects varied from model to model. Although the effect for cognitive skills remained relatively similar to the preferred model, the magnitude of

Table 6
Longitudinal Persistence by Outcome and Intervention Types

Parameter label	Cognitive outcome β (SE) (1)	Social-emotional outcome β (SE) (2)
Panel A: 6- to 12-month follow-up		
Posttest effect size	0.23 (0.13)	0.25 (0.14)
Cog intervention	0.09 (0.05)	
Soc intervention	0.01 (0.06)	0.02 (0.06)
Posttest $\times$ Cog Intervention	0.27 (0.14)	
Posttest $\times$ Soc Intervention	-0.35 (0.13)	0.17 (0.17)
Constant	-0.01 (0.03)	0.04 (0.05)
Sample size (study/intervention/outcomes)	35/48/223	30/37/197
$\tau_{\text{intercept}}$ (null = .17, .27)	.10	.10
τ_{slope}	.16	.31
I^2 (null = 53.60%, 66.24%)	0.00%	48.34%
Panel B: >1 year, up to 2 years follow-up		
Posttest effect size	0.20 (0.13)	0.12 (0.17)
Cog intervention	-0.14 [†] (0.04)	
Soc intervention		0.01 (0.03)
Posttest $\times$ Cog Intervention	0.37 (0.19)	
Posttest $\times$ Soc Intervention		-0.18 (0.28)
Constant	0.05 (0.04)	0.05 (0.01)
Sample size (study/intervention/outcome)	11/11/28	13/14/59
$\tau_{\text{intercept}}$ (null = .16, .00)	.06	.04
τ_{slope}	.15	.40
I^2 (null = 50.92%, 0.00%)	31.44%	0.00%

Note. “Cog Intervention” and “Soc Intervention” are dummy variables for intervention type (reference group = “broad” interventions with both cognitive and social-emotional components). The unit of analysis is “aligned groupings” of posttest and follow-up impacts collected for the same construct using the same measure, subscales, and reporter at posttest and at least one follow-up assessment within an intervention. As such, each regression tests the association between posttest and follow-up effect sizes for the same measure. Panel A presents associations between posttest and follow-up effect sizes collected 6 to 12 months after posttest. Panel B presents associations between posttest and follow-up effect sizes collected greater than 1 year and up to 2 years after posttest. Model parameters were estimated using a random-effects meta-analytic model that included a random effect for study and weights and cluster-robust standard errors (with clustering at the study level). Sample sizes are reported for studies, interventions, and outcomes. Heterogeneity statistics are presented for two null models estimating follow-up effect size magnitude using random-effects metaregression. First, they are presented for the model including only cognitive outcomes and second for the model including only social-emotional outcomes. Negative I^2 values were rounded to zero. *SE* = standard error; Soc = social-emotional; Cog = cognitive.

[†] $p < .10$.

the social-emotional effect varied ($\beta_1 = -0.76$ to -0.06), indicating that this estimate was particularly sensitive to model specifications.

Publication Bias

To probe concerns related to selective publishing of larger treatment impacts at posttest and follow-ups, we ran a series of publication-bias-related tests. On average, approximately 32% of posttest effects were statistically significant at $p < .05$. [Supplemental Figure S4](#) displays funnel plots for the posttest and follow-up effects included in the analyses, averaged by intervention. To statistically test for evidence of publication bias, we conducted a Precision Effect Estimate with Standard Error test. First, we ran our primary null model with the inclusion of standard errors as a predictor.⁷ Consistent with the possibility of publication bias, larger standard errors were predictive of larger effect size magnitudes, particularly at posttest ($\beta_1 = 4.80$; $p < .0001$), but also at 6- to 12-month follow-up ($\beta_1 = 2.03$; $p < .0001$), 1- to 2-year follow-up ($\beta_1 = 1.03$; $p < .005$), and greater than 2 years follow-up ($\beta_1 = 1.40$; $p < .0001$).

However, there are other plausible explanations for an effect size-precision association. For example, interventions from smaller samples may have been more intensive and/or targeted and thus produced larger impacts. We then tested whether the inclusion of the standard errors as a covariate in our primary models changed estimated patterns of persistence. Our results were substantively aligned with the estimates from our primary model ([Supplemental Table S7](#)).

To further probe the possibility of publication bias, we examined the distribution of p values ([Simonsohn et al., 2014](#)). [Supplemental Figure S5](#) presents the relative frequency of p values statistically significant at $p < .05$ for each follow-up assessment wave. The distribution of p values for effect sizes reported at posttest and 6- to 12-month follow-up provided little evidence of p hacking. However, at follow-ups greater than 1 year after posttest, there appeared to be

⁷ In other words, we tested a random-effects metaregression model predicting effect size at posttest, 6- to 12-month follow-up, 1- to 2-year follow-up, and greater than 2 years follow-up, with no independent variables or moderators other than a control for standard errors.

an uptick in p values close to .05, suggesting a greater proportion of estimated impacts may be inflated relative to population values.

We then used an additional model proposed by [Vevea and Woods \(2005\)](#) to test for the consistency of the likelihood of publication bias across assessment periods. [Supplemental Table S8](#) displays the unadjusted and adjusted effect sizes produced by this model, in which effects are weighted according to their likelihood of being published considering their statistical significance. Weights were set to reflect plausible patterns of selective reporting if p values did, indeed, dictate publishing: Effects associated with a $p < .01$ were set to have a weight of 1 (an assumption that 100% of effects of this statistical significance were published, if selective bias was at play), $p < .05$ was set to .90, $p < .50$ was set to .70, and $p < 1$ was set to .50. The results from this analysis suggest that had these biases existed in our data, our meta-analysis would likely have produced overestimated posttest and follow-up effects. Bias-adjusted effects were estimated to be between 60% and 68% the magnitude of unadjusted effects. Although this suggests the likelihood of publication bias, it also suggests that such bias may be relatively similar across follow-up waves. As such, it seems unlikely that selective publication would bias our estimates of conditional persistence rates in one direction or the other.

To avoid estimate inflation due to publication- and reporting-related biases, our primary estimates were generated using models that required the same construct to be measured at posttest and at least one follow-up using the same measure at each assessment wave. Indeed, this approach is limited in that some interventions may have legitimate issues, such as changes in participant age, that require the use of different measures across follow-up assessments. Thus, we ran an alternate model in which we allowed the same construct to be assessed using different measures at posttest and follow-ups.⁸ The conditional persistence rates were relatively similar to those from the preferred aggregation approach ([Supplemental Table S9](#), Column 1).

To check our assumption that using a broad aggregation approach, such as that employed in other studies, would inflate our estimates, we ran a model in which we averaged all cognitive or social-emotional outcomes at posttest and each follow-up assessment, respectively. When the data were aggregated without consideration for measure type or construct alignment ([Supplemental Table S9](#), Column 2), the conditional persistence rates were inflated, as expected, demonstrating greater social-emotional persistence at the 6- to 12-month follow-up and the opposite pattern at the 1- to 2-year follow-up.

Selection Into Follow-Up

One concern in interpreting our findings is that follow-up assessments may be disproportionately collected for outcomes that showed promising treatment impacts at posttest. To evaluate the likelihood of this issue, we first compared the average posttest effect size for all outcomes in our sample to only those that had aligned follow-up. With this, we assessed the extent to which posttests linked with follow-up were larger than posttest impacts without follow-up. As compared to posttests with aligned follow-up ($M = 0.29$ SD), the impact for posttests without aligned follow-up was slightly smaller ($M = 0.20$ SD), suggesting some selection.

Next, we checked the posttest effect sizes associated with the sample of follow-up effects reported at each assessment wave

(see [Supplemental Table S10](#), Columns 2, 4, and 6). We found little evidence for selection into follow-up up to 2 years after posttest, with similar posttest effect sizes for all outcomes with aligned follow-up and outcomes with follow-up up to 2 years after posttest. However, posttest effect sizes at later follow-up waves, particularly for cognitive skills, were generally larger than initial posttest effect sizes. This pattern was generally observed for unweighted posttest estimates and less so for weighted posttest estimates. Given the down-weighting of small-sample interventions in the weighted estimates, this pattern suggests that selection into follow-up may be a more significant issue for smaller sample interventions, which were more common among cognitive outcomes. Accordingly, our estimates of average follow-up impacts (see [Supplemental Table S10](#)) may be inflated, especially for cognitive skills reported more than 2 years after posttest.

Discussion

The present study investigated the extent to which fadeout occurred across a broad set of educational interventions targeting a wide array of child skills, derived from eight preexisting meta-analyses. Specifically, we tested the theory that the impacts of educational intervention on social-emotional skills persist more than impacts on cognitive skills. We used a meta-analytic approach to examine data from educational RCTs that reported posttest and follow-up intervention impacts on the same cognitive and social-emotional outcomes using the same measures over time. We analyzed the follow-up impact data using a metaregression approach, in which we tested the extent to which posttest impacts directly predicted follow-up impacts. We argue that the resulting slope term provides an estimate of the “conditional persistence rate,” as it evaluates the extent to which intervention impacts are stable at posttest and follow-up. This approach also allowed us to observe whether “unmeasured mediators” (i.e., factors affected by the intervention but unrelated to posttest impacts) might also affect follow-up impacts.

Overall, we observed substantial fadeout of posttest impacts in the months and years following the end of the intervention ([Table 4](#), [Figures 5](#) and [6](#)). Across all outcomes tested, we observed a robust conditional persistence rate of approximately 45% between posttest and 6- to 12-month follow-up. This rate further decreased to 18% at the 1- to 2-year follow-up, though this estimate was less precise. Our analyses demonstrated consistent evidence for similar conditional persistence rates for cognitive and social-emotional skills at 6- to 12-month follow-up. We also found suggestive evidence for greater conditional persistence rates for cognitive skills than social-emotional skills at 1- to 2-year follow-up, though this difference was not statistically significant. In addition to the ~45% conditional persistence rate, we observed small unmeasured mediator effects for both outcomes, with slightly larger unmeasured mediator effects for social-emotional skill impacts.

Our metaregression models also allowed us to observe the extent to which follow-up impacts varied across studies. Indeed, we

⁸ Of note, although this approach included more effects from nonaligned measures, this approach did not result in a higher number of aligned groups because for any single construct measured within an intervention, all effects for this construct, regardless of measure, reporter, and subscale were averaged to create a construct-level effect at each follow-up wave. In contrast, in the primary model, each unique construct-measure-subscale-reporter combination was treated as a unique unit of analysis.

observed substantial heterogeneity in follow-up effects, even after accounting for posttest impacts. Across studies, variability in the conditional persistence rate was also large, but this variability was unexplained by outcome type (i.e., cognitive vs. social-emotional skill). These results demonstrate that fadeout is a pervasive phenomenon across a wide range of educational interventions, regardless of skill type. The findings also suggest that theories purporting that intervention impacts on social-emotional skills will persist more in the months and years following intervention end require revision.

Interestingly, our results suggest that treatment impact persistence may vary by posttest effect magnitude. For posttest impacts of the *same magnitude*, conditional persistence rates were similar for cognitive and social-emotional skills at 6- to 12-month follow-up and more favorable for cognitive skills at 1- to 2-year follow-up. Given the presence of small intercept effects for both skill types, when posttest impacts are small, or even zero, our results suggest that there is proportionally *stronger* persistence than when impacts are larger. Although counterintuitive, this pattern may arise because interventions produce positive effects on unmeasured skills that transfer to later measured skills. This could be why average long-term impacts for both skill types appear to asymptote above zero (though such effects are often not statistically distinguishable from zero).

The pattern of smaller initial effects persisting at relatively higher levels weighs on comparisons drawn between cognitive and social-emotional skill impacts. Compared with cognitive skills, social-emotional skills produced much smaller posttest impacts on average (i.e., .14 for social-emotional skills vs. .40 for cognitive skills; Table 3). We also observed some evidence for slightly larger intercept effects (i.e., unmeasured mediators) for social-emotional skills (Table 4). Taken together, this means that the average intervention impact on social-emotional skills would appear to persist *more* than the average intervention impact on cognitive skills (Table 5). However, the strikingly similar conditional persistence rate for the two skill types reveals that this apparent advantage for social-emotional skills would lead to incorrect predictions when forecasting intervention effects in the long term. Upon observing a small and relatively persistent impact on a given social-emotional skill, one might be tempted to argue that intervention developers should try to create programs that produce larger effects on social-emotional skills in the pursuit of meaningful long-term changes to developmental trajectories. Our results suggest that such efforts would likely lead to more fadeout in the long term. As the posttest impact magnitude grows, the relative importance of the small intercept term shrinks, and the conditional persistence rate (~45% at 6 to 12 month) begins to account for most of the follow-up effect.

Thus, the present study did not find evidence to suggest that an intervention would have more persistent effects on targeted skills if it produced, for example, a 1 *SD* gain on social-emotional skills compared to an intervention that produced a 1 *SD* gain on cognitive skills. The similar conditional persistence rate for both skill types suggests that social-emotional skills may be susceptible to similar mechanisms that drive cognitive skill fadeout. For example, intervention-targeted social-emotional skills may be likely to naturally develop in subsequent contexts, facilitating control group catch-up. Educational interventions may also show diminishing impacts on social-emotional skills because they do not overcome the influence of

many individual, contextual, and societal factors that continue to shape skill development when interventions end (see Watts et al., 2017). Given that we observed substantial heterogeneity in follow-up impacts that was unexplained by posttest impacts, it may be the case that traditional skill-building models do not reflect the complexities of skill development.

Interventionists might view this news negatively or positively. On the one hand, it is disappointing that we have yet to identify a large class of skills for which end-of-treatment impacts persist at the same magnitude indefinitely. On the other hand, given many previous findings of positive long-term impacts of educational interventions on important adult outcomes, mediating processes must exist and plausibly include both cognitive and social-emotional skills, despite diminishing impacts after treatment. Although impacts on a focal skill may diminish, long-run impacts may develop through small impacts on a network of unmeasured complementary skills, contexts, and opportunities, with impacts stabilizing at some nonzero (but statistically undetectable) level that is lower than initial impacts. In light of this, predicting *a priori* which skill impacts will show the most persistence may be difficult if persistence is contingent on complex interactions between children and their environments.

Limitations

Several limitations are worth noting. Importantly, despite including a relatively large collection of RCT interventions in the meta-analytic sample, our data contained limited statistical power to predict follow-up impact persistence greater than 2 years after posttest. This imprecision precludes concrete conclusions about the relation between posttest and longer run follow-up impacts. Our reliance on eight published meta-analyses in sourcing studies for our sample could be part of the issue; there may be new RCTs that were published since the publication of these meta-analyses and/or that were not referenced in these meta-analyses. However, this issue is also a symptom of a larger problem of limited grant funding for collecting long-term follow-up data for educational RCTs, and that funding is often allocated to RCTs that demonstrate large posttest impacts (Watts et al., 2019).

This lack of long-run follow-up data constricted our ability to model the functional form of persistence across all follow-up effects. Our descriptive data (see Figures 3 and 4) certainly suggested that this form is nonlinear, a main reason why we fit our regression models separately for the 6- to 12-month and 1- to 2-year periods. Future work should continue to explore the best functional representation of fadeout trends. Beyond nonlinearity in effects across follow-up assessments, it could also be the case that there is nonlinearity in conditional persistence rates by posttest impact magnitude that our models do not fully capture. This could be the case if conditional persistence rates are moderated by posttest impact magnitude. For example, posttest impact magnitude could be indicative of factors relating to the counterfactual condition that affect persistence when interventions end, such as the likelihood of control group catch-up. It is also worth acknowledging that our analytic strategy does not account for measurement error. Correlated measurement error (e.g., stable, erroneously estimated differences between the treatment and control groups) could bias persistence estimates upward. Conversely, random measurement error could bias persistence estimates downward but intercept estimates upward.

It is important to discuss the external validity of our findings. As with any study, issues associated with external validity beyond the

present sample must be considered with caution. Indeed, we did not attempt to search the literature for every educational intervention that has ever reported impacts on cognitive and social-emotional skills. Rather, we sampled RCTs from eight diverse meta-analyses of educational interventions. Although it is likely the case that our approach excludes some interventions that may be of interest to educational researchers (e.g., executive function-focused curricula), we do not have reason to believe that there is some contingency of interventions missing from our sample that would produce categorically different dynamics of fadeout. Had our approach depended on simply presenting average impacts at posttest and various follow-up waves, then the external validity of our results could be limited if other sets of nonobserved interventions produce much larger or smaller impacts. However, the validity of our results rests on the assumption that intervention features affect later skill development through their effects on short-term skill growth. Thus, we assume that differences in intervention features are represented by varying magnitudes of posttest impacts, and it is through these posttest effects that impacts on later skill levels are reached.

As with any study, we cannot be completely confident that our results would translate to all treatments, contexts, or outcomes of interest. Indeed, we did observe significant heterogeneity in conditional persistence rates across studies, suggesting that some studies do produce stronger rates, and others produce weaker rates. We expect that some portion of this heterogeneity is real and worth exploring in the future. However, given that we observed a wide range of intervention features (Table 1), posttest impacts, and follow-up impacts (Figures 3 and 4), we find it unlikely that someone could confidently predict a priori which type of intervention would cause a categorically different relation between posttest and follow-up impacts on a given educational skill of interest. Moreover, of central concern to our research question, our key results consistently suggest that skill type (i.e., cognitive vs. social-emotional) does not meaningfully explain this heterogeneity.

It is also important to note that 80% of cognitive outcomes in our models capture language or literacy outcomes. The social-emotional constructs were more diverse, although these were largely survey-based, which reflect the state of measurement in the field. Indeed, it is possible that measures of social-emotional development are not adept at capturing the sorts of hard-to-measure social-emotional gains that many interventions hope to promote. The slightly larger intercept effects for social-emotional skill impacts may reflect this measurement limitation. Additionally, the majority of studies represented in our meta-analysis occurred in the United States. Future work should examine persistence patterns in low- and middle-income countries where there may be very different counterfactual conditions and opportunities for long-term skill development.

Conclusions

Our findings suggest that fadeout occurs across skill types for a diverse collection of educational interventions. Social-emotional skills may not be the single class of missing mediators of emergent long-run impacts on adult outcomes. Emergence could be consistent with other explanations, such as persistent but hard to measure social-emotional and cognitive impacts (Reynolds & Ou, 2011) and/or impacts on institutional gateways that can generate longer term benefits even in the absence of medium-run skill impacts (see Bailey, Duncan, et al., 2020; Pages et al., 2022). In light of

considerable heterogeneity in follow-up effects and conditional persistence rates that were unexplained by posttest impacts, future work should investigate theoretically motivated intervention- and participant-level characteristics as moderators of persistence and explore the extent to which these moderators and posttest impacts forecast the emergence of long-run outcomes. Deepening our understanding of these fadeout dynamics will further support the development of interventions that have durable long-term effects.

References

References marked with an asterisk indicates studies included in the full meta-analytic sample.

- Abenavoli, R. M. (2019). The mechanisms and moderators of “fade-out”: Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin*, 145(12), 1103–1127. <https://doi.org/10.1037/bul0000212>
- *Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481–1495. <https://doi.org/10.1198/016214508000000841>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, 21(2), 55–97. <https://doi.org/10.1177/1529100620915848>
- Bailey, D. H., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, 10(1), 7–39. <https://doi.org/10.1080/19345747.2016.1232459>
- Bailey, D. H., Jenkins, J. M., & Alvarez-Vargas, D. (2020). Complementarities between early educational intervention and later educational quality? A systematic review of the sustaining environments hypothesis. *Developmental Review*, 56, Article 100910. <https://doi.org/10.1016/j.dr.2020.100910>
- Bailey, D. H., & Weiss, M. J. (2022, April). *Do meta-analyses oversell the longer-term effects of programs? (Part 1)*. MDRC. <https://www.mdrc.org/publication/do-meta-analyses-oversell-longer-term-effects-programs-part-1#:~:text=The%20most%20obvious%20implication%20of,intervention%20on%20longer%20term%20outcomes>
- *Barrett, P. M., Farrell, L. J., Ollendick, T. H., & Dadds, M. (2006). Long-term outcomes of an Australian universal prevention trial of anxiety and depression symptoms in children and youth: An evaluation of the friends program. *Journal of Clinical Child and Adolescent Psychology*, 35(3), 403–411. https://doi.org/10.1207/s15374424jccp3503_5
- *Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., Blair, C., Nelson, K. E., & Gill, S. (2008). Promoting academic and social-emotional school readiness: The head start REDI program. *Child Development*, 79(6), 1802–1817. <https://doi.org/10.1111/j.1467-8624.2008.01227.x>
- *Bierman, K. L., Heinrichs, B. S., Welsh, J. A., & Nix, R. L. (2021). Reducing adolescent psychopathology in socioeconomically disadvantaged children with a preschool intervention: A randomized controlled trial. *The American Journal of Psychiatry*, 178(4), 305–312. <https://doi.org/10.1176/appi.ajp.2020.20030343>
- *Bierman, K. L., Nix, R. L., Heinrichs, B. S., Domitrovich, C. E., Gest, S. D., Welsh, J. A., & Gill, S. (2014). Effects of Head Start REDI on children’s

- outcomes 1 year later in different kindergarten contexts. *Child Development*, 85(1), 140–159. <https://doi.org/10.1111/cdev.12117>
- *Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology*, 96(3), 444–461. <https://doi.org/10.1037/0022-0663.96.3.444>
- *Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, 106(1), 46–57. <https://doi.org/10.1037/a0033663>
- Blanden, J. (2019). *Intergenerational income persistence*. IZA World of Labor. <https://doi.org/10.15185/izawol.176.v2>
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley. <https://doi.org/10.1002/9780470743386>
- *Botvin, G. J., Baker, E., Dusenbury, L., Botvin, E. M., & Diaz, T. (1995). Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. *JAMA*, 273(14), 1106–1112. <https://doi.org/10.1001/jama.1995.03520380042033>
- *Botvin, G. J., Baker, E., Dusenbury, L., Tortu, S., & Botvin, E. M. (1990). Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: Results of a 3-year study. *Journal of Consulting and Clinical Psychology*, 58(4), 437–446. <https://doi.org/10.1037/0022-006X.58.4.437>
- *Botvin, G. J., Baker, E., Filazzola, A. D., & Botvin, E. M. (1990). A cognitive-behavioral approach to substance abuse prevention: One-year follow-up. *Addictive Behaviors*, 15(1), 47–63. [https://doi.org/10.1016/0306-4603\(90\)90006-J](https://doi.org/10.1016/0306-4603(90)90006-J)
- *Botvin, G. J., Baker, E., Renick, N. L., Filazzola, A. D., & Botvin, E. M. (1984). A cognitive-behavioral approach to substance abuse prevention. *Addictive Behaviors*, 9(2), 137–147. [https://doi.org/10.1016/0306-4603\(84\)90051-0](https://doi.org/10.1016/0306-4603(84)90051-0)
- Botvin, G. J., & Griffin, K. W. (2004). Life skills training: Empirical findings and future directions. *The Journal of Primary Prevention*, 25, 211–232. <https://doi.org/10.1023/B:JOPP.0000042391.58573.5b>
- *Botvin, G. J., Griffin, K. W., Diaz, T., & Ifill-Williams, M. (2001). Preventing binge drinking during early adolescence: One- and two-year follow-up of a school-based preventive intervention. *Psychology of Addictive Behaviors*, 15(4), 360–365. <https://doi.org/10.1037/0893-164X.15.4.360>
- *Botvin, G. J., Griffin, K. W., Diaz, T., Scheier, L. M., Williams, C., & Epstein, J. A. (2000). Preventing illicit drug use in adolescents: Long-term follow-up data from a randomized control trial of a school population. *Addictive Behaviors*, 25(5), 769–774. [https://doi.org/10.1016/S0306-4603\(99\)00050-7](https://doi.org/10.1016/S0306-4603(99)00050-7)
- *Botvin, G. J., Schinke, S. P., Epstein, J. A., & Diaz, T. (1994). Effectiveness of culturally focused and generic skills training approaches to alcohol and drug abuse prevention among minority youths. *Psychology of Addictive Behaviors*, 8(2), 116–127. <https://doi.org/10.1037/0893-164X.8.2.116>
- *Botvin, G. J., Schinke, S. P., Epstein, J. A., Diaz, T., & Botvin, E. M. (1995). Effectiveness of culturally focused and generic skills training approaches to alcohol and drug abuse prevention among minority adolescents: Two-year follow-up results. *Psychology of Addictive Behaviors*, 9(3), 183–194. <https://doi.org/10.1037/0893-164X.9.3.183>
- *Bradshaw, C. P., Zmuda, J. H., Kellam, S. G., & Ialongo, N. S. (2009). Longitudinal impact of two universal preventive interventions in first grade on educational outcomes in high school. *Journal of Educational Psychology*, 101(4), 926–937. <https://doi.org/10.1037/a0016586>
- Briley, D. A., & Tucker-Drob, E. M. (2017). Comparing the developmental genetics of cognition and personality over the life span. *Journal of Personality*, 85(1), 51–64. <https://doi.org/10.1111/jopy.12186>
- Burns, M. K., Petersen-Brown, S., Haegele, K., Rodriguez, M., Schmitt, B., Cooper, M., Clayton, K., Hutcheson, S., Conner, C., Hosp, J., & VanDerHeyden, A. M. (2016). Meta-analysis of academic interventions derived from neuropsychological data. *School Psychology Quarterly*, 31(1), 28–42. <https://doi.org/10.1037/spq0000117>
- *Cabell, S. Q., Justice, L. M., Piasta, S. B., Curenton, S. M., Wiggins, A., Turnbull, K. P., & Petscher, Y. (2011). The impact of teacher responsivity education on preschoolers' language and literacy skills. *American Journal of Speech-Language Pathology*, 20(4), 315–330. [https://doi.org/10.1044/1058-0360\(2011/10-0104](https://doi.org/10.1044/1058-0360(2011/10-0104)
- *Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., Barbarin, O. A., Sparling, J. J., & Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology*, 48(4), 1033–1043. <https://doi.org/10.1037/a0026644>
- *Campbell, F. A., & Ramey, C. T. (1990). The relationship between Piagetian cognitive development, mental test performance, and academic achievement in high-risk students with and without early educational experience. *Intelligence*, 14(3), 293–308. [https://doi.org/10.1016/0160-2896\(90\)90020-T](https://doi.org/10.1016/0160-2896(90)90020-T)
- *Campbell, F. A., & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development*, 65(2), 684–698. <https://doi.org/10.2307/1131410>
- *Campbell, F. A., & Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal*, 32(4), 743–772. <https://doi.org/10.3102/00028312032004743>
- *Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education/Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, 6(1), 42–57. https://doi.org/10.1207/S1532480XADS0601_05
- *Campbell, F. A., Wasik, B. H., Pungello, E., Burchinal, M., Barbarin, O., Kainz, K., Sparling, J. J., & Ramey, C. T. (2008). Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly*, 23(4), 452–466. <https://doi.org/10.1016/j.ecresq.2008.03.003>
- *Cardemil, E. V., Reivich, K. J., & Seligman, M. E. P. (2002). The prevention of depressive symptoms in low-income minority middle-school students. *Prevention & Treatment*, 5(7), Article 8. <https://doi.org/10.1037/1522-3736.5.1.58a>
- Cascio, E. U., & Staiger, D. O. (2012). *Knowledge, tests, and fadeout in educational interventions* (No. w18038). National Bureau of Economic Research.
- *Cecchini, T. B. (1997). *An interpersonal and cognitive-behavioral approach to childhood depression: A school-based primary prevention study* (Order No. 9820698, 304420573) [Doctoral dissertation]. ProQuest Dissertations and Theses Global. <https://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/dissertations-theses/interpersonal-cognitive-behavioral-approach/docview/304420573/se-2>
- *Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of reading recovery. *Reading Research Quarterly*, 30(2), 240–263. <https://doi.org/10.2307/748034>
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4), 1593–1660. <https://doi.org/10.1093/qje/qjr041>
- *Cirino, P. T., Vaughn, S., Linan-Thompson, S., Cardenas-Hagan, E., Fletcher, J. M., & Francis, D. J. (2009). One-year follow-up outcomes of Spanish and English interventions for English language learners at risk for reading problems. *American Educational Research Journal*, 46(3), 744–781. <https://doi.org/10.3102/0002831208330214>
- *Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, 21(8), 1106–1116. <https://doi.org/10.1177/0956797610375449>

- *Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal*, 50(4), 812–850. <https://doi.org/10.3102/0002831212469270>
- *Collins, S., Woolfson, M. L., & Durkin, K. (2014). Effects on coping skills and anxiety of a universal school-based mental health intervention delivered in Scottish primary schools. *School Psychology International*, 35(1), 85–100. <https://doi.org/10.1177/0143034312469157>
- Cunha, F., & Heckman, J. (2007). The technology of skill formation. *The American Economic Review*, 97(2), 31–47. <https://doi.org/10.1257/aer.97.2.31>
- *Daunic, A. P., Smith, S. W., Garvan, C. W., Barber, B. R., Becker, M. K., Peters, C. D., Taylor, G. G., Van Loan, C. L., Li, W., & Naranjo, A. H. (2012). Reducing developmental risk for emotional/behavioral problems: A randomized controlled trial examining the Tools for Getting Along curriculum. *Journal of School Psychology*, 50(2), 149–166. <https://doi.org/10.1016/j.jsp.2011.09.003>
- Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3), 111–134. <https://doi.org/10.1257/app.1.3.111>
- Dodge, K. A., Pettit, G. S., McClaskey, C. L., Brown, M. M., & Gottman, J. M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development*, 51(2), 1–85. <https://doi.org/10.2307/1165906>
- Duckworth, A. L., Milkman, K. L., & Laibson, D. (2018). Beyond willpower: Strategies for reducing failures of self-control. *Psychological Science in the Public Interest*, 19(3), 102–129. <https://doi.org/10.1177/1529100618821893>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- *Dunham, R., Kidwell, J., & Portes, P. (1988). Effects of parent-adolescent interaction on the continuity of cognitive development from early childhood to early adolescence. *The Journal of Early Adolescence*, 8(3), 297–310. <https://doi.org/10.1177/0272431688083006>
- *DuPaul, G. J., Jitendra, A. K., Volpe, R. J., Tresco, K. E., Lutz, J. G., Vile, Junod, R. E., Cleary, K. S., Flammer, L. M., & Mannella, M. C. (2006). Consultation-based academic interventions for children with ADHD: Effects on reading and mathematics achievement. *Journal of Abnormal Child Psychology*, 34(5), 635–648. <https://doi.org/10.1007/s10802-006-9046-7>
- *Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*, 58(3), 470–484. <https://doi.org/10.1037/dev0001301>
- *Eccles, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low progress readers. *Computers & Education*, 52(3), 554–561. <https://doi.org/10.1016/j.compedu.2008.10.010>
- *Eckenrode, J., Campa, M., Luckey, D. W., Henderson, C. R., Jr., Cole, R., Kitzman, H., Anson, E., Sidora-Arcoleo, K., Powers, J., & Olds, D. (2010). Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up of a randomized trial. *Archives of Pediatrics & Adolescent Medicine*, 164(1), 9–15. <https://doi.org/10.1001/archpediatrics.2009.240>
- *Eddy, J. M., Reid, J. B., Stoolmiller, M., & Fetrow, R. (2003). Outcomes during middle school for an elementary school-based preventive intervention for conduct problems: Follow-up results from a randomized trial. *Behavior Therapy*, 34(4), 535–552. [https://doi.org/10.1016/S0005-7894\(03\)80034-5](https://doi.org/10.1016/S0005-7894(03)80034-5)
- *Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2016). Early childhood education. In R. A. Moffitt (Ed.), *Economics of means-tested transfer programs in the United States* (Vol. 2, pp. 235–297). University of Chicago Press.
- *Elbro, C., & Peterson, D. K. (2004). Long-term effects of phoneme awareness and letter sound training: An intervention study with children at risk for dyslexia. *Journal of Educational Psychology*, 96(4), 660–670. <https://doi.org/10.1037/0022-0663.96.4.660>
- *Essau, C. A., Conradt, J., Sasagawa, S., & Ollendick, T. H. (2012). Prevention of anxiety symptoms in children: Results from a universal school-based trial. *Behavior Therapy*, 43(2), 450–464. <https://doi.org/10.1016/j.beth.2011.08.003>
- *Faggiano, F., Galanti, M. R., Bohrn, K., Burkhart, G., Vigna-Taglianti, F., Cuomo, L., Fabiani, L., Panella, M., Perez, T., Siliquini, R., van der Kreeft, P., Vassara, M., Wiborg, G., & the European Drug Abuse Prevention trial. (2008). The effectiveness of a school-based substance abuse prevention program: European Drug Abuse Prevention trial cluster randomised controlled trial. *Preventive Medicine*, 47(5), 537–543. <https://doi.org/10.1016/j.ypmed.2008.06.018>
- *Faggiano, F., Richardson, C., Bohrn, K., Galanti, M. R., & the European Drug Abuse Prevention trial Study Group. (2007). A cluster randomized controlled trial of school-based prevention of tobacco, alcohol and drug use: The European Drug Abuse Prevention trial design and study population. *Preventive Medicine*, 44(2), 170–173. <https://doi.org/10.1016/j.ypmed.2006.09.010>
- *Faggiano, F., Vigna-Taglianti, F., Burkhart, G., Bohrn, K., Cuomo, L., Gregori, D., Panella, M., Scatigna, M., Siliquini, R., Varona, L., van der Kreeft, P., Vassara, M., Wiborg, G., Galanti, M. R., & the European Drug Abuse Prevention trial Study Group. (2010). The effectiveness of a school-based substance abuse prevention program: 18-month follow-up of the European Drug Abuse Prevention trial cluster randomized controlled trial. *Drug and Alcohol Dependence*, 108(1–2), 56–64. <https://doi.org/10.1016/j.drugalcdep.2009.11.018>
- *Fälth, L., Gustafson, S., Tjus, T., Heimann, M., & Svensson, I. (2013). Computer-assisted interventions targeting reading skills of children with reading disabilities—A longitudinal study. *Dyslexia*, 19(1), 37–53. <https://doi.org/10.1002/dys.1450>
- *Farrell, A., Meyer, A., Sullivan, T., & Kung, E. (2003). Evaluation of the Responding in Peaceful and Positive Ways Seventh Grade (RIPP-7) universal violence prevention program. *Journal of Child and Family Studies*, 12, 101–120. <https://doi.org/10.1023/A:1021314327308>
- *Field, T., Widmayer, S., Greenberg, R., & Stoller, S. (1982). Effects of parent training on teenage mother and their infants. *Pediatrics*, 69(6), 703–707. <https://doi.org/10.1542/peds.69.6.703>
- *Foshee, V. A., Bauman, K. E., Arriaga, X. B., Helms, R. W., Koch, G. G., & Linder, G. F. (1998). An evaluation of Safe Dates, an adolescent dating violence prevention program. *American Journal of Public Health*, 88(1), 45–50. <https://doi.org/10.2105/AJPH.88.1.45>
- *Foshee, V. A., Bauman, K. E., Ennett, S. T., Suchindran, C., Benefield, T., & Linder, G. F. (2005). Assessing the effects of the dating violence prevention program “Safe Dates” using random coefficient regression modeling. *Prevention Science*, 6(3), 245–258. <https://doi.org/10.1007/s11121-005-0007-0>
- *Foshee, V. A., Bauman, K. E., Greene, W. F., Koch, G. G., Linder, G. F., & MacDougall, J. E. (2000). The Safe Dates program: 1-year follow-up results. *American Journal of Public Health*, 90(10), 1619–1622. <https://doi.org/10.2105/AJPH.90.10.1619>
- *Furr-Holden, C. D., Ialongo, N. S., Anthony, J. C., Petras, H., & Kellam, S. G. (2004). Developmentally inspired drug prevention: Middle school outcomes in a school-based randomized prevention trial. *Drug and Alcohol Dependence*, 73(2), 149–158. <https://doi.org/10.1016/j.drugalcdep.2003.10.002>
- *Gillham, J. E., Reivich, K. J., Freres, D. R., Chaplin, T. M., Shatté, A. J., Samuels, B., Elkon, A. G. L., Litzinger, S., Lascher, M., Gallop, R., & Seligman, M. E. P. (2007). School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the

- Penn Resiliency Program. *Journal of Consulting and Clinical Psychology*, 75(1), 9–19. <https://doi.org/10.1037/0022-006X.75.1.9>
- *Gittelman, R., & Feingold, I. (1983). Children with reading disorders—I. Efficacy of reading remediation. *Journal of Child Psychology and Psychiatry*, 24(2), 167–191. <https://doi.org/10.1111/j.1469-7610.1983.tb00568.x>
- *Gray, S. W., & Klaus, R. A. (1970). The early training project: A seventh-year report. *Child Development*, 41(4), 909–924. <https://doi.org/10.2307/1127321>
- Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2022). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, 138(1), 363–411. <https://doi.org/10.1093/qje/qjac036>
- *Griffin, K. W., Botvin, G. J., & Nichols, T. R. (2006). Effects of a school-based drug abuse prevention program for adolescents on HIV risk behavior in young adulthood. *Prevention Science*, 7(1), 103–112. <https://doi.org/10.1007/s1121-006-0025-6>
- *Güneri, O. Y., & Çoban, R. (2004). The effect of conflict resolution training on Turkish elementary school students: A quasi-experimental investigation. *International Journal for the Advancement of Counseling*, 26(2), 109–124. <https://doi.org/10.1023/B:ADCO.0000027425.24225.87>
- *Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *The Journal of Special Education*, 39(2), 66–85. <https://doi.org/10.1177/00224669050390020301>
- *Gunn, B., Smolkowski, K., & Vadasy, P. (2010). Evaluating the effectiveness of Read Well Kindergarten. *Journal of Research on Educational Effectiveness*, 4(1), 53–86. <https://doi.org/10.1080/19345747.2010.488716>
- *Gustafson, S., Fälth, L., Svensson, I., Tjus, T., & Heimann, M. (2011). Effects of three interventions on the reading skills of children with reading disabilities in grade 2. *Journal of Learning Disabilities*, 44(2), 123–135. <https://doi.org/10.1177/0022219410391187>
- *Hansen, J., Hanewinkel, R., Maruska, K., & Isensee, B. (2011). The 'Eigenständig werden' prevention trial: A cluster randomised controlled study on a school-based life skills programme to prevent substance use onset. *BMJ Open*, 1(2), Article e000352. <https://doi.org/10.1136/bmjopen-2011-000352>
- *Harrington, N. G., Giles, S. M., Hoyle, R. H., Feeney, G. J., & Yungbluth, S. C. (2001). Evaluation of the All Stars character education and problem behavior prevention program: Effects on mediator and outcome variables for middle school students. *Health Education & Behavior*, 28(5), 533–546. <https://doi.org/10.1177/109019810102800502>
- Hart, E. R., Watts, T. W., & Bailey, D. H. (2024). *Meta-analysis of educational RCTs with follow-up*. LDBase. <https://doi.org/10.33009/ldbbase.1719529626.152e>
- *Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry*, 45(2), 338–358. <https://doi.org/10.1111/j.1469-7610.2004.00225.x>
- *Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review*, 103(6), 2052–2086. <https://doi.org/10.1257/aer.103.6.2052>
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451–464. <https://doi.org/10.1016/j.labeco.2012.05.014>
- *Hill, J. L., Brooks-Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology*, 39(4), 730–744. <https://doi.org/10.1037/0012-1649.39.4.730>
- *Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders*, 9(3), 146–161. <https://doi.org/10.1177/106342660100900301>
- *Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology*, 27(5), 599–641. <https://doi.org/10.1023/A:1022137920532>
- IntHout, J., Ioannidis, J. P., Robers, M. M., & Goeman, J. J. (2015). Plea for presenting prediction intervals in meta-analysis. *BMJ Open*, 6(7), Article e010247. <https://doi.org/10.1136/bmjopen-2015-010247>
- *Isensee, B., Hansen, J., Maruska, K., & Hanewinkel, R. (2014). Effects of a school-based prevention programme on smoking in early adolescence: A 6-month follow-up of the 'Eigenständig werden' cluster randomised trial. *BMJ Open*, 4(1), Article e004422. <https://doi.org/10.1136/bmjopen-2013-004422>
- *Johanson, M., Justice, L. M., & Logan, J. (2015). Kindergarten impacts of a preschool language-focused intervention. *Applied Developmental Science*, 20(2), 94–107. <https://doi.org/10.1080/10888691.2015.1074050>
- *Johnson, K. W., Shamblen, S. R., Ogilvie, K. A., Collins, D., & Saylor, B. (2009). Preventing youths' use of inhalants and other harmful legal products in frontier Alaskan communities: A randomized trial. *Prevention Science*, 10(4), 298–312. <https://doi.org/10.1007/s1121-009-0132-2>
- *Johnson, N. C. (2000). *A follow-up study of a primary prevention program targeting childhood depression* (Order No. 1402700, 304634088) [Doctoral dissertation]. ProQuest Dissertations and Theses Global. <https://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/dissertations-theses/follow-up-study-primary-prevention-program/docview/304634088/se-2>
- *Karnes, M. B., Hodgins, A. S., Teska, J. A., & Kirk, S. A. (1969). *Investigations of classroom and at-home interventions: Research and development program on preschool disadvantaged children*. U.S. Department of Health, Education, and Welfare Office of Education. <https://files.eric.ed.gov/fulltext/ED036663.pdf>
- *Klaus, R. A., & Gray, S. W. (1968). The early training project for disadvantaged children: A report after five years. *Monographs of the Society for Research in Child Development*, 33(4), 1–66. <https://doi.org/10.2307/1165717>
- *Kozminsky, L., & Kozminsky, E. (1995). The effects of early phonological awareness training on reading success. *Learning and Instruction*, 5(3), 187–201. [https://doi.org/10.1016/0959-4752\(95\)00004-M](https://doi.org/10.1016/0959-4752(95)00004-M)
- *Kraag, G., Van Breukelen, G. J. P., Kok, G., & Hosman, C. (2009). 'Learn Young, Learn Fair', a stress management program for fifth and sixth graders: Longitudinal results from an experimental study. *Journal of Child Psychology and Psychiatry*, 50(9), 1185–1195. <https://doi.org/10.1111/j.1469-7610.2009.02088.x>
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547–588. <https://doi.org/10.3102/0034654318759268>
- Li, W., Duncan, G. J., Magnuson, K., Schindler, H. S., Yoshikawa, H., & Leak, J. (2020). *Timing in early childhood education: How cognitive and achievement program impacts vary by starting age, program duration, and time since the end of the program* (EdWorkingPaper: 20-201). <https://doi.org/10.26300/5tvgn-nt21>
- *Lie, A. (1991). Effects of a training program for stimulating skills in word analysis in first-grade children. *Reading Research Quarterly*, 26(3), 234–250. <https://doi.org/10.2307/747762>
- *Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, 45(4), 155–176. <https://doi.org/10.1016/j.ecresq.2018.03.005>
- *Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade* (Research report). Peabody Research Institute, Vanderbilt University. <https://files.eric.ed.gov/fulltext/ED566664.pdf>
- *Lock, S., & Barrett, P. M. (2003). A longitudinal study of developmental differences in universal preventive intervention for child anxiety. *Behaviour Change*, 20(4), 183–199. <https://doi.org/10.1375/bech.20.4.183.29383>

- *Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Brooks-Gunn, J., Paulsell, D., Boller, K., Constantine, J., Vogel, C., Fuligni, A. S., & Brady-Smith, C. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, Child Outcomes Research and Evaluation & Administration on Children, Youth, and Families, Head Start Bureau. https://www.acf.hhs.gov/sites/default/files/documents/opre/impacts_vol1.pdf
- *Luna-Adame, M., Carrasco-Giménez, T. J., & Rueda-García, M. M. (2013). Evaluation of the effectiveness of a smoking prevention program based on the 'Life Skills Training' approach. *Health Education Research*, 28(4), 673–682. <https://doi.org/10.1093/her/cyt061>
- *Lyster, S. A. H. (2002). The effects of morphological versus phonological awareness training in kindergarten on reading development. *Reading and Writing*, 15(3–4), 261–294. <https://doi.org/10.1023/A:1015272516220>
- *Lyster, S. H., Lervåg, A. O., & Hulme, C. (2016). Preschool morphological training produces long-term improvements in reading comprehension. *Reading and Writing*, 29(6), 1269–1288. <https://doi.org/10.1007/s11145-016-9636-x>
- *Mantzicopoulos, P., Morrison, D., Stone, E., & Setrakian, W. (1992). Use of the SEARCH/TEACH tutoring approach with middle-class students at risk for reading failure. *The Elementary School Journal*, 92(5), 573–586. <https://doi.org/10.1086/461707>
- Marques, C., & Dias, D. (2005). *Using mean reversion as a measure of persistence*. Banco de Portugal, Economics and Research Department.
- *McCarton, C. M., Brooks-Gunn, J., Wallace, I. F., Bauer, C. R., Bennett, F. C., Bernbaum, J. C., Broyles, R. S., Casey, P. H., McCormick, M. C., Scott, D. T., Tyson, J., Tonascia, J., & Meinert, C. L. (1997). Results at age 8 years of early intervention for low-birth-weight premature infants. The Infant Health and Development Program. *JAMA*, 277(2), 126–132. <https://doi.org/10.1001/jama.1997.03540260040033>
- *McCormick, M. C., Brooks-Gunn, J., Buka, S. L., Goldman, J., Yu, J., Salganik, M., Scott, D. T., Bennett, F. C., Kay, L. L., Bernbaum, J. C., Bauer, C. R., Martin, C., Woods, E. R., Martin, A., & Casey, P. H. (2006). Early intervention in low birth weight premature infants: Results at 18 years of age for the Infant Health and Development Program. *Pediatrics*, 117(3), 771–780. <https://doi.org/10.1542/peds.2005-1316>
- *Merry, S., McDowell, H., Wild, C. J., Bir, J., & Cunliffe, R. (2004). A randomized placebo-controlled trial of a school-based depression prevention program. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(5), 538–547. <https://doi.org/10.1097/00004583-20040500-00007>
- *Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). *Impact findings from the Head Start CARES demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence* (Report No. OPRE 2014-44). U.S. Department of Health and Human Services. <https://files.eric.ed.gov/fulltext/ED546649.pdf>
- *Morris, R. D., Lovett, M. W., Wolf, M., Sevcik, R. A., Steinbach, K. A., Frijters, J. C., & Shapiro, M. B. (2012). Multiple-component remediation for developmental reading disabilities: IQ, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities*, 45(2), 99–127. <https://doi.org/10.1177/0022219409355472>
- Nagaoka, J., Farrington, C. A., Ehrlich, S. B., Heath, R. D., Johnson, D. W., Dickson, S., Tuner, A. C., Mayo, A., & Hayes, K. (2015). *Foundations for young adult success: A developmental framework*. University of Chicago Consortium on Chicago School Research.
- *Nix, R. L., Bierman, K. L., Heinrichs, B. S., Gest, S. D., Welsh, J. A., & Domitrovich, C. E. (2016). The randomized controlled trial of Head Start REDI: Sustained effects on developmental trajectories of social-emotional functioning. *Journal of Consulting and Clinical Psychology*, 84(4), 310–322. <https://doi.org/10.1037/a0039937>
- *Olds, D. L., Henderson, C. R., Jr., Chamberlin, R., & Tatelbaum, R. (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Pediatrics*, 78(1), 65–78. <https://doi.org/10.1542/peds.78.1.65>
- *Olds, D. L., Henderson, C. R., Jr., Cole, R., Eckenrode, J., Kitzman, H., Luckey, D., Pettitt, L., Sidora, K., Morris, P., & Powers, J. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *JAMA*, 280(14), 1238–1244. <https://doi.org/10.1001/jama.280.14.1238>
- *Olds, D. L., Henderson, C. R., Jr., & Kitzman, H. (1994). Does prenatal and infancy nurse home visitation have enduring effects on qualities of parental caregiving and child health at 25 to 50 months of life? *Pediatrics*, 93(1), 89–98. <https://doi.org/10.1542/peds.93.1.89>
- Pages, R., Bailey, D. H., & Duncan, G. J. (2022). *Exploring the "dark matter" of early childhood educational programs: A pattern-of-indirect-effect approach*. PsyArXiv. <https://doi.org/10.31234/osf.io/zubg8>
- *Phillips, L. M., Norris, S. P., & Mason, J. M. (1996). Longitudinal effects of early literacy concepts on reading achievement: A kindergarten intervention and five-year follow-up. *Journal of Literacy Research*, 28(1), 173–195. <https://doi.org/10.1080/10862969609547915>
- *Pössel, P., Adelson, J. L., & Hautzinger, M. (2011). A randomized trial to evaluate the course of effects of a program to prevent adolescent depressive symptoms over 12 months. *Behaviour Research and Therapy*, 49(12), 838–851. <https://doi.org/10.1016/j.brat.2011.09.010>
- *Pössel, P., Horn, A. B., Groen, G., & Hautzinger, M. (2004). School-based prevention of depressive symptoms in adolescents: A 6-month follow-up. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(8), 1003–1010. <https://doi.org/10.1097/01.chi.0000126975.56955.98>
- *Pössel, P., Seemann, S., & Hautzinger, M. (2008). Impact of comorbidity in prevention of adolescent depressive symptoms. *Journal of Counseling Psychology*, 55(1), 106–117. <https://doi.org/10.1037/0022-0167.55.1.106>
- Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, 53, 202–210. <https://doi.org/10.1016/j.intell.2015.10.006>
- Protzko, J. (2017). Raising IQ among school-aged children: Five meta-analyses and a review of randomized controlled trials. *Developmental Review*, 46, 81–101. <https://doi.org/10.1016/j.dr.2017.05.001>
- *Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third grade follow-up to the Head Start Impact Study* (Final report; Report No. OPRE 2012-45). U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://files.eric.ed.gov/fulltext/ED539264.pdf>
- *Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., Jenkins, F., Fletcher, P., Quinn, L., Friedman, J., Ciarico, J., Rohacek, M., Adams, G., & Spier, E. (2010). *Head start impact study: Final report*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. <https://files.eric.ed.gov/fulltext/ED507845.pdf>
- Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, 23(3), 425–438. <https://doi.org/10.1007/s11121-021-01246-3>
- *Quayle, D., Dziurawiec, S., Roberts, C., Kane, R., & Ebsworthy, G. (2001). The effect of an optimism and lifeskills program on depressive symptoms in preadolescence. *Behaviour Change*, 18(4), 194–203. <https://doi.org/10.1375/bech.18.4.194>
- *Redmond, C., Spoth, R. L., Shin, C., Schainker, L. M., Greenberg, M. T., & Feinberg, M. (2009). Long-term protective factor outcomes of evidence-based interventions implemented by community teams through a community-university partnership. *The Journal of Primary Prevention*, 30(5), 513–530. <https://doi.org/10.1007/s10935-009-0189-5>
- *Reitsma, P., & Wesseling, R. (1998). Effects of computer-assisted training of blending skills in kindergarteners. *Scientific Studies of Reading*, 2(4), 301–320. https://doi.org/10.1207/s1532799xssr0204_1

- Reynolds, A. J., & Ou, S.-R. (2011). Paths of effects from preschool to adult well-being: A confirmatory analysis of the child–parent center program. *Child Development*, 82(2), 555–582. <https://doi.org/10.1111/j.1467-8624.2010.01562.x>
- Rieger, S., Göllner, R., Spengler, M., Trautwein, U., Nagengast, B., & Roberts, B. W. (2017). Social cognitive constructs are just as stable as the Big Five between Grades 5 and 8. *AERA Open*, 3(3), 1–9. <https://doi.org/10.1177/2332858417717691>
- *Roberts, C. M., Kane, R., Bishop, B., Cross, D., Fenton, J., & Hart, B. (2010). The prevention of anxiety and depression in children from disadvantaged schools. *Behaviour Research and Therapy*, 48(1), 68–73. <https://doi.org/10.1016/j.brat.2009.09.002>
- *Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P. J., Gathercole, S., Gold, L., Sia, K. L., Mensah, F., Rickards, F., Ainley, J., & Wake, M. (2016). Academic outcomes 2 years after working memory training for children with low working memory: A randomized clinical trial. *JAMA Pediatrics*, 170(5), Article e154568. <https://doi.org/10.1001/jamapediatrics.2015.4568>
- *Rooney, R., Roberts, C., Kane, R., Pike, L., Winsor, A., White, J., & Brown, A. (2006). The prevention of depression in 8- to 9-year-old children: A pilot study. *Australian Journal of Guidance and Counselling*, 16(1), 76–90. <https://doi.org/10.1375/ajgc.16.1.76>
- *Ruiz-Aranda, D., Castillo, R., Salguero, J. M., Cabello, R., Fernández-Berrocal, P., & Balluerka, N. (2012). Short- and midterm effects of emotional intelligence training on adolescent mental health. *The Journal of Adolescent Health*, 51(5), 462–467. <https://doi.org/10.1016/j.jadohealth.2012.02.003>
- *Ryder, J. F., Tunmer, W. E., & Greaney, K. T. (2008). Explicit instruction in phonemic awareness and phonemically based decoding skills as an intervention strategy for struggling readers in whole language classrooms. *Reading and Writing*, 21(4), 349–369. <https://doi.org/10.1007/s1145-007-9080-z>
- *Saine, N. L., Lerkkanen, M., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2010). Predicting word-level reading fluency outcomes in three contrastive groups: Remedial and computer-assisted remedial reading intervention, and mainstream instruction. *Learning and Individual Differences*, 20(5), 402–414. <https://doi.org/10.1016/j.lindif.2010.06.004>
- *Santos, R. G., Chartier, M. J., Whalen, J. C., Chateau, D., & Boyd, L. (2011). Effectiveness of school-based violence prevention for children and youth: Cluster randomized field trial of the Roots of Empathy program with replication and three-year follow-up. *Healthcare Quarterly*, 14(2), 80–91. <https://doi.org/10.12927/hcq.2011.22367>
- *Sarama, J., Lange, A. A., Clements, D. H., & Wolfe, C. B. (2012). The impacts of an early mathematics curriculum on oral language and literacy. *Early Childhood Research Quarterly*, 27(3), 489–502. <https://doi.org/10.1016/j.ecresq.2011.12.002>
- *Sarason, I. G., & Sarason, B. R. (1981). Teaching cognitive and social skills to high school students. *Journal of Consulting and Clinical Psychology*, 49(6), 908–918. <https://doi.org/10.1037/0022-006X.49.6.908>
- *Sawyer, M. G., Harchak, T. F., Spence, S. H., Bond, L., Graetz, B., Kay, D., Patton, G., & Sheffield, J. (2010). School-based prevention of depression: A 2-year follow-up of a randomized controlled trial of the beyondblue schools research initiative. *The Journal of Adolescent Health*, 47(3), 297–304. <https://doi.org/10.1016/j.jadohealth.2010.02.007>
- *Sawyer, M. G., Pfeiffer, S., Spence, S. H., Bond, L., Graetz, B., Kay, D., Patton, G., & Sheffield, J. (2010). School-based prevention of depression: A randomised controlled study of the beyondblue schools research initiative. *Journal of Child Psychology and Psychiatry*, 51(2), 199–209. <https://doi.org/10.1111/j.1469-7610.2009.02136.x>
- *Schacter, J., & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, 28(2), 158–169. <https://doi.org/10.1111/j.1467-9817.2005.00260.x>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666–681. <https://doi.org/10.1177/1745691614553988>
- *Smith, S. W., Daunic, A. P., Barber, B. R., Aydin, B., Van Loan, C. L., & Taylor, G. G. (2014). Preventing risk for significant behavior problems through a cognitive-behavioral intervention: Effects of the tools for getting along curriculum at one-year follow-up. *The Journal of Primary Prevention*, 35(5), 371–387. <https://doi.org/10.1007/s10935-014-0357-0>
- Soland, J., Kuhfeld, M., Wolk, E., & Bi, S. (2019). Examining the state-trait composition of social–emotional learning constructs: Implications for practice, policy, and evaluation. *Journal of Research on Educational Effectiveness*, 12(3), 550–577. <https://doi.org/10.1080/19345747.2019.1615158>
- Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2024). Going beyond traits: Social, emotional, and behavioral skills matter for adolescents' success. *Social Psychological and Personality Science*, 15(1), 33–45. <https://doi.org/10.1177/1948550621127483>
- *Spence, S. H., Sheffield, J. K., & Donovan, C. L. (2003). Preventing adolescent depression: An evaluation of the problem solving for life program. *Journal of Consulting and Clinical Psychology*, 71(1), 3–13. <https://doi.org/10.1037/0022-006X.71.1.3>
- *Spence, S. H., Sheffield, J. K., & Donovan, C. L. (2005). Long-term outcome of a school-based, universal approach to prevention of depression in adolescents. *Journal of Consulting and Clinical Psychology*, 73(1), 160–167. <https://doi.org/10.1037/0022-006X.73.1.160>
- *Spath, R., Randall, G. K., Shin, C., & Redmond, C. (2005). Randomized study of combined universal family and school preventive interventions: Patterns of long-term effects on initiation, regular use, and weekly drunkenness. *Psychology of Addictive Behaviors*, 19(4), 372–381. <https://doi.org/10.1037/0893-164X.19.4.372>
- *Spath, R., Redmond, C., Clair, S., Shin, C., Greenberg, M., & Feinberg, M. (2011). Preventing substance misuse through community-university partnerships: Randomized controlled trial outcomes 4½ years past baseline. *American Journal of Preventive Medicine*, 40(4), 440–447. <https://doi.org/10.1016/j.amepre.2010.12.012>
- *Spath, R., Redmond, C., Shin, C., Greenberg, M., Clair, S., & Feinberg, M. (2007c). Substance-use outcomes at 18 months past baseline: The PROSPER Community-University Partnership Trial. *American Journal of Preventive Medicine*, 32(5), 395–402. <https://doi.org/10.1016/j.amepre.2007.01.014>
- *Spath, R., Redmond, C., Shin, C., Greenberg, M., Feinberg, M., & Schainker, L. (2013). PROSPER community-university partnership delivery system effects on substance misuse through 6 1/2 years past baseline from a cluster randomized controlled intervention trial. *Preventive Medicine*, 56(3–4), 190–196. <https://doi.org/10.1016/j.ypmed.2012.12.013>
- *Spath, R., Redmond, C., Shin, C., Greenberg, M. T., Feinberg, M. E., & Trudeau, L. (2017). PROSPER delivery of universal preventive interventions with young adolescents: Long-term effects on emerging adult substance misuse and associated risk behaviors. *Psychological Medicine*, 47(13), 2246–2259. <https://doi.org/10.1017/S0033291717000691>
- *Spath, R. L., Randall, G. K., Trudeau, L., Shin, C., & Redmond, C. (2008). Substance use outcomes 5½ years past baseline for partnership-based, family–school preventive interventions. *Drug and Alcohol Dependence*, 96(1–2), 57–68. <https://doi.org/10.1016/j.drugalcdep.2008.01.023>
- *Spath, R. L., Redmond, C., Trudeau, L., & Shin, C. (2002). Longitudinal substance initiation outcomes for a universal preventive intervention combining family and school programs. *Psychology of Addictive Behaviors*, 16(2), 129–134. <https://doi.org/10.1037/0893-164X.16.2.129>
- *Stone, W. L., Bendell, R. D., & Field, T. M. (1988). The impact of socioeconomic status on teenage mothers and children who received early intervention. *Journal of Applied Developmental Psychology*, 9(4), 391–408. [https://doi.org/10.1016/0193-3973\(88\)90008-1](https://doi.org/10.1016/0193-3973(88)90008-1)

- *Storr, C. L., Ialongo, N. S., Kellam, S. G., & Anthony, J. C. (2002). A randomized controlled trial of two primary school intervention strategies to prevent early onset tobacco smoking. *Drug and Alcohol Dependence*, 66(1), 51–60. [https://doi.org/10.1016/S0376-8716\(01\)00184-3](https://doi.org/10.1016/S0376-8716(01)00184-3)
- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 49(1), 77–96. <https://doi.org/10.1177/0022219414528540>
- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88(4), 1156–1171. <https://doi.org/10.1111/cdev.12864>
- *Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, 60(1), 40–56. <https://doi.org/10.1007/s11881-009-0032-y>
- *Trudeau, L., Spoth, R., Lillehoj, C., Redmond, C., & Wickrama, K. A. S. (2003). Effects of a preventive intervention on adolescent substance use initiation, expectancies, and refusal intentions. *Prevention Science*, 4(2), 109–122. <https://doi.org/10.1023/A:1022926332514>
- *Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk for reading disabilities. *Journal of Learning Disabilities*, 33(6), 579–590. <https://doi.org/10.1177/002221940003300606>
- *Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology*, 102(4), 786–803. <https://doi.org/10.1037/a0019639>
- *Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, 15(6), 471–497. <https://doi.org/10.1080/10888438.2010.501091>
- *Vadasy, P. F., & Sanders, E. A. (2012). Two-year follow-up of a kindergarten phonics intervention for English learners and native English speakers: Contextualizing treatment impacts by classroom literacy instruction. *Journal of Educational Psychology*, 104(4), 987–1005. <https://doi.org/10.1037/a0028163>
- *Vadasy, P. F., & Sanders, E. A. (2013). Two-year follow-up of a code-oriented intervention for lower-skilled first-graders: The influence of language status and word reading skills on third-grade literacy outcomes. *Reading and Writing*, 26(6), 821–843. <https://doi.org/10.1007/s11145-012-9393-4>
- *Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, 98(3), 508–528. <https://doi.org/10.1037/0022-0666.98.3.508>
- *van der Kooy-Hofland, V. A. C., van der Kooy, J., Bus, A. G., van IJzendoorn, M. H., & Bonsel, G. J. (2012). Differential susceptibility to early literacy intervention in children with mild perinatal adversities: Short- and long-term effects of a randomized control trial. *Journal of Educational Psychology*, 104(2), 337–349. <https://doi.org/10.1037/a0026984>
- *van Otterloo, S. G., van der Leij, A., & Henrichs, L. F. (2009). Early home-based intervention in the Netherlands for children at familial risk of dyslexia. *Dyslexia: An International Journal of Research and Practice*, 15(3), 187–217. <https://doi.org/10.1002/dys.376>
- *Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., Pollard-Durodola, S. D., Fletcher, J. M., & Francis, D. J. (2006). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, 43(3), 449–487. <https://doi.org/10.3102/00028312043003449>
- *Vaughn, S., Cirino, P. T., Tolar, T., Fletcher, J. M., Cardenas-Hagan, E., Carlson, C. D., & Francis, D. J. (2008). Long-term follow-up of Spanish and English interventions for first-grade English language learners at risk for reading problems. *Journal of Research on Educational Effectiveness*, 1(3), 179–214. <https://doi.org/10.1080/19345740802114749>
- *Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., Cardenas-Hagan, E., & Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, 39(1), 56–73. <https://doi.org/10.1177/00222194060390010601>
- *Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S. D., Cardenas-Hagan, E., & Francis, D. J. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal*, 107(2), 153–180. <https://doi.org/10.1086/510653>
- Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, 10(4), 428–443. <https://doi.org/10.1037/1082-989X.10.4.428>
- *Vogel, C. A., Xue, Y., Moiduddin, E. M., Carlson, B. L., & Kisker, E. E. (2010). *Early Head Start children in grade 5: Long-term follow up of the Early Head Start Research and Evaluation Project study sample* (Report No. OPRE 2011-8). Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. <https://www.acf.hhs.gov/sites/default/files/documents/opre/grade5.pdf>
- *Volpe, R. J., DuPaul, G. J., Jitendra, A. K., & Tresco, K. E. (2009). Consultation-based academic interventions for children with attention deficit hyperactivity disorder: Effects on reading and mathematics outcomes at 1-year follow-up. *School Psychology Review*, 38(1), 5–13. <https://doi.org/10.1080/02796015.2009.12087845>
- Wan, S., Bond, T. N., Lang, K., Clements, D. H., Sarama, J., & Bailey, D. H. (2021). Is intervention fadeout a scaling artefact? *Economics of Education Review*, 82, Article 10209. <https://doi.org/10.1016/j.econedu.2021.102090>
- *Wang, Y., Browne, D. C., Petras, H., Stuart, E. A., Wagner, F. A., Lambert, S. F., Kellam, S. G., & Ialongo, N. S. (2009). Depressed mood and the effect of two universal first grade preventive interventions on survival to the first tobacco cigarette smoked among urban youth. *Drug and Alcohol Dependence*, 100(3), 194–203. <https://doi.org/10.1016/j.drugalcdep.2008.08.020>
- *Wasik, B. H., Ramey, C. T., Bryant, D. M., & Sparling, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development*, 61(6), 1682–1696. <https://doi.org/10.2307/1130831>
- Watts, T. W., Bailey, D. H., & Li, C. (2019). Aiming further: Addressing the need for high-quality longitudinal research in education. *Journal of Research on Educational Effectiveness*, 12(4), 648–658. <https://doi.org/10.1080/19345747.2019.1644692>
- Watts, T. W., Bailey, D. H., & Mattera, S. (2022, March). *Exploring patterns of fadeout across multi-site RCTs of an early childhood curriculum intervention* [Conference session]. In 2021 APPAM Fall Research Conference. APPAM.
- Watts, T. W., Clements, D. H., Sarama, J., Wolfe, C. B., Spitler, M. E., & Bailey, D. H. (2017). Does early mathematics intervention change the processes underlying children's learning? *Journal of Research on Educational Effectiveness*, 10(1), 96–115. <https://doi.org/10.1080/19345747.2016.1204640>
- *Weikart, D. P. (1970). *Longitudinal results of the Ypsilanti Perry preschool project*. U.S. Department of Health, Education, and Welfare Office of Education. <https://files.eric.ed.gov/fulltext/ED044536.pdf>
- *Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and

home for children from low-income families. *Developmental Psychology*, 30(5), 679–689. <https://doi.org/10.1037/0012-1649.30.5.679>

*Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology*, 24(4), 552–559. <https://doi.org/10.1037/0012-1649.24.4.552>

*Wolff, U. (2011). Effects of a randomised reading intervention study: An application of structural equation modelling. *Dyslexia*, 17(4), 295–311. <https://doi.org/10.1002/dys.438>

*Wolff, U. (2016). Effects of a randomized reading intervention study aimed at 9-year-olds: A 5-year follow-up. *Dyslexia*, 22(2), 85–100. <https://doi.org/10.1002/dys.1529>

Received June 24, 2023

Revision received July 1, 2024

Accepted July 9, 2024 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <https://my.apa.org/portal/alerts/> and you will be notified by e-mail when issues of interest to you become available!