Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 23-782

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. Hart
Columbia University

Drew H. Bailey
University of California, Irvine

Sha Luo
Columbia University

Pritha Sengupta
Wesleyan University

Tyler W. Watts
Columbia University

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 socialemotional 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 pre-existing meta-analyses, generating a sample of 86 educational RCTs targeting children from infancy through adolescence, together involving 56,662 participants and 450 outcomes measured at post-test and at least one follow-up. Relying on a meta-regression approach for modeling persistence rates, we tested the extent to which post-test impact magnitudes predicted follow-up impact magnitudes. We found that post-test impacts were equally predictive of follow-up impacts for cognitive and social-emotional skills at 6- to 12- months follow-up, indicating similar conditional persistence rates across skill types. At 1- to 2- years 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 post-test impact, indicating that interventions may have long-term effects that are not fully mediated by post-test effects. This pattern of results implied that smaller post-test impacts produced more persistent effects than larger post-test 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.

VERSION: August 2024

Suggested citation: Hart, Emma R., Drew H. Bailey, Sha Luo, Pritha Sengupta, and Tyler W. Watts. (2024). Fadeout and Persistence of Intervention Impacts on Social-Emotional and Cognitive Skills in Children and Adolescents: A Meta-Analytic Review of Randomized Controlled Trials. (EdWorkingPaper: 23-782). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/7j8s-dy98

[MANUSCRIPT ACCEPTED FOR PUBLICATION AT PSYCHOLOGICAL BULLETIN]

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. Hart, Drew H. Bailey, Sha Luo, Pritha Sengupta, Tyler W. Watts*

Corresponding Author:
Tyler W. Watts
462 Grace Dodge Hall
525 W 120th
New York, NY 10027
(212) 678-3095

Acknowledgments

Time spent on this project was supported by the National Institute of Child Health and Human Development (1R01HD095930-01A1 to TW) and the National Science Foundation (DGE-2036197 to EH). 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. We would like to 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 Ofstedal, Fatmanur Ozay, Xinyu Pan, Spruha Reddy, Mindy Rosengarten, John Schupbach, Maddie Scricco, Pritha Sengupta, Jessica Sperber, Devon Turner, Leo Weaver, and Josefa Westerman.

Author Note

The codebooks, analytic syntax, and data necessary to replicate these findings can be found on LDbase: https://doi.org/10.33009/ldbase.1719529626.152e.

Abstract

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 pre-existing meta-analyses, generating a sample of 86 educational RCTs targeting children from infancy through adolescence, together involving 56,662 participants and 450 outcomes measured at post-test and at least one follow-up. Relying on a meta-regression approach for modeling persistence rates, we tested the extent to which post-test impact magnitudes predicted follow-up impact magnitudes. We found that post-test impacts were equally predictive of follow-up impacts for cognitive and social-emotional skills at 6- to 12-months follow-up, indicating similar conditional persistence rates across skill types. At 1- to 2-years 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 post-test impact, indicating that interventions may have long-term effects that are not fully mediated by post-test effects. This pattern of results implied that smaller post-test impacts produced more persistent effects than larger post-test 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.

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

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.

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

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 et al., 2020a). 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., 2022), 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 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 current 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, 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 post-treatment 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 et al., 2020a). 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 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 trait-like 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 carry-over 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 include 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, 2020). In the adolescent intervention example, imagine 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.

[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).

Current Study

The current 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 dataset: 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 pre-existing 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 post-test 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, et al., (2020B); Burns et al., (2016); Kraft et al., (2018); Li et al., (2020); Protzko, (2015, 2017); Suggate, (2016); 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 et al., 2020; Li et al., 2020; Protzko, 2015; Suggate, 2016; Taylor et al., 2017) and meta-analyses that have been widely cited and influential in the field (e.g., Kraft et al., 2018: 1250 citations; Suggate, 2016: 469 citations; Taylor et al., 2017: 2362 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 pre-existing meta-analyses with the goal of generating a varied set of studies with the same underlying causal expectation: Intervention boosts in skills at post-test 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 paper 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 wellbeing, 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 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 post-test, 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 supplement for a complete list). The sample of studies included in the current analysis was further limited by additional criteria (see "Analytic Sample").

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 pre-test, post-test, 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. (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 supplement.

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 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]:

[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 supplement for a model in which only control group post-test standard deviations were used).

$$ES = \frac{M_{tx} - M_{cntrl}}{sd_{cntrl}} \tag{1}$$

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 Bornstein et al., 2009; p. 27):

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

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 (ICC) of 0.10 (see supplement 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 et al., 2020; Bailey & Weiss, 2022; and 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 post-test 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 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 in 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 post-test 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 post-test 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 hereafter) 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 post-test 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 larger post-test 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 post-test 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 post-test 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 post-test and (b) changes in other skills or environments, Y, at post-test. The exact functional form of the decay or persistence of impacts on skill X will depend on several unknown parameters we cannot estimate in the current 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 post-test 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 post-test effect size on skill X. Indeed, a priori, the magnitude of the observed difference between the treatment and control group on skill X at post-test 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 post-test 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 (IRT) based measures of cognitive achievement, Likert-based measures of behavior). It is also reflected in regression-based approaches that relate skill development at one timepoint to skill development at a later timepoint (e.g., Duncan et al., 2007). Our modeling approach follows this conceptualization, as we hypothesize that larger post-test 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 post-test. 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 post-test 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 post-test (i.e., typical omitted variables bias). The slope term is inherently conditioned on post-test 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 post-test intervention effects on skill *X*. The second term of interest is the intercept, which represents the effects of *Y* at post-test on skill *X* at follow-up, for all parts of *Y* that do not

[3] We are not the first to conceptualize of a linear-regression-based slope term as "persistence" (for example, see Dias & Marques, 2005; Blanden, 2019).

reliably change with skill *X* in response to intervention (i.e., changes that are uncorrelated with post-test impacts on skill *X*). To the extent that measures of skill *X* at post-test do not fully capture the underlying skill *X* construct at post-test, 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 post-test impact magnitude. Given our interest in the extent to which *post-test 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 post-test impacts were totally unpredictive 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 post-test 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 post-test 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 post-test impact. For example, a reduction in treatment impacts of 0.05 *SD* is evidence of substantially more fadeout for an intervention with a post-test impact of 0.08 *SD* than a post-test impact of 0.50 *SD*. To further highlight the advantages of this approach, consider an absolute approach that would simply take the difference in SD units between follow-up effect sizes and the post-test impact. A given intervention that produced a post-test 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 post-test 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 one year? Using the same hypothetical example, a relative approach to modeling fadeout would find a rate of 25% conditional persistence from post-test 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 post-test impact of 0.50 *SD* (i.e., 0.50 *SD* post-test impact × 25% 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 post-test impacts, such features are "controlled for" in determining rates of post-test skill persistence using this approach, assuming that the subsequent skill-building dynamics are not influenced by intervention features present before post-test 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 post-test impact (of any magnitude) proportionally persists at follow-up, holding constant absolute differences in post-test 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 non-linearity of treatment impact fadeout across follow-ups. Given the expectation of non-linearity in persistence based on previous fadeout work in Early Childhood Education (ECE) 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 post-test impacts predicted follow-up impacts for each of the aforementioned periods. We binned follow-up waves and treated impacts binned at various timepoints as independent outcomes because we expected that fadeout effects would be non-linear 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 timepoints (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 timepoints in intervention research, as many educational studies follow students on an annual basis during follow-up. Indeed, in our full dataset, a plurality of follow-ups were collected at 12 months after post-test (31%), followed by 10% collected at 24 months after post-test.

Analytic Models

[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 two years" bin was then reconstructed to capture the longest-term follow-up effect in the case that more than one was available.

Using a sample of the total available cognitive and social-emotional post-test impacts carefully linked with follow-up impacts (i.e., same intervention, construct, and measure at both times), we assessed the extent to which post-test effect sizes were predictive of follow-up effect sizes, and the portion of follow-up impacts driven by factors other than skill-specific post-test effects. We then tested whether these parameters varied by skill type (i.e., social-emotional versus cognitive). Because many studies contributed multiple outcomes in our analysis, we address the non-independence of effect sizes from the same study using a random effects model. Given the expectation of non-linearity in persistence, we assessed persistence in post-test 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 post-test impacts predicted follow-up impacts for each of these follow-up periods.

Multi-level random-effects meta-regressions 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:

$$ES_{fsi} = \beta_{0s} + \beta_{1s}ES_{psi} + \varepsilon_{fsi}$$

(3)

Level 2- study:

$$\beta_{0s} = \gamma_{00} + \tau_{0s} \tag{4}$$

$$\beta_{1s} = \gamma_{10} + \tau_{1s} \tag{5}$$

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 post-test assessment wave. Thus, ES_{psi} represents the corresponding effect size for aligned group i from study s at post-test, 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 post-test impacts are zero). The inclusion of this random effect adjusts for the non-independence 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 meta-regression 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 zero. In Figure 2, the blue "100% Persistence" line demonstrates a case where post-test effects do not fade and post-test magnitude perfectly predicts follow-up magnitude (i.e., β_{1s}=1). If this were true, a post-test 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 post-test magnitude, the follow-up effect size is 0. Under this scenario, all intervention impacts fade out regardless of the magnitude of the post-test effect. The green "50% Persistence" line represents a slope of β_{1s}=0.50, in which the follow-up effect is expected to be 50% of the post-test impact. Here, a post-test 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 post-test impact is zero. As such, in this model, the intercept indicates the extent to which factors *uncorrelated* with post-test effects contribute to average follow-up effects. In other words, a non-zero intercept would suggest that follow-up effects are still observed, on average, even when post-test 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 might be consistent with the "dark matter" hypothesis that early interventions may affect other latent skills not captured by post-test impacts (e.g., Elango et al., 2015; 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-post-test 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 post-test.

Likewise, an intercept below 0 would indicate that the follow-up effect is smaller than expected based on the post-test measure, again assuming a linear relation between the post-test 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 test our primary question regarding differences between cognitive and social-emotional impacts:

Level 1- measure/construct groupings:

$$ES_{fsi} = \beta_{0s} + \beta_{1s}ES_{psi} + \beta_2 SOC_{si} + \beta_3 ES_{psi} \times SOC_{si} + \varepsilon_{fsi} \tag{6}$$

Level 2- study:

$$\beta_{0s} = \gamma_{00} + \tau_{0s} \tag{7}$$

$$\beta_{1s} = \gamma_{10} + \tau_{1s} \tag{8}$$

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 post-test 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 (MARS; Appelbaum et al., 2018). The codebooks, analytic syntax, and data necessary to replicate our analyses can be found on 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 post-test and follow-up assessments collected using the same construct and measure over time, the inclusion process—beginning with studies from eight pre-existing meta-analyses—yielded 85 studies with 726 post-test effect sizes and 1,247 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 post-test impacts for 56,662 participants, and 580 follow-up impacts. The supplemental file includes forest plots showing the average cognitive and social-emotional post-test effects for these 86 treatment groups (see 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 wellbeing, 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

[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

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 hours) than social-emotional skills (*M* = 22 hours). 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. 62% 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 5 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 post-test 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 post-test and at least one follow-up. On average, we observed between 1 and 2 follow-up assessments for each aligned group (*M* = 1.41;

intervention with broad targets, the intervention had to target at least one cognitive skill and one social-emotional skill.

range = 1 - 3).[6] The average time elapsed between the post-test 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 post-test and follow-up using the same measure. In alignment with cognitive outcomes, we observed between 1 and 2 follow-up assessments for each grouping (*M* = 1.57; range = 1 - 3), and follow-up measures were collected approximately nine months after the post-test. 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 IntHout 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.

[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 one year; greater than one year, up to two years; greater than two years, up to three years; greater than three years, up to four years; and greater than four years.

Across the board, the RCTs had positive impacts at post-test, though there was considerable range in post-test 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- years follow-up, and 0.05 - 0.14 *SD* ($p = .14 - .65$) at subsequent follow-ups conducted at least two years after post-test. Social-emotional impacts were smaller at post-test (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 around zero: 0.05 to 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.

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

Modeled Persistence

To examine patterns of fadeout more rigorously, and with the same outcomes matched over time, we fit meta-regression models to test the extent to which follow-up effects were a function of post-test 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-ups and 1- to 2-year follow-ups can be found in Table 4 and in Figures 5 and 6. Results for estimates greater than two years after post-test exhibited low estimation precision and, as such, are detailed in the supplement (see 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 ($p < 0.001$), and did not differ by outcome type (i.e., see interaction in 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 post-test impact. For 1- to 2-year follow-up, we observed a statistically non-significant overall effect of post-test on follow-up, suggesting little predictive association between post-test and follow-up impacts across both skill types ($\beta_1 = 0.18$, $p = .15$). However, the post-test impacts on cognitive skills appeared to be more predictive of follow-up effects than were post-test 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 Column 1, we observed an intercept across skill types of 0.07 and 0.02 at the follow-up timepoints 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-years 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 ($p = .14$ to $.20$).

Heterogeneity in Persistence

At the 6- to 12-months follow-up, we observed substantial heterogeneity in follow-up effects between studies (I^2= 61%; $\tau_{intercept}$ = 0.23). Consistent with our results, introducing post-test effect size and outcome type into the model considerably reduced the model heterogeneity ($\tau_{intercept}$ = 0.11 and I^2= 36%), but the inclusion of the interaction between post-test and outcome type did not ($\tau_{intercept}$ = 0.11 and I^2= 33%). The inclusion of random slopes in the model indicated substantial heterogeneity in the association between post-test and follow-up effect size that was similarly not explained by the inclusion of outcome type in the model (τ_{slope}= 0.23 to 0.25 across models).

We observed less between-study heterogeneity in 1- to 2-year follow-up effects (null model: $\tau_{intercept}$ = 0.07; I^2 = 22%). This variability was minimally explained by post-test effect size ($\tau_{intercept}$ = 0.04; I^2= 20%) and skill type ($\tau_{intercept}$ = 0.04; I^2= 20%), and marginally explained by the interaction between post-test and skill type ($\tau_{intercept}$= 0.03; I^2= 5%). Notably, there was substantial heterogeneity in the association between post-test and follow-up effects (τ_{slope}= 0.36), partially explained by the interaction between post-test and skill type (τ_{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 post-test 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 post-test for social-emotional skills in the sample, and 0.40 *SD* was the average for cognitive skills.

We note a few important take-aways 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 post-test 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 post-test impacts showed greater persistence, driven by the intercept effect; when a post-test 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 post-test 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 months 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 discussion). At 1- to 2-years follow-up, overall persistence was greater for cognitive skills than social-emotional skills across almost all post-test intervention impact magnitudes due to the conditional persistence rate. Indeed, social-emotional skill follow-up impacts were nearly entirely predicted by the intercept at the 1- to 2-years 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 post-test 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 post-test and follow-up. Overall, this model produced estimates that were substantively aligned with those from the primary model, though less precise (see Table S4, Column 1).

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 post-test 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 (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 hours of content and then, further, into a subset involving less than 100 hours. This reduced the average hours of intervention for cognitive outcomes to 49.42 and 32.08 hours, respectively (social-emotional average = 14.30 hours). Interventions that did not report intensity were still included in these two models. Again, these checks produced largely consistent results (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 post-test 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-months follow-up, these intervention features did not appear to confound modeled persistence and differences across social-emotional and cognitive skills (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 which 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 Table S5, Column 1). Interestingly, the model indicated that at 6- to 12-months follow-up, intercept-level variation was largely explained by within-study/intervention group variance ($\tau = .09$), rather than between-group differences ($\tau = .00$ for study and intervention). At 1- to 2-years 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/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 Table S5, Column 2). At 6- to 12-months, 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-years 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: post-test effect size. However, there was more heterogeneity in the meta-analysis-level random slopes at 6- to 12-months follow-up ($\tau = .23$) and 1- to 2-years 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 post-test 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, etc.). Insofar as these larger cluster-based RCTs generate smaller post-test impacts, we expect that these differences should not affect our key estimated parameters as the slope is conditioned on post-test impact magnitude. However, there may exist concerns that the dynamics of fadeout (i.e., association between post-test and follow-up impacts) 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 (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 (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 (Table S5, Column 5; see Table S6 for meta-analytic averages of unweighted effects at each timepoint). 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 (Table S6).

Next, we set out to examine the validity of the observed negative post-test impacts. First, we returned to the original reports to check that all negative and statistically significant post-test impacts were correctly coded (see supplement for more discussion). Second, given that theory concerning fadeout generally operates from the assumption that interventions have *positive* post-test impacts, we ran models in which we dropped all outcomes that had negative post-test 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 (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 supplement). 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 (Table S5, Column 7).

Finally, many have expressed concern that fadeout may be the result of changes in variance across assessment waves (see Wan et al., 2021 and Cascio & Staiger, 2012, 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 re-scaled follow-up effect sizes using post-test 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 post-test standard deviations, respectively, for cognitive and social-emotional skills and by follow-up assessment wave. (The ratio of follow-up SD to post-test SD for cognitive skills was: 6- to 12-months: 1.17, 1- to 2-years: .97; for social-emotional skills: 6- to 12-months: 1.05, 1- to 2-years: 1.00.) Then, we tested our primary model using these re-scaled 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-months 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 Table S5, Column 8).

Summary. Across these robustness checks, our primary findings were largely corroborated. For the 6- to 12-month follow-ups, the sensitivity checks provided estimates that were consistent with estimates from our preferred model. For the 1- to 2-year follow-ups, 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 model-to-model. Although the effect for cognitive skills remained relatively similar to the preferred model, the magnitude of the social-emotional effect varied

(β_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 post-test and follow-ups, we ran a series of publication-bias-related tests. On average, approximately 32% of post-test effects were statistically significant at $p < .05$. Figure S4 displays funnel plots for the post-test 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 (PEESE) 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 post-test (β_1= 4.80; $p < .0001$), but also at 6- to 12-month follow-up (β_1= 2.03; $p < .0001$), 1- to 2-year follow-up (β_1= 1.03; $p < .005$), and greater than 2-year follow-up (β_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 (Table S7).

To further probe the possibility of publication bias, we examined the distribution of p values (Simonsohn et al., 2014). Figure S5 presents the relative frequency of p values statistically significant at $p < .05$ for each follow-up assessment wave. The distribution of p

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

values for effect sizes reported at post-test and 6- to 12-month follow-up provided little evidence of *p*-hacking. However, at follow-ups greater than one year after post-test, there appeared to be 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. 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 $p < 0.01$ 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 over-estimated post- and follow-up effects. Bias-adjusted effects were estimated to be between 60-68% 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 post-test 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 post-test and follow-ups.[8] The conditional persistence rates were relatively similar to those from the preferred aggregation approach (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 post-test and each follow-up assessment, respectively. When the data were aggregated without consideration for measure type or construct alignment (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 post-test. To evaluate the likelihood of this issue, we first compared the average post-test effect size for all outcomes in our sample to only those that had aligned follow-up. With this we assessed the extent to which post-tests linked with follow-up were larger than post-test impacts without follow up. As compared to post-tests for with aligned follow-up (M = 0.29 *SD*), the impact for post-tests without aligned follow-up was slightly smaller (M = 0.20 *SD*), suggesting some selection.

[8] Of note, although this approach included more effects, from non-aligned 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.

Next, we checked the post-test effect sizes associated with the sample of follow-up effects reported at each assessment wave (see Table S10, Columns 2, 4, and 6). We found little evidence for selection into follow-up up to 2 years after post-test, with similar post-test effect sizes for all outcomes with aligned follow-up and outcomes with follow-up up to 2 years after post-test. However, post-test effect sizes at later follow-up waves, particularly for cognitive skills, were generally larger than initial post-test effect sizes. This pattern was generally observed for unweighted post-test estimates and less-so for weighted post-test 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 Table S10) may be inflated, especially for cognitive skills reported more than two-years after post-test.

Discussion

The current 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 pre-existing 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 post-test 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 meta-regression approach, in which we tested the extent to which post-test 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 post-test and follow-up. This approach also allowed us to observe whether "unmeasured mediators" (i.e., factors affected by the intervention but unrelated to post-test impacts) might also affect follow-up impacts.

Overall, we observed substantial fadeout of post-test 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 post-test and 6- to 12-month follow-up. This rate further decreased to 18% at the 1- to 2-years 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 meta-regression models also allowed us to observe the extent to which follow-up impacts varied across studies. Indeed, we observed substantial heterogeneity in follow-up effects, even after accounting for post-test 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 post-test effect magnitude. For post-test impacts of the *same magnitude*, conditional persistence rates were similar for cognitive and social-emotional skills at 6- to 12-months follow-up, and more favorable for cognitive skills at 1- to 2-years follow-up. Given the presence of small intercept effects for both skill types, when post-test impacts are small, or even zero, our results suggest that there is proportionally *stronger* persistence than when impacts are larger. Although counter-intuitive, 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 post-test impacts on average (i.e., .14 for social-emotional skills versus .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 post-test impact magnitude grows, the relative importance of the small intercept term shrinks, and the conditional persistence rate (~45% at 6- to 12-months) begins to account for most of the follow-up effect.

Thus, the current 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 post-test 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 non-zero (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 two years after post-test. This imprecision precludes concrete conclusions about the relation between post-test 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 randomized control trials 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 post-test 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 non-linear, 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 non-linearity in effects across follow-up assessments, it could also be the case that there is non-linearity in conditional persistence rates by post-test impact magnitude that our models do not fully capture. This could be the case if conditional persistence rates are moderated by post-test impact magnitude. For example, post-test 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 post-test and various follow-up waves, then the external validity of our results could be limited if other sets of non-observed 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 post-test impacts, and it is through these post-test 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), post-test 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 post-test 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 versus 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 reflects 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 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 et al., 2020; Pages et al., 2022). In light of considerable heterogeneity in follow-up effects and conditional persistence rates that were unexplained by post-test 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 post-test 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

* Indicates papers 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

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

*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

Bailey, D., 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., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020a). 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., Jenkins, J. M., & Alvarez-Vargas, D. (2020b). Complementarities between early educational intervention and later educational quality? A systematic review of the sustaining environments hypothesis. *Developmental Review*, *56*, 100910. https://doi.org/10.1016/j.dr.2020.100910

Bailey, D., & 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%2Dterm%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 & 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. *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.tc.idm.oclc.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.tc.idm.oclc.org/10.1037/a0033663

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

*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. *The Journal of the American Medical Association, 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

*Botvin, G. J., Baker, E., Renick, N. L., Filazzola, A. D., & Botvin, E. M. (1984).. *Addictive Behaviors, 9*(2), 137-147. https://doi.org/10.1016/0306-4603(84)90051-0

*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

*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

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

*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), 926-927. 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)

*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

*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., 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. https://doi.org/10.1037/a0026644

*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

*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

*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/10.1111/j.1467-8624.1994.tb00777.x

*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). https://doi-org.tc.idm.oclc.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). Available from ProQuest Dissertations & Theses Global. (304420573).

http://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.

*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

*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.

*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.

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), i–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

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

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... &

Japel, C. (2007). School readiness and later achievement. *Developmental psychology*, *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. *Journal of Early Adolescence, 8*(3), 297-310. https://doi-org.tc.idm.oclc.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), 635-48. 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

*Ecalle, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low progress readers. *Computers and 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., Cole, R., Kitsman, H., Arson, E., Sidora, K., Powers, P., & Olds, D. L. (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. A*rchives of Pediatric 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

*Elango, S., García J., L., Heckman, J. J., Hojman, A. (2016). *Early childhood education*. In Economics of Means-Tested Transfer Programs in the United States, volumn 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. *Behavioral 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., (2008). The effectiveness of a school-based substance abuse prevention program: EU-Dap cluster randomized 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., & EU-Dap Study Group (2007). A cluster randomized controlled trial of school-based prevention of tobacco, alcohol and drug use: The EU-Dap 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., Rosaria Galanti, M., & the EU-Dap Study Group (2010). The effectiveness of a school-based substance abuse prevention program: 18-month follow-up of the EU-Dap 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 mothers 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 and 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., 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

*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*, 245-258. https://doi.org/10.1007/s11121-005-0007-0

*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.tc.idm.oclc.org/10.1037/0022-006X.75.1.9

*Gittelman, R., & Feingold, I. (1983). Children with reading disorders: Efficacy of reading remediation. *Journal of Child Psychology & 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*, 103-112. https://doi.org/10.1007/s11121-006-0025-6

*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 Counselling, 26*, 109-124. https://doi.org/10.1023/B:ADCO.0000027425.24225.87

*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

*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. *Journal of Special Education, 39*(2), 66–85. https://doi.org/10.1177/00224669050390020301

*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), 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 and 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/ldbase.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 Psychiatry, 45*(2), 338–358. https://doi.org/10.1111/j.1469-7610.2004.00225.x

* Head Start Classroom-based Approaches and Resources for Emotion and Social Skill Promotion Administration for Children and Families Office of Planning, Research, and Evaluation. (2014). *Impact findings from the Head Start CARES demonstration: National evaluation of three approaches to improving preschoolers' social and emotional competence* (P. Morris, S. K. Mattera, N. Castells, M. Bangser, K. Bierman, & C. Raver, Authors; Research Report No. OPRE 2014-44). U.S. Department of Health and Human Services. https://files.eric.ed.gov/fulltext/ED546649.pdf

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

*Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review, 103*(6), 2052-2086. https://doi.org/10.1257/aer.103.6.2052

*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. http://dx.doi.org/10.1037/0012-1649.39.4.730

*Horn, A.B., Goren, G., Hautzinger, M. (2004). School-based prevention of depressive symptoms in adolescents: a 6-month follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(8), 1003-1010. https://doi.org/10.1097/01.chi.0000126975.56955.98

*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

*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 & Behavioral Disorders, 9*(3), 146-161.

IntHout, J. Ioannidis, J. P. A, Robers, M. M., Goeman J. J. (2015). Plea for presenting prediction intervals in meta-analysis. BJM Open, *6*, 1-6. https://doi.og/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), 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.tc.idm.oclc.org/10.1080/10888691.2015.1074050

*Johnson, K. W., Shamblen, S., 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*, 298-312. https://doi.org/10.1007/s11121-009-0132-2

*Johnson, N. C. (2000). *A follow-up study of a primary prevention program targeting childhood depression* (Order No. 1402700). Available from ProQuest Dissertations & Theses Global. (304634088). http://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.pdfAQ68

*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

* 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

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/10.26300/5tvg-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.pdfAQ69

* Luna-Adame, M., Carrasco-Giménez, T., & del Mar Rueda-García, 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. A. 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

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

*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

*Vaughn, S., Mathes, P., Linian-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. *Elementary School Journal*, *107*(2), 153–180. https://doi.org/10.1086/510653

*McCarton, Brooks-Gunn, Wallace, et al. (1997). Results at age 8 years of early intervention for low-birth-weight premature infants: The infant health and development program. *The Journal of the American Medical Association, 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., Julliet, B., & Cunliffe, R. (2004). A randomized placebo-controlled trial of a school-based depression prevention program. *Journal of American Academy of Child & Adolescent Psychiatry, 43*(5), 538-547. https://doi.org/10.1097/00004583-200405000-00007

*Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). I*mpact 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.pdfAQ70

*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., 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., & Kitzmna, 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

*Olds, D. L., Henderson, C. R., 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

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., 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.pdfAQ72

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. *Behavior 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), 109. 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. *Behavior 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., 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) 1-10. 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 Psychologists and Counsellors in Schools, 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 mid-term effects of emotional intelligence training on adolescent mental health. *Journal of Adolescence Health, 51*(5), 462–467. https://dx.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/s11145-007-9080-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), 908–919. 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. *Journal Of Adolescent Health, 47*(3), 297-304. doi: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 randomized 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, 2*(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. (2022). Going beyond traits: Social, emotional, and behavioral skills matter for adolescents' success. *Social Psychological and Personality Science*, 19485506221127484. https://doi.org/10.1177/19485506221127483

*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

*Spence, S. H., Sheffield, J., & 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

* Spoth, 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

* Spoth, 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

*Spoth, 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

* Spoth, 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

*Spoth, 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

*Spoth, R., Redmond, C., Shin, C., Greenberg, M., Feinberg, M., & Schainker, L. (2013). PROSPER community–university partnership delivery system effects on substance misuse through 6 ½ 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

* Spoth, R., 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/10.1016/01933973(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

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*, 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*, 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/10.1037/0022-0663.98.3.508

*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.

*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.tc.idm.oclc.org/10.1080/19345740802114749

*Vaughn, S., Linian-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

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.

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*, 10209. https://doi.org/10.1016/j.econedurev.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

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., Bailey, D. H., & Mattera, S. (2022, March). Exploring Patterns of Fadeout across Multi-Site RCTs of an Early Childhood Curriculum Intervention. In 2022 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

*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/10.1111/j.1467-8624.1990.tb03559.x

*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., 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

*Whitehurst, G.J., Arnold, D.S., Epstein, J.N., & Angell, A.L. (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.tc.idm.oclc.org/10.1037/0012-1649.30.5.679

*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

Figure 1

Flow of Reports and Studies into the Meta-Analysis

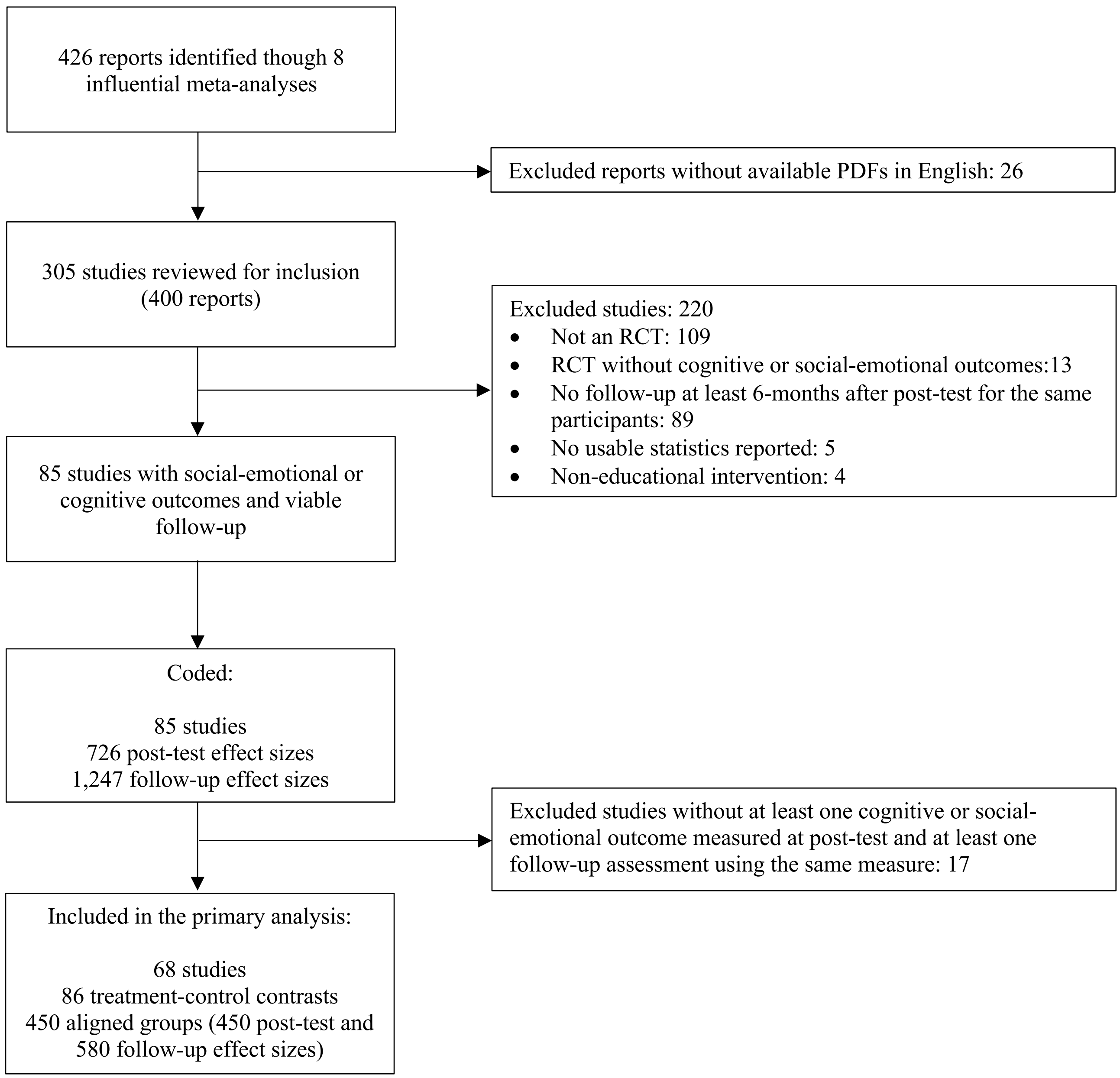
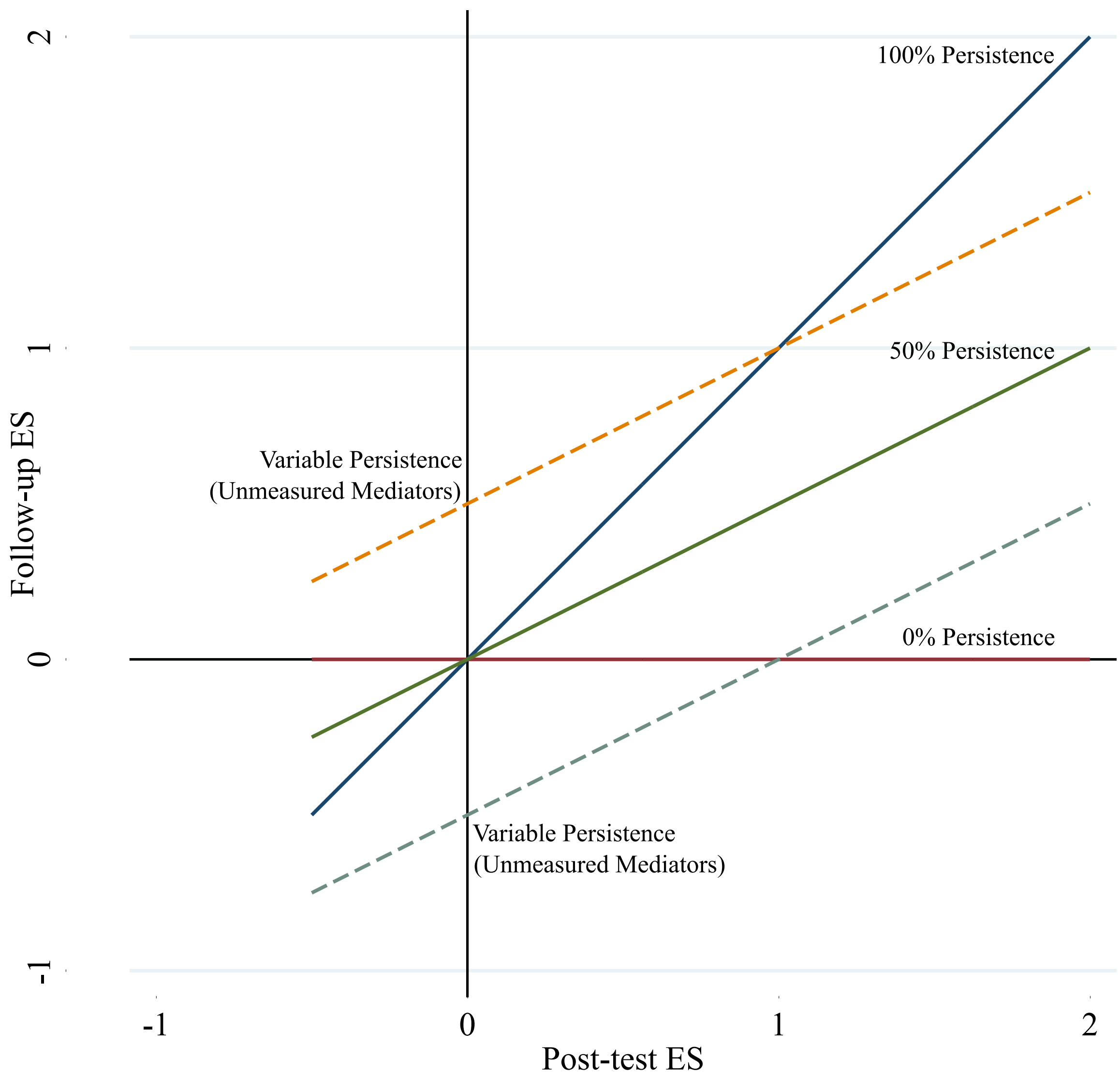


Figure 2

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: 1) y =.5 + .5x where the intercept and slope are .50 (orange); and 2) y = -.5 + .5x where the intercept is -.50 and the slope is .50 (gray). "ES" indicates effect size.

Table 1

Intervention and Participant Characteristics (mean [minimum- maximum])

Intervention/Participant Characteristics	All Outcomes		Social-emotional Outcomes		Cognitive Outcomes	
	Average (1)	% non-missing (2)	Average (3)	% non-missing (4)	Average (5)	% non-missing (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 [4-1075]	65%	21.73 [6-163]	50%	115.22 [4-1075]	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 Wellbeing	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 n	5.23 [1-33]	100%	5.95 [1-19]	100%	4.93 [1-33]	100%
Participant n (at post-test)	727.71 [24-10170]	100%	1332.33 [42-10170]	100%	375.17 [24-3928]	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 what 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.

Table 2
Cognitive and Social-Emotional Outcome Descriptives

Outcome/Treatment Focus	ESs from group (%)	Treatment Groups (#)	Aligned Groupings (#)	*M* # of Follow-ups	*M* Months since Post-test	*M* Post-test ES, weighted	*M* Post-test ES (SE), unweighted	Post-test 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%**	**1392**
Social-Emotional- General	42%	28	89	1.57	8.38	0.17 (0.06)*	0.21 (0.09)*	25%	1474
Substance Use	29%	11	63	1.05	6.90	0.13 (0.06)+	0.18 (0.06)+	30%	2262
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%	1413
Broad TXs	10%	11	47	2.51	17.98	0.35 (0.10)**	0.29 (0.11)*	45%	1047

+ $p < 0.10$ * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
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 post-test). Timing of assessments refers to the average number of months that elapsed between post-test at follow-up assessment(s). The number of aligned groups refers to the number of groupings in which a post-test 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.

Table 3

Average Effect Sizes Across Post-test and Follow-up Assessment Waves

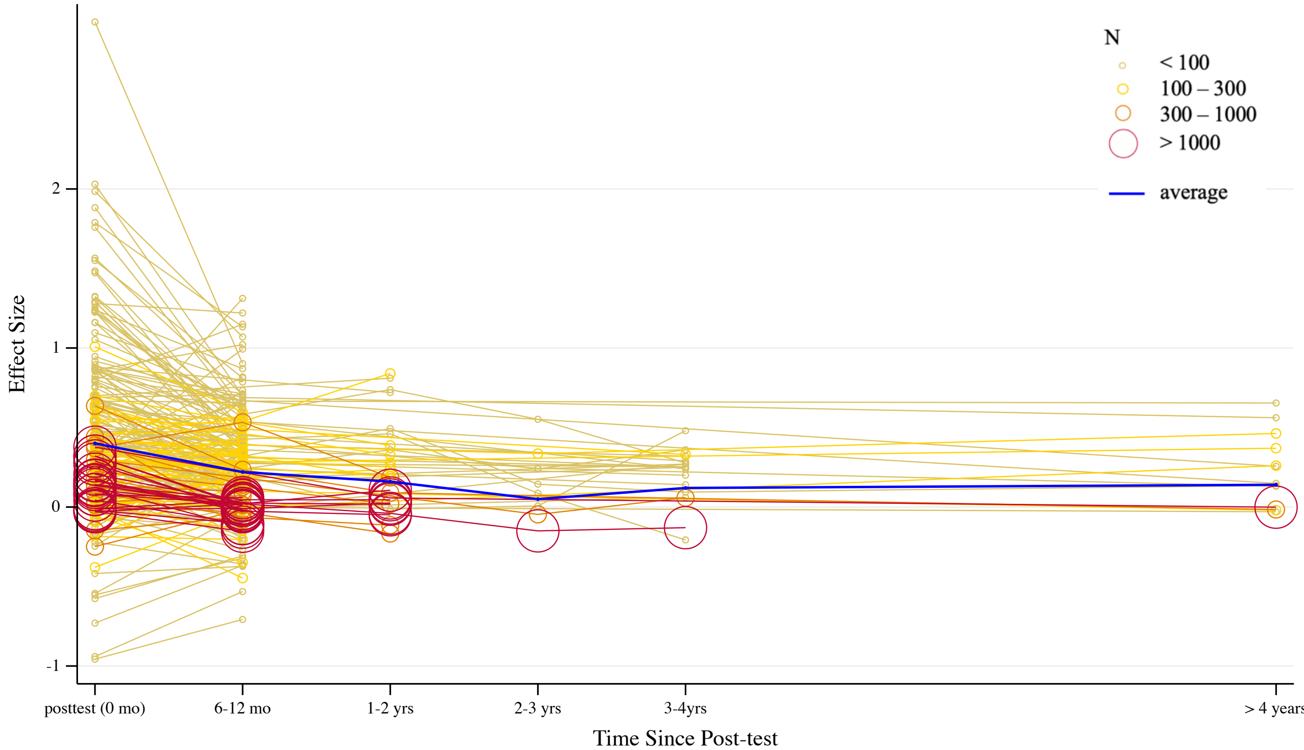
Assessment Wave	All Outcomes		Social-Emotional Outcomes		Cognitive Outcomes	
	M (SE) (1)	n	*M* (SE) (2)	n	*M* (SE) (3)	n
Post-test	0.29 (0.04)***	450	0.14 (0.04)**	214	0.40 (0.06)***	236
	[-0.48, 1.06]		[-0.39, 0.67]		[-0.42, 1.22]	
6 months to 1 year	0.21 (0.03)***	420	0.16 (0.05)**	197	0.22 (0.04)***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

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. "ES" = Effect size. "n" refers to the number of aligned groups. 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 IntHout et al., 2015). The analytic sample was constituted by "aligned groupings" which included a post-test 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).

Figure 3

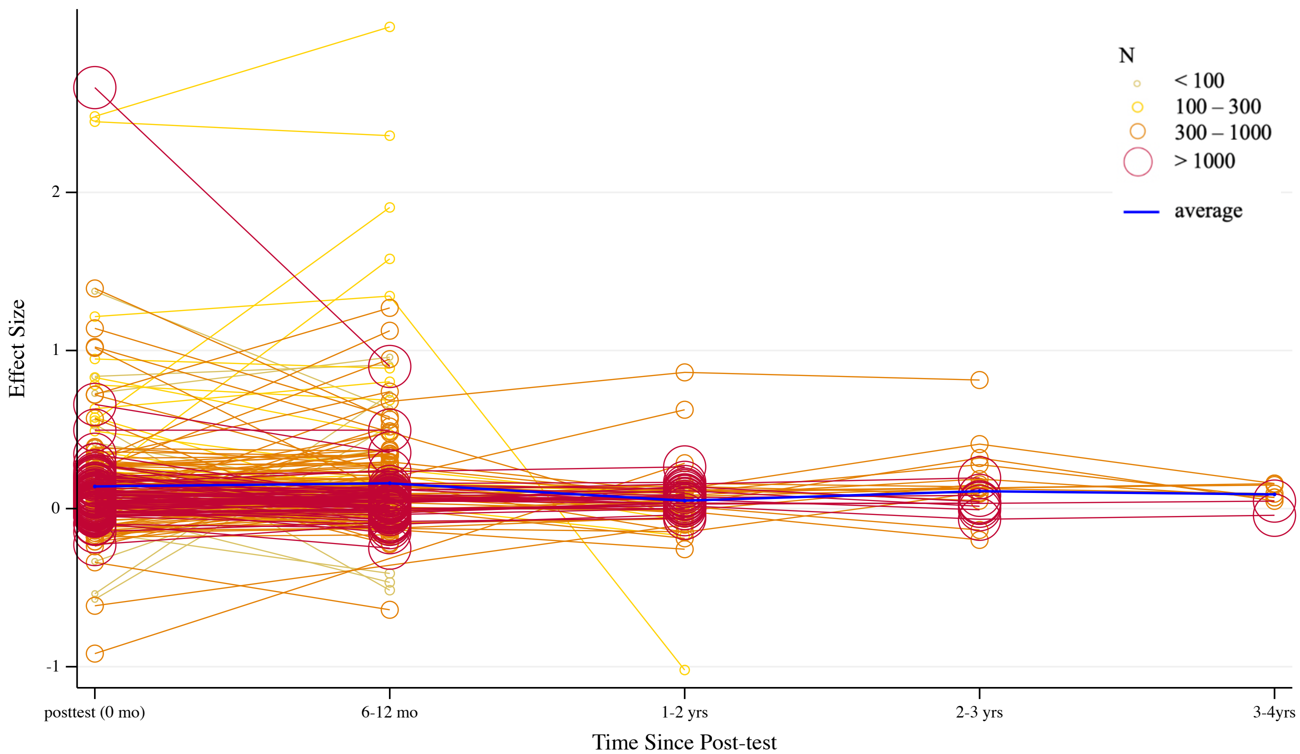
Effect Size Trajectories- Cognitive Outcomes



Note. Each line represents a cognitive construct that was measured at post-test 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 post-test 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 standard deviations at post-test are not displayed (n = 1 group).

Figure 4

Effect Size Trajectories- Social-Emotional Outcomes



Note. Each line represents a social-emotional construct that was measured at post-test 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 post-test sample size (larger coordinates represent estimates from larger samples) with aligned color-coding (darker colors represent larger sample sizes).

Table 4

Regression-Based Persistence Estimates for Social-Emotional and Cognitive Outcomes (β (se))

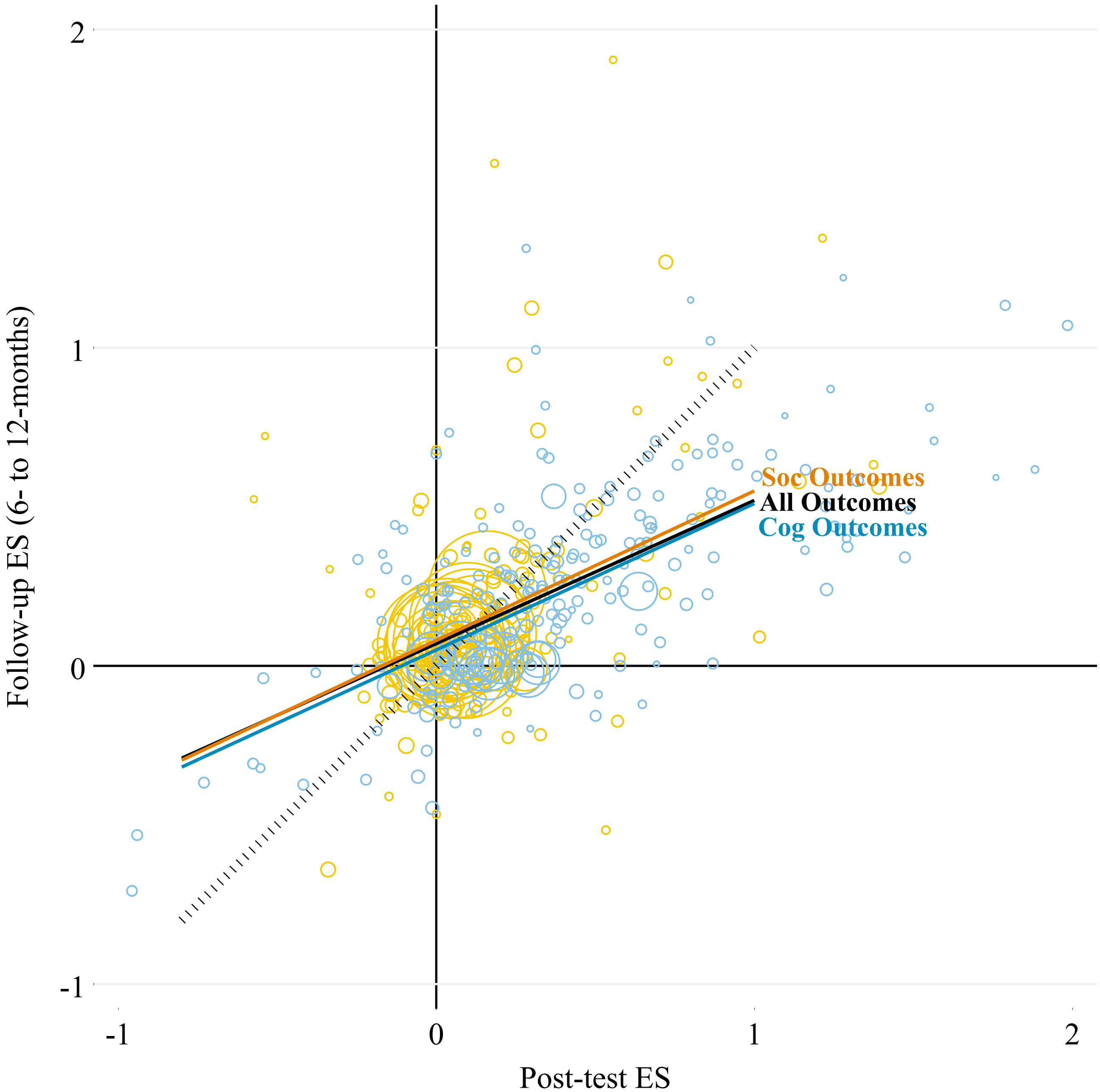
Parameter Labels	All Outcomes			Split Outcomes	
	Baseline (1)	Main Effect (2)	Interaction (3)	Cognitive (4)	Social-Emotional (5)
Panel A: 6- to 12-months Follow-up					
Post-test Effect	0.45*** (0.05)	0.46*** (0.05)	0.46*** (0.06)	0.45*** (0.06)	0.41*** (0.09)
Soc (vs. cog) Outcome		0.03 (0.01)	0.03 (0.01)		
Post-test x 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/intervention/outcome)	60 / 77 / 420	60 / 77 / 420	60 / 77 / 420	35 / 48 /223	30 / 37 / 197
$\tau_{intercept}$ (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					
Post-test 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)		
Post-test x 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/intervention/outcome)	23 / 24 / 87	23 / 24 / 87	23 / 24 / 87	11 / 11 / 28	13/ 14 / 59
$\tau_{intercept}$ (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%

+ $p < 0.10$ * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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 post-test and follow-up impacts collected for the same construct using the same measure, subscales, and reporter at post-test and at least one follow-up assessment within an intervention. As such, each regression tests the association between post-test and follow-up effect sizes for the same measure. The conditional persistence rate is captured in the estimates of the post-test effect, and the interaction between post-test and the social-emotional skill indicator. Panel A presents associations between post-test and follow-up effect sizes collected 6 to 12 months after post-test. Panel B presents associations between post-test and follow-up collected greater that 1 year and up to 2 years after post-test. Model parameters were estimated using a random effects meta-analytic model that included a random effect for study, random slopes for post-test 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 meta-regression. 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; Int = Intervention; Soc = social-emotional; Cog = cognitive.

Figure 5

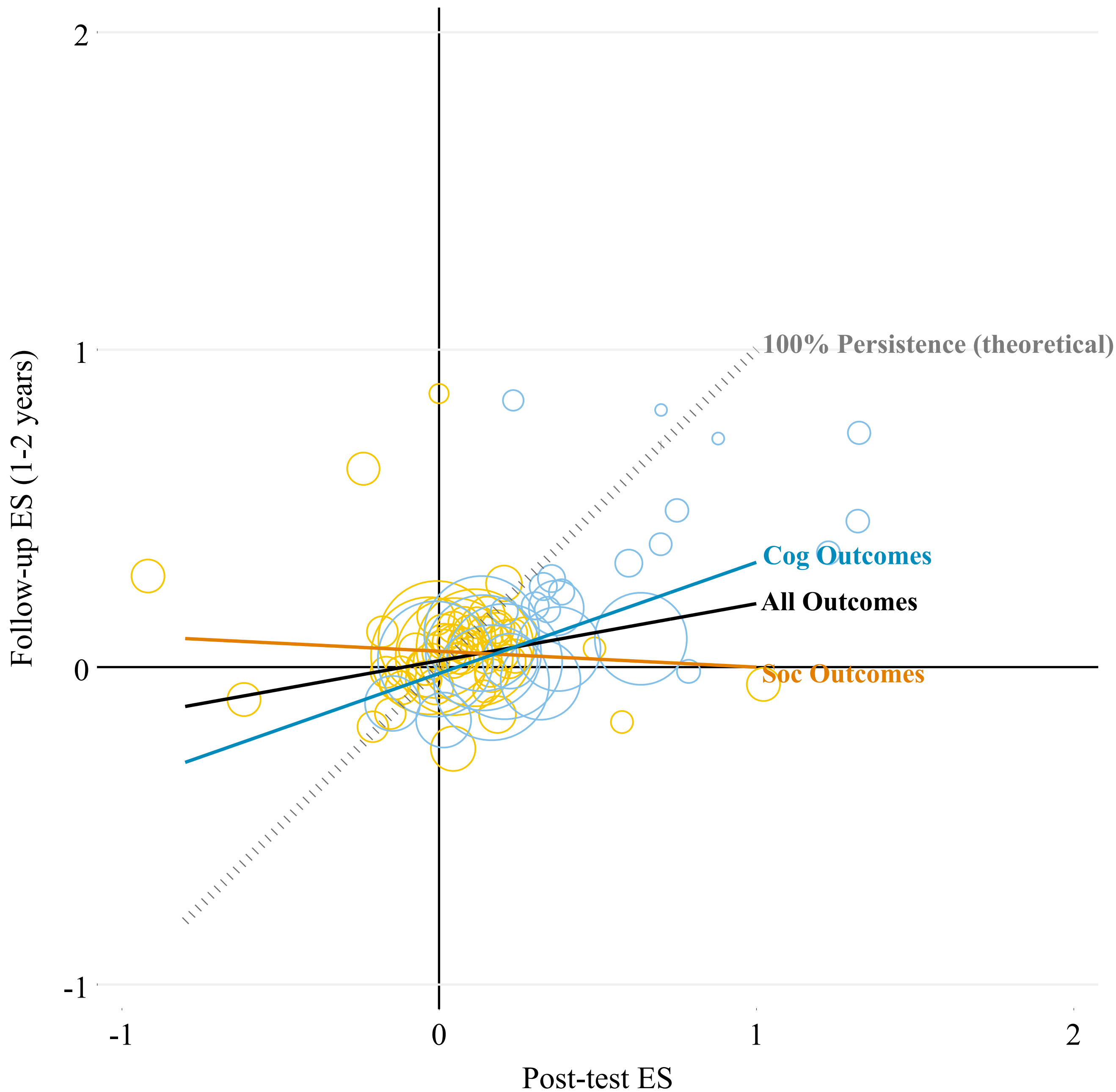
Persistence Patterns: Post-test to 6- to 12-month Follow-up



Note. Each coordinate represents an impact on a cognitive or social-emotional construct that was measured at post-test and 6- to 12- months after post-test (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-months follow-up). The gray dashed "100% Persistence" line was included for reference. For display purposes, post-test 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.

Figure 6

Persistence Patterns: Post-test to 1- to 2-year Follow-up



Note. Each coordinate represents an impact on a cognitive or social-emotional construct that was measured at post-test and 1 to 2 years after post-test (i.e., one "aligned group"). Yellow coordinates represent social-emotional impacts and blue coordinates represent cognitive constructs. 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-years follow-up). The gray dashed "100% Persistence" line was included for reference. For display purposes, post-test 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.

Table 5
Forecasted Follow-up Effects and Persistence for Various Plausible Post-test Effects

Parameter Labels	Social-Emotional Skills				Cognitive Skills			
Hypothetical Post-test Effect	0.14^	0.25	0.40	0.50	0.14	0.25	0.40^	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 SD	-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 SD	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 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 post-test 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). ^ indicates that the post-test effect matches the meta-analytic average for the respective skill type (0.14 *SD* for social-emotional skills, 0.40 *SD* for cognitive skills).

Table 6

Longitudinal Persistence by Outcome and Intervention Types (β (se))

Parameter Labels	Intervention Type Interactions	
	Cognitive Outcomes (1)	Social-Emotional Outcomes (2)
Panel A: 6- to 12-months Follow-up		
Post-test Effect Size	0.23	0.25
	(0.13)	(0.14)
Cog Intervention	0.09	
	(0.05)	
Soc Intervention	0.01	0.02
	(0.06)	(0.06)
Post-test x Cog Intervention	0.27	
	(0.14)	
Post-test x Soc Intervention	-0.35	0.17
	(0.13)	(0.17)
Constant	-0.01	0.04
	(0.03)	(0.05)
Sample size (study/intervention/outcomes)	35 / 48 /223	30 / 37 / 197
$\tau_{intercept}$ (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		
Post-test Effect Size	0.20	0.12
	(0.13)	(0.17)
Cog Intervention	-0.14+	
	(0.04)	
Soc Intervention		0.01
		(0.03)
Post-test x Cog Intervention	0.37	
	(0.19)	
Post-test x Soc Intervention		-0.18
		(0.28)
Constant	0.05	0.05
	(0.04)	(0.01)
Sample size (study/intervention/outcome)	11 / 11 / 28	13/ 14 / 59
$\tau_{intercept}$ (null = .16, .00)	.06	.04
τ_{slope}	.15	.40
I^2 (null = 50.92%, 0.00%)	31.44%	0.00%

+ $p < 0.10$

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 post-test and follow-up impacts collected for the same construct using the same measure, subscales, and reporter at post-test and at least one follow-up assessment within an intervention. As such, each regression tests the association between post-test and follow-up effect sizes for the same measure. Panel A presents associations between post-test and follow-up effect sizes collected 6 to 12 months after post-test. Panel B presents associations between post-test and follow-up collected greater than 1 year and up to 2 years after post-test. 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 meta-regression. 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.

Supplemental File for:

Fadeout and Persistence of Intervention Impacts on Social-Emotional and Cognitive Skills in Children and Adolescents: A Meta-Analysis of Educational Randomized Controlled Trials with Follow-Up

Additional Methodological Details

The full meta-analysis protocol is available on LDbase (https://doi.org/10.33009/ldbase.1719529626.152e). The protocol provides additional details on the study inclusion/exclusion, coding, data cleaning, and effect size calculation processes.

Inclusion Determinations

Note that there were 10 duplicate reports (i.e., papers) included in the original sample of reports from the eight meta-analyses (i.e., there were 436 unique reports pre-removal).

RCT with Social-Emotional or Cognitive Outcomes

The first author and at least one Masters-level student independently reviewed each intervention to determine whether each was an RCT (first) and then, of these, whether each RCT reported social-emotional or cognitive outcomes.[1] Discrepancies in determinations were resolved by the first author and a project PI.

Follow-up at Least 6 months after Post-test on the Same Sample

First, for each intervention, a Ph.D. student determined the report that presented "initial impacts" (i.e., the first report with intervention impacts). Second, at least two research assistants used these initial impact reports to conduct a Google Scholar search to identify and gather all additional reports documenting treatment impacts for that intervention. Next, at least two research assistants independently reviewed all of the gathered reports for each intervention and determined whether the intervention contained adequate follow-up. The first author reviewed all decisions and resolved discrepancies.

Usable Statistics

The first author reviewed all of the reports for the availability of usable statistics. A Masters-level student reviewed cases that the first author deemed as lacking usable statistics prior to formal exclusion.

In one case, the authors had access to participant-level data for a study that was screened for inclusion (TRIAD study; Clements et al., 2013). Because this study contained several waves of data with consistently observed measures that were not clearly reported in prior published work, the study authors generated standardized treatment impact estimates ourselves using the participant dataset.

Coding

The coding protocol can be found in the larger meta-analysis protocol on LDbase (https://doi.org/10.33009/ldbase.1719529626.152e). The protocol details guiding principles of the coding process as well as code-level details that informed how the coders made determinations.

Construct Categorizations

Following coding, the first author and a study PI independently conceptually categorized the author-reported constructs for each reported treatment impact. The team reviewed the constructs and derived categories that conceptually captured the key constructs present in the data. The first author and study PI then independently categorized each construct according to the following options: achievement composite, attendance, criminality, educational attainment, externalizing behaviors, general cognition, grades/GPA, internalizing symptoms, language and literacy, learning skills, math, mixed composite (i.e., a measure that combined cognitive and

---

[1] In several cases there was more than one report that presented treatment impacts on the same intervention. In these cases, all reports on a particular intervention were reviewed to determine intervention inclusion or exclusion.

social-emotional skills), other academic ability, retention, general social-emotional skills, special education designation, and substance use. Some outcomes did not fall within these categories. For these outcomes, the construct categorization was set to missing.

For the purposes of the current analysis, outcomes were classified as "cognitive" or "social-emotional" if they clearly fit into one category or the other. Thus, the following outcomes were classified as "cognitive": achievement composites, general cognition, language and literacy, math, and other academic abilities. The following outcomes were classified as "social-emotional": crime, externalizing behaviors, internalizing symptoms, general social-emotional skills, or substance use. As such, outcomes that fell into the following categories were not included in the current analysis: attendance, educational attainment, grades/GPA, learning skills, mixed composites, retention, special education designation.

Intervention Target Categorizations

Intervention targets were coded based on skills that the study authors explicitly stated that the intervention aimed to improve (see coding protocol for details). To be categorized as a cognitive intervention for the purposes of the exploratory outcome by intervention type interaction analyses, an intervention had to target no social-emotional skills and at least one of the following cognitive skills: math, language and literacy, executive function, general cognition, or science. Alternately, to be categorized as a social-emotional intervention the intervention had to target no cognitive skills and 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.

Effect Size and Standard Error Calculations

The first author worked closely with the study PIs to calculate effect sizes. In keeping with our "double coding" process, an additional Ph.D. student checked all calculations. Figure S1 details the formulas used to calculate effect sizes based on the available, reported results.

The ultimate goal of the effect size calculation process was to identify one effect size for each coded outcome. While the standard protocol was to calculate effect sizes according to the formula detailed in the manuscript, or to use a viable author-reported effect sizes when these were available, there were many cases in which additional decision criteria were used to determine which effect size to use, or to calculate the effect size.

Adjustments for Effect Sizes Calculated using SEs, t statistics, and f statistics

In cases when standard deviations were not provided and viable reported effect sizes were not available, reported standard errors, *t* statistics, and *f* statistics were used to derive effect sizes (see Figure S1). In the case that any of these statistics were used to calculate effect sizes for a given outcome, the first author returned to the original report to check whether the statistic appeared to have been calculated in a model with the inclusion of the pre-test control. In these cases, an adjustment was made when calculating the effect size given the likelihood that standard errors may have been reduced as a result of the inclusion of this control, thus biasing the effect sizes calculated using these estimates. In the cases that this control was included, the standard errors calculated from the available statistics were divided by the square root of 1 minus R^2 (assuming an R^2 between pre- and post-test measures of .50) in the effect size calculation process (using the formulas outlined in Figure S1). Thus, adjustments were made by dividing standard errors by .87 in these cases to ensure that the standard errors were not inaccurately small in the effect size calculation process.

Importantly, in many cases, these adjusted effect sizes were then used to estimate an accompanying standard error for use in our models (i.e., to weight more heavily studies with greater precision). To ensure that these estimated standard errors used were not inaccurately large in our meta-analytic models due to the .87 effect size adjustment, estimated standard errors were multiplied by .87.

Calculating ES using p values

In the case that no alternative statistics were available to use for calculating effect sizes, the last resort was to estimate an effect size using reported *p* values. If precise *p* values were reported (e.g., ".002"), then *t* statistics were calculated from these *p* values and the formulas detailed in Figure S1 were then used to convert *t* values to effect sizes.

If relatively precise *p* values were reported (e.g., "< .05"), we found the smallest difference between means for each measure within a given study and assumed this *p* value was the largest possible associated *p* value (e.g., .05). For these cases, we then converted the *p* value to a *t* value using the "invt" function in Stata, assuming the p-value came from a two-tailed test (i.e., we divided the *p* value by 2 given "invt" function assumptions). Next, we calculated the effect size from this *t* value (as described above), and recovered a SD from this calculated effect size. For the cases in which the same measure was available within a study but did not qualify as having the smallest difference between means, the recovered SD was then used to calculate these effect sizes.

In the case that *p* values were only reported to be statistically non-significant, with no precise value associated, we found the largest difference between means for each measure within a given study and assumed that this *p* value was .10. We then converted the *p* value to a *t* value using the same procedure described above for relatively precise *p* values and recovered a SD that was then used to calculate the effect size for the other cases within a study that had smaller differences between the means for each measure.

In the cases where treatment and control group means were not provided for an outcome, and the treatment impact was noted to be statistically non-significant, *p* values were assumed to be .10 and *t* statistics were calculated from these *p* values. Because means were not available, an alternate equation was used to convert *t* values to effect sizes (see next section).

For all of these aforementioned processes, we made the .87 pre-test covariate adjustment when it appeared that the *p* value came from a model including a pre-test control (see previous section for more details).

Calculating ES from f and t statistics when Means were Not Reported.

When treatment and control group means were not provided, and effect sizes were estimated using *t* statistics (only in the case of *p* value conversions) or *f* statistics (in the case of one study), the following equations were used (Higgins et al., 2023):

$$ES = t \times \sqrt{\frac{n_{tx} + n_{cntrl}}{n_{tx} * n_{cntrl}}}$$

$$ES = \sqrt{f} \times \sqrt{\frac{n_{tx} + n_{cntrl}}{n_{tx} * n_{cntrl}}}$$

Choosing between Using Author-Reported or Calculated Effect Sizes

In cases when both author-reported effect sizes and calculated effect sizes were available for an outcome, we opted for consistency in using either reported or calculated effect sizes for all outcomes in a report, if possible. For example, if a particular report provided means and standard deviations for 20 outcomes that we used to calculate effect sizes, and also reported viable effect size estimates for 10 of those outcomes, we opted to use our calculated effect sizes for all outcomes because these were available consistently.

In cases when within-report consistency was not an issue, we then checked for differences in reported effect sizes and calculated effect sizes. If the difference in estimates was less than 1 *SE* for all effect sizes within a report, we opted to use the reported effect size because this estimate was, presumably, more precise if authors incorporated controls for baseline covariates or other relevant covariates in their estimations. If the difference in estimates was greater than 1 SE for any outcome within a report, the first author checked whether issues related to valence (see next section) may have driven differences in the final reported and calculated effects. The first author also determined whether there were any issues (e.g., longitudinal effects were modeled linearly in a growth curve model, interaction terms were included in the model, etc.) in the estimation strategies used to calculate the author-reported effect size that the coders missed in the coding process (i.e., only "viable" effect sizes should have been coded). The first author reviewed decisions with one of the study PIs to arrive at final determinations about whether to use the reported or calculated effect sizes. So long as there were no estimation issues with the reported effect sizes, these were used with the assumption that such effects should be more precise due to the inclusion of covariates when modeling the estimates.

Calculating Standard Errors for Odds Ratios, Log Odds Ratios, Proportions, and Percentages

To calculate standard errors for effect sizes derived from odds ratios, log odds ratios, proportions, and percentages, we used the standard error equation presented in the manuscript, plugging in the effect size calculated using the methods detailed in Figure S1. These standard error estimates are likely slightly downwardly biased (we estimate by ~13-14%) as we were unable to use the variance associated with the original author-reported statistics (as suggested here by Hasselblad & Hedges, 1995) as this variance information was not consistently reported.

Determining Effect Size Valence

Unfortunately, we failed to code for effect size valence in the primary coding process. Thus, a post-coding process was initiated to identify the valence of each effect size included in the meta-analysis. For each effect size, the first author and a study PI independently determined whether each effect size should be multiplied by 1 or -1 to capture that a higher score on the construct was positive (e.g., math scores) or negative (e.g., depressive symptoms), respectively. With the addition of the second study PI, the team reviewed all discrepant cases and resolved discrepancies. For the effect sizes that the team could not reach a resolution on, two research assistants, at least at the Masters-student level, reviewed each case by returning to the respective report and gathering evidence for a valence determination. The first author reviewed these cases and made final determinations. The study PIs were consulted for complicated cases.

Calculated effect sizes were multiplied by the valence. In the case that a reported effect size was used in analyses, however, an additional round of valence coding was initiated to identify whether the reported effect sizes were already re-valanced (e.g., if a study found reduced behavioral problems (lower mean), they reported a positive treatment impact), or whether effect sizes were presented as expected given their measure valence (i.e., if a study found reduced behavioral problems (lower mean), they reported a negative treatment impact). Three Masters-level research assistants reviewed all of the reported effect sizes that were suspected to have a high likelihood of valence-related issues (e.g., social-emotional outcomes). Two Masters-level research assistants reviewed all of the reported effect sizes that were not likely to have valence-related issues (e.g., cognitive outcomes). The first author reviewed these cases and made final determinations.

Negative Post-test Effect Sizes

As an additional check, the team reviewed the valence of outcomes for which the post-test effect size was negative and statistically significant after valence adjustments were made. Given the unlikelihood that treatments produced a negative, statistically significant post-test impact, we hoped that this check would catch errors in valence coding. There were 57 cases of statistically significant, negative post-test effects. Either two Masters-level research assistants and one Ph.D.-student-level research assistant, or the first-author and one Ph.D.-student-level research assistant, reviewed these cases. For each case, the reviewers indicated when valence should be re-coded and effect sizes should be adjusted, accordingly. The first author reviewed their determinations and resolved discrepancies as needed. From this additional checking process, 7 cases were identified as needing valence re-coding and were re-coded.

Results Presented for Subsamples & Multiple Treatment Groups

Notably, there were cases when statistics were reported separately for different sub-samples within a study (e.g., for boys and girls, for “low-risk” and “high-risk” participants, etc.). For these cases, we derived a main treatment effect by averaging the effect size estimates for each group, weighted by the group sample size. The same weighted averaging was used for standard errors and *p* values. Critically, if the treatment effect was only reported for one sub-sample (e.g., only boys, only “low-risk” participants, etc.), then the effects were dropped from the meta-analysis so that each estimate in the sample represented a main treatment impact of the original random assignment to treatment or control.

Results were also commonly reported for multiple treatment groups formed via random assignment within a study. We opted to leave effect sizes presented separately by treatment group when possible since the effects reflected experimental treatment impacts. However, there were some instances when effect sizes were reported for each treatment group separately at earlier assessment waves (e.g., pre-test, post-test, 6-12-month follow-up), and in aggregated form at later assessment waves (e.g., 3-year follow-up). In these cases, treatment-specific effect sizes, standard errors, and *p* values were averaged to form an average treatment effect that could be investigated in alignment with the effect sizes from later assessment waves.

Additional Analytic Details

Sensitivity Analyses

Below we detail additional analytic details for two sensitivity analyses presented in the manuscript. The first pertains to an analysis that adjusted standard errors to account for cluster-based randomization, and the second is for an analysis that removed effect sizes calculated using a heavy reliance on estimation.

Cluster-Related Standard Error Adjustment

We adjusted standard errors to account for clustering concerns. To do so, we multiplied the standard errors by the square root of the variance inflation factor, acknowledging that in some cases such an adjustment *would* be appropriate (i.e., when cluster adjustments were not yet made), and in other cases the adjustment would be too conservative (i.e., when cluster adjustments were already made and/or when pre-test covariates were accounted for in calculating treatment impacts). The VIF was calculated as follows:

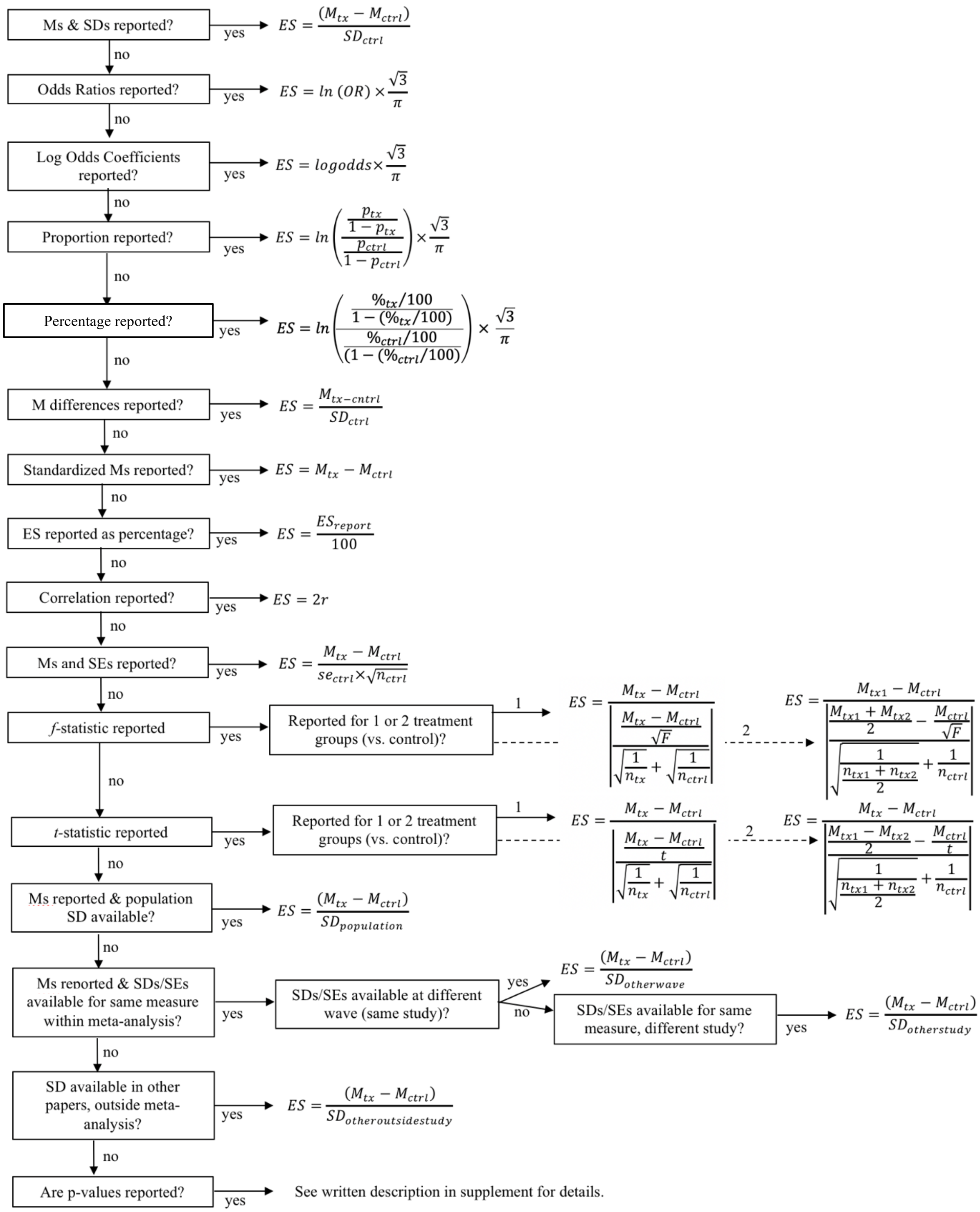
$$VIF = 1 + ICC * (m - 1)$$

where ICC = 0.10 and m represented the number of participants within a cluster. Because we did not code for how many clusters were randomly assigned, we assumed 20 clusters.

Removing "Estimated" Effects

We performed an analysis in which we removed effect sizes that required more-than-typical estimation in the calculation process (see Table S5, Column 7). First, we dropped effect sizes calculated for dichotomous outcomes because the conversion from odds to standard deviation units depends on the distribution underlying the dichotomous outcome, which may violate normality in some cases. This included effect sizes calculated through transforming raw statistics reported as percentages and proportions as well as through transforming effect sizes reported as Odds Ratios and Log Odds Coefficients. Second, we dropped effect sizes that were calculated from imprecise *p* values (see "Calculating ES using P values"). Finally, we dropped cases in which standard deviations were estimated from: a) population level SDs, 2) SDs reported for the same measure within our meta-analytic, or 3) SDs reported in other reports outside of the meta-analytic sample. Effect sizes were dropped prior to forming the aligned analytic groupings for use in analyses.

Figure S1
Effect Size Calculation Flow Chart



Note. This flow chart details the formulas used to calculate effect sizes and the decision-making process for deciding which calculation to use. Additional details relevant to this process, such as adjustments to standard errors when these were estimated in models controlling for pre-test scores, and the procedure used to calculate effect sizes from *p* values (if no other information was provided) are included in the supplemental text.

Table S1
Included Interventions Targeting Social-Emotional Skills, Cognitive Skills, or Both (“Broad”)

Social-Emotional Interventions	
	All Stars
	Aussie Optimism Program
	Beyondblue Schools Research Initiative
	Childhood Depression Prevention
	Cognitive-Behavioral Approach to Drug Abuse Prevention ~ Both Implementation Arms
	Conflict-Resolution Training Program
	Ease of Handling Social Aspects in Everyday of Life Training
	Emotional Intelligence Training Program
	Head Start REDI~ Incredible Years Teacher Training Group
	Head Start REDI~ Promoting Alternative Thinking Strategies Group
	Head Start REDI~ Tools of the Mind Group
	LARS & LISA (Lust An Realistischer Sicht & Leichtigkeit Im Sozialen Alltag)
	Learn Young, Learn Fair
	Life Skills Training
	Life Skills Training in Minority Youth
	Linking the Interests of Families and Teachers
	Optimism and Lifeskills Program
	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group
	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group
	Penn Resiliency Program #1
	Penn Resiliency Program #2~ Penn Enhancement Group
	Penn Resiliency Program #2~ Penn Resiliency Group
	Positive Thinking Program
	Problem Solving for Life
	Resourceful Adolescence Program (Kiwi)
	Responding in Peaceful and Positive Ways (Seventh Grade)
	Roots of Empathy
	Safe Dates
	Strengthening Families Program & Life Skills Training #1~ Both
	Strengthening Families Program & Life Skills Training #1~ LST Only
	Strengthening Families Program and Life Skills Training #2
	Think Smart
	Tools for Getting Along
	Universal School-based Mental Health Intervention~ Psychologist-Led Group
	Universal School-based Mental Health Intervention~ Teacher-Led Group
	Unplugged~ All Groups
Cognitive Interventions	
	Building Blocks
	Building Blocks plus TRIAD Follow-Through Group
	Classroom and At-Home Preschool Interventions
	Code-Oriented Reading Instruction
	Cogmed Working Memory Training
	Computer-Assisted Blending Skill Training
	Computer-Assisted Learning Program
	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group
	Computer-Assisted Reading Intervention~ Read, Write, and Type Group
	Computer-Assisted Remedial Reading Intervention
	Dialogic Reading #1~ School Reading Group
	Dialogic Reading #1~ School plus Home Reading Group
	Dialogic Reading #2
	English Reading Intervention for English Language Learners
	Explicit Phonological Awareness Instruction
	Home-Based Dyslexia Prevention

	Living Letters #2~ Combined Text Comprehension and Oral Language Group
	Living Letters #2~ Oral Language Group
	Living Letters #2~ Text Comprehension Group
	Multi-Component Reading Remediation~ PHAB + RAVE-O
	Multi-Component Reading Remediation~ PHAB + WIST
	Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group
	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension
	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group
	Phonics-Based Instruction for First Graders
	Phonological/Early Reading Skills
	Read Well Kindergarten
	Reading Recovery
	Reading Remediation
	Reading Remediation for Children with Reading Disorders
	Reading with Rhyme, Reading with Phoneme~ Reading with Phoneme Group
	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme Group
	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme and Phoneme Group
	SEARCH Screening Test and TEACH Tutoring~ Phonetic Group
	SEARCH Screening Test and TEACH Tutoring~ TEACH Tutoring Group
	Spanish Reading Intervention for English Language Learners
	Supplemental Phonics-based Instruction
	Swedish Phonics-based Intervention
	Teacher Responsivity Education
Broad Interventions (Both Cognitive and Social-Emotional)	
	Abecedarian~ All Pre-K (tx + tx; tx + cntrl) Group
	Classroom-Centered & School-Family Partnership~ Classroom Group
	Classroom-Centered & School-Family Partnership~ Family-School Group
	Head Start Impact Study
	Head Start Research-Based, Developmentally-Informed Program
	Infant Health and Development Program
	Parent Training for Teenage Moms
	Perry Preschool
	Supplemental Reading Instruction
	Tennessee Pre-K
	The Early Training Project~ 3-year and 2-year Intervention Group

Note. Each intervention is presented above and classified as a treatment targeting just social-emotional skills, just cognitive skills, or both (“broad”).

Table S2
Construct Categorization Examples

	Example Construct(s)	Example Measure(s)
Cognitive		
Achievement Composite	Pre-Academic Achievement	Composite of Woodcock Johnson subscales (letter-word identification, spelling, and applied problems)
General Cognition	IQ Verbal Short-term Memory	Stanford-Binet Intelligence Scale Automated Working Memory Assessment
Language and Literacy	Vocabulary Auditory-Vocal Association	Peabody Picture Vocabulary Test Illinois Test of Psycholinguistic Abilities
Math	Arithmetic Calculation	Wechsler Woodcock Johnson
Other Academic Abilities	Science Social Studies	Tennessee Comprehensive Assessment Program Stanford Achievement Tests
Social-Emotional		
Crime	Lifetime Violent Arrests Convictions	Study-created measures
Externalizing Behaviors	Aggressive Behaviors Disruptive Behavior	Child Behavior Checklist Finn Disruptive Behavior Scale
Internalizing Symptoms	Anxiety Depression	Children’s Manifest Anxiety Scale Child Depression Inventory
General Social-Emotional Skills	Social Skills Attributional Style	Social Skills Rating Scale Children’s Attributional Style Questionnaire
Substance Use	Alcohol Consumption Anti-Marijuana-Use Attitudes	Study-created measure Teenager’s Self Test

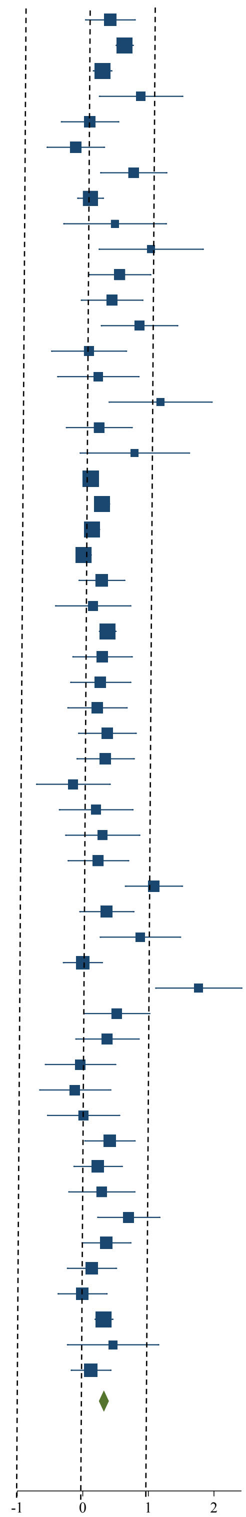
Note. Examples of the constructs and measures that were categorized in each of the construct categories.

Figure S2

Forest Plot: Average Cognitive Outcomes at Post-test (unweighted)

Study	Effect size with 95% CI	Weight (%)
Abecedarian~ All Pre-K (tx + tx; tx + cntrl) Group (N = 111)	0.42 [0.04, 0.79]	2.00
Building Blocks (N = 834)	0.64 [0.51, 0.76]	3.59
Building Blocks plus TRIAD Follow-Through Group (N = 741)	0.30 [0.16, 0.44]	3.52
Classroom and At-Home Preschool Interventions (N = 49)	0.89 [0.25, 1.52]	1.05
Classroom-Centered & School-Family Partnership~ Classroom Group (N = 820)	0.11 [-0.33, 0.54]	1.72
Classroom-Centered & School-Family Partnership~ Family-School Group (N = 818)	-0.11 [-0.54, 0.33]	1.72
Code-Oriented Reading Instruction (N = 67)	0.77 [0.27, 1.28]	1.44
Cogmed Working Memory Training (N = 414)	0.12 [-0.08, 0.31]	3.17
Computer-Assisted Blending Skill Training (N = 53)	0.49 [-0.29, 1.27]	0.76
Computer-Assisted Learning Program (N = 28)	1.04 [0.25, 1.83]	0.74
Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group (N = 74)	0.56 [0.10, 1.03]	1.58
Computer-Assisted Reading Intervention~ Read, Write, and Type Group (N = 73)	0.44 [-0.02, 0.91]	1.58
Computer-Assisted Remedial Reading Intervention (N = 50)	0.86 [0.28, 1.44]	1.19
Dialogic Reading #1~ School Reading Group (N = 48)	0.09 [-0.48, 0.66]	1.22
Dialogic Reading #1~ School plus Home Reading Group (N = 41)	0.23 [-0.38, 0.85]	1.09
Dialogic Reading #2 (N = 30)	1.18 [0.40, 1.97]	0.75
English Reading Intervention for English Language Learners (N = 93)	0.25 [-0.25, 0.75]	1.46
Explicit Phonological Awareness Instruction (N = 24)	0.79 [-0.04, 1.62]	0.68
Head Start Impact Study (N = 3933)	0.12 [0.03, 0.21]	3.78
Head Start REDI~ Incredible Years Teacher Training Group (N = 1323)	0.29 [0.18, 0.40]	3.69
Head Start REDI~ Promoting Alternative Thinking Strategies Group (N = 1290)	0.14 [0.03, 0.25]	3.69
Head Start REDI~ Tools of the Mind Group (N = 1299)	0.01 [-0.10, 0.12]	3.69
Head Start Research-Based, Developmentally-Informed Program (N = 336)	0.29 [-0.06, 0.63]	2.17
Home-Based Dyslexia Prevention (N = 48)	0.15 [-0.42, 0.73]	1.22
Infant Health and Development Program (N = 1082)	0.38 [0.25, 0.50]	3.61
Living Letters #2~ Combined Text Comprehension and Oral Language Group (N = 77)	0.30 [-0.15, 0.75]	1.65
Living Letters #2~ Oral Language Group (N = 75)	0.27 [-0.18, 0.73]	1.63
Living Letters #2~ Text Comprehension Group (N = 77)	0.22 [-0.23, 0.67]	1.66
Multi-Component Reading Remediation~ PHAB + RAVE-O (N = 127)	0.37 [-0.06, 0.81]	1.71
Multi-Component Reading Remediation~ PHAB + WIST (N = 129)	0.34 [-0.09, 0.78]	1.72
Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group (N = 50)	-0.15 [-0.71, 0.41]	1.25
Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension (N = 50)	0.20 [-0.36, 0.76]	1.25
Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group (N = 50)	0.30 [-0.26, 0.86]	1.24
Parent Training for Teenage Moms (N = 110)	0.23 [-0.22, 0.69]	1.61
Perry Preschool (N = 95)	1.08 [0.65, 1.51]	1.73
Phonics-Based Instruction for First Graders (N = 187)	0.36 [-0.05, 0.77]	1.84
Phonological/Early Reading Skills (N = 46)	0.88 [0.27, 1.48]	1.11
Read Well Kindergarten (N = 1427)	-0.00 [-0.30, 0.29]	2.49
Reading Recovery (N = 52)	1.77 [1.11, 2.42]	1.00
Reading Remediation (N = 66)	0.52 [0.02, 1.02]	1.46
Reading Remediation for Children with Reading Disorders (N = 56)	0.37 [-0.11, 0.85]	1.53
Reading with Rhyme, Reading with Phoneme~ Reading with Phoneme Group (N = 202)	-0.04 [-0.57, 0.50]	1.33
Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme Group (N = 195)	-0.12 [-0.66, 0.42]	1.32
Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme and Phoneme Group (N = 191)	0.01 [-0.54, 0.56]	1.29
SEARCH Screening Test and TEACH Tutoring~ Phonetic Group (N = 109)	0.41 [0.03, 0.79]	1.98
SEARCH Screening Test and TEACH Tutoring~ TEACH Tutoring Group (N = 116)	0.23 [-0.13, 0.60]	2.06
Spanish Reading Intervention for English Language Learners (N = 93)	0.29 [-0.21, 0.79]	1.45
Supplemental Phonics-based Instruction (N = 148)	0.70 [0.23, 1.17]	1.56
Supplemental Reading Instruction (N = 230)	0.36 [-0.01, 0.73]	2.04
Swedish Phonics-based Intervention (N = 112)	0.14 [-0.24, 0.51]	2.03
Teacher Responsivity Education (N = 229)	-0.01 [-0.38, 0.36]	2.04
Tennessee Pre-K (N = 1076)	0.32 [0.19, 0.45]	3.55
The Early Training Project~ 3-year and 2-year Intervention Group (N = 58)	0.46 [-0.23, 1.15]	0.92
Tools for Getting Along (N = 1162)	0.12 [-0.18, 0.42]	2.45
Overall	0.32 [0.25, 0.40]	

Heterogeneity: $\tau^2 = 0.04$, $I^2 = 67.78\%$, $H^2 = 3.10$

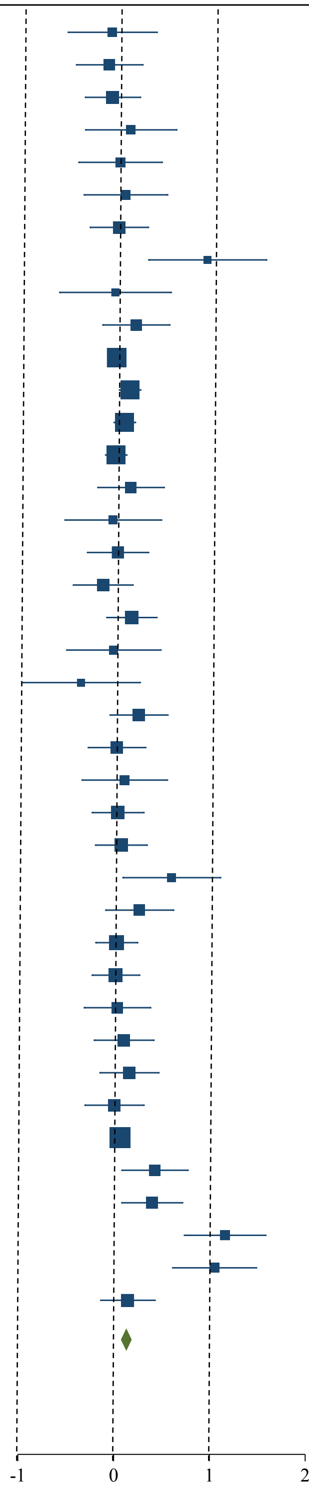


Note. Unweighted average post-test effect sizes for all cognitive outcomes included in the analytic sample, averaged within each treatment-control group contrast. Total sample sizes are indicated for each contrast.

Figure S3

Forest Plot: Average Social-Emotional Outcomes at Post-test (unweighted)

Study	Effect size with 95% CI	Weight (%)
All Stars (N = 1584)	-0.01 [-0.47, 0.45]	1.23
Aussie Optimism Program (N = 370)	-0.04 [-0.39, 0.30]	1.96
Beyondblue Schools Research Initiative (N = 4011)	-0.01 [-0.29, 0.28]	2.57
Childhood Depression Prevention (N = 96)	0.18 [-0.29, 0.66]	1.18
Classroom-Centered & School-Family Partnership~ Classroom Group (N = 820)	0.07 [-0.36, 0.50]	1.37
Classroom-Centered & School-Family Partnership~ Family-School Group (N = 818)	0.13 [-0.31, 0.56]	1.36
Cognitive-Behavioral Approach to Drug Abuse Prevention ~ Both Implementation Arms (N = 4464)	0.06 [-0.24, 0.36]	2.39
Conflict-Resolution Training Program (N = 54)	0.98 [0.37, 1.59]	0.75
Ease of Handling Social Aspects in Everyday of Life Training (N = 347)	0.02 [-0.56, 0.60]	0.83
Emotional Intelligence Training Program (N = 479)	0.24 [-0.11, 0.58]	1.94
Head Start Impact Study (N = 3920)	0.03 [-0.06, 0.12]	6.49
Head Start REDI~ Incredible Years Teacher Training Group (N = 1323)	0.17 [0.06, 0.28]	6.05
Head Start REDI~ Promoting Alternative Thinking Strategies Group (N = 1290)	0.11 [0.01, 0.22]	6.02
Head Start REDI~ Tools of the Mind Group (N = 1299)	0.03 [-0.08, 0.13]	6.03
Head Start Research-Based, Developmentally-Informed Program (N = 339)	0.18 [-0.16, 0.52]	1.97
LARS & LISA (Lust An Realistischer Sicht & Leichtigkeit Im Sozialen Alltag) (N = 289)	-0.01 [-0.51, 0.50]	1.07
Learn Young, Learn Fair (N = 1359)	0.04 [-0.27, 0.36]	2.23
Life Skills Training (N = 938)	-0.11 [-0.42, 0.20]	2.30
Life Skills Training in Minority Youth (N = 2800)	0.19 [-0.07, 0.45]	2.91
Linking the Interests of Families and Teachers (N = 348)	0.00 [-0.49, 0.49]	1.11
Optimism and Lifeskills Program (N = 42)	-0.34 [-0.95, 0.27]	0.75
Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group (N = 711)	0.26 [-0.04, 0.56]	2.40
Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group (N = 711)	0.03 [-0.26, 0.33]	2.42
Penn Resiliency Program #1 (N = 145)	0.11 [-0.33, 0.56]	1.32
Penn Resiliency Program #2~ Penn Enhancement Group (N = 425)	0.05 [-0.22, 0.31]	2.79
Penn Resiliency Program #2~ Penn Resiliency Group (N = 427)	0.08 [-0.19, 0.35]	2.80
Positive Thinking Program (N = 118)	0.61 [0.10, 1.11]	1.05
Problem Solving for Life (N = 1246)	0.27 [-0.08, 0.62]	1.90
Resourceful Adolescence Program (Kiwi) (N = 332)	0.03 [-0.18, 0.25]	3.61
Responding in Peaceful and Positive Ways (Seventh Grade) (N = 350)	0.02 [-0.22, 0.27]	3.11
Roots of Empathy (N = 352)	0.04 [-0.31, 0.38]	1.97
Safe Dates (N = 1294)	0.11 [-0.20, 0.42]	2.30
Strengthening Families Program & Life Skills Training #1~ Both (N = 869)	0.16 [-0.14, 0.47]	2.34
Strengthening Families Program & Life Skills Training #1~ LST Only (N = 883)	0.01 [-0.30, 0.31]	2.34
Strengthening Families Program and Life Skills Training #2 (N = 10170)	0.07 [0.03, 0.11]	7.54
Think Smart (N = 401)	0.43 [0.09, 0.77]	1.98
Tools for Getting Along (N = 1149)	0.40 [0.09, 0.72]	2.24
Universal School-based Mental Health Intervention~ Psychologist-Led Group (N = 200)	1.16 [0.74, 1.59]	1.42
Universal School-based Mental Health Intervention~ Teacher-Led Group (N = 176)	1.05 [0.62, 1.49]	1.36
Unplugged~ All Groups (N = 6370)	0.15 [-0.13, 0.43]	2.62
	0.13 [0.08, 0.19]	



Note. Unweighted average post-test effect sizes for all social-emotional outcomes included in the analytic sample, averaged within each treatment-control group contrast. Total sample sizes are indicated for each contrast.

Table S3

Longitudinal Persistence Rates by Outcome for Long-Term Follow-up Greater than 2 Years after Post-test (β(se))

	RE, weighted (1)
> 2 years Follow-up	
Post-test Effect	0.07
	(0.12)
Soc (vs. cog) Outcome	-0.03
	(0.09)
Post-test x Soc (vs. cog) Outcome	0.03
	(0.53)
Constant	0.10
	(0.06)
Sample size (study/intervention/outcome)	18 / 21 / 56
$\tau_{intercept}$ (null = .10)	.02
τ_{slope}	.27
I^2 (null model= 0.00%)	0.00%

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

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 post-test and follow-up impacts collected for the same construct using the same measure, subscales, and reporter at post-test and at least one follow-up assessment within an intervention. As such, each regression tests the association between post-test and follow-up effect sizes for the same measure. Model parameters 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). Sample sizes are reported for studies, interventions, and outcomes. Negative I^2 values were rounded to zero.

Table S4
Robustness Checks Probing Intervention-Related Differences

	Fixed Effects (1)	Key Covariates (2)	Intensity (< 200 hours) (3)	Intensity (< 100 hours) (4)	All Covariates (5)
Panel A: 6- to 12-month Follow-up					
Post-test Effect	0.44*	0.54***	0.48***	0.50***	0.41***
	(0.07)	(0.06)	(0.06)	(0.07)	(0.05)
Soc (vs. cog) Outcome	0.06	0.05+	0.03	0.05+	0.03
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Post-test x Soc (vs. cog) Outcome	-0.01	-0.13	-0.01	-0.05	0.07
	(0.09)	(0.08)	(0.09)	(0.1)	(0.09)
Constant	--	0.04+	0.05*	0.03	0.05+
		(0.02)	(0.02)	(0.03)	(0.02)
Sample size (study/int/outcome)	60 / 77 / 420	60 / 77 / 420	57 / 74 / 412	53 / 68 / 360	60 / 77 / 420
$\tau_{intercept}$	--	.11	.11	.11	.11
τ_{slope}	--	.21	.25	.26	.18
I^2	.62%	31.65%	32.98%	36.28%	25.86%
Panel B: 1- to 2-year Follow-up					
Post-test Effect	0.35	0.50	0.40+	0.40+	0.06
	(0.25)	(0.30)	(0.17)	(0.17)	(0.27)
Soc (vs. cog) Outcome	0.07	0.08	0.07	0.07	0.03
	(0.03)	(0.09)	(0.04)	(0.04)	(0.02)
Post-test x Soc (vs. cog) Outcome	-0.50	-0.63	-0.43	-0.43	-0.06
	(0.26)	(0.52)	(0.26)	(0.26)	(0.29)
Constant	--	-0.02	-0.02	-0.02	0.08
		(0.06)	(0.03)	(0.03)	(0.08)
Sample size (study/int/outcome)	23 / 24 / 87	23 / 24 / 87	21 / 22 / 81	21 / 22 / 81	23 / 24 / 87
$\tau_{intercept}$	--	.06	.03	.03	.19
τ_{slope}	--	.30	.29	.29	.39
I^2	0.00%	8.42%	3.69%	3.69%	1.79%

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. The unit of analysis is "aligned groupings" of post-test and follow-up impacts collected for the same construct using the same measure, subscales, and reporter at post-test and at least one follow-up assessment within an intervention. Column 1 presents a fixed effects meta-analytic model with an econometric fixed effect for study and weights. Column 2 presents the primary random effects meta-analytic model with a covariate for participant age at post-test and duration of intervention in months, and the interactions between these variables and post-test effect size. Columns 3 and 4 present the primary model with analytic samples limited to interventions with fewer than 200, or fewer than 100 hours of intervention. Column 5 presents the primary model with covariates for several intervention characteristics, and the interactions between these and post-test effect size. All models use cluster-robust standard errors (clustered at the study level). All covariates were standardized. Sample sizes are reported for studies, interventions ("int"), and outcomes. Negative I^2 values were rounded to zero.

Table S5

Robustness Checks Probing Alternate Analytic Approaches and Concerns

	CHE Model (1)	Meta-level RE (2)	Cluster Covariate (3)	No Cluster-Adj SE (4)	Unweighted Model (5)	No Neg Post-tests (6)	No Est. Effects (7)	Post-test SDs Used (8)
Panel A: 6- to 12-month Follow-up								
Post-test Effect	0.45***	0.37**	0.45***	0.43***	0.42*	0.47***	0.47***	0.62***
	(0.06)	(0.12)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)	(0.08)
Soc (vs. cog) Outcome	0.04	0.02	0.03	0.03	-0.03	0.04+	0.02	0.05+
	(0.02)	(0.02)	(0.01)	(0.02)	(0.06)	(0.02)	(0.02)	(0.02)
Post-test x Soc (vs. cog) Outcome	-0.01	0.08	0.02	0.07	0.26	-0.01	0.03	-0.13
	(0.09)	(0.16)	(0.09)	(0.10)	(0.28)	(0.12)	(0.10)	(0.11)
Constant	0.01	0.06*	0.06	0.05*	0.08	0.05	0.04	0.03
	(0.02)	(0.03)	(0.04)	(0.02)	(0.05)	(0.03)	(0.02)	(0.02)
Sample size (study/int/outcome)	60 / 77 / 420	60 / 77 / 420	60 / 77 / 420	60 / 77 / 420	60 / 77 / 420	59 / 75 / 332	57 / 74 / 366	60 / 77 / 420
$\tau_{intercept}$	.00 / .00 / .09	.01 / .11	.11	.14	.11	.15	.09	.11
τ_{slope}	.18	.23 / .19	.24	.28	.24	.27	.25	.34
I^2	53.73%	32.47%	32.52%	66.01%	32.47%	37.84%	25.04%	45.84%
Panel B: 1- to 2-year Follow-up								
Post-test Effect	0.34*	0.35*	0.34*	0.37+	0.45	0.41*	0.34*	0.37*
	(0.12)	(0.14)	(0.14)	(0.16)	(0.22)	(0.15)	(0.13)	(0.15)
Soc (vs. cog) Outcome	0.08	0.07+	0.05	0.07	0.03	0.12	0.05	0.07
	(0.04)	(0.03)	(0.02)	(0.04)	(0.08)	(0.06)	(0.03)	(0.04)
Post-test x Soc (vs. cog) Outcome	-0.46+	-0.39+	-0.40	-0.25	-0.76	-0.65+	-0.25	-0.40
	(0.23)	(0.23)	(0.24)	(0.24)	(0.33)	(0.29)	(0.19)	(0.25)
Constant	-0.02	-0.02	-0.02	-0.01	0.03	-0.04	-0.03	-0.02
	(0.04)	(0.03)	(0.03)	(0.03)	(0.07)	(0.04)	(0.04)	(0.03)
Sample size (study/int/outcome)	23 / 24 / 87	23 / 24 / 87	23 / 24 / 87	23 / 24 / 87	23 / 24 / 87	23 / 24 / 66	19 / 20 / 67	23 / 24 / 87
$\tau_{intercept}$	.05/.00 /.06	.01 / .03	.04	.03	.03	.03	.03	.03
τ_{slope}	.28	.15 / .23	.31	.42	.27	.26	.19	.30
I^2	42.69%	5.21%	6.06%	72.48%	5.21%	3.34%	0.00%	9.05%

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. Column 1 presents the results for the Correlated-Hierarchical Model, with rho set to .60 and random effects with treatment-control contrast groups nested within trials (presented as: study / intervention / outcome). Column 2 presents the primary model with a study-level random effects nested with origin meta-analysis random effects (presented as: meta-analysis / study). Column 3 presents the primary model with a covariate for whether cluster randomization was used. Column 4 presents the primary model with no standard error adjustment for clustering. Column 5 presents the primary model with no weighting by effect precision. Column 6 presents the primary model with negative post-test effects dropped. Column 7 presents the primary model with effect sizes that relied on estimation in the calculation process dropped. Column 8 presents the primary model with effect sizes calculated using post-test, instead of concurrent, control group standard deviations (and a scaling factor when post-test standard deviations were unavailable). All models use cluster-robust standard errors (clustered at the study level). Sample sizes are reported for trials, interventions ("int"), and outcomes. Negative I^2 values were rounded to zero.

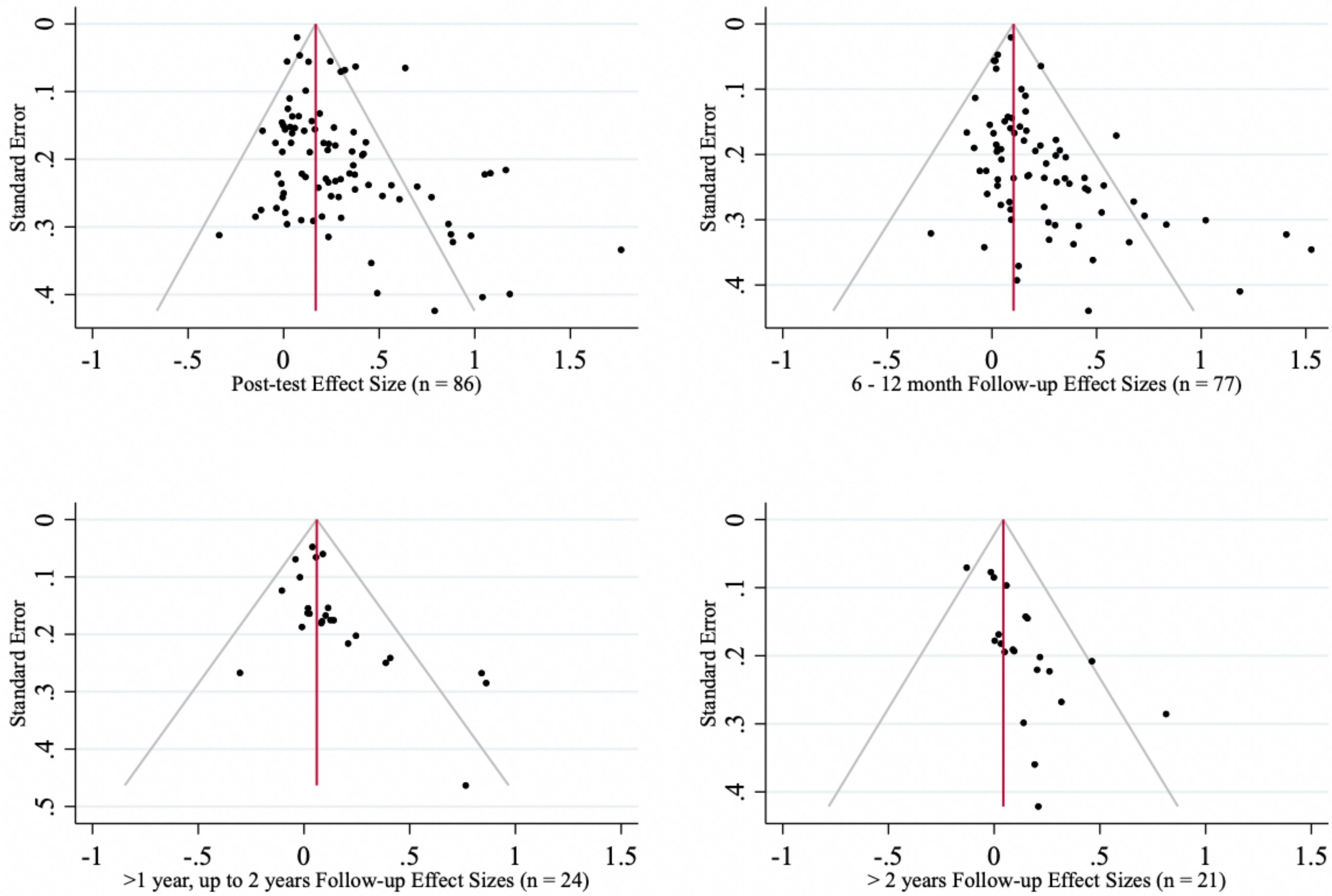
Table S6
Unweighted Average Effect Sizes Across Post-test and Follow-up Assessment Waves

	All Outcomes		Social-Emotional Outcomes		Cognitive Outcomes	
	M (SE) (1)	n	*M (SE)* (2)	n	*M (SE)* (3)	n
Post-test	0.28 (0.04)***	450	0.17 (0.05)**	214	0.39 (0.06)***	236
6 months to 1 year	0.21 (0.04)***	420	0.18 (0.07)*	197	0.25 (0.04)***	223
> 1 year, up to 2 years	0.10 (0.04)*	87	0.04 (0.03)	59	0.24 (0.07)*	28
> 2 years, up to 3 years	0.14 (0.05)*	31	0.13 (0.06)+	22	0.16 (0.06)	9
> 3 years, up to 4 years	0.17 (0.03)*	29	0.08 (0.03)	9	0.20 (0.02)*	20
> 4 years	0.22 (0.05)*	13			0.22 (0.05)*	13

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. “ES” = Effect size. 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 analytic sample was constituted by “aligned groupings” that included a post-test and at least one follow-up effect size for the same construct measured using the same measure, subscales, and reporter within a treatment-control contrast. Average effect sizes were estimated using a random effects meta-analytic model that included a random effect for study, no weights for impact precision, and cluster-robust standard errors (clustered at the study level).

Figure S4
Funnel Plot



Note. Gray lines represent 95% confidence intervals. Each coordinate represents the average effect size for each treatment-control group contrast contributing aligned constructs to the analytic sample (for which the same construct was measured using the same measure at post-test and at least one follow-up assessment).

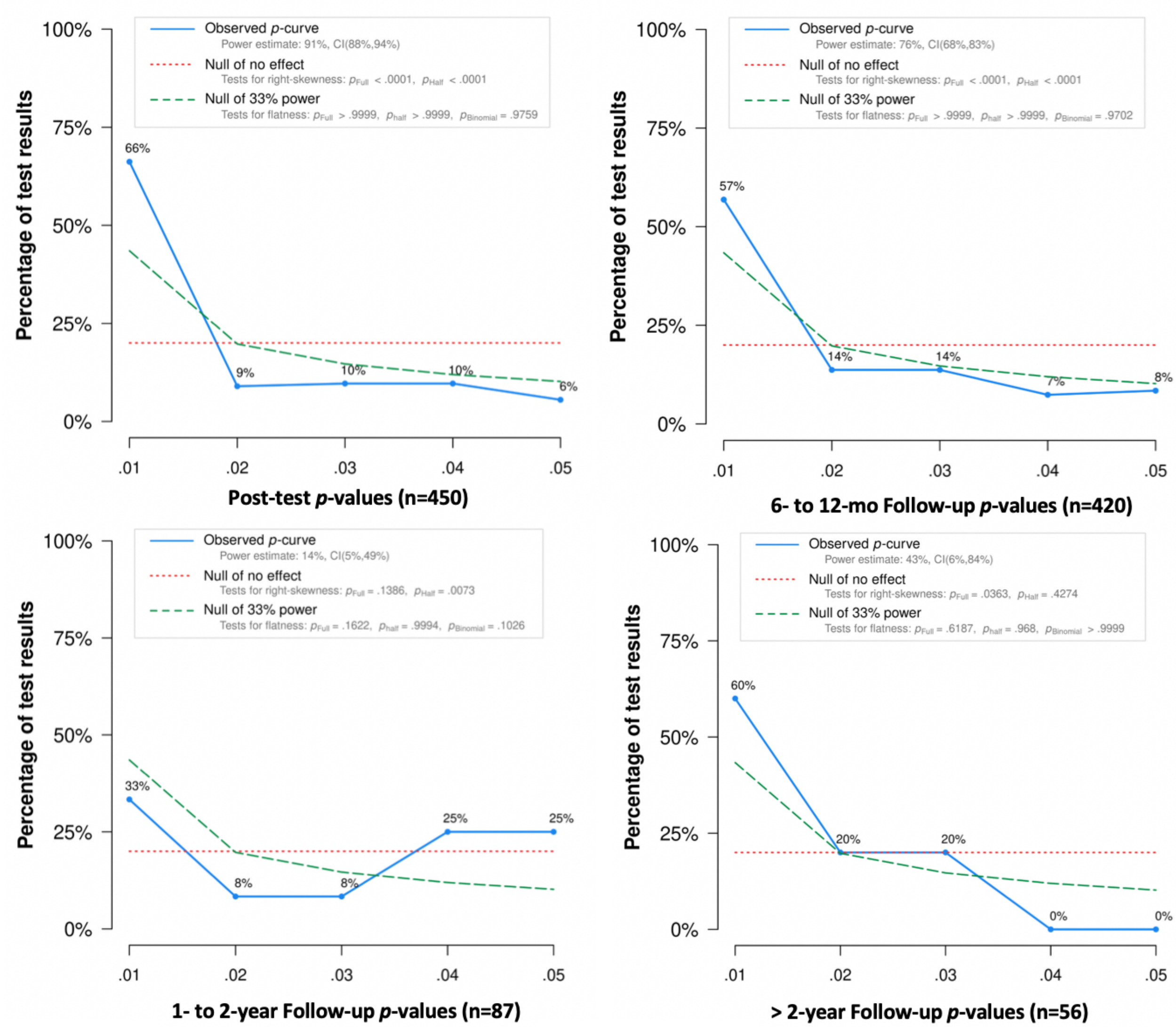
Table S7
PEESE Test (β(se))

	PEESE Test (1)
Panel A: 6- to 12-months Follow-up	
Post-test Effect	0.43***
	(0.06)
Soc (vs. cog) Outcome	0.03+
	(0.01)
Post-test x Soc (vs. cog) Outcome	0.02
	(0.09)
Standard Error	0.83*
	(0.30)
Constant	-0.11+
	(0.05)
Sample size (study/intervention/outcome)	60 / 77 / 420
$\tau_{intercept}$	.11
τ_{slope}	.22
I^2	30.38%
Panel B: > 1 year, up to 2 years Follow-up	
Post-test Effect	0.27*
	(0.07)
Soc (vs. cog) Outcome	0.03
	(0.02)
Post-test x Soc (vs. cog) Outcome	-0.50+
	(0.23)
Standard Error	1.09+
	(0.46)
Constant	-0.12
	(0.07)
Sample size (study/intervention/outcome)	23 / 24 / 87
$\tau_{intercept}$	.08
τ_{slope}	.26
I^2	2.88%

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. This table presents alternate grouping approaches. “Soc (vs. cog) Outcome” is a dummy variable for outcome type (0 = cognitive outcome, 1= social-emotional outcome). Parameters were estimated using a random effects meta-analytic model that included a random effect for trial, weights, and cluster-robust standard errors (clustered at the trial level).

Figure S5
Relative Frequency of *p* values < .05



Note. Each figure contains all of the *p* values included in the analytic sample at post-test and each follow-up wave. Post-tests and follow-up impacts are taken from the larger analytic sample that was constituted of "aligned groupings" in which the same construct was measured using the same measure, subscales, and reporter at post-test and at least one follow-up assessment within a study. *P* curve figures were created on p-curve.com (Simonsohn, Nelson, & Simmons, 2015).

Table S8
Vevea-Hedges Selection Model

Assessment Wave	Unadjusted ES Estimate (1)	Adjusted ES Estimate (2)	Adjusted ES /Unadjusted ES (3)
Post-test	0.257	0.153	60%
6- to 12-mo Follow-up	0.166	0.107	64%
1- to 2-year Follow-up	0.050	0.034	68%

Note. This table reports unadjusted and adjusted effect sizes estimated using a weight-function model with cut-points and weights established by Vevea and Woods (2005). In this model, effects are weighted according to their likelihood of being published on account of the p value associated with the treatment impact. Weights were set as follows to reflect patterns of selective reporting if p values dictated publishing: effects associated with a $p < 0.01$ were set to have a weight of 1 (an assumption that 100% of effects of this statistical significance are published if selection biases are at play), $p < .05$ is set to .9, $p < .50$ is set to .70, and $p < 1$ is set to .50. Column 3 represents the percentage of the unadjusted effect size that is estimated in the adjusted effect size, suggesting the extent to which estimates may be biased in our sample.

Table S9
Longitudinal Persistence Rates by Outcome using Alternate Grouping Approaches (β(se))

	Alternate Grouping #1 (1)	Alternate Grouping #2 (2)
Panel A: 6- to 12-months Follow-up		
Post-test Effect	0.44***	0.52***
	(0.06)	(0.07)
Soc (vs. cog) Outcome	0.02	0.00
	(0.02)	(0.03)
Post-test x Soc (vs. cog) Outcome	0.03	0.24
	(0.09)	(0.22)
Constant	0.06*	0.01
	(0.02)	(0.02)
Sample size (study/int/outcome)	62 / 80 / 418	69 / 90 / 98
$\tau_{intercept}$	.11	.02
τ_{slope}	.22	.21
I^2	32.87%	0.00%
Panel B: > 1 year, up to 2 years Follow-up		
Post-test Effect	0.32+	0.39*
	(0.14)	(0.14)
Soc (vs. cog) Outcome	0.06	0.07
	(0.04)	(0.04)
Post-test x Soc (vs. cog) Outcome	-0.33	-0.67+
	(0.24)	(0.29)
Constant	-0.01	0.00
	(0.03)	(0.05)
Sample size (study/int/outcome)	24 / 25 / 87	29 / 32 / 33
$\tau_{intercept}$	.03	.06
τ_{slope}	.31	.17
I^2	6.22%	36.39%

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note. This table presents alternate grouping approaches. In (1), the analytic sample was constituted of the same "aligned groups," but measure, subscale, and reporter were allowed to vary across waves. In (2), the analytic sample was constituted of all social-emotional and all cognitive outcomes averaged together at each wave (e.g., charting the average of all social-emotional impacts at post-test to the average of all social-emotional impacts at follow-up). "Soc (vs. cog) Outcome" is a dummy variable for outcome type (0 = cognitive outcome, 1= social-emotional outcome). Parameters 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). Sample sizes are reported for trials, interventions ("int"), and outcomes. Negative I^2 values were rounded to zero.

Table S10
Average Post-test Effect Sizes for Effect Sizes Reported at Each Follow-up Assessment Wave

	All Outcomes			Social-Emotional Outcomes			Cognitive Outcomes		
	M (SE) (1)	M_{POST} (SE) (2)	n	*M (SE)* (3)	M_{POST} (SE) (4)	n	*M (SE)* (5)	M_{POST} (SE) (6)	n
Panel A: Weighted									
Post-test	0.29 (0.04)***	0.29 (0.04)***	450	0.14 (0.04)**	0.14 (0.04)**	214	0.40 (0.06)***	0.40 (0.06)***	236
6 months to 1 year	0.21 (0.03)***	0.30 (0.05)***	420	0.16 (0.05)**	0.15 (0.04)**	197	0.22 (0.04)***	0.41 (0.06)***	223
> 1 year, up to 2 years	0.07 (0.02)**	0.25 (0.06)***	87	0.05 (0.01)*	0.07 (0.04)+	59	0.16 (0.06)*	0.44 (0.09)***	28
> 2 years, up to 3 years	0.08 (0.05)	0.26 (0.09)*	31	0.11 (0.05)+	0.10 (0.03)*	22	0.05 (0.10)	0.52 (0.17)*	9
> 3 years, up to 4 years	0.10 (0.05)+	0.35 (0.10)**	29	0.09 (0.03)	0.15 (0.03)*	9	0.12 (0.08)	0.46 (0.13)*	20
> 4 years	0.14 (0.08)	0.49 (0.08)**	13				0.14 (0.08)	0.49 (0.08)**	13
Panel B: Unweighted									
Post-test	0.28 (0.04)***	0.28 (0.04)***	450	0.17 (0.05)**	0.17 (0.05)**	214	0.39 (0.06)***	0.39 (0.06)***	236
6 months to 1 year	0.21 (0.04)***	0.29 (0.04)***	420	0.18 (0.07)*	0.18 (0.05)**	197	0.25 (0.04)***	0.39 (0.06)***	223
> 1 year, up to 2 years	0.10 (0.04)*	0.20 (0.07)*	87	0.04 (0.03)	0.08 (0.05)	59	0.24 (0.07)*	0.46 (0.14)*	28
> 2 years, up to 3 years	0.14 (0.05)*	0.29 (0.16)	31	0.13 (0.06)+	0.10 (0.03)*	22	0.16 (0.06)	0.77 (0.28)	9
> 3 years, up to 4 years	0.17 (0.03)*	0.40 (0.14)+	29	0.08 (0.03)	0.15 (0.03)*	9	0.20 (0.02)*	0.51 (0.19)+	20
> 4 years	0.22 (0.05)*	0.51 (0.11)*	13				0.22 (0.05)*	0.51 (0.11)*	13

* p < 0.05, ** p <0.01, *** p<0.001
Note. "ES" = Effect size. 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 analytic sample was constituted by "aligned groupings" that included a post-test and at least one follow-up effect size for the same construct measured using the same measure, subscale, and reporter within a treatment-control contrast. In Columns 1, 3, and 5, average effect sizes were estimated using a random effects meta-analytic model that included a random effect for trial, weights, and cluster-robust standard errors (clustered at the trial level). To evaluate the possibility of selection into longer-run follow-up assessments, Columns 2, 4, and 6 present average post-test effects for the outcomes collected at each follow-up wave.

Table S11
Effect sizes and Standard Errors

Viable Group	Trial name~ Intervention group name	Construct	Skill type (1=soc; 0=cog)	ES (post-test)	SE (post-test)	ES (6-12 mo)	SE (6-12 mo)	ES (1-2 yr)	SE (1-2 yr)
392	Abecedarian~ All Pre-K (tx + tx; tx + cntrl) Group	iq	0	0.42	0.19				
449	All Stars	beliefs in conventional norms	1	0.01	0.24	0.05	0.24		
450	All Stars	bonding with school	1	-0.03	0.24	0.05	0.24		
451	All Stars	positive ideals	1	-0.02	0.24	0.00	0.24		
452	All Stars	strength of commitment	1	0.00	0.24	0.02	0.24		
24	Aussie Optimism Program	externalizing problems	1	0.24	0.19	0.15	0.20	0.11	0.21
25	Aussie Optimism Program	anxiety	1	-0.21	0.17	0.00	0.18	-0.19	0.18
26	Aussie Optimism Program	depressive symptoms	1	-0.14	0.17	-0.12	0.18	-0.05	0.18
27	Aussie Optimism Program	internalizing problems	1	0.27	0.19	0.17	0.20	0.11	0.21
28	Aussie Optimism Program	attributional style for negative events	1	-0.15	0.17	-0.13	0.18	-0.15	0.18
29	Aussie Optimism Program	attributional style for positive events	1	-0.12	0.17	0.01	0.18	-0.01	0.18
30	Aussie Optimism Program	poor social skills	1	-0.18	0.17	0.06	0.18	0.11	0.18
78	Beyondblue Schools Research Initiative	depression	1	-0.04	0.15	0.00	0.15	0.00	0.15
79	Beyondblue Schools Research Initiative	family social support	1	-0.05	0.15	-0.04	0.15	0.00	0.15
80	Beyondblue Schools Research Initiative	friends social support	1	-0.01	0.15	-0.08	0.15	-0.06	0.15
81	Beyondblue Schools Research Initiative	interpersonal competence	1	0.05	0.15	0.04	0.15	0.08	0.15
82	Beyondblue Schools Research Initiative	negative coping	1	-0.02	0.15	0.00	0.15	0.02	0.15
83	Beyondblue Schools Research Initiative	optimistic thinking style	1	0.04	0.15	0.04	0.15	0.08	0.15
84	Beyondblue Schools Research Initiative	positive coping	1	0.02	0.15	0.06	0.15	0.08	0.15
85	Beyondblue Schools Research Initiative	significant other social support	1	-0.05	0.15	-0.10	0.15	-0.04	0.15
393	Building Blocks	math achievement	0	0.64	0.07	0.23	0.06	0.09	0.06
394	Building Blocks plus TRIAD Follow-Through Group	math achievement	0	0.30	0.07				
295	Childhood Depression Prevention	depression	1	-0.15	0.24	-0.41	0.30		
296	Childhood Depression Prevention	depression	1	0.11	0.24	-0.15	0.27		
297	Childhood Depression Prevention	happiness	1	0.53	0.24	-0.52	0.31		
298	Childhood Depression Prevention	popularity	1	0.00	0.24	-0.47	0.30		
299	Childhood Depression Prevention	social skills	1	0.42	0.25	0.08	0.42		
311	Classroom and At-Home Preschool Interventions	iq	0	0.54	0.31	0.27	0.31		
312	Classroom and At-Home Preschool Interventions	visual perception	0	1.23	0.33	0.56	0.31		
304	Classroom-Centered & School-Family Partnership~ Classroom Group	reading achievement	0	0.09	0.22	-0.06	0.22		

306	Classroom-Centered & School-Family Partnership~ Classroom Group	math achievement	0	0.13	0.22	0.01	0.23		
308	Classroom-Centered & School-Family Partnership~ Classroom Group	classroom social adaptation	1	0.07	0.22	-0.12	0.22		
305	Classroom-Centered & School-Family Partnership~ Family-School Group	reading achievement	0	0.03	0.22	0.07	0.22		
307	Classroom-Centered & School-Family Partnership~ Family-School Group	math achievement	0	-0.25	0.22	-0.01	0.23		
309	Classroom-Centered & School-Family Partnership~ Family-School Group	classroom social adaptation	1	0.13	0.22	-0.14	0.22		
211	Code-Oriented Reading Instruction	comprehension	0	0.29	0.25	0.24	0.30		
212	Code-Oriented Reading Instruction	developmental spelling	0	0.65	0.25	-0.12	0.30		
213	Code-Oriented Reading Instruction	oral reading fluency	0	1.29	0.27	0.40	0.30		
214	Code-Oriented Reading Instruction	reading accuracy	0	1.16	0.26	0.36	0.30		
215	Code-Oriented Reading Instruction	reading efficacy	0	0.48	0.25	0.47	0.31		
218	Cogmed Working Memory Training	verbal short term memory	0	0.01	0.10	-0.03	0.10	-0.17	0.10
219	Cogmed Working Memory Training	verbal working memory	0	0.22	0.10	0.13	0.10	0.02	0.10
220	Cogmed Working Memory Training	visuospatial short term memory	0	0.37	0.10	0.53	0.10	0.19	0.10
221	Cogmed Working Memory Training	visuospatial working memory	0	-0.15	0.10	-0.06	0.10	-0.11	0.10
407	Cognitive-Behavioral Approach to Drug Abuse Prevention ~ Both Implementation Arms	drunkenness frequency	1	0.04	0.15				
408	Cognitive-Behavioral Approach to Drug Abuse Prevention ~ Both Implementation Arms	frequency of marijuana use	1	0.08	0.15				
216	Computer-Assisted Blending Skill Training	blending	0	0.68	0.40	0.44	0.39		
217	Computer-Assisted Blending Skill Training	segmentation	0	0.30	0.39	-0.20	0.39		
116	Computer-Assisted Learning Program	word reading- irregular words	0	1.28	0.41	1.22	0.41		
117	Computer-Assisted Learning Program	word recognition	0	0.80	0.39	1.15	0.41		
134	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	passage comprehension	0	0.47	0.24	0.41	0.24		
136	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	phonemic decoding efficiency	0	0.81	0.24	0.46	0.24		
138	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	phonological awareness- blending words	0	0.44	0.24	0.20	0.23		
140	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	phonological awareness- elision	0	0.61	0.24	0.39	0.23		
142	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	phonological awareness- segmenting words	0	0.87	0.24	0.54	0.24		
144	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	rapid digit naming	0	0.33	0.23	0.25	0.23		
146	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	rapid letter naming	0	0.00	0.23	0.67	0.24		
148	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	word analysis	0	0.95	0.25	0.63	0.24		
150	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	word efficiency	0	0.52	0.24	0.40	0.24		

152	Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group	word identification	0	0.64	0.24	0.47	0.24
133	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	passage comprehension	0	0.33	0.24	0.24	0.24
135	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	phonemic decoding efficiency	0	0.26	0.24	0.19	0.24
137	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	phonological awareness- blending words	0	0.70	0.24	0.07	0.23
139	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	phonological awareness- elision	0	0.28	0.24	0.18	0.24
141	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	phonological awareness- segmenting words	0	0.64	0.24	0.11	0.23
143	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	rapid digit naming	0	0.67	0.24	0.25	0.24
145	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	rapid letter naming	0	0.33	0.24	0.67	0.24
147	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	word analysis	0	0.59	0.24	0.24	0.24
149	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	word efficiency	0	0.22	0.24	0.28	0.24
151	Computer-Assisted Reading Intervention~ Read, Write, and Type Group	word identification	0	0.41	0.24	0.27	0.24
273	Computer-Assisted Remedial Reading Intervention	reading fluency	0	0.86	0.30	1.02	0.30
446	Conflict-Resolution Training Program	overall conflict resolution- “name calling” conflict	1	1.38	0.33	0.63	0.30
447	Conflict-Resolution Training Program	overall conflict resolution- “queuing” conflict	1	0.84	0.31	0.91	0.31
448	Conflict-Resolution Training Program	overall conflict resolution- “taking turns” conflict	1	0.73	0.30	0.96	0.31
374	Dialogic Reading #1~ School plus Home Reading Group	expressive vocabulary	0	0.43	0.32	0.18	0.37
376	Dialogic Reading #1~ School plus Home Reading Group	receptive vocabulary	0	0.24	0.31	0.01	0.37
378	Dialogic Reading #1~ School plus Home Reading Group	verbal fluency	0	0.03	0.31	0.20	0.37
373	Dialogic Reading #1~ School Reading Group	expressive vocabulary	0	0.18	0.29	0.18	0.34
375	Dialogic Reading #1~ School Reading Group	receptive vocabulary	0	0.13	0.29	-0.21	0.34
377	Dialogic Reading #1~ School Reading Group	verbal fluency	0	-0.03	0.29	-0.08	0.35
379	Dialogic Reading #2	expressive vocabulary	0	1.10	0.39	0.79	0.44
380	Dialogic Reading #2	receptive vocabulary	0	0.69	0.38	0.01	0.43
381	Dialogic Reading #2	verbal expressiveness	0	1.76	0.43	0.59	0.45
63	Ease of Handling Social Aspects in Everyday of Life Training	depression	1	0.10	0.30	0.37	0.30
64	Ease of Handling Social Aspects in Everyday of Life Training	dysfunctional thoughts	1	-0.21	0.29	0.23	0.30
65	Ease of Handling Social Aspects in Everyday of Life Training	frequency of seeking assistance from social network	1	0.22	0.29	-0.14	0.30
66	Ease of Handling Social Aspects in Everyday of Life Training	satisfaction with network’s support	1	0.16	0.30	0.17	0.30

67	Ease of Handling Social Aspects in Everyday of Life Training	size of social supports network	1	-0.18	0.30	-0.17	0.30		
35	Emotional Intelligence Training Program	anxiety	1	0.18	0.18	0.18	0.18		
36	Emotional Intelligence Training Program	depression	1	0.32	0.18	0.33	0.18		
37	Emotional Intelligence Training Program	sense of incapacity	1	0.26	0.18	0.35	0.18		
38	Emotional Intelligence Training Program	atypicality	1	0.38	0.18	0.36	0.18		
39	Emotional Intelligence Training Program	external locus	1	0.17	0.18	0.35	0.18		
40	Emotional Intelligence Training Program	mental health	1	0.20	0.18	0.19	0.18		
41	Emotional Intelligence Training Program	negative affect	1	0.04	0.18	0.34	0.18		
42	Emotional Intelligence Training Program	social stress	1	0.27	0.18	0.38	0.18		
43	Emotional Intelligence Training Program	somatization	1	0.29	0.18	0.27	0.18		
251	English Reading Intervention for English Language Learners	oral language composite- english	0	0.02	0.25	0.24	0.19		
252	English Reading Intervention for English Language Learners	oral language composite- spanish	0	-0.07	0.24	-0.13	0.19		
253	English Reading Intervention for English Language Learners	passage comprehension- english	0	0.37	0.25	0.31	0.19		
254	English Reading Intervention for English Language Learners	passage comprehension- spanish	0	0.23	0.25	0.28	0.19		
255	English Reading Intervention for English Language Learners	reading fluency- english	0	0.31	0.25	0.36	0.19		
256	English Reading Intervention for English Language Learners	reading fluency- spanish	0	0.04	0.25	-0.01	0.20		
257	English Reading Intervention for English Language Learners	word attack- english	0	0.67	0.27	0.45	0.19		
258	English Reading Intervention for English Language Learners	word attack- spanish	0	0.40	0.27	0.16	0.20		
222	Explicit Phonological Awareness Instruction	context-free word recognition	0	0.88	0.43			0.72	0.46
223	Explicit Phonological Awareness Instruction	word recognition accuracy	0	0.70	0.42			0.81	0.47
363	Head Start Impact Study	pre-academic skills	0	0.21	0.05	0.03	0.05	0.02	0.05
364	Head Start Impact Study	behavior problem	1	0.11	0.05	0.04	0.05	0.06	0.05
365	Head Start Impact Study	elison	0	0.08	0.05	0.06	0.05		
366	Head Start Impact Study	letter word identification- spanish	0	0.13	0.05	0.02	0.05	0.11	0.05
367	Head Start Impact Study	oral comprehension	0	-0.01	0.05	-0.01	0.05	0.03	0.05
368	Head Start Impact Study	receptive vocabulary	0	0.14	0.05	0.05	0.05	0.05	0.05
369	Head Start Impact Study	receptive vocabulary- spanish	0	0.17	0.05	-0.03	0.05	-0.05	0.05
370	Head Start Impact Study	positive parent/student relationships	1	0.05	0.05	0.05	0.05	0.03	0.05
371	Head Start Impact Study	social competencies	1	-0.03	0.05	-0.01	0.05	0.04	0.05
372	Head Start Impact Study	social skills and positive approaches to learning	1	-0.01	0.05	0.08	0.05	0.09	0.05
347	Head Start REDI~ Incredible Years Teacher Training Group	language and literacy	0	0.27	0.06	-0.02	0.06		

350	Head Start REDI~ Incredible Years Teacher Training Group	mathematical thinking	0	0.32	0.06	0.01	0.06		
353	Head Start REDI~ Incredible Years Teacher Training Group	general knowledge	0	0.29	0.06	-0.03	0.06		
356	Head Start REDI~ Incredible Years Teacher Training Group	behavior problems: total score	1	0.06	0.06	0.12	0.06		
359	Head Start REDI~ Incredible Years Teacher Training Group	social skills	1	0.28	0.06	-0.01	0.06		
348	Head Start REDI~ Promoting Alternative Thinking Strategies Group	language and literacy	0	0.17	0.06	-0.01	0.06		
351	Head Start REDI~ Promoting Alternative Thinking Strategies Group	mathematical thinking	0	0.14	0.06	-0.01	0.06		
354	Head Start REDI~ Promoting Alternative Thinking Strategies Group	general knowledge	0	0.11	0.06	0.01	0.06		
357	Head Start REDI~ Promoting Alternative Thinking Strategies Group	behavior problems: total score	1	0.04	0.06	0.07	0.06		
360	Head Start REDI~ Promoting Alternative Thinking Strategies Group	social skills	1	0.19	0.06	-0.01	0.06		
349	Head Start REDI~ Tools of the Mind Group	language and literacy	0	0.08	0.06	0.02	0.06		
352	Head Start REDI~ Tools of the Mind Group	mathematical thinking	0	-0.01	0.06	0.04	0.06		
355	Head Start REDI~ Tools of the Mind Group	general knowledge	0	-0.03	0.06	0.02	0.06		
358	Head Start REDI~ Tools of the Mind Group	behavior problems: total score	1	-0.02	0.06	-0.01	0.06		
361	Head Start REDI~ Tools of the Mind Group	social skills	1	0.07	0.06	0.01	0.06		
227	Head Start Research-Based, Developmentally-Informed Program	aggression	1	0.10	0.18	0.20	0.18		
228	Head Start Research-Based, Developmentally-Informed Program	aggression	1	0.18	0.18	0.28	0.18		
229	Head Start Research-Based, Developmentally-Informed Program	aggressive problem solving	1	0.21	0.18	0.02	0.18		
230	Head Start Research-Based, Developmentally-Informed Program	aggressive-oppositional behavior	1	0.18	0.18	0.27	0.18	0.08	0.18
231	Head Start Research-Based, Developmentally-Informed Program	attention problems	1	0.30	0.18	0.20	0.18		
232	Head Start Research-Based, Developmentally-Informed Program	attention problems	1	0.08	0.17	0.16	0.18	0.08	0.18
233	Head Start Research-Based, Developmentally-Informed Program	blending/phoneme decoding	0	0.39	0.18	0.14	0.18		
234	Head Start Research-Based, Developmentally-Informed Program	elison/sight word reading	0	0.44	0.18	-0.08	0.18		
235	Head Start Research-Based, Developmentally-Informed Program	print awareness/letter word	0	0.19	0.18	0.01	0.18		
236	Head Start Research-Based, Developmentally-Informed Program	vocabulary	0	0.14	0.18	0.10	0.18		
237	Head Start Research-Based, Developmentally-Informed Program	competent problem solving	1	0.39	0.18	0.27	0.18		
238	Head Start Research-Based, Developmentally-Informed Program	peer rejection	1	0.06	0.17	0.14	0.18	0.03	0.18

239	Head Start Research-Based, Developmentally-Informed Program	social competence	1	0.09	0.18	0.14	0.18		
240	Head Start Research-Based, Developmentally-Informed Program	social competence	1	0.21	0.18	0.30	0.18	0.10	0.18
241	Head Start Research-Based, Developmentally-Informed Program	student-teacher closeness	1	0.18	0.18	0.14	0.18	0.12	0.18
274	Home-Based Dyslexia Prevention	phoneme blending	0	0.04	0.29	0.24	0.31		
275	Home-Based Dyslexia Prevention	phoneme segmentation	0	-0.17	0.29	0.35	0.31		
276	Home-Based Dyslexia Prevention	receptive letter knowledge	0	0.59	0.30	0.32	0.31		
310	Infant Health and Development Program	receptive language	0	0.38	0.06			0.06	0.07
224	LARS & LISA (Lust An Realistischer Sicht & Leichtigkeit Im Sozialen Alltag)	conduct problems	1	-0.03	0.26	-0.10	0.26		
225	LARS & LISA (Lust An Realistischer Sicht & Leichtigkeit Im Sozialen Alltag)	depression	1	0.02	0.25	0.06	0.26		
1	Learn Young, Learn Fair	anxiety	1	-0.02	0.15	-0.02	0.16		
2	Learn Young, Learn Fair	depression	1	0.00	0.15	0.02	0.16		
3	Learn Young, Learn Fair	physiological stress symptoms	1	-0.03	0.21	0.12	0.21		
4	Learn Young, Learn Fair	psychological stress symptoms	1	-0.10	0.15	-0.12	0.16		
5	Learn Young, Learn Fair	stress awareness	1	0.66	0.15	0.35	0.16		
6	Learn Young, Learn Fair	stress symptoms total	1	-0.04	0.15	-0.02	0.16		
7	Learn Young, Learn Fair	emotion-focused coping	1	0.21	0.19	0.07	0.19		
8	Learn Young, Learn Fair	maladaptive coping	1	-0.15	0.15	-0.09	0.16		
9	Learn Young, Learn Fair	problem solving	1	-0.09	0.15	-0.25	0.16		
10	Learn Young, Learn Fair	social support skills	1	0.02	0.15	0.01	0.16		
13	Life Skills Training	daily smokers	1	1.02	0.16			-0.05	0.17
14	Life Skills Training	monthly smokers	1	-0.24	0.16			0.63	0.17
15	Life Skills Training	non-repeat smoking triers	1	-0.12	0.15			-0.07	0.17
16	Life Skills Training	non-smokers	1	0.17	0.15			-0.03	0.17
17	Life Skills Training	smoking triers	1	-0.92	0.16			0.29	0.17
18	Life Skills Training	sporadic smokers	1	-0.07	0.15			0.05	0.17
19	Life Skills Training	weekly smokers	1	-0.61	0.16			-0.10	0.17
300	Life Skills Training in Minority Youth	binge drinking	1	0.50	0.15	0.50	0.15		
301	Life Skills Training in Minority Youth	drinking knowledge	1	0.07	0.13	0.00	0.13		
302	Life Skills Training in Minority Youth	peer drinking norms	1	0.10	0.13	0.09	0.13		
303	Life Skills Training in Minority Youth	pro-drinking attitudes	1	0.08	0.13	0.06	0.13		
320	Linking the Interests of Families and Teachers	police arrest	1	0.00	0.25	0.68	0.27	0.86	0.28
104	Living Letters #2~ Combined Text Comprehension and Oral Language Group	reading comprehension	0	0.13	0.23	0.23	0.23		
107	Living Letters #2~ Combined Text Comprehension and Oral Language Group	reading comprehension	0	0.46	0.23	0.26	0.23		

110	Living Letters #2~ Combined Text Comprehension and Oral Language Group	vocabulary	0	0.23	0.23	0.08	0.23
113	Living Letters #2~ Combined Text Comprehension and Oral Language Group	arithmetic	0	0.38	0.23	0.12	0.23
105	Living Letters #2~ Oral Language Group	reading comprehension	0	0.04	0.23	0.23	0.24
108	Living Letters #2~ Oral Language Group	reading comprehension	0	0.35	0.23	0.65	0.24
111	Living Letters #2~ Oral Language Group	vocabulary	0	0.37	0.23	0.37	0.24
114	Living Letters #2~ Oral Language Group	arithmetic	0	0.32	0.23	0.14	0.23
106	Living Letters #2~ Text Comprehension Group	reading comprehension	0	0.12	0.23	0.11	0.23
109	Living Letters #2~ Text Comprehension Group	reading comprehension	0	0.38	0.23	0.39	0.23
112	Living Letters #2~ Text Comprehension Group	vocabulary	0	0.09	0.23	0.20	0.23
115	Living Letters #2~ Text Comprehension Group	arithmetic	0	0.30	0.23	0.00	0.23
189	Multi-Component Reading Remediation~ PHAB + RAVE-O	oral reading quotient	0	0.34	0.22	0.21	0.24
191	Multi-Component Reading Remediation~ PHAB + RAVE-O	passage comprehension	0	0.23	0.22	0.28	0.24
193	Multi-Component Reading Remediation~ PHAB + RAVE-O	reading	0	0.45	0.22	0.56	0.25
195	Multi-Component Reading Remediation~ PHAB + RAVE-O	spelling	0	0.13	0.22	0.28	0.24
197	Multi-Component Reading Remediation~ PHAB + RAVE-O	word attack	0	0.55	0.22	0.56	0.25
199	Multi-Component Reading Remediation~ PHAB + RAVE-O	word identification	0	0.29	0.22	0.42	0.24
201	Multi-Component Reading Remediation~ PHAB + RAVE-O	word reading efficiency nonwords	0	1.01	0.23	0.60	0.25
203	Multi-Component Reading Remediation~ PHAB + RAVE-O	word reading efficiency real words	0	0.26	0.22	0.44	0.24
205	Multi-Component Reading Remediation~ PHAB + RAVE-O	arithmetic	0	0.08	0.22	-0.02	0.24
190	Multi-Component Reading Remediation~ PHAB + WIST	oral reading quotient	0	-0.02	0.22	0.17	0.24
192	Multi-Component Reading Remediation~ PHAB + WIST	passage comprehension	0	0.29	0.22	0.23	0.24
194	Multi-Component Reading Remediation~ PHAB + WIST	reading	0	0.64	0.22	0.39	0.24
196	Multi-Component Reading Remediation~ PHAB + WIST	spelling	0	0.43	0.22	0.35	0.24
198	Multi-Component Reading Remediation~ PHAB + WIST	word attack	0	0.43	0.22	0.34	0.24
200	Multi-Component Reading Remediation~ PHAB + WIST	word identification	0	0.25	0.22	0.36	0.24
202	Multi-Component Reading Remediation~ PHAB + WIST	word reading efficiency nonwords	0	0.63	0.22	0.35	0.24
204	Multi-Component Reading Remediation~ PHAB + WIST	word reading efficiency real words	0	0.28	0.22	0.26	0.24

206	Multi-Component Reading Remediation~ PHAB + WIST	arithmetic	0	0.18	0.22	0.34	0.24
119	Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group	non-word reading	0	-0.09	0.28	0.10	0.28
122	Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group	reading comprehension	0	-0.01	0.28	-0.09	0.28
125	Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group	segment subtraction	0	-0.55	0.29	-0.32	0.28
128	Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group	sight word reading	0	0.03	0.29	0.33	0.28
131	Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group	word recognition	0	-0.10	0.28	0.43	0.29
118	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension	non-word reading	0	0.34	0.28	0.82	0.29
121	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension	reading comprehension	0	0.47	0.29	0.34	0.28
124	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension	segment subtraction	0	-0.13	0.28	0.44	0.29
127	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension	sight word reading	0	0.04	0.28	0.73	0.29
130	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension	word recognition	0	0.28	0.28	1.31	0.31
120	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group	non-word reading	0	-0.02	0.28	0.19	0.28
123	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group	reading comprehension	0	0.86	0.30	0.51	0.29
126	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group	segment subtraction	0	0.20	0.28	0.50	0.29
129	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group	sight word reading	0	0.15	0.29	0.43	0.29
132	Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group	word recognition	0	0.31	0.28	0.99	0.30
20	Optimism and Lifeskills Program	depressive symptoms	1	-0.57	0.31	0.52	0.36
21	Optimism and Lifeskills Program	loneliness	1	-0.33	0.31	0.30	0.36

22	Optimism and Lifeskills Program	attributional style	1	0.10	0.31	0.38	0.36		
23	Optimism and Lifeskills Program	self worth	1	-0.54	0.31	0.72	0.37		
391	Parent Training for Teenage Moms	mental score	0	0.23	0.23			0.84	0.27
409	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	assertiveness	1	0.11	0.15	-0.06	0.18		
411	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	general influenceability	1	0.05	0.15	-0.09	0.18		
413	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	locus of control	1	0.14	0.15	0.06	0.18		
415	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	self-esteem	1	0.12	0.15	-0.05	0.18		
417	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	social anxiety	1	-0.14	0.15	-0.04	0.18		
419	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	alcohol attitudes	1	0.27	0.15	0.11	0.18		
421	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	alcohol consumption volume	1	0.16	0.15	0.14	0.18		
422	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	alcohol knowledge	1	0.27	0.15	0.23	0.18		
424	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	daily cigarette smoking	1	0.57	0.17	-0.17	0.21		
426	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	frequency of drunkenness	1	-0.11	0.15	0.03	0.18		
428	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	marijuana attitudes	1	0.27	0.15	0.06	0.18		
430	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	marijuana knowledge	1	0.27	0.15	0.03	0.18		
432	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	monthly cigarette smoking	1	0.23	0.17	-0.23	0.21		
434	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	monthly marijuana smoking	1	0.72	0.17	0.23	0.21		
436	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	smoking influenceability	1	0.25	0.15	0.00	0.18		
438	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	tobacco attitudes	1	0.21	0.15	0.08	0.18		
440	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	tobacco knowledge	1	0.27	0.15	0.34	0.19		
442	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	weekly cigarette smoking	1	0.33	0.17	-0.22	0.21		
444	Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	weekly marijuana smoking	1	1.02	0.18	0.09	0.21		
410	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	assertiveness	1	-0.03	0.15	0.06	0.18		
412	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	general influenceability	1	0.06	0.15	-0.14	0.18		
414	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	locus of control	1	-0.08	0.15	0.01	0.18		
416	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	self-esteem	1	0.00	0.15	0.14	0.18		

418	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	social anxiety	1	-0.18	0.15	0.02	0.18		
420	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	alcohol attitudes	1	0.01	0.15	0.06	0.18		
423	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	alcohol knowledge	1	0.05	0.15	0.05	0.18		
425	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	daily cigarette smoking	1	0.17	0.17	0.10	0.21		
427	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	frequency of drunkenness	1	-0.11	0.15	0.14	0.18		
429	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	marijuana attitudes	1	-0.03	0.15	-0.02	0.18		
431	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	marijuana knowledge	1	0.14	0.15	-0.10	0.18		
433	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	monthly cigarette smoking	1	-0.03	0.17	-0.09	0.21		
435	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	monthly marijuana smoking	1	0.00	0.17	0.10	0.21		
437	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	smoking influenceability	1	0.09	0.15	0.06	0.18		
439	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	tobacco attitudes	1	-0.01	0.15	0.17	0.18		
441	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	tobacco knowledge	1	0.27	0.15	0.29	0.19		
443	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	weekly cigarette smoking	1	0.05	0.17	0.00	0.21		
445	Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	weekly marijuana smoking	1	0.24	0.17	-0.08	0.21		
313	Penn Resiliency Program #1	automatic thoughts	1	0.01	0.23	-0.06	0.25		
314	Penn Resiliency Program #1	depressive symptoms	1	0.36	0.21	0.09	0.23		
315	Penn Resiliency Program #1	feeling of hopelessness	1	0.05	0.23	-0.10	0.25		
316	Penn Resiliency Program #1	attributional style	1	0.15	0.23	0.16	0.25		
317	Penn Resiliency Program #1	self esteem	1	0.00	0.23	0.04	0.26		
318	Penn Resiliency Program #2~ Penn Enhancement Group	depressive symptoms	1	0.05	0.14	0.09	0.16	0.13	0.18
319	Penn Resiliency Program #2~ Penn Resiliency Group	depressive symptoms	1	0.08	0.14	0.13	0.16	0.14	0.18
382	Perry Preschool	iq	0	0.75	0.21	0.32	0.20	0.49	0.24
383	Perry Preschool	non-verbal ability	0	0.79	0.21	0.19	0.20	-0.01	0.24
384	Perry Preschool	auditory-vocal association	0	1.32	0.23	0.58	0.21	0.74	0.25
385	Perry Preschool	language deficiencies	0	1.23	0.22	0.24	0.20	0.36	0.24
386	Perry Preschool	vocabulary	0	1.32	0.23	0.43	0.20	0.46	0.24
389	Phonics-Based Instruction for First Graders	spelling	0	0.39	0.21	0.19	0.21	0.24	0.22
390	Phonics-Based Instruction for First Graders	word reading	0	0.34	0.21	0.33	0.21	0.18	0.22
208	Phonological/Early Reading Skills	oral passage reading	0	0.51	0.30	-0.09	0.33		

210	Phonological/Early Reading Skills	word attack	0	1.24	0.32	0.87	0.35		
31	Positive Thinking Program	anxiety	1	0.14	0.23	0.22	0.23	-0.08	0.25
32	Positive Thinking Program	depressive symptoms	1	0.49	0.23	0.25	0.23	0.06	0.25
33	Positive Thinking Program	proportion with depressive diagnosis	1	1.21	0.34	1.34	0.33	-1.02	0.32
34	Positive Thinking Program	attributions	1	0.58	0.23	0.02	0.23	-0.17	0.25
74	Problem Solving for Life	depression	1	0.24	0.15	-0.03	0.16	0.03	0.16
75	Problem Solving for Life	avoidant problem solving	1	0.28	0.21	0.08	0.23		
76	Problem Solving for Life	negative problem orientation	1	0.34	0.21	0.05	0.23		
77	Problem Solving for Life	total problem solving	1	0.22	0.15	-0.02	0.17	0.02	0.16
186	Read Well Kindergarten	receptive language	0	-0.03	0.15	-0.15	0.17		
187	Read Well Kindergarten	word attack	0	0.04	0.15	-0.11	0.17		
188	Read Well Kindergarten	word id	0	-0.02	0.15	-0.10	0.17		
86	Reading Recovery	book level	0	3.05	0.41	0.90	0.34		
87	Reading Recovery	passage reading	0	1.57	0.32	0.71	0.34		
88	Reading Recovery	phonemic awareness	0	0.79	0.29	0.37	0.33		
89	Reading Recovery	phonological recoding	0	1.88	0.34	0.62	0.33		
90	Reading Recovery	reading ability	0	2.03	0.34	0.70	0.33		
91	Reading Recovery	spelling	0	1.48	0.32	0.49	0.33		
92	Reading Recovery	word reading	0	1.55	0.32	0.81	0.34		
282	Reading Remediation	basic skills cluster	0	0.92	0.27	0.69	0.26		
283	Reading Remediation	nonword repetition	0	0.16	0.24	-0.03	0.24		
284	Reading Remediation	oral reading quotient	0	0.66	0.26	0.39	0.26		
285	Reading Remediation	phonological awareness	0	0.39	0.24	0.27	0.24		
286	Reading Remediation	rapid naming of letters	0	0.21	0.24	0.20	0.24		
287	Reading Remediation	spelling	0	0.87	0.25	0.71	0.25		
288	Reading Remediation	spelling	0	0.87	0.26	0.67	0.26		
289	Reading Remediation	word attack	0	0.89	0.27	0.54	0.26		
290	Reading Remediation	word id	0	0.82	0.26	0.67	0.26		
291	Reading Remediation	word reading	0	0.67	0.25	0.66	0.25		
292	Reading Remediation	word reading efficiency	0	0.69	0.25	0.71	0.25		
293	Reading Remediation	math applied problems	0	-0.17	0.25	0.00	0.25		
294	Reading Remediation	math calculations	0	-0.25	0.25	0.33	0.26		
153	Reading Remediation for Children with Reading Disorders	basic reading	0	-0.54	0.24	-0.04	0.23		
154	Reading Remediation for Children with Reading Disorders	blend score	0	-1.84	0.28	-0.96	0.25		

155	Reading Remediation for Children with Reading Disorders	common words recognition	0	0.87	0.24	0.01	0.23
156	Reading Remediation for Children with Reading Disorders	consonant blends beginning recognition	0	1.29	0.26	0.37	0.24
157	Reading Remediation for Children with Reading Disorders	consonant blends ending recognition	0	1.25	0.25	0.44	0.24
158	Reading Remediation for Children with Reading Disorders	consonant combination	0	0.87	0.24	0.34	0.23
159	Reading Remediation for Children with Reading Disorders	graded phonetically complex word recognition	0	0.76	0.24	0.63	0.24
160	Reading Remediation for Children with Reading Disorders	language	0	0.15	0.23	-0.06	0.23
161	Reading Remediation for Children with Reading Disorders	nonsense syllables recognition	0	1.99	0.28	1.07	0.25
162	Reading Remediation for Children with Reading Disorders	phoneme discrimination	0	-0.94	0.25	-0.53	0.24
163	Reading Remediation for Children with Reading Disorders	phonetically simple word recognition	0	1.05	0.25	0.66	0.24
164	Reading Remediation for Children with Reading Disorders	polysyllabic- phonetically simple word recognition	0	0.67	0.24	0.43	0.24
165	Reading Remediation for Children with Reading Disorders	reading ability	0	0.85	0.24	0.23	0.23
166	Reading Remediation for Children with Reading Disorders	reading ability	0	0.48	0.24	0.15	0.23
167	Reading Remediation for Children with Reading Disorders	reading age	0	-0.96	0.25	-0.71	0.24
168	Reading Remediation for Children with Reading Disorders	reading comprehension	0	0.50	0.24	-0.16	0.23
169	Reading Remediation for Children with Reading Disorders	reading quotient	0	-0.73	0.24	-0.37	0.24
170	Reading Remediation for Children with Reading Disorders	reversible words recognition	0	1.23	0.25	0.50	0.24
171	Reading Remediation for Children with Reading Disorders	rule of silent e	0	1.16	0.25	0.62	0.24
172	Reading Remediation for Children with Reading Disorders	short vowels	0	1.79	0.28	1.13	0.25
173	Reading Remediation for Children with Reading Disorders	single consonant sounds	0	1.47	0.26	0.34	0.23
174	Reading Remediation for Children with Reading Disorders	spelling	0	0.11	0.23	0.01	0.23
175	Reading Remediation for Children with Reading Disorders	spelling	0	0.36	0.24	0.16	0.23
176	Reading Remediation for Children with Reading Disorders	syllabification	0	-0.16	0.23	0.31	0.23
177	Reading Remediation for Children with Reading Disorders	vocabulary	0	0.39	0.24	0.07	0.23
178	Reading Remediation for Children with Reading Disorders	vowel combination	0	0.55	0.24	0.21	0.23
179	Reading Remediation for Children with Reading Disorders	word study skills	0	0.58	0.24	0.00	0.23

180	Reading Remediation for Children with Reading Disorders	arithmetic	0	-0.03	0.23	-0.27	0.23		
181	Reading Remediation for Children with Reading Disorders	math application	0	-0.42	0.24	-0.37	0.24		
182	Reading Remediation for Children with Reading Disorders	math computation	0	-0.58	0.24	-0.31	0.23		
183	Reading Remediation for Children with Reading Disorders	math concepts	0	0.06	0.23	-0.18	0.23		
184	Reading Remediation for Children with Reading Disorders	science	0	0.29	0.23	0.02	0.23		
185	Reading Remediation for Children with Reading Disorders	social studies	0	-0.22	0.23	-0.36	0.23		
395	Reading with Rhyme, Reading with Phoneme~ Reading with Phoneme Group	letter identification	0	0.03	0.27	0.04	0.27		
398	Reading with Rhyme, Reading with Phoneme~ Reading with Phoneme Group	nonword reading	0	0.20	0.27	0.30	0.28		
401	Reading with Rhyme, Reading with Phoneme~ Reading with Phoneme Group	word reading	0	0.00	0.27	0.02	0.27		
404	Reading with Rhyme, Reading with Phoneme~ Reading with Phoneme Group	number	0	-0.38	0.27	-0.02	0.27		
397	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme and Phoneme Group	letter identification	0	0.01	0.28	0.18	0.28		
400	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme and Phoneme Group	nonword reading	0	0.09	0.28	0.36	0.28		
403	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme and Phoneme Group	word reading	0	0.02	0.28	0.19	0.28		
406	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme and Phoneme Group	number	0	-0.09	0.28	0.27	0.28		
396	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme Group	letter identification	0	-0.17	0.27	0.14	0.28		
399	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme Group	nonword reading	0	-0.03	0.27	0.21	0.28		
402	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme Group	word reading	0	-0.09	0.27	0.03	0.27		
405	Reading with Rhyme, Reading with Phoneme~ Reading with Rhyme Group	number	0	-0.19	0.28	-0.20	0.28		
61	Resourceful Adolescence Program (Kiwi)	depression	1	0.04	0.11	-0.13	0.11	-0.26	0.12
62	Resourceful Adolescence Program (Kiwi)	depression	1	0.02	0.11	-0.03	0.11	0.05	0.12
336	Responding in Peaceful and Positive Ways (Seventh Grade)	delinquent behavior frequency	1	0.06	0.11	0.12	0.14		
337	Responding in Peaceful and Positive Ways (Seventh Grade)	in-school suspensions	1	0.00	0.17	-0.04	0.23		
338	Responding in Peaceful and Positive Ways (Seventh Grade)	nonphysical aggression	1	-0.03	0.11	0.12	0.14		
339	Responding in Peaceful and Positive Ways (Seventh Grade)	out-of-school suspensions	1	-0.06	0.17	0.49	0.24		
340	Responding in Peaceful and Positive Ways (Seventh Grade)	violent behavior	1	0.14	0.17	0.48	0.24		
341	Responding in Peaceful and Positive Ways (Seventh Grade)	violent behavior frequency	1	-0.11	0.11	0.06	0.14		

342	Responding in Peaceful and Positive Ways (Seventh Grade)	manifest anxiety	1	-0.05	0.11	0.03	0.14		
343	Responding in Peaceful and Positive Ways (Seventh Grade)	attitudes towards supporting non violence	1	-0.01	0.11	0.11	0.14		
344	Responding in Peaceful and Positive Ways (Seventh Grade)	attitudes towards supporting violence	1	-0.06	0.11	0.04	0.14		
345	Responding in Peaceful and Positive Ways (Seventh Grade)	37ip knowledge	1	0.36	0.11	0.33	0.14		
346	Responding in Peaceful and Positive Ways (Seventh Grade)	drug use frequency	1	0.00	0.11	0.09	0.14		
68	Roots of Empathy	indirect aggression	1	0.16	0.18	0.09	0.17	-0.01	0.18
69	Roots of Empathy	indirect aggression	1	0.18	0.17	0.16	0.16	-0.15	0.15
70	Roots of Empathy	physical aggression	1	-0.17	0.18	0.04	0.17	-0.02	0.18
71	Roots of Empathy	physical aggression	1	0.01	0.17	0.20	0.16	0.11	0.15
72	Roots of Empathy	pro-social behavior	1	0.05	0.18	0.09	0.17	0.01	0.18
73	Roots of Empathy	pro-social behavior	1	-0.01	0.17	0.05	0.16	0.16	0.15
242	Safe Dates	moderate physical perpetration	1	0.09	0.15	0.07	0.15	0.06	0.18
243	Safe Dates	severe physical perpetration	1	0.09	0.15	0.05	0.15	0.11	0.18
244	Safe Dates	sexual perpetration	1	0.12	0.15	0.04	0.15	0.14	0.18
245	Safe Dates	destructive anger response	1	0.09	0.18	0.09	0.15		
246	Safe Dates	psychological perpetration	1	0.15	0.15			0.03	0.18
247	SEARCH Screening Test and TEACH Tutoring~ Phonetic Group	reading achievement	0	0.28	0.19	0.13	0.19		
249	SEARCH Screening Test and TEACH Tutoring~ Phonetic Group	word attack	0	0.54	0.20	0.52	0.20		
248	SEARCH Screening Test and TEACH Tutoring~ TEACH Tutoring Group	reading achievement	0	0.16	0.19	0.21	0.19		
250	SEARCH Screening Test and TEACH Tutoring~ TEACH Tutoring Group	word attack	0	0.30	0.19	0.25	0.19		
259	Spanish Reading Intervention for English Language Learners	passage comprehension- english	0	0.06	0.25	0.18	0.20		
260	Spanish Reading Intervention for English Language Learners	passage comprehension- spanish	0	0.45	0.24	0.49	0.20		
261	Spanish Reading Intervention for English Language Learners	reading fluency- english	0	0.13	0.27	-0.01	0.20		
262	Spanish Reading Intervention for English Language Learners	reading fluency- spanish	0	0.50	0.26	0.39	0.20		
263	Spanish Reading Intervention for English Language Learners	word attack- english	0	-0.04	0.27	0.24	0.20		
264	Spanish Reading Intervention for English Language Learners	word attack- spanish	0	0.62	0.25	0.54	0.20		
12	Strengthening Families Program & Life Skills Training #1~ Both	substance initiation index	1	0.16	0.16	0.09	0.14		
11	Strengthening Families Program & Life Skills Training #1~ LST Only	substance initiation index	1	0.01	0.16	0.08	0.14		

93	Strengthening Families Program and Life Skills Training #2	assertiveness	1	0.03	0.02	0.07	0.02		
94	Strengthening Families Program and Life Skills Training #2	association with antisocial peers	1	0.11	0.02	0.12	0.02		
95	Strengthening Families Program and Life Skills Training #2	parent-child affective quality – toward father	1	0.05	0.02	0.02	0.02		
96	Strengthening Families Program and Life Skills Training #2	parent-child affective quality – toward mother	1	0.09	0.02	0.02	0.02		
97	Strengthening Families Program and Life Skills Training #2	problem solving	1	0.13	0.02	0.10	0.02		
98	Strengthening Families Program and Life Skills Training #2	attitude toward substance use	1	0.05	0.02	0.08	0.02		
99	Strengthening Families Program and Life Skills Training #2	perceived substance use norms	1	0.16	0.02	0.24	0.02		
100	Strengthening Families Program and Life Skills Training #2	substance refusal efficacy	1	0.00	0.02	0.07	0.02		
101	Strengthening Families Program and Life Skills Training #2	substance refusal intentions	1	0.03	0.02	0.09	0.02		
102	Strengthening Families Program and Life Skills Training #2	substance use expectancies	1	0.10	0.02	0.14	0.02		
103	Strengthening Families Program and Life Skills Training #2	substance use plans	1	0.02	0.02	0.05	0.02		
207	Supplemental Phonics-based Instruction	word reading	0	0.70	0.24	0.53	0.25	0.39	0.25
277	Supplemental Reading Instruction	comprehension	0	0.30	0.20			0.19	0.20
278	Supplemental Reading Instruction	letter name identification	0	0.35	0.18			0.28	0.20
279	Supplemental Reading Instruction	oral reading fluency	0	0.33	0.18			0.25	0.20
280	Supplemental Reading Instruction	vocabulary	0	0.21	0.20			0.18	0.20
281	Supplemental Reading Instruction	word attack	0	0.60	0.18			0.33	0.20
265	Swedish Phonics-based Intervention	phoneme deletion	0	-0.06	0.19	-0.35	0.19		
267	Swedish Phonics-based Intervention	reading speed	0	0.13	0.19	0.00	0.19		
268	Swedish Phonics-based Intervention	reversed spoonerism	0	-0.01	0.19	-0.45	0.19		
269	Swedish Phonics-based Intervention	spelling	0	0.23	0.19	0.10	0.19		
270	Swedish Phonics-based Intervention	spelling	0	0.38	0.19	0.33	0.19		
271	Swedish Phonics-based Intervention	spoonerism	0	0.01	0.19	-0.15	0.19		
272	Swedish Phonics-based Intervention	word decoding	0	0.27	0.19				
226	Teacher Responsivity Education	expressive vocabulary	0	-0.01	0.19	0.16	0.11		
362	Tennessee Pre-K	cognitive achievement	0	0.32	0.07	0.02	0.07	-0.04	0.07
387	The Early Training Project~ 3-year and 2-year Intervention Group	iq	0	0.47	0.37				
388	The Early Training Project~ 3-year and 2-year Intervention Group	vocabulary	0	0.44	0.33	0.27	0.33		
52	Think Smart	assertiveness skills	1	0.32	0.17	0.74	0.17		
53	Think Smart	cultural identity	1	-0.34	0.17	-0.64	0.17		

54	Think Smart	alcohol use	1	-0.05	0.17	0.52	0.17		
55	Think Smart	hlp use	1	0.30	0.17	1.12	0.18		
56	Think Smart	inhalant use	1	0.72	0.18	1.27	0.18		
57	Think Smart	knowledge of drug-related consequences	1	1.14	0.18	0.58	0.17		
58	Think Smart	otc use	1	0.25	0.17	0.95	0.17		
59	Think Smart	prescription use	1	0.14	0.17	0.25	0.17		
60	Think Smart	tobacco use	1	1.39	0.19	0.56	0.17		
44	Tools for Getting Along	metacognition	0	0.12	0.15	-0.13	0.20		
45	Tools for Getting Along	anger out	1	0.08	0.15	-0.07	0.21		
46	Tools for Getting Along	proactive aggression	1	0.10	0.15	-0.03	0.20		
47	Tools for Getting Along	trait anger	1	0.01	0.15	-0.15	0.21		
48	Tools for Getting Along	behavioral regulation	1	0.02	0.15	-0.07	0.20		
49	Tools for Getting Along	positive problem orientation	1	0.16	0.15	0.01	0.21		
50	Tools for Getting Along	problem solving knowledge	1	2.66	0.21	0.90	0.22		
51	Tools for Getting Along	rational problem solving	1	-0.23	0.15	-0.10	0.21		
321	Universal School-based Mental Health Intervention~ Psychologist-Led Group	anxiety	1	0.78	0.20	0.69	0.31		
323	Universal School-based Mental Health Intervention~ Psychologist-Led Group	avoidance	1	0.83	0.20	0.47	0.30		
325	Universal School-based Mental Health Intervention~ Psychologist-Led Group	problem solving	1	2.48	0.26	3.05	0.42		
327	Universal School-based Mental Health Intervention~ Psychologist-Led Group	seeking social support	1	0.56	0.20	1.90	0.35		
322	Universal School-based Mental Health Intervention~ Teacher-Led Group	anxiety	1	0.63	0.21	0.80	0.30		
324	Universal School-based Mental Health Intervention~ Teacher-Led Group	avoidance	1	0.95	0.21	0.89	0.30		
326	Universal School-based Mental Health Intervention~ Teacher-Led Group	problem solving	1	2.45	0.27	2.36	0.37		
328	Universal School-based Mental Health Intervention~ Teacher-Led Group	seeking social support	1	0.18	0.20	1.58	0.33		
329	Unplugged~ All Groups	any episode of drunkenness	1	0.18	0.14			0.12	0.15
330	Unplugged~ All Groups	any smoking	1	0.07	0.14			0.03	0.15
331	Unplugged~ All Groups	any use of cannabis	1	0.14	0.14			0.10	0.15
332	Unplugged~ All Groups	daily smoking	1	0.20	0.14			0.05	0.15
333	Unplugged~ All Groups	frequent cannabis	1	0.15	0.14			0.17	0.15
334	Unplugged~ All Groups	frequent drunkenness	1	0.20	0.14			0.26	0.15
335	Unplugged~ All Groups	frequent smoking	1	0.08	0.14			0.06	0.15

Note. Due to the volume of our data, this table only reports the post-test, 6- to 12-month, and 1-to-2 year follow-up effects included in our primary analyses (i.e., post-test effect sizes without matched follow-up effects in this table have an aligned effect size(s) reported at a later follow-up wave). The complete dataset is posted online for public use. Note that various steps were taken to produce the documented statistics (e.g., collapsing across moderator groups, converting reported statistics into standardized effects, averaging effects across papers reporting on the same outcome). Please reference the details in this supplemental file and the meta-analytic protocol and readme posted in conjunction with the data to understand the steps taken to produce these effect sizes and standard errors.

References for Reports Included in the Full MERF Sample

Learn Young, Learn Fair

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. (Campbell et al., 2012)

Strengthening Families Program and Life Skills Training #1

Spoth, R., 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

Spoth, 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

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*, 109-122. https://doi.org/10.1023/a:1022926332514

Spoth, 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

Life Skills Training

Luna-Adame, M., Carrasco-Giménez, T., & del Mar Rueda-García, 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

Optimism and Lifeskills Program

Quayle, D., Dziurawiec, S., Roberts, C., Kane, R., & Ebsworthy, G. (2001). The effect of an optimism and lifeskills program on depressive symptoms in preadolescence. *Behavior Change, 18*(4), 194-203. https://doi.org/10.1375/bech.18.4.194

Aussie Optimism Program

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. *Behavior Research and Therapy, 48*(1), 68-73. https://doi.org/10.1016/j.brat.2009.09.002

Positive Thinking Program

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 Psychologists and Counsellors in Schools, 16*(1), 76-90. https://doi.org/10.1375/ajgc.16.1.76

Emotional Intelligence Training Program

Ruiz-Aranda, D., Castillo, R., Salguero, J. M., Cabello, R., Fernández-Berrocal, P., & Balluerka, N. (2012). Short- and mid-term effects of emotional intelligence training on adolescent mental health. *Journal of Adolescence Health, 51*(5), 462–467. https://dx.doi.org/10.1016/j.jadohealth.2012.02.003.

Tools for Getting Along

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

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

Think Smart

Johnson, K. W., Shamblen, S., 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*, 298-312. https://doi.org/10.1007/s11121-009-0132-2

RAP-Kiwi; Resourceful Adolescent Program (RAP)

Merry, S., McDowell, H., Wild, C. J., Julliet, B., & Cunliffe, R. (2004). A randomized placebo-controlled trial of a school-based depression prevention program. *Journal of American Academy of Child & Adolescent Psychiatry, 43*(5), 538-547. https://doi.org/10.1097/00004583-200405000-00007

Ease of Handling Social Aspects in Everyday of Life Training

Pössel, P., Horn, A.B., Goren, G., Hautzinger, M. (2004). School-based prevention of depressive symptoms in adolescents: a 6-month follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*(8), 1003-1010. https://doi.org/10.1097/01.chi.0000126975.56955.98

Cognitive and Social Skills Program

Sarason, I. G., & Sarason, B. R. (1981). Teaching cognitive and social skills to high school students. *Journal of Consulting and Clinical Psychology, 49*(6), 908–919. https://doi.org/10.1037/0022-006X.49.6.908

Roots of Empathy

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

Problem Solving for Life

Spence, S. H., Sheffield, J., & 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

Beyondblue Schools Research Initiative

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 randomized 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

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. *Journal Of Adolescent Health, 47*(3), 297-304. doi:10.1016/j.jadohealth.2010.02.007

Reading Recovery

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

Strengthening Families Program and Life Skills Training #2

Spoth, 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

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

Spoth, 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

Spoth, R., Redmond, C., Shin, C., Greenberg, M., Feinberg, M., & Schainker, L. (2013). PROSPER community–university partnership delivery system effects on substance misuse through 6 ½ 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

Spoth, 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

Living Letters #1

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

Living Letters #2

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

Computer-Assisted Learning Program

Ecalle, J., Magnan, A., & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low progress readers. *Computers and Education, 52*(3), 554–561. https://doi.org/10.1016/j.compedu.2008.10.010

Phonemic Awareness and Letter Sound Training

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

Omega-Interactive Sentences, Computerized Phonological Training

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

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

Phonological Awareness Training

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

Morphological versus Phonological Awareness Training

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. A. 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

Computer-Assisted Reading Intervention for Children at Risk of Dyslexia

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

Reading Summer Day-Camp

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

Reading Remediation for Children with Reading Disorders

Gittelman, R., & Feingold, I. (1983). Children with reading disorders: Efficacy of reading remediation. *Journal of Child Psychology & Psychiatry, 24*(2), 167–191. https://doi.org/10.1111/j.1469-7610.1983.tb00568.x

Read Well Kindergarten

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

Multi-Component Reading Remediation

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

Early Literacy Concepts

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

Supplemental Phonics-based Instruction

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. (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

Phonological/Early Reading Skills

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

Code-Oriented Reading Instruction

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/10.1037/0022-0663.98.3.508

Computer-Assisted Blending Skill Training

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

Cogmed Working Memory Training

Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P. J., Gathercole, S., Gold, L., Sia, K., 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) 1-10. https://doi.org/10.1001/jamapediatrics.2015.4568

Explicit Phonological Awareness Instruction

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/s11145-007-9080-z

Project Care

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/10.1111/j.1467-8624.1990.tb03559.x

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

LARS & LISA (Lust An Realistischer Sicht & Leichtigkeit Im Sozialen Alltag)

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., 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

Consultation-based Intervention

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), 635-48. https://doi.org/10.1007/s10802-006-9046-7

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.

Teacher Responsivity Education

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)

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.tc.idm.oclc.org/10.1080/10888691.2015.1074050

Head Start Research-Based, Developmentally-Informed Program

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., 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

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

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. *American Journal of Psychiatry, 178*(4), 305-312. https://doi.org/10.1176/appi.ajp.2020.20030343

Safe Dates

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 and 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., 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

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*, 245-258. https://doi.org/10.1007/s11121-005-0007-0

SEARCH Screening Test and TEACH Tutoring

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

English Reading Intervention for English Language Learners

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.

Vaughn, S., Mathes, P., Linian-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. *Elementary School Journal*, *107*(2), 153–180. https://doi.org/10.1086/510653

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

Spanish Reading Intervention for English Language Learners

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., Linian-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., 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.tc.idm.oclc.org/10.1080/19345740802114749

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

Swedish Phonics-based Intervention

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

Computer-Assisted Remedial Reading Intervention

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

Home-Based Dyslexia Prevention

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

Supplemental Reading Instruction

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. *Journal of Special Education, 39*(2), 66–85. https://doi.org/10.1177/00224669050390020301

Word Analysis Skills Training

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

Reading Remediation

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.tc.idm.oclc.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.tc.idm.oclc.org/10.1037/a0033663

Childhood Depression Prevention

Cecchini, T. B. (1997). *An interpersonal and cognitive-behavioral approach to childhood depression: A school-based primary prevention study* (Order No. 9820698). Available from ProQuest Dissertations & Theses Global. (304420573). http://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/dissertations-theses/interpersonal-cognitive-behavioral-approach/docview/304420573/se-2

Johnson, N. C. (2000). *A follow-up study of a primary prevention program targeting childhood depression* (Order No. 1402700). Available from ProQuest Dissertations & Theses Global. (304634088). http://ezproxy.cul.columbia.edu/login?url=https://www.proquest.com/dissertations-theses/follow-up-study-primary-prevention-program/docview/304634088/se-2

FRIENDS Program #4

Essau, C. A., Conradt, J., Sasagawa, S., & Ollendick, T. H. (2012). Prevention of anxiety symptoms in children: Results from a universal school-based trial. *Behavioral Therapy, 43*(2), 450-464. https://doi.org/10.1016/j.beth.2011.08.003

Life Skills Training in Minority Youth

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

Classroom-Centered and School-Family Partnership Interventions

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

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 & Behavioral Disorders, 9*(3), 146-161.

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

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

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

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), 926-927. https://doi.org/10.1037/a0016586

Parent-Adolescent Interaction

Dunham, R., Kidwell, J., & Portes, P. (1988). Effects of parent-adolescent interaction on the continuity of cognitive development from early childhood to early adolescence. *Journal of Early Adolescence, 8*(3), 297-310. https://doi-org.tc.idm.oclc.org/10.1177/0272431688083006

Infant Health and Development Program

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. http://dx.doi.org/10.1037/0012-1649.39.4.730

McCarton, Brooks-Gunn, Wallace, et al. (1997). Results at age 8 years of early intervention for low-birth-weight premature infants: The infant health and development program. *The Journal of the American Medical Association, 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

Classroom and At-Home Preschool Interventions

U.S. Department of Health, Education, and Welfare Office of Education. (1969). *Investigations of classroom and at-home interventions: Research and development program on preschool disadvantaged children* (M. B. Karnes, A. S. Hodgins, J. A. Teska, & S. A. Kirk, Authors; Research Report No. ED 036 663). https://files.eric.ed.gov/fulltext/ED036663.pdf

Penn Resiliency Program #1

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). https://doi-org.tc.idm.oclc.org/10.1037/1522-3736.5.1.58a

Penn Resiliency Program #2

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.tc.idm.oclc.org/10.1037/0022-006X.75.1.9

Linking the Interests of Families and Teachers

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

Universal School-based Mental Health Intervention

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

Unplugged

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., (2008). The effectiveness of a school-based substance abuse prevention program: EU-Dap cluster randomized 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., & EU-Dap Study Group (2007). A cluster randomized controlled trial of school-based prevention of tobacco, alcohol and drug use: The EU-Dap 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., Rosaria Galanti, M., & the EU-Dap Study Group (2010). The effectiveness of a school-based substance abuse prevention program: 18-month follow-up of the EU-Dap cluster randomized controlled trial. *Drug and Alcohol Dependence, 108*(1-2), 56-64. https://doi.org/10.1016/j.drugalcdep.2009.11.018

Responding in Peaceful and Positive Ways (Seventh Grade)

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

Head Start Classroom-based Approaches and Resources for Emotion and Social Skill Promotion

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

Early Head Start

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

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

Tennessee Pre-K

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

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

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

Head Start Impact Study

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

Dialogic Reading #1

Whitehurst, G.J., Arnold, D.S., Epstein, J.N., & Angell, A.L. (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.tc.idm.oclc.org/10.1037/0012-1649.30.5.679

Nurse Home Visitation Program

Olds, D. L., Henderson, C. R., 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., & Kitzmna, 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

Olds D. L., Henderson, C. R., 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

Eckenrode J., Campa M., Luckey D.W., Henderson, C. R., Cole, R., Kitsman, H., Arson, E., Sidora, K., Powers, P., & Olds, D. L. (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. A*rchives of Pediatric Adolescent Medicine, 164*(1), 9–15. https://doi.org/10.1001/archpediatrics.2009.240

Dialogic Reading #2

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

Perry Preschool

Weikart, D. P. (1970). Longitudinal Results of the Ypsilanti Perry Preschool Project. Final Report. Volume II of 2 Volumes.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review, 103*(6), 2052-2086. https://doi.org/10.1257/aer.103.6.2052

The Early Training Project

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

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

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

Phonics-Based Instruction for First Graders

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. (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*, 821–843. https://doi.org/10.1007/s11145-012-9393-4

Parent Training for Teenage Moms

Field, T., Widmayer, S., Greenberg, R., Stoller, S. (1982). Effects of parent training on teenage mothers and their infants. *Pediatrics, 69*(6), 703-707. https://doi.org/10.1542/peds.69.6.703

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.

Abecedarian

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

Elango, S., García J., L., Heckman, J. J., Hojman, A. (2016). *Early childhood education.* NBER 2.pdf

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., 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. https://doi.org/10.1037/a0026644

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., & 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/10.1111/j.1467-8624.1994.tb00777.x

Technology-Enhanced, Research-Based, Instruction, Assessment, and Professional Development Scale-up Model

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

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

Reading with Rhyme, Reading with Phoneme

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 Psychiatry, 45*(2), 338–358. https://doi.org/10.1111/j.1469-7610.2004.00225.x

FRIENDS Program #3

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

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 & Adolescent Psychology, 35*(3) 403-411. https://doi.org/10.1207/s15374424jccp3503_5

Cognitive-Behavioral Approach to Drug Abuse Prevention

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., 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. *The Journal of the American Medical Association, 273*(14), 1106-1112. https://doi.org/10.1001/jama.1995.03520380042033

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

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*, 103-112. https://doi.org/10.1007/s11121-006-0025-6

Peer-Led Life Skills Training

Botvin, G. J., Baker, E., Renick, N. L., Filazzola, A. D., & Botvin, E. M. (1984).. *Addictive Behaviors, 9*(2), 137-147. https://doi.org/10.1016/0306-4603(84)90051-0

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

Culturally-Focused Substance Use Prevention

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

Conflict-Resolution Training Program

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 Counselling, 26*, 109-124. https://doi.org/10.1023/B:ADCO.0000027425.24225.87

All Stars

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 and Behavior, 28*(5), 533-546. https://doi.org/10.1177/109019810102800502

Eigenständig Werden (Fifth and Sixth Grade)

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), e000352. https://doi.org/10.1136/bmjopen-2011-000352

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), e004422. https://doi.org/10.1136/bmjopen-2013-004422

References

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological bulletin*, *117*(1), 167.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better *P*-curves: Making *P*-curve analysis more robust to errors, fraud, and ambitious *P*-hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General, 144*(6), 1146–1152. https://doi.org/10.1037/xge0000104

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