Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1366

How General is Educational Intervention Fadeout? A Meta-Analysis of Educational RCTs with Follow-Up

Tyler W. Watts
Teachers College, Columbia University

Emma R. Hart
Boston College

Drew H. Bailey
University of California, Irvine

Researchers and policymakers pursue educational interventions with the goal of altering children's long-term trajectories. However, many effects fade quickly after interventions end. Researchers have sought to address the fadeout problem by identifying characteristics of interventions that lead to persistent effects, though reliable answers have been elusive. We present evidence from 87 randomized controlled trials of educational interventions targeted at a diverse array of skills across developmental stages. We examined whether salient intervention features consistently explain medium-term persistence rates, testing widely held theories about skill building. We observed evidence of fadeout for most interventions. Although persistence rates varied, salient features of interventions explained only a small portion of this variation.

VERSION: December 2025

Suggested citation: Watts, Tyler W., Emma R. Hart, and Drew H. Bailey. (2025). How General is Educational Intervention Fadeout? A Meta-Analysis of Educational RCTs with Follow-Up. (EdWorkingPaper: 25-1366). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/hxhd-cv70

How General is Educational Intervention Fadeout?
A Meta-Analysis of Educational RCTs with Follow-Up

Tyler W. Watts[1*], Emma R. Hart[2], Drew H. Bailey[3]

1. Teachers College, Columbia University
2. Boston College
3. University of California, Irvine

* *Corresponding Author:*
Teachers College, Columbia University
525 W 120th St
New York, New York 10027
212-678-3095
tww2108@tc.columbia.edu

Acknowledgements

We would like to thank Jens Detrichson, Greg Duncan, James Pustejovsky, Mikkel Vembye, and Mike Weiss for their helpful comments on this work. We would also like to acknowledge the contributions of the following research assistants (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.

Abstract

Researchers and policymakers pursue educational interventions with the goal of altering children's long-term trajectories. However, many effects fade quickly after interventions end. Researchers have sought to address the fadeout problem by identifying characteristics of interventions that lead to persistent effects, though reliable answers have been elusive. We present evidence from 87 randomized controlled trials of educational interventions targeted at a diverse array of skills across developmental stages. We examined whether salient intervention features consistently explain medium-term persistence rates, testing widely held theories about skill building. We observed evidence of fadeout for most interventions. Although persistence rates varied, salient features of interventions explained only a small portion of this variation.

Keywords: intervention, academic achievement, social-emotional skills, longitudinal, meta-analysis, evaluation

How General is Educational Intervention Fadeout?

A Meta-Analysis of Educational RCTs with Follow-Up

Understanding the effects of educational interventions is key to both educational research and practice. Billions of public dollars have been invested in educational intervention evaluations in the U.S. over the past 25 years, with the Institute of Educational Sciences overseeing a massive expansion of evaluation and experimentation since 2002[i]. Educational interventions are often built on the premise that childhood and adolescence constitute sensitive periods marked by substantial malleability (e.g., Bornstein, 1989), during which interventions can target critical skills with long-reaching benefits. This expectation rests on the intuitive assumption that "skills beget skills;" earlier skill gains will lead to future skill advancements (Cunha & Heckman, 2007; Masten & Cicchetti, 2010). Well-controlled longitudinal studies have also underwritten the promise of skill-building theories, as individual differences in early skills are often highly predictive of later differences in skills (e.g., Duncan et al., 2007; Watts et al., 2014). The predictive validity of childhood skills even extends into adulthood, with many studies reporting that various measures of childhood cognitive (e.g., Ritchie & Bates, 2013; Watts, 2020) and social-emotional (e.g., Moffitt et al., 2012; Koepp et al., 2022) skills strongly predict key measures of adult functioning, like economic attainment and health. Together, this research literature has built the foundation for the science of educational interventions with a convincing logic: educational interventions improve important skills, and such skill gains set children on an advanced trajectory that translates to better outcomes over time and into adulthood. Indeed, several well-known early childhood interventions run in the mid-20th century appeared to produce this exact pattern of results (see review in Elango et al., 2015).

However, despite impressive advancements in rigorous evaluations of educational programs and policies in recent years, the field has struggled to generate a reliable intervention science that provides sure-fire recommendations for producing the kinds of persistent effects predicted by our theoretical models (Lortie-Forgues & Inglis, 2019). Accumulating evidence of intervention fadeout, whereby skill impacts observed at the end of an intervention program diminish over time, intuitively appears to conflict with the skill-building model of educational intervention[ii]. In particular, fadeout has been a major source of concern in the early childhood education (ECE) literature over the past decade (Bailey et al., 2017). Although fadeout has received the most attention in the ECE literature, we have little reason to believe that the phenomenon is limited to ECE interventions. Indeed, fadeout has been documented in several K-12 studies of various educational programs (see review in Bailey et al., 2020).

Fadeout is often viewed as an aberrant phenomenon for researchers to overcome through “better” intervention design and implementation. Yet, the field lacks a framework for understanding when fadeout should be expected across various models of educational interventions. Skill-building interventions can vary widely in the skills targeted, intervention components, and theories of change. Multiple reviews articulate potential sources of heterogeneity that might explain whether an educational intervention effect is likely to persist or fade (e.g., Abenavoli, 2019; Bailey et al., 2020), though these theories have been challenging to test empirically. It remains an open question whether we can use researchers’ theories about impact persistence to predict a priori whether a given intervention is likely to produce an enduring or fading effect.

In the current study, we created a meta-analytic dataset of intervention impacts from randomized controlled trials (RCTs) of a wide variety of educational interventions. These interventions targeted a range of child and adolescent skills and behaviors, including literacy, mathematics, general cognitive ability, social-emotional functioning, and adolescent risk-taking. We began by broadly examining the patterns of persistence and fadeout across all 87 RCTs included in our dataset. We then tested whether salient intervention features were associated with increased intervention persistence rates across studies. For these tests, we operationalized the issue as a forecasting problem: if we observe two end-of-treatment effect sizes of, say, 0.30 SDs produced by two different interventions, can we identify intervention features that would allow us to predict which intervention will generate the larger impact at follow-up? Ultimately, we were interested in examining whether the literature has produced any class of intervention—categorized by easily definable features—for which fadeout was *not* observed.

Below, we review the literature and theories on the features of interventions that could matter for persistence, before detailing the meta-analytic data, results, and implications for the field moving forward.

Defining Fadeout and Persistence

We use the terms "persistence" and "fadeout" to describe two sides of the same coin: the former describes the extent to which intervention effects on the same construct hold over time, while the latter describes the extent to which these effects diminish. In our conception, intervention effects are typically operationalized as standardized mean differences between an intervention group and a comparison group. So, an intervention that produced a .30 SD effect on mathematics achievement at intervention end, followed by a .10 mathematics effect at 1-year follow-up, could be characterized by 33% persistence (i.e., .10/.30) or 67% fadeout (i.e., (.30-.10)/.30). Note that in our definitions, persistence and fadeout are captured by comparing follow-up effect sizes against the magnitudes of the initial effects. Also crucial, fadeout and persistence describe effect patterns on the same construct over time. Holding the construct constant is important because it helps distinguish fadeout from transfer issues, which concern the production of intervention impacts that generalize to other skill domains (e.g., Sala & Gobet, 2019).

Why might some intervention effects persist more than others?

Skills

Educational interventions target a wide variety of skills. Some skills might be particularly prone to fadeout, whereas others could provide the necessary elements for persistence. Bailey et al. (2017) argued that interventions should target "trifecta" skills to produce longer-lasting impacts. They described trifecta skills as skills that are malleable, fundamental, and unlikely to be learned under the counterfactual condition. Though theoretically convincing, the field has struggled to identify a priori intervention targets that will meet the definition of "trifecta" skills. Most notably, researchers have argued that longer-term effects for educational programs are likely to be driven by persistent effects on social-emotional skills rather than cognitive skills (Heckman & Kautz, 2013). This theory has been fueled by some interventions that apparently produced long-term effects on adult outcomes, even when short-term effects on measured cognitive skills faded (Chetty et al., 2011; Deming, 2009). Indeed, much of the research on early intervention fadeout has focused on cognitive skill impacts (e.g., Bailey et al., 2017). Social-emotional skills appear to meet the "trifecta" definition in that they show sensitivity to interventions (e.g., Durlak et al., 2011) and strong validity in predicting adult outcomes (Moffitt et al., 2011). Additionally, children may be less likely to encounter direct social-emotional instruction in typical school environments.

However, recent meta-analyses have cast doubt on the simplest versions of these skill-specific theories. The Meta-Analysis of Educational RCTs with Follow-Up (MERF), which provides the data for the current study, has suggested that fadeout trajectories are similar across social-emotional and cognitive skills (Hart et al., 2024). Persistence rates varied considerably within and across studies, yet targeted skills appeared to explain little of this variation. The ubiquity of fadeout across different skill types raises the question of whether intervention features might play a more important role in shaping persistence.

Intervention Features

Intervention implementation quality is often evoked in reasoning about what kinds of interventions should generate more persistent effects. Interventions that are based on sound theory, target important developmental constructs, and are implemented more carefully may generate more persistent effects. In contrast, interventions that are executed haphazardly or generate large effects by “teaching to the test” might produce impacts that are “hollow” (e.g., the “overalignment” problem; Cheung & Slavin, 2016), and, thus, effects that fade quickly. Standout programs have contributed to these expectations. For example, researchers often evoke the exceptional intensity and quality of the Perry Preschool and Abecedarian early childhood programs – especially in contrast with the difficult circumstances poor children faced in the absence of intervention in the 1960s and 70s – when explaining their impressive effects on child and adult outcomes (Elango et al., 2015). It may follow that the fadeout effects observed in modern studies are caused by the implementation of lower-quality programs that provide marginal benefits that are necessarily transitory (Barnett et al., 2017; Elango et al., 2015; Stipek, 2017a).

Beyond intervention quality, the field does not lack for other potential candidates to explain heterogeneity in program effects. In Table S1, we provide direct quotes from articles that discussed many such theories, with much of the speculation appearing in “discussion sections.”

For example, the developmental timing of the intervention might be a primary source of variation worth considering. The "earlier is better" assumption has long held sway across social sciences (e.g., Heckman, 2006; Institute of Medicine & National Research Council, 2000). The intensity and duration of the intervention could also matter, as more substantial investments likely produce longer-lasting impacts (Barnett et al., 2017), and lengthier programs could potentially provide support over key 'make or break' developmental transition points (Herrera et al., 2013). Programs that provide support during after-school hours, rather than as a substitute for school time, may yield benefits by providing additional time in supportive and instructional contexts (Herrera et al., 2013).

Developmental theory has long emphasized the importance of engaging multiple systems of development, or "proximal processes," if environmental interventions are likely to take root (Bronfenbrenner & Morris, 2006). Perhaps interventions that create a "mosaic" of supports across multiple contexts or environments both inside of and outside of school (e.g., like the home environment vis à vis parents) might have stronger and longer-lasting effects by changing highly influential outside-of-school contexts in enduring ways (Garcia et al., 2020; Spoth et al., 2009; Stipek, 2017b). Likewise, targeting multiple skills through multi-component programs may be advantageous when the underlying theory of change is highly complex or when it is unclear *what* will work and *why* (Herrera et al., 2013; Klaus & Gray, 1968; Spoth et al., 2009). Alternatively, perhaps targeting the *right* skill that is foundational for subsequent developmental cascades may be more critical than breadth (e.g., targeting specific reading skills; Blachman et al., 2014). Narrowly scoped programs could also generate more replicable and consistent patterns of effects (Dynarski et al., 2013).

Finally, the broader environmental contexts that surround children targeted by the intervention could also carry weight. The famous early childhood intervention studies described above recruited children from extremely disadvantaged backgrounds during times of limited social services (Whitaker et al., 2025). Indeed, educational intervention evaluations have commonly shown that interventions targeting children from disadvantaged backgrounds produce longer-lasting effects (Watts et al., 2023; Yoshikawa et al., 2013). Changing cultural milieus could also impact persistence patterns (e.g., the legalization and decriminalization of marijuana in the case of adolescent substance-use prevention programs; Spoth et al., 2022).

Current Study

Taken together, the predictions regarding longer-term impact variation across intervention features and skills converge on a basic question that we operationalize as a forecasting problem: Using the myriad theories offered for why intervention impacts might fade or persist, can we predict a priori which interventions will produce longer-lasting effects? Because impact persistence and fadeout must be understood relative to initial impacts, we ask whether two initial intervention impacts of equal magnitude diverge over the long term based on observable intervention characteristics. We addressed these questions using meta-analytic data and straightforward prediction models, which we describe below. Because of the wide-ranging theories that have been proposed and the lack of systematic empirical investigation to date, we considered our empirical models largely exploratory. We had no confident a priori predictions about which intervention features would explain heterogeneity in follow-up impacts.

Method

Additional details regarding the data and measures can be found in the online supplementary file, and Hart et al. (2024) describe the study at length. The full codebook for the meta-analysis can also be found online: https://doi.org/10.33009/ldbase.1719529626.152e. The data and analytic code for this paper will be posted prior to publication. Here, we provide a brief overview of key details.

Data

To sample studies for inclusion in our meta-analysis, we drew from eight educational meta-analysis studies: Bailey et al. (2020); Burns et al. (2016); Kraft et al. (2018); Li et al. (2020); Protzko (2015, 2017); Suggate (2016); and Taylor et al. (2017). These meta-analyses were chosen because they covered a wide range of educational programs and interventions, including early childhood education, phonics for reading remediation, adolescent risky behaviors, social-emotional learning, and teacher professional development (among other topics). Unlike most meta-analyses, we were not interested in estimating the average effect for each intervention type. Instead, we were interested in examining longitudinal effects across a wide range of programs. The sampled meta-analyses provided 400 study reports in English.

Figure 1 presents the inclusion criteria that led to the analytic sample of effect sizes used in the current study. In an attempt to include studies with high internal validity, we only included RCTs (109 studies were excluded). We included only studies that reported effects on at least one cognitive or social-emotional outcome (broadly defined) at least 6 months after the end of the intervention and that used the same sample across time (102 studies were excluded). After eliminating a small number of studies that had no usable statistics reported (n = 5) or no educational intervention evaluated (n = 4), we were left with a sample of 85 studies with 110 unique treatment-control contrasts (i.e., “interventions”) that reported 829 end-of-treatment (i.e., “endline;” “posttest”) effect sizes and 1,459 follow-up effect sizes.

All elements of the data were double-coded by a then-doctoral student (second author) and at least one other graduate student coder. Initial reliability was established across the three primary coders, with agreement ranging from 82% to 89%. During the coding process, each paper was coded independently by two coders, and any discrepancies were then resolved via a meeting that often involved the study PI (doctoral-level researcher; first author). The coding process was extensive, with coders pulling information on study features, intervention characteristics, statistical reporting, and most importantly, effect sizes (the full coding protocol can be found here: https://doi.org/10.33009/ldbase.1719529626.152e).

Measures

Effect Sizes and Standard Errors

Our key measures of interest were the intervention impacts on child outcomes in standard deviation units, along with the corresponding standard errors to account for statistical uncertainty. Thus, for each intervention, we attempted to obtain an effect size comparable to the standardized mean difference on a given outcome between the intervention and control groups. Owing to concerns that interventions might affect the variance of scores in an intervention group, when possible, we preferred Glass's Delta, which represents the standardized mean difference in control-group standard deviations (i.e., $ES = (M_{tx} - M_{ctrl}) / SD_{ctrl}$). When coding effect sizes, we used author-reported impacts from the featured statistical model, unless the model did not allow for an interpretable average treatment effect (e.g., a model with treatment group interactions or treatment mediators). If necessary, we also used the descriptive statistics to calculate effect sizes. In some cases, we had to use extensive calculations to derive an effect size from reported statistics, including for binary outcomes (e.g., odds ratios) that had to be translated into an estimate of the standardized mean difference. In all cases, we scaled effect sizes so that higher values represented more socially desirable outcomes for the intervention group (e.g., a drop in the likelihood of drinking was rescaled as a "positive" effect).

As with effect sizes, we also coded author-reported standard errors to determine the precision of each impact estimate. However, in many cases, we calculated standard errors using descriptive statistics when no author-reported standard errors could be located (see the supplement for details). In several studies, randomization was conducted at a cluster level (e.g., schools or classrooms). If authors reported impacts from a statistical model using a cluster RCT, we assumed they used a statistical correction for clustering (e.g., hierarchical linear modeling). For the effect sizes and standard errors we calculated from descriptive statistics, we inflated the standard errors by a variance inflation factor to account for clustering (see supplemental file).

For all effect sizes, we also coded details about the construct and measure. Our preferred models included only effect sizes that could be linked using the same construct and measure over time. Thus, the key unit of observation in our analyses was an "aligned outcome," which consisted of a series of longitudinal impacts for a given intervention reported for a consistently measured construct over time (e.g., the impact for "Study A" on language ability measured via the PPVT at posttest, 1-year follow-up, 2-year follow-up, etc.). Hart et al. (2024) fit models that relaxed the grouping criteria to flexibly allow for changes in measure, subscale, and reporter across follow-up waves and found that primary estimates from the regression-based modeling (detailed below) were robust. Supplemental Figure S1 lists each intervention that contributed at least one aligned group to our sample and, hence, was included in our models. References to all reports coded to create the larger MERF sample are listed at the end of the supplement.

Intervention Characteristics

We aimed to code the salient features of interventions in our sample. For the purposes of the current paper, we focused on characteristics referenced in discussions of persistence and those consistently reported (or reasonably inferred) in studies in our sample. We coded a number of variables related to intervention structure, including: whether children in the counterfactual condition received any inputs (versus "business as usual"), the intended length of the program, and whether the program added time to a child's school experience rather than replacing time with a new curriculum or program. We additionally coded participant characteristics, including age at baseline and whether participants were included in the sample based on an ability characteristic (e.g., children with learning disabilities).

We also employed non-exclusive codes to capture the intervention targets. We coded whether the intervention targeted various child skills and proficiencies, among which language and literacy, social-emotional skills, substance use, and psychological well-being were the most common and, thus, examined in the present study. Based on these codes, we created a coarse indicator for the intervention's breadth. Programs that targeted both social-emotional and cognitive skills were considered "broad," whereas programs that targeted only one were considered "narrow." We also coded whether the intervention targeted parents.

Analysis

Because we operationalized the problem of whether intervention characteristics explain intervention effect persistence and fadeout as a forecasting problem, we fit straightforward predictive models to our meta-analytic data. These analyses were not pre-registered.

Level 1:

$$EffectSize_{fs} = \beta_{0s} + \beta_{1s} EffectSize_{ps} + \varepsilon_{fs}$$

Level 2:

$$\beta_{0s} = \gamma_{00} + u_{0s}$$

$$\beta_{1s} = \gamma_{10} + u_{1s}$$

Here, we regress follow-up effect sizes ($EffectSize_{fs}$) on posttest (i.e., endline) effect sizes ($EffectSize_{ps}$) to generate an average persistence rate (β_{1s}) across the meta-analytic sample. We term the resulting slope term as the "conditional persistence rate," which can be interpreted as the predicted follow-up effect (in SD units) that would result from a 1-SD increase in the posttest effect. The intercept term (β_{0s}) captures the extent to which factors unrelated to posttest impacts predict follow-up effects (i.e., the average follow-up effect when posttest effects are zero; "unmeasured mediators"). Because we use a 2-level random effects model to account for study-level clustering of effect sizes, β_{0s} and β_{1s} are composed of two components, respectively: 1) an average effect across all studies (i.e., γ_{00}, γ_{10}); and 2) a study-specific random effect (u_{0s}, u_{1s}).

We used this base model to then examine the extent to which observable intervention features forecast longer-term intervention effects. We examined each intervention characteristic as a binary indicator so that the resulting meta-regression main effects and interactions could be compared across diverse features (see supplement for details). For each intervention feature, we first ran "split sample" models in which we estimated the constant and slope for only interventions that fit the intervention feature of focus (e.g., only ECE programs in one model and only non-ECE programs in a second model). By doing so, we were able to examine whether persistence was systematically different across key intervention features. To test for statistical differences in persistence, we ran models using the full sample in which we entered the main effect of the intervention characteristic and the interaction between the intervention characteristic and posttest (the p-value for the interaction term reflects the difference in the slope, or "persistence" by the intervention feature).

We also fit models in which we sequentially entered all of the intervention characteristics, and their interactions with posttest, to examine how the random effect components of our model varied with the inclusion of intervention features. By doing so, we were able to assess the extent to which each intervention feature explained variance in follow-up effects, over and above what was explained by the posttest impact magnitude alone.

All meta-regression models were computed in R using the *metafor* package, with inverse variance weights (Viechtbauer, 2010). We used the *clubsandwich* package to then employ robust variance estimation with study-level clustering given the non-independence of impacts (i.e., multiple impacts per study) with small sample size ("CR2") adjustments (Hedges, Tipton, & Johson, 2010; Pustejovsky, 2023; Pustejovsky & Tipton, 2018; Tipton, 2015). In sensitivity analyses, we combined RVE with the Correlated-and-Hierarchical (CHE) modeling approach (Pustejovsky & Tipton, 2022) that further accounts for the non-independence of effects by modeling within-study variance and assuming that effects from the same study were correlated at r = .60 (see supplemental file).

Results

Table 1 provides descriptive statistics for the sample of intervention RCTs included in our meta-analytic dataset. At the top of the table, we present the intervention features that we used in our analytic models to examine predictors of heterogeneity in persistence. As noted above, we limited our analyses to only interventions that provided at least one "aligned group," in which impacts were provided on a consistently measured construct at posttest and at least one follow-up. Of note, the ages of children targeted by the interventions in our sample ranged from 0 months old (i.e., recruitment at birth) to 14 years old (M = 94.67 months). On average, interventions lasted approximately 7.40 months, though this varied considerably (range: 1 to 36 months). The interventions also varied in their apparent theories of change and targeted skills. Approximately 53% of interventions targeted language or literacy skills, while 52% targeted social-emotional functioning. A smaller set of interventions (22%) attempted to involve parents, and 14% added instructional time to student experiences. Table S2 presents the correlations among intervention characteristics. Some intervention features were highly correlated (r = -.77 to .64).

The second panel of Table 1 also provides additional details regarding the sample. The average study had just over 5 aligned outcomes (i.e., a series of impacts on consistently measured constructs) that we were able to include in our analyses. Unfortunately, key demographic characteristics were reported inconsistently or not at all, which prevented us from including sample-level demographic indicators as potential predictors of heterogeneity. Importantly, the interventions included in our sample reported impacts on a range of skills (see Hart et al., 2024 for additional details). 51% of the "aligned groups" in our sample were cognitive skills (e.g., language, math, IQ), 46% were social-emotional (e.g., externalizing, internalizing, substance use), and 2% were undefined (i.e., measures that could not be categorized; see supplement for details).

Figure 2 provides meta-analytic averages for the aligned groups in our data over time (see also Table S3 and Figure S2). In our data, the average posttest impact for any outcome with a linked follow-up impact was 0.30 SDs ($p < .001$). To provide longitudinal meta-analytic averages, we further binned the data at various follow-up points. At 6 to 12 months after the posttest, the average impact fell to .21 SDs ($p < .001$), and further to .07 SDs ($p < .01$) at follow-up of over 1 to 2 years. The figure shows a relatively stable asymptote for further follow-ups, hovering around .10 SDs, which was imprecisely estimated due to the lack of studies reporting follow-ups several years past intervention end. Importantly, sample selection appeared to play an increasingly important role as follow-up extended; studies reporting follow-up at later timepoints had larger initial effect sizes than those that stopped reporting (see Table S2). Given selection issues and the lack of data at further follow-up assessment waves, we present analyses for only the 6- to 12-month follow-up period (see supplement for 1- to 2-year results).

Table 2 provides average effect sizes at posttest and 6- to 12-months follow-up, split across each intervention feature. With the exception of substance-use interventions, every category shows some degree of fadeout. However, if fadeout is understood relative to initial effect size, interventions that produced smaller initial effects tended to produce more persistence. For example, interventions that targeted older children (i.e., non-ECE interventions) produced initial effects of .21 SDs ($p < .001$), and this faded to only .19 SDs ($p < .01$) by follow-up. In contrast, ECE interventions produced initial effects of .42 ($p < .001$) that faded to .24 at follow-up ($p < .001$). A similar pattern of smaller initial effects followed by more persistence was observed for interventions targeting social-emotional skills, interventions begun after the year 2000, interventions targeting substance use, and interventions that did not include selection criteria (i.e., specific screening of children).

Although these descriptive patterns are interesting, they do not directly answer our key questions, as they do not fully take advantage of the aligned outcomes by asking how each respective impact persists or fades over time (e.g., the impacts driving up the average at 6 to 12 months for one type of intervention may not be the same impacts that drove up the average at posttest). Moreover, simply observing meta-analytic averages fails to equate posttest impacts of different magnitudes.

Consequently, in Table 3 we present our key meta-regression results, which are also displayed in Figure 3. We found that in our bivariate forecasting model, posttest impacts predicted 6- to 12-month follow-up impacts with a slope term of .46 (SE = .05; $p < .001$). This indicated that educational intervention impacts persisted at a rate of 46% by 6-12 months follow-up (or faded by 56%). We observed a small, non-zero, intercept effect (.06, SE = .02, p = .003), indicating that, on average, a small portion of the meta-analytic follow-up effect was unexplained by posttest impacts on the same skill.

We next turned to our tests of heterogeneity by intervention features. In Table 3, we present "split sample" results for each intervention feature. These results show the resulting slope (i.e., persistence rate) and constant term for each set of aligned groups reported for the various intervention types. In the p-value column, we provide results from a series of models that tested the main effect and interaction with posttest impact for each intervention feature. Figure 3 then depicts the interaction effects as the gray lines that vary around the overall slope of .46 for the full sample. As Table 3 and Figure 3 reflect, we did not observe full persistence for any class of intervention. Modeled persistence rates varied from .15 ($p = .28$) for interventions that broadly targeted multiple constructs to .56 ($p < .001$) for interventions that targeted older children (i.e., non-ECE studies).

As depicted in Panel B of Figure 3, two sets of interactions produced statistically significant effects: ECE vs. non-ECE interventions and "broad" vs. "narrow" interventions. We found that non-ECE interventions produced a higher persistence rate than ECE interventions ($p = .03$), and interventions that narrowly targeted one domain of child development also produced a higher persistence rate than interventions that were broader in focus (p = .05). Notably, ECE studies generated a larger constant term, suggesting a higher portion of the follow-up effects for ECE studies was unaccounted for by posttest effect sizes. Given the number of exploratory tests we ran, none of these interaction terms would withstand a conservative adjustment for multiple testing (e.g., a Bonferroni alpha for the interaction terms would be .004 (.05/12 interaction tests)).

Finally, Panel A of Table 4 demonstrates the extent to which intervention features explained heterogeneity in follow-up effects and persistence rates. When no predictors were included in the meta-regression model, follow-up effects varied substantially across studies (τ_{0s}= .23 SDs). When posttest impacts were entered as predictors, between-study variation dropped by over half (τ_{0s} = .11 SDs), and we observed substantial variation in the persistence rate across studies (τ_{1s} = .24 SDs). However, when we added the set of intervention features (Column 3), and their interactions with posttest (Column 4), we found no reduction of the intercept (τ_{0s} = .12 SDs), and only a minimal reduction of the conditional persistence rate (τ_{1s} = .18 SDs). Finally, we adopted a "kitchen sink" approach, also including outcome-related factors (see the supplement for a full list). Again, variance estimates did not decrease. Panel B of Table 4 provides the estimates from the CHE model that additionally estimated within-study variance, showing that we observed considerable within-study variation in the intercepts across outcomes, in some cases more than the variation estimated at the between-study level. Again, this was minimally explained by intervention and outcome features (initial τ = .04, final τ = .03).

Supplemental Analyses

1- to 2-year follow-up results are presented in the supplement in Tables S4 and S5. Results for the 1- to 2-year follow-up period included far fewer aligned groups (n = 90), making estimates much noisier. Although a few interactions met statistical significance thresholds, they did not align with any of the interactions observed for 6 to 12 months. Notably, in contrast to what was observed at 6- to 12-months follow-up, ECE interventions actually showed stronger persistence than non-ECE interventions (p = .09). The supplemental text details the meta-analytic averages and persistence produced by the CHE model, suggesting that the overall estimates of persistence at 1- to 2- ear follow-up for the full sample were less robust than those for 6- to 12-months follow-up (which were highly robust), casting further caution in interpreting 1- to 2-year results. We additionally tested the robustness of our 6- to 12-month persistence estimates using sample size weighting as an alternative method that relies on less computational complexity than our standard errors did (see Hart et al., 2024 for details). As reported in Table S6, the overall intercept and slope (slope = .46, intercept = .08) were very similar to the estimates from our primary model. Using sample-size weighting, we only observed one statistically significant difference across intervention characteristics: narrow interventions showed greater conditional persistence, as measured by the slope, than broad interventions.

The supplement also includes a forest plot of average posttest impacts for each intervention (Figure S1) and examinations of the reliability of this literature, including p-curves and funnel plots (see Figure S3 and S4). In general, we did not observe substantial evidence of p-hacking at posttest, 6-12 month follow-up, or 1-2year follow-up. However, as reported in the supplemental text, we observed that impacts with larger standard errors were also larger in magnitude. Of note, it could be the case that smaller interventions were able to provide higher intensity programming that produced larger effect sizes.

Discussion

Although educational interventions are often implemented with the goal of affecting child skills and behaviors beyond the end of the intervention, relatively few rigorous evaluations of report follow-up effects (Watts et al., 2019). When effects have been reported, like in the case of ECE interventions, many studies have reported substantial fadeout, whereby initial benefits for the intervention group diminish in the years following the end of the intervention (Bailey et al., 2017). Although educational researchers have produced convincing theories describing when effects might persist or fade (e.g., Bailey et al., 2020), the field has relatively few systematic examinations of educational intervention fadeout. Moreover, despite a proliferation of theories regarding what interventions will produce more persistent effects, it remains unclear whether we can predict a priori which interventions will persist or fade. In the current study, we used meta-analytic data to operationalize this issue as a forecasting problem. We posed a simple question: could we predict whether two interventions that produced initial effects of the same size would persist or fade at different rates based on observable factors of the interventions?

Looking broadly across a diverse set of educational interventions, we found that while posttest impacts were strong forecasters of follow-up impacts, initial intervention impacts faded considerably. Our predictive models suggested that by the 6- to 12-month follow-up, only about 46% of initial impacts persisted. Although we observed substantial heterogeneity in persistence rates, after accounting for posttest effect size, salient intervention features explained very little of this heterogeneity. Thus, if a researcher knew an intervention's posttest effects, having the additional information included in our meta-analysis would be unhelpful in further predicting follow-up effects.

Our results suggest that fadeout, at least in the short-term, is likely to be ubiquitous across broad categories of educational interventions. If one were forecasting what the 1-year follow-up effects for a given intervention might be, a reasonable "best" guess would probably come by just cutting the post-test effect in half. More formally, you could multiply the effect by .46 before .06 (i.e., the intercept term; the "unmeasured mediator" effect). However, some important caveats are worth noting. As our descriptive findings suggested, many of the interventions that produced smaller initial effects (e.g., interventions targeting older children, interventions targeting social-emotional skills) also produced less absolute fadeout (Table 2). While interesting, it would be misleading to use these descriptive findings to assert that such interventions produce persistent effects. Because posttest and follow-up effect sizes are strongly related, accounting for the posttest effect size is crucial when making inferences across different types of interventions. Thus, our regression results suggest that once we place, say, interventions targeting social-emotional skills and interventions targeting cognitive skills on similar footing by controlling for initial posttest magnitude, they produce very similar persistence rates at one year follow-up. In other words, our results imply that increasing the posttest effect size for the interventions that produced smaller initial effects would likely lead to more fadeout.

Our regression results provide some clues as to why we observed that smaller initial effects tended to persist more in absolute terms. Note that in our overall meta-regression model, we observed a slope of approximately .46 and a constant term of .06. Recall that the constant term can be interpreted as the average predicted follow-up effect when posttest effects are zero. In other words, this term represents the potential for interventions to affect longer-term outcomes through channels that are not captured by posttest impacts on the same outcome. If one were to use our regression model to forecast follow-up impacts for medium- to large initial effects (e.g., .30 *SD*), they would find that the constant term has little bearing on the forecasted effect (i.e., the slope term dictates fadeout). However, for smaller initial effects (e.g., .15 *SD*), the constant term will constitute a larger proportion of the forecasted effect (see further explanation in Hart et al., 2024). Although it is somewhat unsatisfying that we have not identified meta-analytic moderators that account for different persistence rates across interventions, the nonzero intercept may be practically useful. One possible interpretation is that, on average, interventions affect outcomes through multiple pathways, including measured and unmeasured ones. Influential intercept terms indicate that auto-regressive skill-building processes (i.e., “skills beget skills”) are unlikely to be the sole mediators of persistent effects. Insofar as the intercept captures idiosyncratic study-specific factors and error, persistence by way of the constant term may be less likely to replicate when taken to new contexts or when implemented at scale.

Why did the intervention features provide such little predictive validity, even when we observed substantial heterogeneity in follow-up effects? Maybe our list of intervention features was simply not up to the task. Perhaps the field has identified other observable characteristics of interventions that would have fared better in our empirical tests. Yet, a simple descriptive look at our data suggests we were unlikely to find a way to split the data that would have produced a group of interventions producing persistent effects. To gauge the prevalence of intervention impacts that did not fade, we counted the number of observed intervention impacts above 0.20 SDs at posttest that fully persisted by 1-year follow-up.[iii] Only 19% (i.e., 37/199) of intervention effects larger than .20 did not decrease by one year. These intervention impacts are listed in Table S7 of the supplement. When reading through the list, it is hard to identify any pattern that might explain what links this group of impacts. Moreover, although we may have missed important intervention features, any unobserved features correlated with our measured set should also have contributed to the decomposition effects shown in Table 4. If, say, ECE interventions differ from non-ECE interventions in ways we did not measure, those differences should still be reflected in our decomposition efforts. Indeed, Table S2 suggests that some intervention features correlate quite strongly, indicating that these features were not merely picking up unsystematic differences between studies.

Unfortunately, our intervention science may simply lack strong theoretical models that can be systematically falsified. In our data, substantial heterogeneity remained, even after accounting for the post-test effect size. However, such heterogeneity might be idiosyncratic and difficult to reproduce. Although some individual studies have found more persistent effects on the basis of differences in participants, treatments, and settings (e.g., Mattera et al., 2025), our results suggest that we will need to turn these post-hoc tests into a priori predictions to know if such sources of heterogeneity can be used to guide intervention development moving forward. Some researchers may be disappointed that we did not examine more specific and sophisticated skill-building theories (e.g., perhaps complex interactions between skills and environments produce persistent impacts). Although highly complex theories may better capture how development unfolds, the field largely lacks the data available to test these theories. At present, we lack a rich understanding of the basic developmental processes that our educational interventions often target. However, it is worth noting that some moderation analyses were significantly underpowered (e.g., only 5% of aligned groups were from studies that added school time); future meta-analytic replication work in larger samples will be an important next step.

Finally, we should caution that evidence of fadeout is not a reason to “give up” on educational interventions. Bailey et al. (2024) describe how short-term fading effects on targeted skills could lead to longer-term effects on key adult outcomes, and at least several interventions have found follow-up impacts even when no medium-term effects on observed skills were detected (e.g., Gray-Lobe et al., 2022). Yet completely dismissing fadeout will not help us move the field forward, nor should we assume that fadeout is a “problem” that only plagues “bad” interventions. Instead, our results suggest that most educational interventions should expect to observe fadeout at follow-up, making the promise of persistent impacts on targeted skills unlikely. Instead of dismissing evidence of fadeout with post hoc explanations, we should use our current science of educational interventions to improve our theories of skill building using a priori predictions and experimental testing, and we should rethink why educational investments might provide longer-term benefits for children by means other than persistent skill impacts.

Finally, our results also call for epistemic humility; identifying generalizable developmental phenomena that can be tested and that can inform intervention science is no easy task. That follow-ups are rarely conducted (Watts et al., 2019), and that firm a priori predictions are rarely made, makes it even more challenging for the field to progress. If we are to identify intervention features that make a difference for intervention impact persistence, following up on initial intervention impacts and testing solidified a priori theories will be critical.

References

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.

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.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*(2), 55-97.

Bailey, D. H., Watts, T. W., Hart, E. R., & Yu, M. J. (2024). Learning about development from interventions. *Annual Review of Developmental Psychology*, *6*(1), 251-272.

Barnett, W. S., & Frede, E. C. (2017). Long-term effects of a system of high-quality universal preschool education in the United States. In *Childcare, Early Education and Social Inequality* (pp. 152-172). Edward Elgar Publishing.

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

Bornstein, M. H. (1989). Sensitive periods in development: structural characteristics and causal interpretations. *Psychological Bulletin*, *105*(2), 179.

Bronfenbrenner U, Morris PA. 2006. The bioecological model of human development. In *Handbook of Child Psychology*, *Vol. 1: Theoretical Models of Human Development*, ed. RM Lerner, W Damon, pp. 793–828. Hoboken, NJ: Wiley. 6th ed.

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.

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283-292.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31-47.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, *1*(3), 111-134.

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.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405-432.

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, *32*(4), 692-717. https://doi.org/10.1002/pam.21715

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2015). *Early childhood education* (NBER Working Paper No. 21766). Cambridge, MA: National Bureau of Economic Research.

Garcia, I., Grossman, J. B., Herrera, C. Linden, L. L. (2020). *The Impact of an Intensive Year-Round Middle School Program on College Attendance.* MDRC.

Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, *138*(1), 363-411.

Hart, E. R., Bailey, D. H., Luo, S., Sengupta, P., & Watts, T. W. (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. *Psychological Bulletin*, *150*(10), 1207–1236. https://doi.org/10.1037/bul0000450

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900-1902.

Heckman, J. J. & Kautz, T. (2013). *Fostering and Measuring Skills: Interventions That Improve Character and Cognition.* NBER Working Paper 19656. https://doi.org/10.3386/w19656.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Herrera, C., Grossman, J. B., Linen, L. L. (2013). *Staying on track: Testing Higher Achievement's long-term impact on academic outcomes and high school choice.* MDRC.

Institute of Medicine & National Research Council. (2000). *From neurons to neighborhoods: The science of early childhood development.* National Academies Press. https://doi.org/10.17226/9824

Klaus R. A., & Gray S. W. (1986). The Early Training Project for disadvantaged children: A report after five years. *Monographs of the Society in Child Development, 33*(4), 1-66.

Koepp, A. E., Watts, T. W., Gershoff, E. T., Ahmed, S. F., Davis-Kean, P., Duncan, G. J., ... & Vandell, D. L. (2023). Attention and behavior problems in childhood predict adult financial status, health, and criminal activity: A conceptual replication and extension of Moffitt et al.(2011) using cohorts from the United States and the United Kingdom. *Developmental Psychology*, *59*(8), 1389.

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

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491-495.

Mattera, S., Jacob, R., Morris-Perez, P., & Asher, C. (2025). The Impact of Making PreK Count on Short and Longer-Term Outcomes. *Journal of Research on Educational Effectiveness*, 1-28.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693-2698.

Pustejovsky, J. E. (2023). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections.* R package version 0.5.10. https://CRAN.R-project.org/package=clubSandwich

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics, 36*(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004

Pustejovsky, J., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*. https://doi.org/10.1007/s11121-021-01246-3

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science, 24*(7), 1301-1308.

Sala, G., & Gobet, F. (2019). Cognitive training does not enhance general cognition. *Trends in Cognitive Sciences*, *23*(1), 9-20.

Spoth, R., Trudeau, L., Guyll, M., Shin, C., & Redmond, C. (2009). Universal intervention effects on substance use among young adults mediated by delayed adolescent substance initiation. *Journal of Consulting and Clinical Psychology*, *77*(4), 620. https://doi.org/10.1037/a0016029

Spoth, R., Redmond, C., Shin, C., Trudeau, L., Greenberg, M. T., Feinberg, M. E., & Welsh, J. (2022). Applying the PROSPER prevention delivery system with middle schools: Emerging adulthood effects on substance misuse and conduct problem behaviors through 14 years past baseline. *Child Development*, *93*(4), 925-940. https://doi.org/10.1111/cdev.13746

Stipek, D. (2017a). *The value of early education: Why the fade-out effect does not mean we should give up on preschool.* DREME. https://dreme.stanford.edu/news/the-value-of-early-education-why-the-fade-out-effect-does-not-mean-we-should-give-up-on-preschool/

Stipek, D. (2017b). *The preschool fade-out effect is not inevitable.* Education Week. https://www.edweek.org/teaching-learning/opinion-the-preschool-fade-out-effect-is-not-inevitable/2017/03

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3), 375–393. https://doi.org/10.1037/met0000011

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, *43*(7), 352-360.

Watts, T. W. (2020). Academic achievement and economic attainment: Reexamining associations between test scores and long-run earnings. *Aera Open*, *6*(2), 2332858420928985.

Watts, T. W., Jenkins, J. M., Dodge, K. A., Carr, R. C., Sauval, M., Bai, Y., ... & Ananat, E. (2023). Understanding heterogeneity in the impact of public preschool programs. *Monographs of the Society for Research in Child Development*, *88*(1), 7-182.

Whitaker, A. A., Burchinal, M., Jenkins, J. M., Bailey, D. H., Watts, T. W., Duncan, G. J., ... & Peisner-Feinberg, E. (2026). Why are preschool programs becoming less effective? *Journal of Policy Analysis and Management*, *45*(1), e70031.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., ... & Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education. *Society for Research in Child Development*.

Endnotes

i. The annual budget from IES hovered around $800M between 2002 and 2024 (see https://usafacts.org/explainers/what-does-the-us-government-do/subagency/institute-of-education-sciences/).

ii. We view fadeout as wholly consistent with these broad theoretical claims and empirical regularities. Such claims are rarely accompanied by explicit quantitative predictions, but attempts to translate these theories into such predictions frequently result in models that imply that some fadeout is likely (Bailey et al., 2020; Campbell and Frey, 1970).

iii. Many of the intervention impacts in our dataset were near zero at both posttest and follow-up, so we set an arbitrary threshold of 0.20 for determining which posttest effects were large enough to warrant examination in this descriptive exercise. Kraft et al. (2020) argue that effects bigger than 0.20 should be considered “large” in educational research, and we agree. Yet, the discussion about fadeout and persistence hinges upon the presence of an effect large enough at posttest that we can reliably capture changes to the effect size at follow-up. Although we agree that many worthwhile interventions may produce effects that are much smaller than 0.20, our studies are almost never powered to detect small changes near zero.

Figure 1
Flow of Reports and Studies into the Meta-Analytic Sample

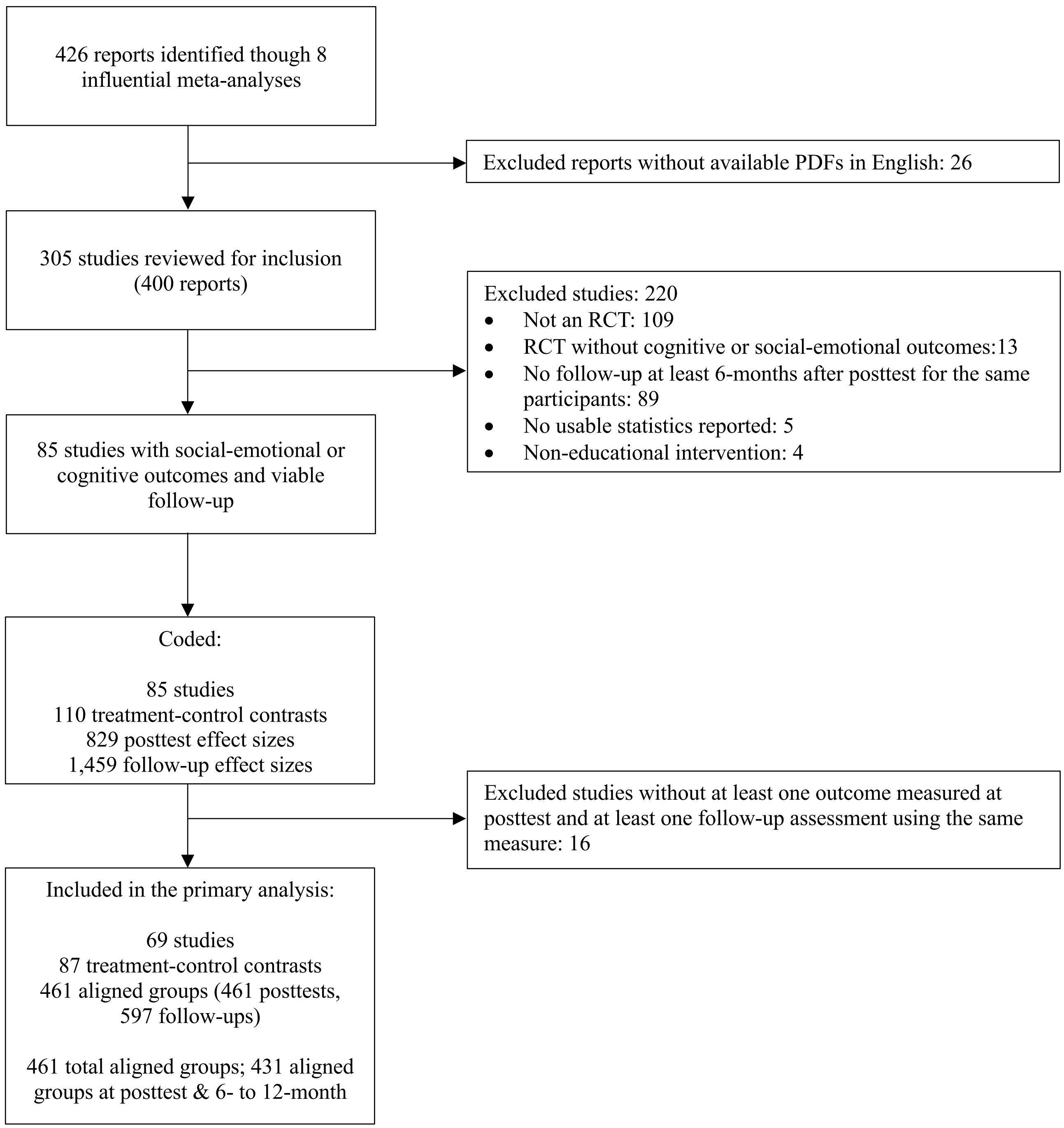


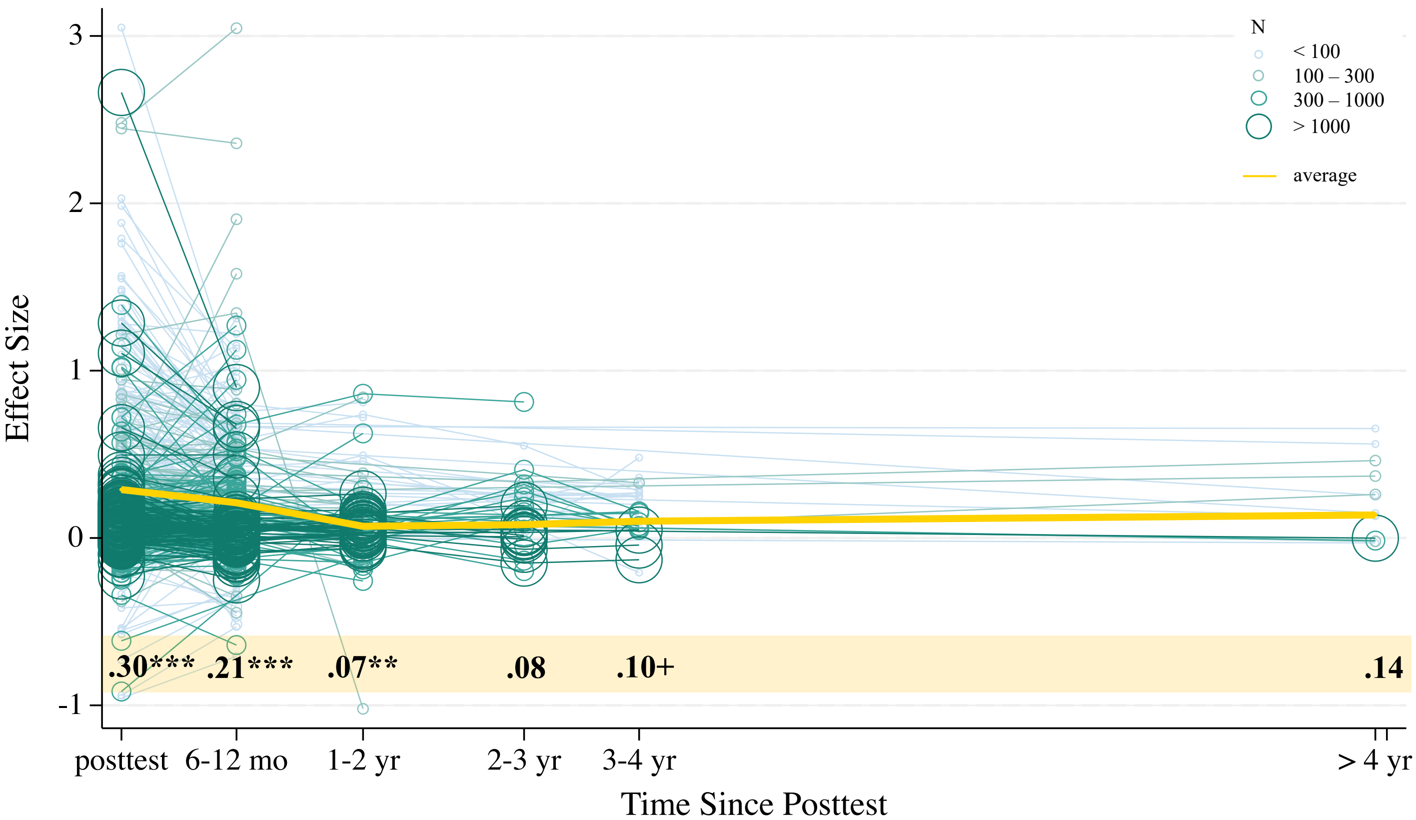
Table 1
Intervention, Participant, and Outcome Characteristics

	M / %	Min	Max	Obs	% Nonmissing
Focal Intervention Characteristics					
Baseline Age (months)	94.67	0	170	86	99%
Treatment Duration (months)	7.40	1	36	71	82%
Baseline Treatment Year	1998.62	1962	2013	86	99%
Parent Involvement	21.84	0	100	87	100%
Time in School	13.95	0	100	86	99%
Broad Treatment	12.64	0	100	87	100%
Sample Selection	35.63	0	100	87	100%
Counterfactual Provided Something	29.63	0	100	81	93%
Language/Literacy Targeted	52.87	0	100	87	100%
Social-Emotional Skills Targeted	51.72	0	100	87	100%
Substance Use Targeted	12.64	0	100	87	100%
Psychological Wellbeing Targeted	10.34	0	100	87	100%
Additional Intervention Characteristics					
Aligned Groups n (at posttest)	5.30	1	33	87	100%
Participant n (at posttest)	722.59	24	10170	87	100%
Other Intervention Targets					
Teachers Targeted	55.17	0	100	87	100%
Math Targeted	8.05	0	100	87	100%
Science Targeted	1.15	0	100	87	100%
General Cognition Targeted	5.75	0	100	87	100%
Executive Functioning Targeted	1.15	0	100	87	100%
Learning Skills Targeted	1.15	0	100	87	100%
Participant Demographics					
Asian	13.99	2	45	15	17%
Black	41.34	2	100	42	48%
White	54.85	2	98	37	43%
Hispanic	25.29	2	100	29	33%
Female	47.09	26	100	73	84%
Intervention Location					
United States	60.92	0	100	87	100%
Europe	22.99	0	100	87	100%
Oceania	10.34	0	100	87	100%
Other	5.75	0	100	87	100%
Key Outcome Characteristics					
Cognitive	51.19	0	100	461	100%
Social-emotional	46.42	0	100	461	100%
Undefined	2.39	0	100	461	100%

Note: Intervention-level characteristics for all interventions that contained at least one aligned group (same construct measured consistently at posttest and follow-up) at any follow-up wave are presented.

Figure 2

Fadeout Trajectories for Aligned Groups



+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: Each line represents a construct that was measured at posttest and at least one follow-up assessment using the same measure for the same treatment-control group contrast (i.e., aligned group). The average effect size trajectory is displayed in gold with corresponding meta-analytic averages at each assessment wave provided in the yellow-shaded box. Meta-analytic averages were computed using a random effects meta-analytic model that included a random effect for study, weighting by the inverse variances, and robust standard errors (clustered at the study level). As detailed in the key, coordinates were weighted by the posttest sample size (larger circles represent estimates from larger samples) with aligned color-coding (darker colors represent larger sample sizes). The points on the x-axis are spaced according to the average time since posttest for each respective bin (e.g., for 6-12 months follow-up, the average time since posttest was 10.27 months). For display purposes, one aligned group with a posttest effect size less than -1 is not presented in this figure.

Table 2

Average Effect Sizes at Posttest and 6- to 12-month Follow-up Assessments

	Moderator = 1		**Moderator = 0**		
	Avg. ES (SE) (1)	**% in group (2)**	**Avg. ES (SE) (3)**	**% in group (4)**	**Total Obs. (5)**
ECE	***ECE***		***Non-ECE***		
Posttest	0.42 (0.08)***	41%	0.21 (0.05)***	59%	425
6- to 12- month Follow-up	0.24 (0.04)***		0.19 (0.05)**		
TX Duration	***Long***		***Short***		
Posttest	0.26 (0.06)***	52%	0.38 (0.08)***	48%	341
6- to 12- month Follow-up	0.16 (0.03)***		0.25 (0.05)***		
Parent Involvement	***Parents***		***No Parents***		
Posttest	0.26 (0.10)*	18%	0.31 (0.05)***	82%	431
6- to 12- month Follow-up	0.13 (0.04)**		0.22 (0.04)***		
Time in School	***More Time***		***Curricular***		
Posttest	0.41 (0.16)*		0.29 (0.05)***	95%	431
6- to 12- month Follow-up	0.13 (0.06)+	5%	0.22 (0.04)***		
TX year > 2000	***> 2000***		***<= 2000***		
Posttest	0.23 (0.05)***	50%	0.40 (0.08)***	50%	431
6- to 12- month Follow-up	0.22 (0.05)***		0.21 (0.04)***		
Broad TX	***Broad***		***Narrow***		
Posttest	0.35 (0.16)+	9%	0.30 (0.05)***	91%	431
6- to 12- month Follow-up	0.10 (0.06)		0.23 (0.04)***		
Sample Selection	***Selection Criteria***		***No Criteria***		
Posttest	0.50 (0.09)***	39%	0.21 (0.04)***	61%	431
6- to 12- month Follow-up	0.30 (0.04)***		0.16 (0.04)***		
Counterfactual	***Something Provided***		***Nothing Provided***		
Posttest	0.26 (0.08)**	27%	0.33 (0.06)***	73%	408
6- to 12- month Follow-up	0.16 (0.04)**		0.23 (0.04)***		
Language/Literacy	***Targeted***		***Not Targeted***		
Posttest	0.43 (0.07)***	53%	0.19 (0.05)***	47%	431
6- to 12- month Follow-up	0.25 (0.04)***		0.18 (0.05)**		
Social-Emotional Skills	***Targeted***		***Not Targeted***		
Posttest	0.20 (0.05)***	55%	0.43 (0.07)***	45%	431
6- to 12- month Follow-up	0.18 (0.05)**		0.25 (0.04)***		
Substance Use	***Targeted***		***Not Targeted***		
Posttest	0.15 (0.06)*	16%	0.33 (0.05)***	84%	431
6- to 12- month Follow-up	0.16 (0.09)		0.22 (0.04)***		
Psych Wellbeing	***Targeted***		***Not Targeted***		
Posttest	0.12 (0.03)*	13%	0.32 (0.05)***	87%	431
6- to 12- month Follow-up	0.05 (0.02)		0.23 (0.04)***		
Full Sample Posttest Effect: 0.31 (0.04)***					431
Full Sample 6- to 12-month Follow-up Effect: 0.21 (0.03)***					

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: This table presents the meta-analytic averages of effects measured at the posttest and 6- to 12-month follow-up waves from our primary random-effects meta-regression with random slopes and intercepts for study, weights for effect precision, and robust standard errors (with study-level clustering). Each unit of observation is one aligned group, in which the same construct was assessed with the same measure at posttest and follow-up within the same study at 6- to 12-months follow-up specifically.

Table 3

Persistence Rates by Intervention-Related Characteristic at 6- to 12-months Follow-up (Est (SE))

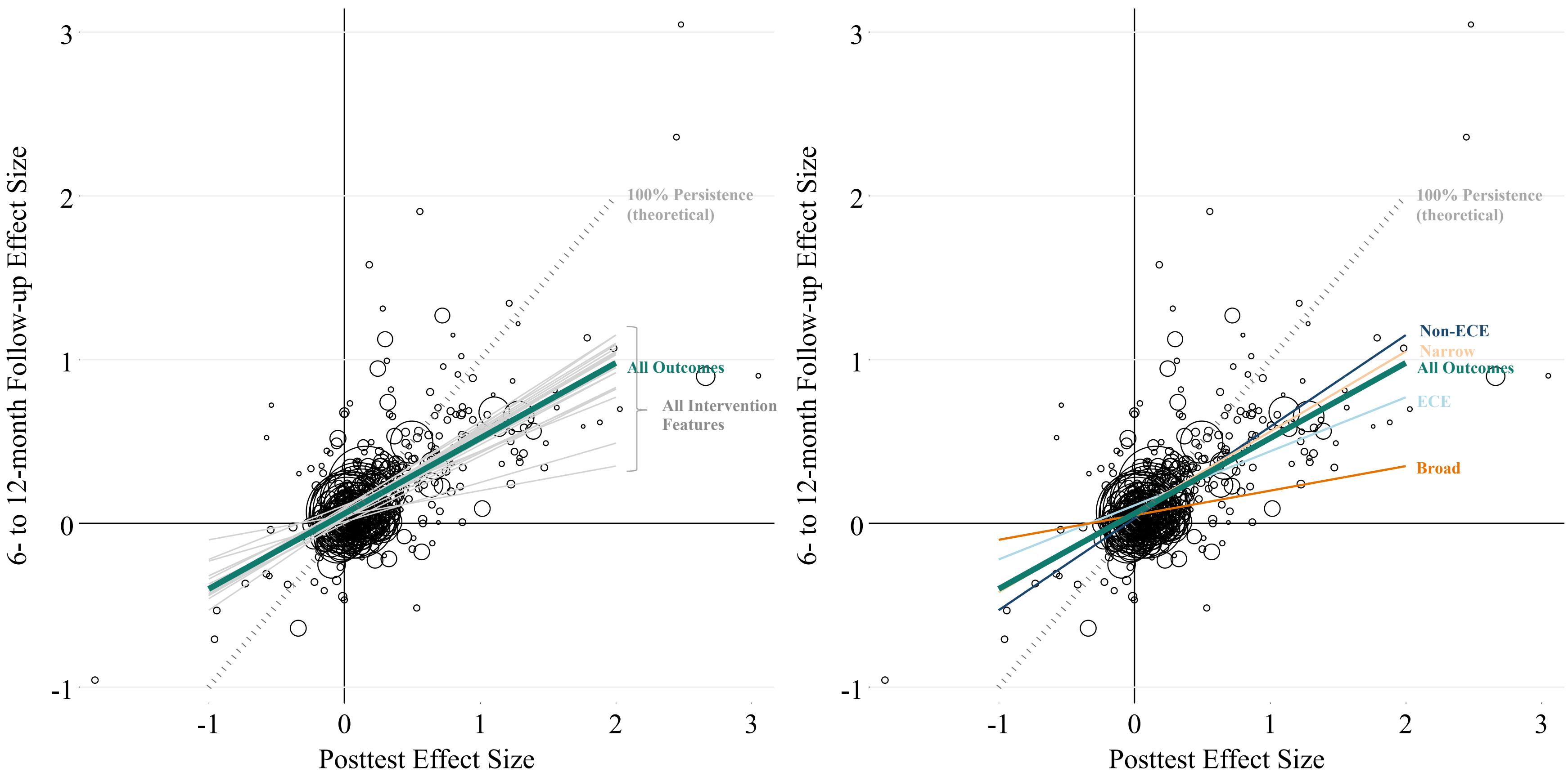
	Split Sample Models		**Interaction Models**	
	Moderator=1	**Moderator=0**	**p-value**	**Obs**
	(1)	**(2)**	**(3)**	**(4)**
ECE	***ECE***	***Non-ECE***		
Constant	0.11 (0.03)**	0.03 (0.02)	0.07 +	425
Posttest Effect Size	0.33 (0.07)***	0.56 (0.07)***	0.03 *	
Observations	176	249		
TX Duration	***Long***	***Short***		
Constant	0.03 (0.00)**	0.07 (0.04)	0.94	341
Posttest Effect Size	0.47 (0.08)***	0.51 (0.06)***	0.31	
Observations	176	165		
Parent Involvement	***Parents***	***No Parents***		
Constant	0.03 (0.00)*	0.07 (0.03)*	0.96	431
Posttest Effect Size	0.40 (0.10)**	0.48 (0.06)***	0.38	
Observations	78	353		
Time in School	***More Time***	***Curricular***		
Constant	0.01 (0.01)	0.07 (0.02)**	0.49	431
Posttest Effect Size	0.24 (0.11)	0.48 (0.05)***	0.11	
Observations	22	409		
TX year > 2000	***> 2000***	***<= 2000***		
Constant	0.08 (0.03)*	0.04 (0.03)	0.39	431
Posttest Effect Size	0.48 (0.09)***	0.44 (0.05)***	0.74	
Observations	216	215		
Broad TX	***Broad***	***Narrow***		
Constant	0.05 (0.03)	0.07 (0.02)**	0.77	431
Posttest Effect Size	0.15 (0.10)	0.49 (0.05)***	0.05 +	
Observations	40	391		
Sample Selection	***Selection Criteria***	***No Criteria***		
Constant	0.10 (0.05)*	0.06 (0.02)*	0.59	431
Posttest Effect Size	0.50 (0.08)***	0.38 (0.07)***	0.19	
Observations	168	263		
Counterfactual	***Something Provided***	***Nothing Provided***		
Constant	0.04 (0.03)	0.06 (0.02)**	0.57	408
Posttest Effect Size	0.50 (0.02)**	0.45 (0.07)***	0.61	
Observations	111	297		
Language/Literacy	***Targeted***	***Not Targeted***		
Constant	0.08 (0.03)*	0.05 (0.03)*	0.57	431
Posttest Effect Size	0.42 (0.05)***	0.45 (0.09)***	0.90	
Observations	227	204		
Social-Emotional Skills	***Targeted***	***Not Targeted***		
Constant	0.06 (0.02)*	0.07 (0.04)+	0.99	431
Posttest Effect Size	0.38 (0.08)***	0.50 (0.06)***	0.24	
Observations	237	194		
Substance Use	***Targeted***	***Not Targeted***		
Constant	0.09 (0.06)	0.06 (0.02)**	0.68	431
Posttest Effect Size	0.53 (0.19)+	0.45 (0.05)***	0.77	
Observations	67	364		
Psych Wellbeing	***Targeted***	***Not Targeted***		
Constant	0.00 (0.02)	0.07 (0.02)**	0.14	431
Posttest Effect Size	0.41 (0.24)	0.46 (0.05)***	0.83	
Observations	58	373		
Full Sample Constant: 0.06 (0.02)** **Full Sample Posttest Effect Size (Slope):** 0.46 (0.05) ***				431

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: This table presents estimates from our primary random-effects meta-regression with random slopes and intercepts for study, inverse variance weights, and robust standard errors (with study-level clustering). Aligned groups are the unit of observation. Columns 1 and 2 present estimates from split-sample models, in which the persistence rate and constant were estimated using the sample of results for the intervention feature of focus. In Columns 3 and 4, estimates of the statistical difference in rates and constants are reported, from model that tested the main effect and interaction with posttest impact for each intervention feature.

Figure 3

Persistence Patterns: Posttest to 6- to 12-month Follow-up



Note: Each coordinate represents an impact on an outcome measured consistently at posttest and 6- to 12-months follow-up. Coordinates were weighted by the inverse variances such that larger coordinates reflect estimates with smaller standard errors at 6- to 12-months follow-up. The hypothetical "100% Persistence" line was included as a gray dashed line. The other lines reflect the estimated association between posttest and follow-up effects (see Table 3). The figure in Panel A (left) shows the slope (i.e., persistence rate) and constant term for each set of aligned groups reported for the various intervention types. The estimated slope and intercept for all outcomes are presented in green, and the estimates for all of the intervention features are presented in grey. The figure in Panel B (right) depicts the associations for broad versus narrow and early versus later childhood interventions, the two intervention features for which conditional persistence rates were at least marginally statistically significantly different. For display purposes, one aligned group with a posttest effect size less than -1 is not presented.

Table 4

Variance Decomposition for 6- to 12-month Follow-up Effects

	Null (1)	**Base (2)**	**Main Effects (3)**	**Interactions (4)**	**"Kitchen Sink" (5)**
Panel A: Random Effects Models					
Study-level Random Intercept (τ_0)	0.23	0.11	0.12	0.12	0.13
Study-level Random Slope (τ_1)	--	0.24	0.26	0.18	0.22
I^2 (%)	60.81	35.14	9.38	0	0
Panel B: Correlated-and-Hierarchical Model					
Study-level Random Intercept (τ_{B0})	0.20	0.00	0.05	0.05	0.09
Study-level Random Slope (τ_{B1})	--	0.19	0.28	0.23	0.24
Within-study Random Intercept (τ_{W0})	0.04	0.09	0.03	0.03	0.03
I^2 (%)	72.63	55.79	48.36	38.24	30.37
Posttest Effect Size		x	x	x	x
Intervention Features			x	x	x
Posttest x Features				x	x
Posttest x All Available Features					x
Observations	431	431	319	319	319

Note: The models in this table together evaluate how study-level random intercept and slope variance changed with the inclusion of theoretically salient follow-up impact moderators. Random effects meta-regression models with inverse variance weights, a study-level random intercept and slope, and robust standard errors were used in both Panels A and B. In Panel B, the Correlated-and-Hierarchical Model was additionally used with an assumed correlation of r = .60 among outcomes from the same study (Pustejovsky & Tipton, 2022). Model 1 presents the null model with no predictors (note: random slopes were not estimated as posttest effect size was not yet included in the model). Model 2 presents the null model with the addition of posttest effect size as a predictor of follow-up effect size. Model 3 presents estimates from a model in which posttest effect size and dummy-variable indicators for intervention characteristics (i.e., those included in the primary models) were added to the model. Model 4 presents variance estimates from a model in which posttest effect size, intervention-related indicators, and the interaction between posttest effect size and intervention indicators were included. Model 5 shows variance estimates from a model that includes all of the parameters as Model 4 with the addition of indicators for other coded intervention features (reporter type, source meta-analysis, whether outcome was cognitive or social-emotional, outcome construct category, and what calculations were required to estimate the effect size) and the interaction between these and posttest effect size. Negative I^2 values were set to zero.

Supplemental File for:
“The Ubiquity of Educational Intervention Fadeout: A Meta-Analysis of Educational RCTs with Follow-Up”

Additional Methodological Details

Intervention Feature Indicators

Three of the intervention feature indicators were originally coded as continuous variables. These included measures of age at baseline, intervention year, and intervention duration. We dichotomized the measures in the following ways: We categorized interventions as ECE if they targeted children at or before age 7 and as non-ECE if they targeted children after age 7. We split "older" and more recent interventions at the year 2000 (i.e., programs that targeted children at or before 2000 versus programs that targeted children after 2000). Finally, we split the intervention duration by the average (7 months). Programs that were intended to last 7 or more months were coded as "longer" programs, and those lasting less than 7 months were coded as "shorter programs.

Outcome Details

Hart et al., (2024) provides in-depth details on the social-emotional and cognitive outcomes that comprised the bulk (98%) of our sample. To include as many aligned groups as possible in the present analyses focused on intervention characteristics, we additionally included 11 aligned groups that we could not clearly categorize as definitively cognitive or social-emotional using reported information about the measure. We had determined these non-categorizable outcomes as having been tangential enough to social-emotional or cognitive functioning to have coded them (one of our inclusion criteria). These "undefined" constructs included: moderate physical victimization, work-related skills (3 occurrences), total social and cognitive knowledge, perception of school climate, learning engagement, learning behaviors, critical thinking, awareness of victim services, and awareness of perpetrator services.

Additional Results Details

Kitchen Sink Model

Column 5 of Table 4 reports the results from a "kitchen sink" approach to modeling how much intervention- and outcome-related features explained heterogeneity in the persistence of initial impacts. In addition to including the intervention-characteristics of focus in this paper, the kitchen sink model additionally included fixed effects for various outcome-related measures: 1) outcome reporter type (e.g., self-report, direct assessment), 2) which of the original eight meta-analyses the study was sourced from, 2) whether the outcome was social-emotional, cognitive or undefined, and 3) what calculation steps were used to estimate the effect size. The model included both the main effect of the intervention and outcome-related variables, as well as the interaction between the variables and posttest.

Correlated-and-Hierarchical Model

Table 4 reports the heterogeneity-related statistics from the Correlated-and-Hierarchical model (CHE; Pustejovsky & Tipton, 2022). CHE accounts for the non-independence of observations from the same study by modeling the within-study variance in effects, and imposing the assumption that effects from the same study are correlated at $r = .60$. Here, we additionally reported estimates of the meta-analytic average at posttest and 6- to 12-month follow-up from CHE. Additionally, we reported the regression-based estimates from the CHE model for the full sample at 6-to 12-month follow-up and 1- to 2-year follow-up.

The CHE model estimated meta-analytic averages at posttest (.26, *SE* = .03, $p < .0001$) and 6- to 12-month follow-ups (.15, *SE* = .03, $p < .0001$) that were smaller than those estimated through our primary models at posttest and (.30 and .21, respectively). The 1- to 2-year follow-up estimate from CHE (.09, *SE* = .03, $p = .03$) was slightly larger than that from the primary model (.07). At the 6- to 12-month follow-up, the CHE model estimated a slope of .45 (*SE* = .04, $p < .0001$) and an intercept of .04 (*SE* = .01, $p = .02$), in line with the primary estimates. At 1- to 2-years follow-up, the CHE model produced a smaller slope (.13, *SE* = .13, $p = .33$) and intercept (.03, *SE* = .03, $p = .28$) than our primary model.

Correlation between Standard Errors and Effect Sizes

We ran models to probe the extent to which larger (and less precise) standard errors were predictive of larger effect sizes. For these models, we entered standard errors as a predictor of effects using our primary model specifications. Standard errors were consistently predictive of effects such that less precise standard errors predicted larger effects. Standard errors were predictive of effects at posttest ($\beta = 5.64$, $p < .001$), 6- to 12-months follow-up ($\beta = 2.05$, $p < .001$), and 1- to 2-years follow-up ($\beta = 1.00$, $p = .005$).

Table S1
Supporting Evidence for Select Claims Regarding Intervention Features that may Moderate Persistence

Our Summary of Intervention Feature Claims	Supporting Evidence
“Lengthier programs could potentially provide support over key ‘make or break’ developmental transition points (Herrera et al., 2013).”	Herrera et al., 2013: “OST programs may be most effective when taking a long-term approach. Higher Achievement represents a comprehensive, long-term investment in children’s lives. Its intensive and rigorous academic environment improved scholars’ academic performance—but not until after the first year of program involvement. Similarly, the decision to attend a competitive high school is not simply made in the eighth grade but rather is a culmination of experiences and choices (such as course selection) throughout the middle school years (Eccles et al. 2004). Thus, the fact that Higher Achievement is able to work with and support youth throughout middle school is probably critical to its success. These effects may also compound over time, even after program participation ends. Many of the youth in the treatment group are now attending academically rigorous high schools better prepared than they would have been otherwise. These students (as well as those who matriculated to less competitive schools but were likewise better prepared academically) may well end up attending better colleges and ultimately having higher-paying jobs and careers. They and their parents are now familiar with a long set of choices and steps that are similar to those they will face in the college application process. Funders and policymakers who hope that OST programs can foster substantial changes in youth’s trajectories may need to consider longer-term investments—like Higher Achievement—to produce benefits similar to those seen in this study.”
“Programs that provide support during after-school hours, rather than as a substitute for school time, may yield benefits by providing additional time in supportive and instructional contexts (Herrera et al., 2013).”	Herrera et al., 2013: “Higher Achievement may have boosted youth’s test scores, at least in part, by simply providing more academic instruction and experiences than youth would have had otherwise."
“Perhaps interventions that create a “mosaic” of supports across multiple contexts or environments both inside of and outside of school (e.g., like the home environment vis à vis parents) might have deeper and longer-lasting effects by changing highly influential outside-of-school contexts in enduring ways (Garcia et al., 2020; Spoth et al., 2009; Stipek, 2017b).”	Garcia et al., 2020: “Still, it is extremely rare for short-term youth programs by themselves — even multiyear, intensive programs like Higher Achievement — to have long-term impacts. Rather, programs serving young people in under-resourced neighborhoods need to help them and their families build a mosaic of strong in- and out-of-school experiences with caring adults that shift with each child’s interests and developmental needs. Selected activities or programs should be challenging, yet engaging, stretching the young person in developmentally appropriate ways. It is this type of rich, developmentally responsive environment that, over time, supports young people to achieve their full potential.” Spoth et al., 2009: “Many of the risk and protective factors for adolescent substance misuse originate in the family environment (Hawkins, Catalano, & Miller, 1992; Wood, Read, Mitchell, & Brand, 2004).” Stipek, 2017b: “In addition to curricular continuity, continuity in parent involvement can affect children’s learning. Parents play a significant role in the preschool programs that have produced the best-documented sustained

	effects on children, such as the Abecedarian Project in North Carolina, the Perry Preschool Project, and the Child-Parent Centers in Chicago. These programs encouraged parents to be involved in their children's education and they gave parents instruction and tools to do so. Most elementary schools invest much less in parents. As a result, parent involvement, which is known to have considerable impact on children's learning, is not sustained. Again, an important benefit of quality preschool is lost."
"Likewise, targeting many different skills through multi-component programs may be advantageous when the underlying theory of change is highly complex or when it is unclear *what* will work and *why* it will work (Herrera et al., 2013; Klaus & Gray, 1968; Spoth et al., 2009)."	Herrera et al., 2013: "Second, while we have been able to identify some of the elements of Higher Achievement that likely contributed to youth's gains, we do not know which are necessary and which (if any) could be streamlined without diminishing the effectiveness of the program. Future research that rigorously examines the mechanisms underlying OST program benefits—starting with some of those outlined here—would provide important guidance about how to get the greatest return on OST investments." Klaus & Gray, 1968: "The intervention program of the Early Training Project was based upon the fact that at the time we began, and indeed today, there is little clear evidence that, short of a complete change of milieu in infancy, it is possible to offset to any practical extent the progressive retardation with which we were concerned. We attempted, therefore, to develop an intervention "package," consisting of manipulations of those variables which, from research on social class, cognitive development, and motivation, seemed most likely to be influential in later school performance. At the same time, these were to be variables for which we could hope to effect some changes. We also tried to construct the package within a framework that could be employed on a wide scale, should our intervention program be successful." Spoth et al., 2009: "The developmental model examined herein posits that (a) the range of previously demonstrated proximal effects of the tested universal interventions (intervention-related effects on parenting, such as parental monitoring, along with effects on adolescents' intention to use, attitudes, or skills; e.g., Spoth, Redmond, & Shin, 1998) likely delays substance initiation or slows its rate of increase across the adolescent years and decreases the average level of initiation (e.g., Spoth et al., 2001, 2004), and (b) those effects on substance initiation are the primary means by which long-term effects into young adulthood are produced."
"Alternatively, perhaps targeting the *right* skill that is foundational for subsequent developmental cascades may be more critical than breadth (e.g., targeting specific reading skills; Blachman et al., 2014)."	Blachman et al., 2014: "proficiency in decoding acts as a self-teaching mechanism to facilitate retaining words and spellings in memory (Cunningham, Perry, Stanovich, & Share, 2002) and contributes to the acquisition of sight words (Ehri, 2005)."
"Narrowly scoped programs could also generate more replicable and consistent patterns of effects (Dynarski et al., 2013)."	Dynarski et al., 2013: "Our analysis identifies the effect of manipulating a single policy-relevant educational input on adult educational attainment. By contrast, the early-childhood interventions for which researchers have identified lifetime effects (e.g., Head Start, Abecedarian) are multipronged, including home visits, parental coaching, and vaccinations in addition to time in a preschool classroom. We cannot distinguish which dimensions of these

	treatments generate short-term effects on test scores, and whether they differ from the dimensions that generate long-term effects on adult well-being… By contrast, the effects we measured for this paper, both short-term and long-term, can be attributed to a well-defined and replicable intervention: reduced class size.”
“Changing cultural milieus could also impact persistence patterns (e.g., the legalization and decriminalization of marijuana in the case of adolescent substance-use prevention programs; Spoth et al., 2022).”	Spoth et al., 2022: “There were no point-in-time effects on alcohol use or drunkenness, unlike what had been demonstrated at younger ages. This could be attributed to the typical age-related patterns of alcohol use in the general population, given that drinking alcohol is normative among emerging adults at the legal drinking age or older. In the case of marijuana use, the legalization, decriminalization, and lowered penalties for marijuana use in many states during the PROSPER study may have contributed to more widespread use by emerging adults in both the intervention and the control groups.”

Table S2
Correlations among Intervention Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1) ECE	1.00											
(2) TX Duration	0.16	1.00										
(3) Targeted parents	0.06	0.14	1.00									
(4) Time in School	0.08	0.12	0.28*	1.00								
(5) TX year > 2000	0.17	0.00	-0.17	-0.20+	1.00							
(6) Broad TX	0.31**	0.30*	0.44***	0.49***	-0.07	1.00						
(7) Sample Selection	0.21+	-0.10	-0.21+	-0.16	0.05	-0.14	1.00					
(8) Counterfactual	-0.02	0.00	0.27*	-0.01	-0.26*	-0.20+	-0.15	1.00				
(9) Targeted Language/Literacy	0.56***	0.07	0.03	-0.03	-0.03	0.20+	0.64***	0.16	1.00			
(10) Targeted Social-Emotional Skills	-0.48***	0.03	0.27*	0.17	-0.10	0.22*	-0.69***	-0.19+	-0.77***	1.00		
(11) Targeted Substance Use	-0.26*	0.14	0.17	0.03	-0.20+	-0.11	-0.25*	0.02	-0.37**	0.34**	1.00	
(12) Targeted Psych Wellbeing	-0.34**	0.01	-0.05	0.16	-0.20+	-0.11	-0.25*	-0.04	-0.37**	0.34**	0.58***	1.00

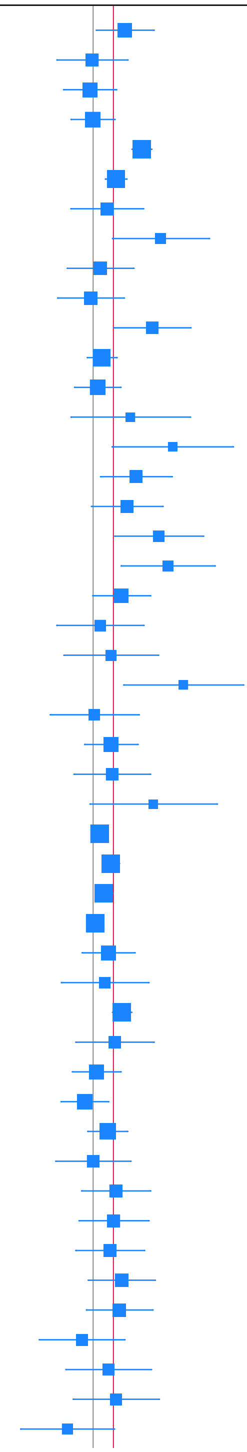
+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: This correlation matrix shows the intervention-level associations among intervention characteristics.

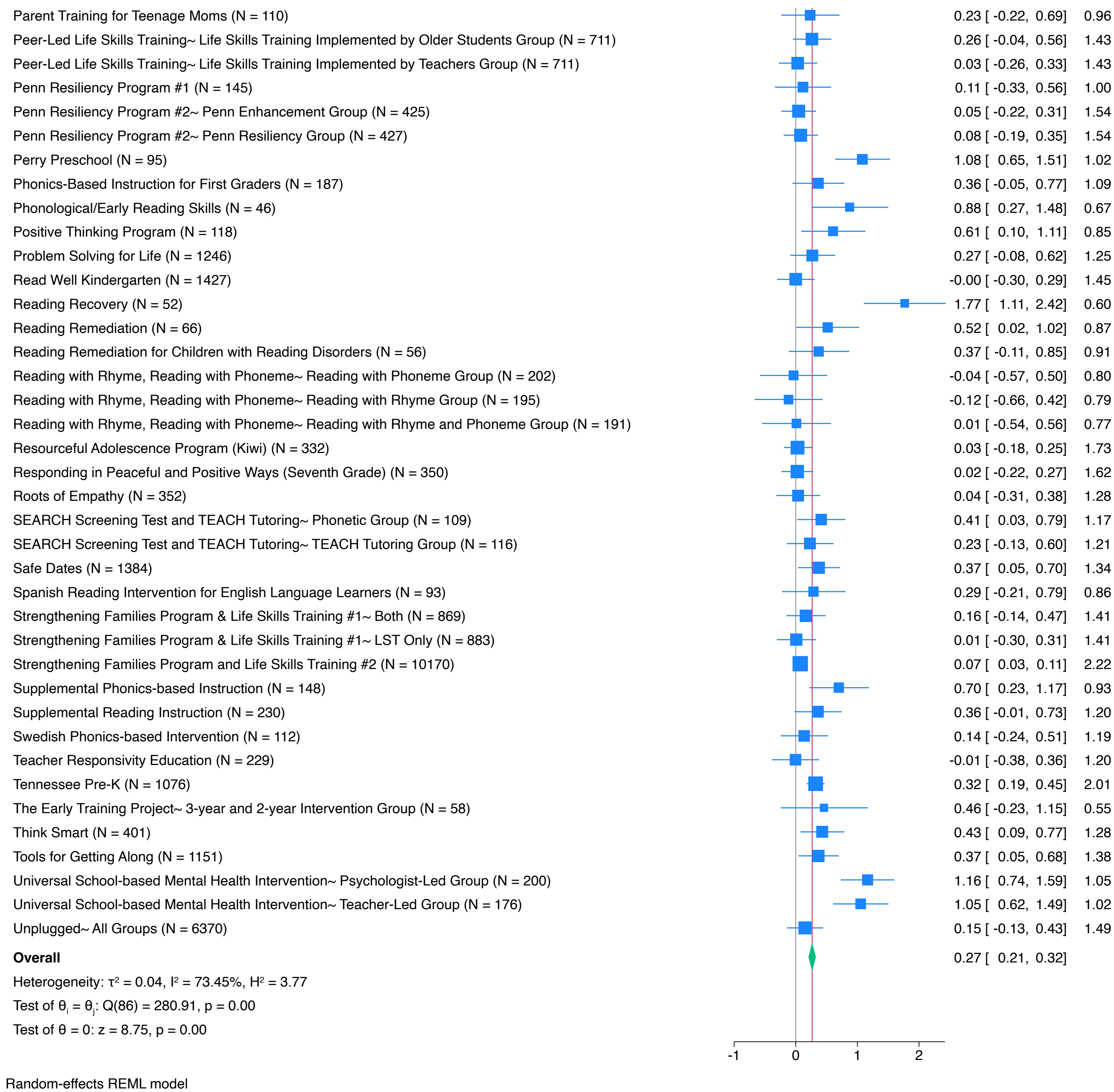
Figure S1

Forest Plot of Average Posttest Impact for Each Intervention

Intervention-level Average Posttest Effects



Study	Effect size with 95% CI	Weight (%)
Abecedarian~ All Pre-K (tx + tx; tx + cntrl) Group (N = 111)	0.42 [0.04, 0.79]	1.18
All Stars (N = 1584)	-0.01 [-0.47, 0.45]	0.95
Aussie Optimism Program (N = 370)	-0.04 [-0.39, 0.30]	1.28
Beyondblue Schools Research Initiative (N = 4011)	-0.00 [-0.29, 0.28]	1.48
Building Blocks (N = 834)	0.64 [0.51, 0.76]	2.03
Building Blocks plus TRIAD Follow-Through Group (N = 741)	0.30 [0.16, 0.44]	2.00
Childhood Depression Prevention (N = 96)	0.18 [-0.29, 0.66]	0.92
Classroom and At-Home Preschool Interventions (N = 49)	0.89 [0.25, 1.52]	0.63
Classroom-Centered & School-Family Partnership~ Classroom Group (N = 820)	0.10 [-0.34, 0.53]	1.02
Classroom-Centered & School-Family Partnership~ Family-School Group (N = 818)	-0.03 [-0.47, 0.40]	1.02
Code-Oriented Reading Instruction (N = 67)	0.77 [0.27, 1.28]	0.86
Cogmed Working Memory Training (N = 414)	0.12 [-0.08, 0.31]	1.81
Cognitive-Behavioral Approach to Drug Abuse Prevention ~ Both Implementation Arms (N = 4464)	0.06 [-0.24, 0.36]	1.42
Computer-Assisted Blending Skill Training (N = 53)	0.49 [-0.29, 1.27]	0.46
Computer-Assisted Learning Program (N = 28)	1.04 [0.25, 1.83]	0.45
Computer-Assisted Reading Intervention~ Lindamood Phoneme Sequencing Group (N = 74)	0.56 [0.10, 1.03]	0.94
Computer-Assisted Reading Intervention~ Read, Write, and Type Group (N = 73)	0.44 [-0.02, 0.91]	0.94
Computer-Assisted Remedial Reading Intervention (N = 50)	0.86 [0.28, 1.44]	0.72
Conflict-Resolution Training Program (N = 54)	0.98 [0.37, 1.59]	0.66
Consultation-based Intervention (N = 221)	0.37 [-0.01, 0.75]	1.18
Dialogic Reading #1~ School Reading Group (N = 48)	0.09 [-0.48, 0.66]	0.73
Dialogic Reading #1~ School plus Home Reading Group (N = 41)	0.23 [-0.38, 0.85]	0.66
Dialogic Reading #2 (N = 30)	1.18 [0.40, 1.97]	0.46
Ease of Handling Social Aspects in Everyday of Life Training (N = 347)	0.02 [-0.56, 0.60]	0.71
Emotional Intelligence Training Program (N = 479)	0.24 [-0.11, 0.58]	1.27
English Reading Intervention for English Language Learners (N = 93)	0.25 [-0.25, 0.75]	0.87
Explicit Phonological Awareness Instruction (N = 24)	0.79 [-0.04, 1.62]	0.42
Head Start Impact Study (N = 3928)	0.08 [-0.01, 0.17]	2.13
Head Start REDI~ Incredible Years Teacher Training Group (N = 1323)	0.23 [0.12, 0.34]	2.09
Head Start REDI~ Promoting Alternative Thinking Strategies Group (N = 1290)	0.14 [0.03, 0.25]	2.08
Head Start REDI~ Tools of the Mind Group (N = 1299)	0.03 [-0.08, 0.13]	2.09
Head Start Research-Based, Developmentally-Informed Program (N = 339)	0.20 [-0.15, 0.54]	1.28
Home-Based Dyslexia Prevention (N = 48)	0.15 [-0.42, 0.73]	0.73
Infant Health and Development Program (N = 1082)	0.38 [0.25, 0.50]	2.04
LARS & LISA (Lust An Realistischer Sicht & Leichtigkeit Im Sozialen Alltag) (N = 287)	0.28 [-0.23, 0.79]	0.84
Learn Young, Learn Fair (N = 1359)	0.04 [-0.27, 0.36]	1.37
Life Skills Training (N = 938)	-0.11 [-0.42, 0.20]	1.39
Life Skills Training in Minority Youth (N = 2800)	0.19 [-0.07, 0.45]	1.57
Linking the Interests of Families and Teachers (N = 348)	0.00 [-0.49, 0.49]	0.89
Living Letters #2~ Combined Text Comprehension and Oral Language Group (N = 77)	0.30 [-0.15, 0.75]	0.98
Living Letters #2~ Oral Language Group (N = 75)	0.27 [-0.18, 0.73]	0.97
Living Letters #2~ Text Comprehension Group (N = 77)	0.22 [-0.23, 0.67]	0.98
Multi-Component Reading Remediation~ PHAB + RAVE-O (N = 127)	0.37 [-0.06, 0.81]	1.01
Multi-Component Reading Remediation~ PHAB + WIST (N = 129)	0.34 [-0.09, 0.78]	1.02
Omega-Interactive Sentences, Computerized Phonological Training~ Comprehension Training Group (N = 50)	-0.15 [-0.71, 0.41]	0.75
Omega-Interactive Sentences, Computerized Phonological Training~ Phonological & Comprehension (N = 50)	0.20 [-0.36, 0.76]	0.75
Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group (N = 50)	0.30 [-0.26, 0.86]	0.75
Optimism and Lifeskills Program (N = 42)	-0.34 [-0.95, 0.27]	0.66



Note: This forest plot depicts the intervention-average impact for each intervention. The plot was generated in Stata. The "overall" meta-analytic estimate, intervention-specific confidence intervals, and weights were computed using the default Stata model (random effects, REML). The overall model heterogeneity estimates correspond with the default Stata model (not our primary meta-analytic model specifications). The weights indicator provides a sense of the relative weighting of estimates, in the context of the default Stata model (not our primary meta-analytic model specifications).

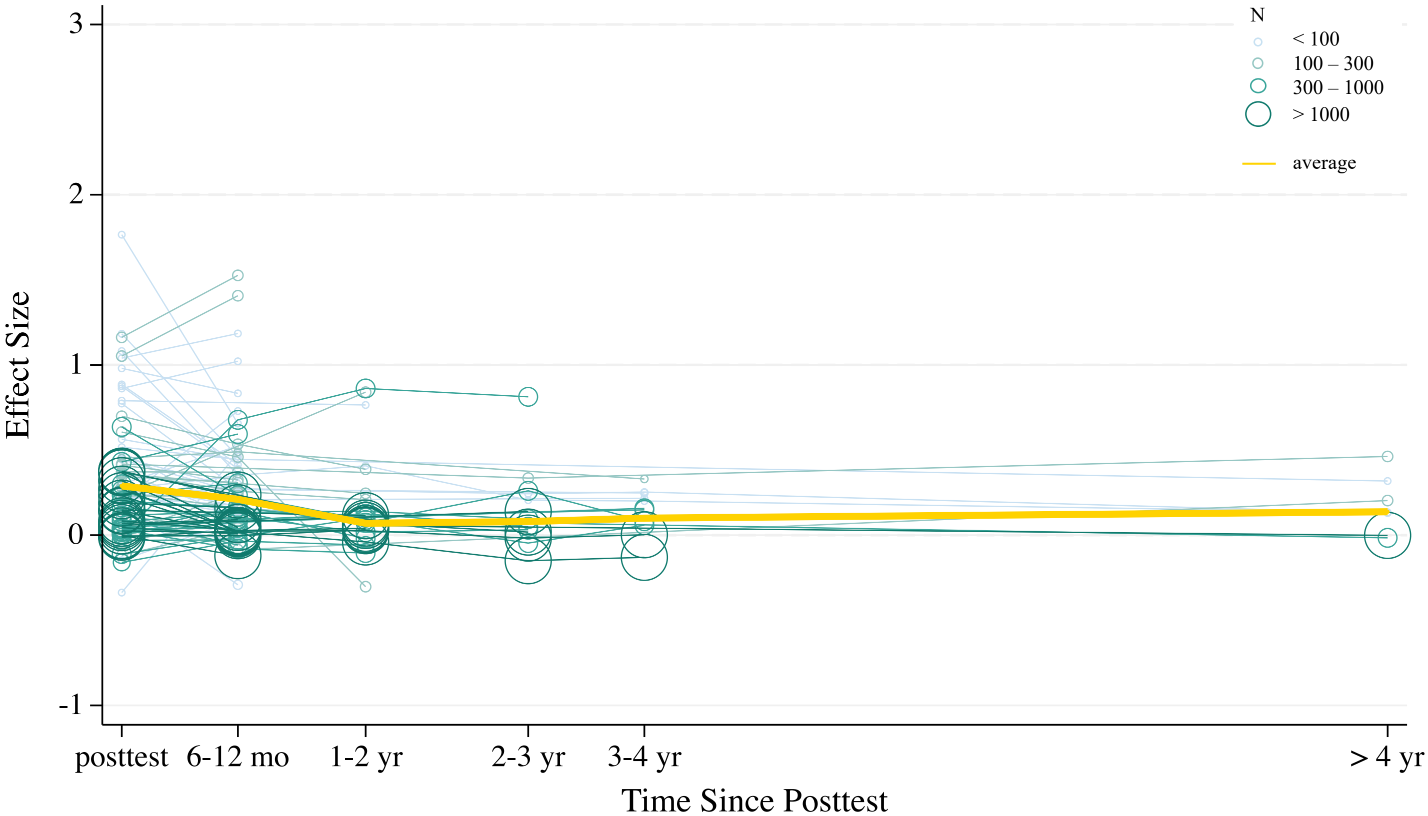
Table S3
Average Effect Sizes Across Posttest and Follow-up Assessment Waves for Aligned Outcomes (ES (SE)

	Avg. ES /wave (1)	Avg. Posttest ES for obs. /wave (2)	n
Panel A: Weighted Average Effect Sizes			
Posttest	0.30 (0.04)***	0.30 (0.04)***	461
6 months to 1 year	0.21 (0.03)***	0.31 (0.04)***	431
> 1 year, up to 2 years	0.07 (0.02)**	0.25 (0.06)***	90
> 2 years, up to 3 years	0.08 (0.05)	0.26 (0.09)*	33
> 3 years, up to 4 years	0.10 (0.05)+	0.35 (0.10)**	30
> 4 years	0.14 (0.08)	0.49 (0.08)**	13
Panel B: Unweighted Average Effect Sizes			
Posttest	0.29 (0.04)***	0.29 (0.04)***	461
6 months to 1 year	0.21 (0.04)***	0.29 (0.04)***	431
> 1 year, up to 2 years	0.10 (0.04)*	0.20 (0.07)*	90
> 2 years, up to 3 years	0.14 (0.05)*	0.28 (0.15)	33
> 3 years, up to 4 years	0.16 (0.03)*	0.39 (0.15)+	30
> 4 years	0.22 (0.05)*	0.51 (0.11)*	13
Panel C: Sample Size Weighting			
Posttest	0.33 (0.04)***	0.33 (0.04)***	461
6 months to 1 year	0.25 (0.04)***	0.34 (0.05)***	431
> 1 year, up to 2 years	0.16 (0.06)*	0.26 (0.06)***	90
> 2 years, up to 3 years	0.15 (0.07)+	0.26 (0.09)*	33
> 3 years, up to 4 years	0.11 (0.04)*	0.36 (0.10)**	30
> 4 years	0.17 (0.07)+	0.50 (0.07)***	13

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: "ES" = Effect size. Effect sizes are presented in standard deviation units. The analytic sample was constituted by "aligned groupings" which included a posttest and at least one follow-up effect size for the same construct measured using the same measure, subscales, and reporter within a treatment-control contrast. In Panel A, effects were estimated using a random effects meta-analytic model that included a random effect for study, weights, and robust standard errors (clustered at the study level). In Panel B, average effect sizes were estimated using a fixed effects meta-analytic model with robust standard errors (clustered at the study level), but no random effect for study and no weighting. In Panel C, effects were estimated using the same random effects approach as in Panel A, but with weighting by the total sample size at the concurrent assessment wave. In Column 1, average effect sizes are presented for the posttest assessment wave and all follow-up assessment waves. To evaluate the possibility of selection into longer-run follow-up assessments, Column 2 presents average posttest effects for the outcomes collected at each follow-up wave.

Figure S2
Fadeout Trajectories for Aligned Groups (by Intervention Group)



Note: Each line represents the average effect size trajectory of outcomes that were measured at posttest and at least follow-up using the same measure for a single intervention. The average effect size trajectory (calculated on the underlying sample of aligned groups) is displayed in gold. Meta-analytic averages were computed using a random effects meta-analytic model that included a random effect for study, weighting by the inverse variances, and robust standard errors (clustered at the study level). As detailed in the key, coordinates were weighted by the posttest sample size (larger circles represent estimates from larger samples) with aligned color-coding (darker colors represent larger sample sizes).

Table S4

Average Effect Sizes at Posttest and 1- to 2-year Follow-up Assessments

	Moderator = 1		**Moderator = 0**		
	Avg. ES (SE) (1)	**% in group (2)**	**Avg. ES (SE) (3)**	**% in group (4)**	**Total Obs. (5)**
ECE	***ECE***		***Non-ECE***		
Posttest	0.41 (0.09)***	45%	0.08 (0.05)	55%	84
1- to 2- year Follow-up	0.14 (0.05)*		0.04 (0.03)		
TX Duration	***Long***		***Short***		
Posttest	0.31 (0.12)*	52%	0.21 (0.07)*	48%	75
1- to 2- year Follow-up	0.11 (0.04)*		0.05 (0.03)		
Parent Involvement	***Parents***		***No Parents***		
Posttest	0.26 (0.10)*	54%	0.23 (0.08)**	46%	90
1- to 2- year Follow-up	0.15 (0.05)*		0.02 (0.02)		
Time in School	***More Time***		***Curricular***		
Posttest	0.36 (0.13)*	27%	0.19 (0.06)**	73%	90
1- to 2- year Follow-up	0.17 (0.07)+		0.05 (0.02)*		
TX year > 2000	***> 2000***		***<= 2000***		
Posttest	0.20 (0.06)**	66%	0.30 (0.12)*	34%	90
1- to 2- year Follow-up	0.04 (0.01)*		0.19 (0.09)+		
Broad TX	***Broad***		***Narrow***		
Posttest	0.37 (0.13)*	31%	0.18 (0.06)*	69%	90
1- to 2- year Follow-up	0.16 (0.07)+		0.05 (0.02)+		
Sample Selection	***Selection Criteria***		***No Criteria***		
Posttest	0.51 (0.13)**	22%	0.14 (0.05)*	78%	90
1- to 2- year Follow-up	0.21 (0.08)*		0.04 (0.01)*		
Counterfactual	***Something Provided***		***Nothing Provided***		
Posttest	0.08 (0.07)	29%	0.30 (0.08)**	71%	85
1- to 2- year Follow-up	0.03 (0.02)		0.09 (0.03)**		
Language/Literacy	***Targeted***		***Not Targeted***		
Posttest	0.45 (0.12)**	34%	0.15 (0.06)*	66%	90
1- to 2- year Follow-up	0.18 (0.06)*		0.04 (0.02)		
Social-Emotional Skills	***Targeted***		***Not Targeted***		
Posttest	0.18 (0.07)*	87%	0.41 (0.10)**	13%	90
1- to 2- year Follow-up	0.07 (0.02)*		0.21 (0.11)		
Substance Use	***Targeted***		***Not Targeted***		
Posttest	0.02 (0.13)	16%	0.27 (0.06)***	84%	90
1- to 2- year Follow-up	0.11 (0.01)*		0.08 (0.03)*		
Psych Wellbeing	***Targeted***		***Not Targeted***		
Posttest	0.16 (0.09)	3%	0.25 (0.06)***	97%	90
1- to 2- year Follow-up	0.06 (0.05)		0.08 (0.02)**		
Full Sample Posttest: 0.25 (0.06)***					90
Full Sample 1- to 2-year Follow-up: 0.07 (0.02)**					

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: This table presents the meta-analytic averages of effects at posttest and 6- to 12-month follow-up waves from our primary random-effects meta-regression with random slopes and intercepts for study, weights for effect precision, and robust standard errors (with study-level clustering). Each unit of observation is one aligned group, in which the same construct was assessed with the same measure at posttest and follow-up within the same study at 1- to 2-years follow-up specifically.

Table S5

Persistence Rates by Intervention-Related Characteristic at 1- to 2-year Follow-up (Est. (SE))

	Split Sample Models		Full Interaction Models		
	Moderator=1	**Moderator=0**	**p-value**		**Total Obs**
	(1)	**(2)**	**(3)**		**(4)**
ECE	***ECE***	***Non-ECE***			
Constant	-0.01 (0.03)	0.07 (0.04)	0.09	+	84
Posttest Effect Size	0.32 (0.13)*	-0.03 (0.27)	0.18		
Observations	38	46			
TX Duration	***Long***	***Short***			
Constant	0.05 (0.01)+	0.02 (0.05)	0.65		75
Posttest Effect Size	0.06 (0.10)	0.27 (0.27)	0.43		
Observations	39	36			
Parent Involvement	***Parents***	***No Parents***			
Constant	0.05 (0.01)*	0.00 (0.03)	0.46		90
Posttest Effect Size	0.26 (0.12)+	0.14 (0.17)	0.42		
Observations	49	41			
Time in School	***More Time***	***Curricular***			
Constant	0.08 (0.02)+	0.03 (0.03)	0.87		90
Posttest Effect Size	0.13 (0.12)	0.17 (0.16)	0.99		
Observations	24	66			
TX year > 2000	***> 2000***	***<= 2000***			
Constant	0.02 (0.02)	0.17 (0.08)	0.20		90
Posttest Effect Size	0.16 (0.12)	-0.09 (0.30)	0.61		
Observations	59	31			
Broad TX	***Broad***	***Narrow***			
Constant	0.04 (0.00)+	0.03 (0.03)	0.71		90
Posttest Effect Size	0.16 (0.14)	0.15 (0.16)	0.86		
Observations	28	62			
Sample Selection	***Selection Criteria***	***No Criteria***			
Constant	-0.07 (0.04)	0.05 (0.01)*	0.09	+	90
Posttest Effect Size	0.52 (0.10)+	-0.06 (0.12)	0.01	**	
Observations	20	70			
Counterfactual	***Something Provided***	***Nothing Provided***			
Constant	0.01 (0.03)	0.02 (0.02)	0.97		85
Posttest Effect Size	0.19 (0.13)	0.22 (0.12)	0.90		
Observations	25	60			
Language/Literacy	***Targeted***	***Not Targeted***			
Constant	0.02 (0.01)	0.05 (0.03)	0.17		90
Posttest Effect Size	0.33 (0.14)+	-0.02 (0.16)	0.10		
Observations	31	59			
Social-Emotional Skills	***Targeted***	***Not Targeted***			
Constant	0.04 (0.01)*	-0.01 (0.21)	0.13		90
Posttest Effect Size	0.07 (0.14)	0.50 (0.27)	0.08	+	
Observations	78	12			
Substance Use	***Targeted***	***Not Targeted***			
Constant	0.11 (0.02)+	0.02 (0.02)	0.21		90
Posttest Effect Size	-0.09 (0.18)	0.19 (0.12)	0.63		
Observations	14	76			
Psych Wellbeing	***Targeted***	***Not Targeted***			
Constant	0.15 (0.03)	0.02 (0.01)	0.15		90
Posttest Effect Size	-0.52 (0.11)	0.18 (0.12)	0.14		
Observations	3	87			
Full Sample Constant: 0.02 (0.01) **Full Sample Posttest Effect Size:** 0.17 (0.11)					90

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: This table presents estimates from our primary random-effects meta-regression with random slopes and intercepts for study, inverse variance weights, and robust standard errors (with study-level clustering). Aligned groups are the unit of observation. Columns 1 and 2 present estimates from split-sample models, in which the persistence rate and constant were estimated using the sample of results for the intervention feature of focus. In Columns 3 and 4, estimates of the statistical difference in rates and constants are reported, from model that tested the main effect and interaction with posttest impact for each intervention feature.

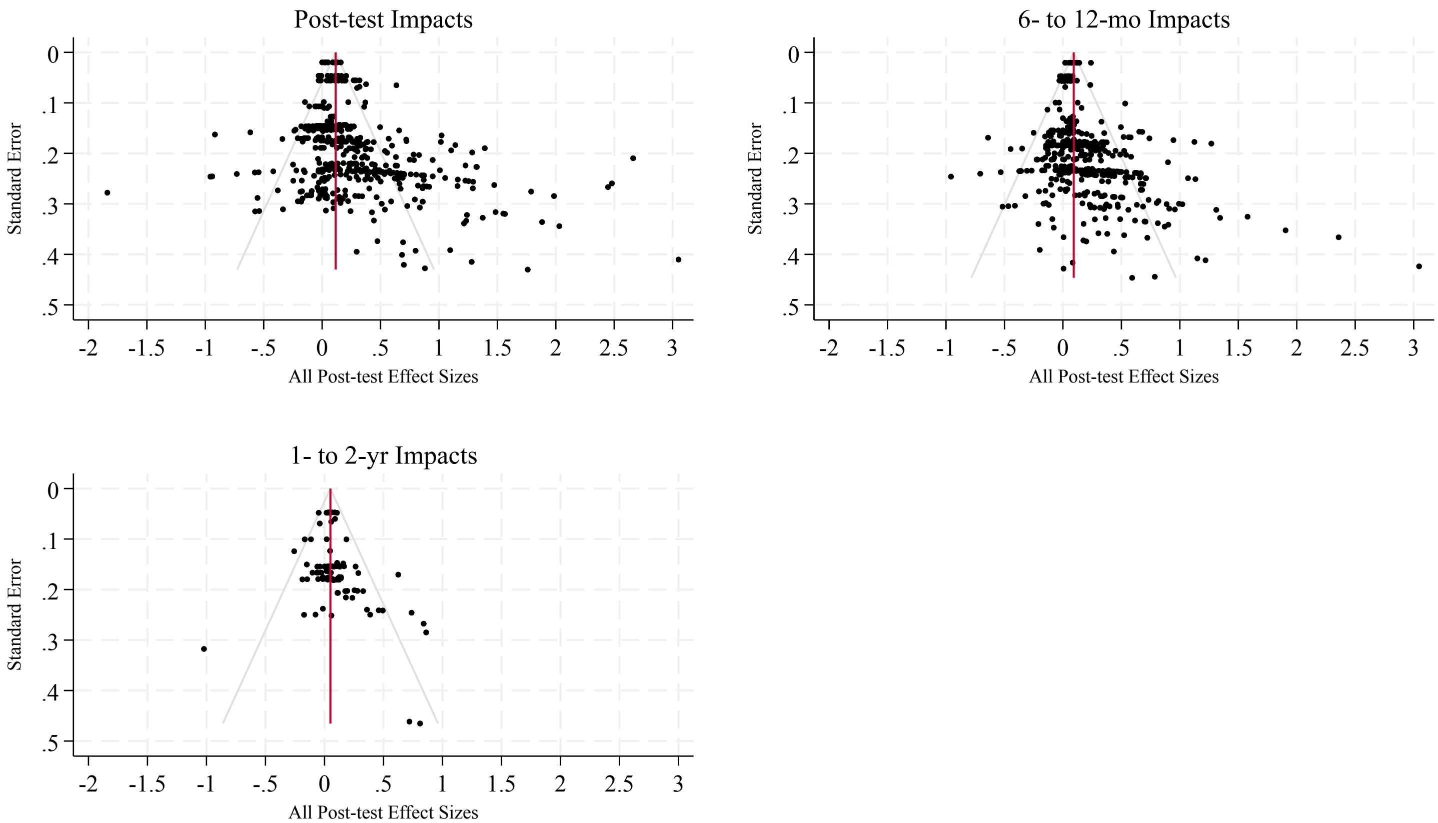
Table S6
Sensitivity Check: Persistence Rates by Intervention-Related Characteristic at 6- to 12-months Follow-up using Sample Size Weighting (Est (SE))

	Split Sample Models		**Interaction Models**	
	Moderator=1	**Moderator=0**	**p-value**	**Obs**
	(1)	**(2)**	**(3)**	**(4)**
ECE	***ECE***	***Non-ECE***		
Constant	0.10 (0.04)**	0.07 (0.05)	0.69	425
Posttest Effect Size	0.37 (0.08)***	0.53 (0.08)***	0.19	
Observations	176	249		
TX Duration	***Long***	***Short***		
Constant	0.08 (0.04)*	0.06 (0.05)	0.75	341
Posttest Effect Size	0.41 (0.09)***	0.56 (0.09)***	0.28	
Observations	176	165		
Parent Involvement	***Parents***	***No Parents***		
Constant	0.11 (0.06)+	0.06 (0.03)+	0.19	431
Posttest Effect Size	0.20 (0.16)	0.53 (0.06)***	0.10	
Observations	78	353		
Time in School	***More Time***	***Curricular***		
Constant	0.05 (0.01)*	0.09 (0.03)**	0.38	431
Posttest Effect Size	0.21 (0.08)+	0.48 (0.06)***	0.16	
Observations	22	409		
TX year > 2000	***> 2000***	***<= 2000***		
Constant	0.10 (0.04)**	0.05 (0.05)	0.37	431
Posttest Effect Size	0.48 (0.08)***	0.44 (0.08)***	0.62	
Observations	216	215		
Broad TX	***Broad***	***Narrow***		
Constant	0.06 (0.05)	0.08 (0.03)*	0.84	431
Posttest Effect Size	0.01 (0.11)	0.51 (0.06)***	0.02 *	
Observations	40	391		
Sample Selection	***Selection Criteria***	***No Criteria***		
Constant	0.09 (0.05)+	0.07 (0.04)+	0.75	431
Posttest Effect Size	0.58 (0.10)***	0.39 (0.07)***	0.13	
Observations	168	263		
Counterfactual	***Something Provided***	***Nothing Provided***		
Constant	0.08 (0.06)	0.09 (0.03)*	0.89	408
Posttest Effect Size	0.52 (0.08)***	0.44 (0.07)***	0.71	
Observations	111	297		
Language/Literacy	***Targeted***	***Not Targeted***		
Constant	0.07 (0.04)+	0.09 (0.05)+	0.86	431
Posttest Effect Size	0.46 (0.08)***	0.46 (0.08)***	0.98	
Observations	227	204		
Social-Emotional Skills	***Targeted***	***Not Targeted***		
Constant	0.09 (0.04)*	0.05 (0.04)	0.53	431
Posttest Effect Size	0.37 (0.08)***	0.58 (0.08)***	0.08 +	
Observations	237	194		
Substance Use	***Targeted***	***Not Targeted***		
Constant	0.09 (0.07)	0.08 (0.03)*	0.93	431
Posttest Effect Size	0.52 (0.19)*	0.45 (0.06)***	0.72	
Observations	67	364		
Psych Wellbeing	***Targeted***	***Not Targeted***		
Constant	0.00 (0.02)	0.09 (0.03)**	0.23	431
Posttest Effect Size	0.38 (0.22)	0.45 (0.06)***	0.85	
Observations	58	373		
Full Sample Constant: 0.08 (0.03)** **Full Sample Posttest Effect Size:** 0.46 (0.06)***				431

+ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note: This table presents estimates from a sensitivity check in which we used our primary random-effects meta-regression with random slopes and intercepts for study, robust standard errors (with study-level clustering), and sample size weighting instead of inverse variance weighting. Aligned groups are the unit of observation. Columns 1 and 2 present estimates from split-sample models, in which the persistence rate and constant were estimated using the sample of results for the intervention feature of focus. In Columns 3 and 4, estimates of the statistical difference in rates and constants are reported, from model that tested the main effect and interaction with posttest impact for each intervention feature.

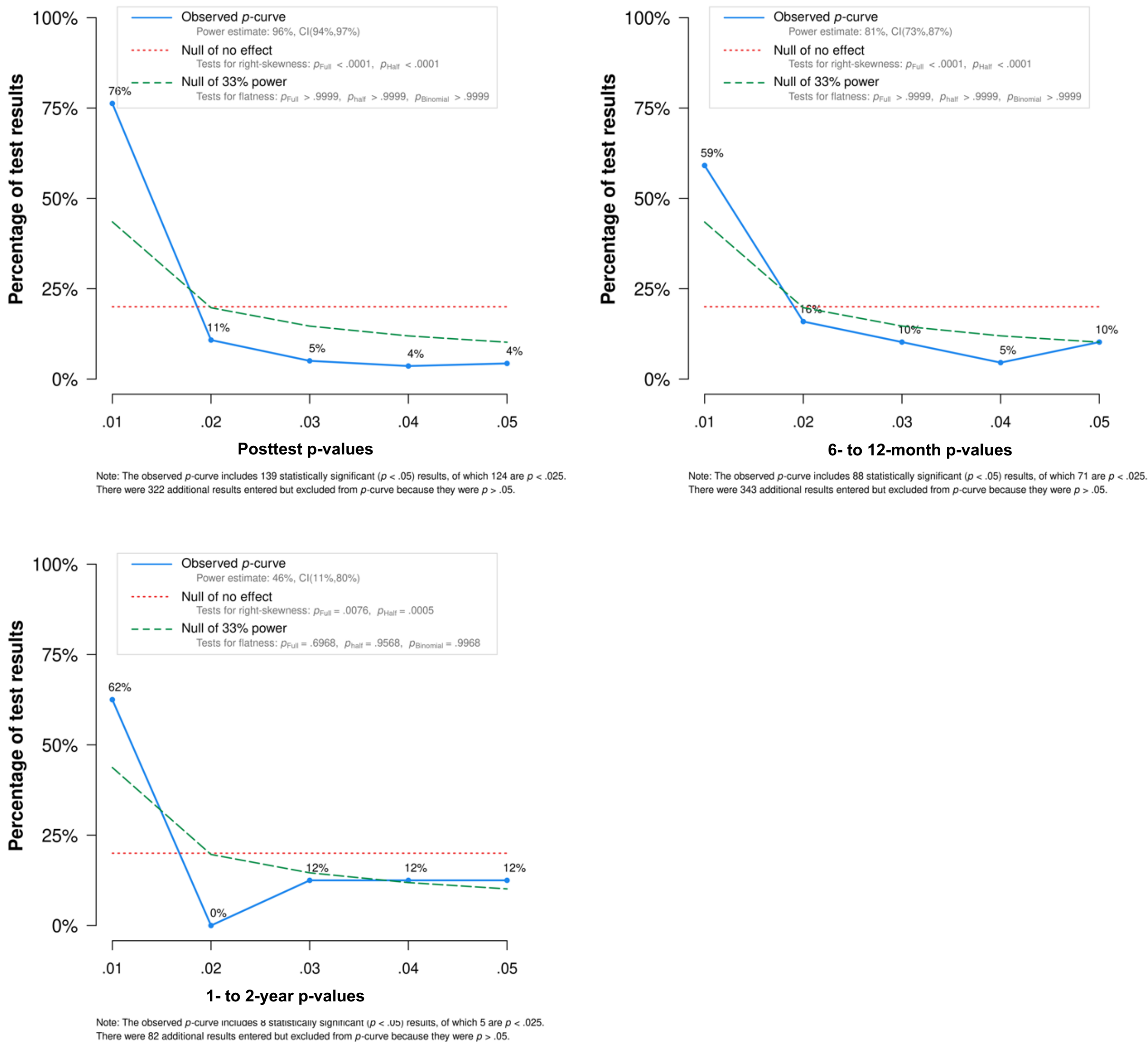
Figure S3
Funnel Plots



Note: Each coordinate reflects an effect from aligned groups at the respective assessment wave (posttest, 6- to 12-month follow-up, 1- to 2-years follow-up).

Figure S4

P-curves at Posttest, 6- to 12-month follow-up, and 1- to 2-year follow-up



Note: We generated these p-curve models using all effects from aligned groups at the respective assessment wave (posttest, 6- to 12-month follow-up, 1- to 2-years follow-up). The models were created on p-curve.com (Simonsohn, Nelson, & Simmons, 2015).

Table S7
Details on the Outcomes that Showed 100% or More Absolute Persistence at the 6- to 12-month Follow-up Assessment

Intervention Name	**Construct Category**	**Specific Construct**	**Measure**	**Follow-up (months)**	**Posttest ES**	**Follow-up ES**	**ECE (=1)**	**Long Duration (=1)**	**Parent (=1)**	**More Time in School (=1)**	**TX year > 2000 (=1)**	**Broad TX (=1)**	**Selection Criteria (=1)**	**Active Control (=1)**	**Lang./Lit. (=1)**	**Soc.-emo. (=1)**	**Subs. (=1)**	**Psych. Well. (=1)**	**Posttest Sample Size**
Positive Thinking Program	internalizing	proportion with depressive diagnosis	diagnostic interview	9	1.21	1.34	0	0	0	0	0	0	0		0	1	0	0	120
Emotional Intelligence Training Program	internalizing	depression	behavior assessment system for children and adolescents	6	0.32	0.33	0	1	0	0	1	0	0	0	0	1	0	0	479
Emotional Intelligence Training Program	internalizing	sense of incapacity	behavior assessment system for children and adolescents	6	0.26	0.35	0	1	0	0	1	0	0	0	0	1	0	0	479
Emotional Intelligence Training Program	social-emotional	social stress	behavior assessment system for children and adolescents	6	0.27	0.38	0	1	0	0	1	0	0	0	0	1	0	0	479
Think Smart	social-emotional	assertiveness skills	study-created measure	6	0.32	0.74	1	0	0	0	1	0	0		0	1	1	0	401
Think Smart	substance use	hlp use	study-created measure	6	0.30	1.12	1	0	0	0	1	0	0		0	1	1	0	401
Think Smart	substance use	inhalant use	study-created measure	6	0.72	1.27	1	0	0	0	1	0	0		0	1	1	0	401
Think Smart	substance use	otc use	study-created measure	6	0.25	0.95	1	0	0	0	1	0	0		0	1	1	0	401
Living Letters #2~ Oral Language Group	language and literacy	reading comprehension	wechsler	11	0.35	0.65	0		0	0	1	0	1	0	1	0	0	0	75
Living Letters #2~ Text Comprehension Group	language and literacy	reading comprehension	wechsler	11	0.38	0.39	0	0	0	0	1	0	1	0	1	0	0	0	77
Computer-Assisted Learning Program	language and literacy	word recognition	time2	9	0.80	1.15	1	0	0	0	1	0	1		1	0	0	0	28
Omega-Interactive Sentences, Computerized Phonological Training~ Combined Phonological and	language and literacy	non-word reading	study-created measure	9	0.34	0.82	1	0	0	0	1	0	1	0	1	0	0	0	50

Table S7
Details on the Outcomes that Showed 100% or More Absolute Persistence at the 6- to 12-month Follow-up Assessment

Intervention Name	Construct Category	Specific Construct	Measure	Follow-up (months)	Posttest ES	Follow-up ES	ECE (=1)	Long Duration (=1)	Parent (=1)	More Time in School (=1)	TX year > 2000 (=1)	Broad TX (=1)	Selection Criteria (=1)	Active Control (=1)	Lang./Lit. (=1)	Soc.-emo. (=1)	Subs. (=1)	Psych. Well. (=1)	Posttest Sample Size
Comprehension Training Group																			
Omega-Interactive Sentences, Computerized Phonological Training~ Combined Phonological and Comprehension Training Group	language and literacy	word recognition	word-chains test	9	0.28	1.31	1	0	0	0	1	0	1	0	1	0	0	0	50
Omega-Interactive Sentences, Computerized Phonological Training~ Phonological Training Group	language and literacy	word recognition	word-chains test	9	0.31	0.99	1	0	0	0	1	0	1	0	1	0	0	0	50
Computer-Assisted Reading Intervention for Children at Risk of Dyslexia~ Read, Write, and Type Group	language and literacy	rapid letter naming	comprehensive test of phonological processing	12	0.33	0.67	1	1	0	0	0	0	1	0	1	0	0	0	73
Computer-Assisted Reading Intervention for Children at Risk of Dyslexia~ Read, Write, and Type Group	language and literacy	word efficiency	test of word reading efficiency	12	0.22	0.28	1	1	0	0	0	0	1	0	1	0	0	0	73
Multi-Component Reading Remediation~ Phonological Analysis and Blending/Direct Instruction and Retrieval, Automaticity, Vocabulary, Engagement with Language, and Orthography Group	language and literacy	passage comprehension	woodcock reading mastery test	12	0.23	0.28	0	1	0	0	0	0	1		1	0	0	0	127
Multi-Component Reading Remediation~ Phonological Analysis and Blending/Direct Instruction and Retrieval, Automaticity, Vocabulary, Engagement with	language and literacy	reading	wide-range achievement test	12	0.45	0.56	0	1	0	0	0	0	1		1	0	0	0	127

Table S7
Details on the Outcomes that Showed 100% or More Absolute Persistence at the 6- to 12-month Follow-up Assessment

Intervention Name	**Construct Category**	**Specific Construct**	**Measure**	**Follow-up (months)**	**Posttest ES**	**Follow-up ES**	**ECE (=1)**	**Long Duration (=1)**	**Parent (=1)**	**More Time in School (=1)**	**TX year > 2000 (=1)**	**Broad TX (=1)**	**Selection Criteria (=1)**	**Active Control (=1)**	**Lang./Lit. (=1)**	**Soc.-emo. (=1)**	**Subs. (=1)**	**Psych. Well. (=1)**	**Posttest Sample Size**
Language, and Orthography Group																			
Multi-Component Reading Remediation~ Phonological Analysis and Blending/Direct Instruction and Retrieval, Automaticity, Vocabulary, Engagement with Language, and Orthography Group	language and literacy	word attack	woodcock reading mastery test	12	0.55	0.56	0	1	0	0	0	0	1		1	0	0	0	127
Multi-Component Reading Remediation~ Phonological Analysis and Blending/Direct Instruction and Retrieval, Automaticity, Vocabulary, Engagement with Language, and Orthography Group	language and literacy	word identification	woodcock reading mastery test	12	0.29	0.42	0	1	0	0	0	0	1		1	0	0	0	127
Multi-Component Reading Remediation~ Phonological Analysis and Blending/Direct Instruction and Word Identification Strategy Training Group	language and literacy	word identification	woodcock reading mastery test	12	0.25	0.36	0	1	0	0	0	0	1		1	0	0	0	129
Multi-Component Reading Remediation~ Phonological Analysis and Blending/Direct Instruction and Retrieval, Automaticity, Vocabulary, Engagement with Language, and Orthography Group	language and literacy	word reading efficiency real words	test of word reading efficiency	12	0.26	0.44	0	1	0	0	0	0	1		1	0	0	0	127

Table S7
Details on the Outcomes that Showed 100% or More Absolute Persistence at the 6- to 12-month Follow-up Assessment

Intervention Name	Construct Category	Specific Construct	Measure	Follow-up (months)	Posttest ES	Follow-up ES	ECE (=1)	Long Duration (=1)	Parent (=1)	More Time in School (=1)	TX year > 2000 (=1)	Broad TX (=1)	Selection Criteria (=1)	Active Control (=1)	Lang./Lit. (=1)	Soc.-emo. (=1)	Subs. (=1)	Psych. Well. (=1)	Posttest Sample Size
Cogmed Working Memory Training	cognitive	visuospatial short term memory	automated working memory assessment	6	0.37	0.53	1	0	0	0	1	0	1	0	0	0	0	0	414
Head Start Research-Based, Developmentally-Informed Program	social-emotional	social competence	social competence scale	11.75	0.21	0.30	1		1	0	1	1	0	0	1	1	0	0	343
English Reading Intervention for English Language Learners~ Study 1 & 4	language and literacy	passage comprehension-spanish	woodcock language proficiency battery	12	0.23	0.28	1	1	0	0	1	0	1	0	1	0	0	0	95.5
English Reading Intervention for English Language Learners~ Study 1 & 4	language and literacy	reading fluency-english	dynamic indicators of basic early literacy skills	12	0.31	0.36	1	1	0	0	1	0	1	0	1	0	0	0	93.5
Spanish Reading Intervention for English Language Learners~ Study 2 & 3	language and literacy	passage comprehension-spanish	woodcock language proficiency battery	12	0.45	0.49	1	1	0	0	1	0	1	0	1	0	0	0	102.5
Computer-Assisted Remedial Reading Intervention	language and literacy	reading fluency	lukilasse graded fluency test	12	0.86	1.02	0	0	0	0	1	0	1	0	1	0	0	0	50
Reading Remediation	language and literacy	word reading efficiency	test of word reading efficiency	9	0.69	0.71	0	1	0	0	0	0	1	0	1	0	0	0	69
Life Skills Training in Minority Youth	substance use	binge drinking	study-created measure	12	0.50	0.50	0	1	0	0	0	0	0	0	0	1	1	1	2982
Universal School-based Mental Health Intervention~ Teacher-Led Group	internalizing	anxiety	spence children's anxiety scale	6	0.63	0.80	0		0	0	1	0	0	0	0	1	0	0	176
Universal School-based Mental Health Intervention~ Psychologist-Led Group	social-emotional	problem solving	coping strategy indicator	6	2.48	3.05	0		0	0	1	0	0	0	0	1	0	0	200
Universal School-based Mental Health Intervention~ Psychologist-Led Group	social-emotional	seeking social support	coping strategy indicator	6	0.56	1.90	0		0	0	1	0	0	0	0	1	0	0	200

Table S7
Details on the Outcomes that Showed 100% or More Absolute Persistence at the 6- to 12-month Follow-up Assessment

Intervention Name	**Construct Category**	**Specific Construct**	**Measure**	**Follow-up (months)**	**Posttest ES**	**Follow-up ES**	**ECE (=1)**	**Long Duration (=1)**	**Parent (=1)**	**More Time in School (=1)**	**TX year > 2000 (=1)**	**Broad TX (=1)**	**Selection Criteria (=1)**	**Active Control (=1)**	**Lang./Lit. (=1)**	**Soc.-emo. (=1)**	**Subs. (=1)**	**Psych. Well. (=1)**	**Posttest Sample Size**
Peer-Led Life Skills Training~ Life Skills Training Implemented by Older Students Group	substance use	tobacco knowledge	study-created measure	12	0.27	0.34	0		0	0	0	0	0	0	0	1	1	1	711
Peer-Led Life Skills Training~ Life Skills Training Implemented by Teachers Group	substance use	tobacco knowledge	study-created measure	12	0.27	0.29	0		0	0	0	0	0	0	0	1	1	1	711
Conflict-Resolution Training Program	social-emotional	overall conflict resolution- "queuing" conflict	the conflict scenario written measure	6	0.84	0.91	0	0	0	0	0	0	0	0	0	1	0	0	54
Conflict-Resolution Training Program	social-emotional	overall conflict resolution- "taking turns" conflict	the conflict scenario written measure	6	0.73	0.96	0	0	0	0	0	0	0	0	0	1	0	0	54

References for Reports Included in the Full MERF Sample
(See Figure S1 for Interventions Included in the Present Study)

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

<u>Reading Recovery</u>

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

<u>Strengthening Families Program and Life Skills Training #2</u>

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

<u>Living Letters #1</u>

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

<u>Living Letters #2</u>

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

<u>Computer-Assisted Learning Program</u>

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

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

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, *32*(4), 692-717. https://doi.org/10.1002/pam.21715

Garcia, I., Grossman, J. B., Herrera, C. Linden, L. L. (2020). *The Impact of an Intensive Year-Round Middle School Program on College Attendance.* MDRC.

Hart, E. R., Bailey, D. H., Luo, S., Sengupta, P., & Watts, T. W. (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. *Psychological Bulletin*, *150*(10), 1207–1236. https://doi.org/10.1037/bul0000450

Herrera, C., Grossman, J. B., Linen, L. L. (2013). *Staying on track: Testing Higher Achievement's long-term impact on academic outcomes and high school choice.* MDRC.

Klaus R. A., & Gray S. W. (1986). The Early Training Project for disadvantaged children: A report after five years. *Monographs of the Society in Child Development, 33*(4), 1-66.

Pustejovsky, J., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*. https://doi.org/10.1007/s11121-021-01246-3

Spoth, R., Redmond, C., Shin, C., Trudeau, L., Greenberg, M. T., Feinberg, M. E., & Welsh, J. (2022). Applying the PROSPER prevention delivery system with middle schools: Emerging adulthood effects on substance misuse and conduct problem behaviors through 14 years past baseline. *Child Development*, *93*(4), 925-940. https://doi.org/10.1111/cdev.13746

Spoth, R., Trudeau, L., Guyll, M., Shin, C., & Redmond, C. (2009). Universal intervention effects on substance use among young adults mediated by delayed adolescent substance initiation. *Journal of Consulting and Clinical Psychology*, *77*(4), 620. https://doi.org/10.1037/a0016029

Stipek, D. (2017b). *The preschool fade-out effect is not inevitable.* Education Week. https://www.edweek.org/teaching-learning/opinion-the-preschool-fade-out-effect-is-not-inevitable/2017/03