Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1369

Using experimental variation to examine the (co-)development of cognitive and social-emotional skills in early childhood

Emma R. Hart
Teachers College, Columbia University

Caroline M. Botvin
Teachers College, Columbia University

Drew H. Bailey
University of California, Irvine

Tyler W. Watts
Teachers College, Columbia University

Questions about the stability of psychological constructs, skill generalization, and transfer have long motivated psychological research. Despite a proliferation of theory, the field has rarely established causal effects. We employed a novel approach to test the stability and codevelopment of cognitive and social-emotional skills in early childhood using longitudinal randomized controlled trial data from the nationally representative Head Start Impact Study (n = 4,667). Capitalizing on the study's clustered design, we computed treatment effects on both skills for each cluster (k = 84). Using meta-analytic techniques, we found that changes to children's cognitive skills persisted at a rate of approximately 40% one year after program end and 30% two years after program end. Changes to social-emotional skills persisted at a rate of approximately 20% at both timepoints, though estimates were statistically non-significant. We observed more consistent, but not statistically significant, support for cognitive to socialemotional skill transfer. While models relying on exogenous variation attenuated traditional correlational estimates of same-skill associations, correlational estimates of cross-skill associations appeared to be less biased.

VERSION: January 2026

Suggested citation: Hart, Emma R., Caroline M. Botvin, Drew H. Bailey, and Tyler W. Watts. (2026). Using experimental variation to examine the (co-)development of cognitive and social-emotional skills in early childhood. (EdWorkingPaper: 25-1369). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/sb70-rd25

Using experimental variation to examine the (co-)development of cognitive and social-emotional skills in early childhood

Emma R. Hart, Caroline M. Botvin, Drew H. Bailey, Tyler W. Watts*

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

Acknowledgments

Thanks to the NICHD (1R01HD095930-01A1) and NSF (DGE-2036197) for support and Jens Dietrichson, Jade Jenkins, and Stephen Raudenbush, and Michael Weiss for feedback on earlier versions of this work.

Author Note

The data needed to reproduce the analyses presented in this manuscript (United States Department of Health and Human Services, 2018) are available, with approval, from the Inter-university Consortium for Political and Social Research (ICPSR) at: https://www.icpsr.umich.edu/web/ICPSR/studies/36968. The analytic syntax necessary to replicate these findings will be publicly posted, and linked here, prior to publication. This study was not pre-registered.

Abstract

Questions about the stability of psychological constructs, skill generalization, and transfer have long motivated psychological research. Despite a proliferation of theory, the field has rarely established causal effects. We employed a novel approach to test the stability and co-development of cognitive and social-emotional skills in early childhood using longitudinal randomized controlled trial data from the nationally representative Head Start Impact Study (n = 4,667). Capitalizing on the study's clustered design, we computed treatment effects on both skills for each cluster (k = 84). Using meta-analytic techniques, we found that changes to children's cognitive skills persisted at a rate of approximately 40% one year after program end and 30% two years after program end. Changes to social-emotional skills persisted at a rate of approximately 20% at both timepoints, though estimates were statistically non-significant. We observed more consistent, but not statistically significant, support for cognitive to social-emotional skill transfer. While models relying on exogenous variation attenuated traditional correlational estimates of same-skill associations, correlational estimates of cross-skill associations appeared to be less biased.

Keywords: skill building, social-emotional, cognitive, transfer, early childhood, randomized controlled trial, meta-analysis

Using experimental variation to examine the (co-)development of cognitive and social-emotional skills in early childhood

Causal dynamics between the same and different psychological constructs are a central focus of research. Across sub-disciplines of psychology, researchers have employed a variety of techniques to investigate the causal dynamics that link performance in one area to later performance in the same, and other, domains. Researchers have long (e.g., Thorndike, 1924) used experiments to test whether cognitive skill training effects transfer to broader abilities (see Gobet & Sala, 2023; Green et al., 2019). Longitudinal correlational studies have also been used to examine the co-development of competencies over months and years (e.g., Napolitano et al., 2025; Roemer et al., 2022; Peng et al., 2019) and to investigate how early skills shape life outcomes decades later (Burchinal & Vandell, 2025; Duncan et al., 2007; Moffitt et al., 2011). An adjacent body of experimental and correlational research has examined the stability of skills across development, often with an interest in isolating the causal role that earlier skills play in shaping later skills in the same domain (Bailey et al., 2018; Breit et al., 2024; Perry et al., 2018; Watts et al., 2017).

Developmental theory supports the expectation that skills develop through complex interactions with other skills, contexts, and relationships (Bronfenbrenner, 1992). Psychologists and economists alike have argued that ‘skills beget skills’ through self- and cross-productivities, cascades, and positive feedback loops (Cunha & Heckman, 2007; Masten & Cicchetti, 2010). Interdependent developmental processes may function in highly complex ways (Trapp et al., 2019; van Geert, 1994). The mystery of these interrelated developmental dynamics is often implicated when studies face difficulty predicting life outcomes using earlier measures of functioning (Liou et al., 2023; Lundberg et al., 2024).

In developmental psychology, a great deal of attention has been directed towards understanding the development of cognitive and social-emotional skills. Identifying how to best promote early development in both domains has been a high priority given that persistent disparities in student achievement are present by school entry (Reardon, 2011). Advocates often argue that early social-emotional development is key to optimal long-run outcomes across the board (e.g., CASEL, 2025). It is unclear, however, what early skills interventions should target for maximum impact.

Theories regarding the development of cognitive and social-emotional skills generally assert that rudimentary skills developed earlier in life set the stage for the acquisition of sophisticated skills later in development. For example, earlier math competencies are thought to lay the groundwork for more advanced math skills (Sarama & Clements, 2009). Likewise, social (mal-)adjustment at one stage of development is expected to cascade forward and shape subsequent social functioning (Dodge et al., 1986).

In the case of cross-domain development, researchers have proposed a variety of specific theories that predict the co-development of social-emotional and cognitive skills. Stronger self-regulation may equip children to learn in distraction-ridden environments (Blair & Raver, 2015; McClelland et al., 2007). At the same time, cognitive processes, like executive functioning, may lay the foundation for emotional self-regulation by providing children with the cognitive capacity to modulate their behavior (Li et al., 2023; Ursache et al., 2012). Social-emotional skills may support children to form positive relationships with peers and teachers, increasing their enjoyment of school and, in turn, motivation to learn (Zins, 2004). Likewise, succeeding academically may increase enjoyment (Miles & Stipek, 2006), and stronger language abilities may enable positive social interactions (Chow & Wehby, 2018; Stansbury & Zimmermann, 1999; Vygotsky, 1986). Strong self-esteem may support students to seek academic opportunities (Denham & Brown, 2010).

Historically, most empirical investigations of cognitive and social-emotional skill development have relied on correlational designs. Although correlational methods often produce evidence of auto-regressive (e.g., Breit et al., 2024) and cross-lagged effects (e.g., Hübner et al., 2022), it is impossible to fully disentangle whether observed associations reflect the causal effects of the skills of interest or the influence of other stable or dynamic child and contextual characteristics (i.e., confounding; Bailey et al., 2018; Berry & Willoughby, 2017; Rohrer & Lucas, 2020). Accordingly, methodologists have developed techniques that disaggregate the portion of auto-regressive and cross-lagged associations that are explained by stable between-child factors from the portion that reflects within-child dynamics (Hamaker et al., 2015). Building from traditional Cross-Lagged Panel Modeling—developmental psychology's "workhorse" (Berry & Willoughby, 2017)—such advancements have attempted to approximate the causal dynamics of skill development with stronger controls for confounding. However, the extent to which the estimates produced from such models are causal remains controversial (Hamaker, 2023; Lüdtke & Robitzsch, 2022).

Longitudinal evaluations of intervention effects may be informative for advancing our understanding of skill development (Bailey et al., 2024). If functioning in one domain shapes functioning in the same domain and another domain, then intervening to change skill one should yield benefits to both skill one and skill two (Bailey et al., 2020; Protzko, 2017). Indeed, by examining the dynamics of experimentally generated variation in children's skills following randomization to receive, or not receive, an intervention, it may be possible to overcome the confounding that has limited correlational work on skill development. Hart et al. (2024) recently meta-analyzed the stability of RCT-induced changes in children's skills and found similar rates of stability for both social-emotional and cognitive skills that were lower than what correlational evidence often predicts (~45% of initial effects persisted 6 to 12 months after interventions ended). Cipriano et al. (2023) probed the possibility of cross-domain effects in SEL RCT and quasi-experiments, finding a ~.10 *SD* impact on academic ability. However, given that treatments are often broad in focus, it is challenging to isolate whether impacts on the skills primarily targeted by interventions are necessarily the cause of effects on non-targeted skills (Eronen, 2020). Thus, traditional experimental evidence can only yield limited inferences regarding the mechanisms underlying effects (Rohrer & Lucas, 2023). Differences in the targets and content of treatments further complicate between-study meta-analyses of transfer dynamics. Given the challenges of estimating the causal associations within and between psychological constructs, it may be useful to triangulate across studies using a variety of experimental and non-experimental methods.

Current Study

The current study aimed to provide a novel examination of the development of social-emotional and cognitive skills using RCT data from the Head Start Impact Study (HSIS; Puma et al., 2012). In the HSIS, children were randomized to a year of Head Start preschool services within centers that were nested within 84 larger "grantee" groups. Early childhood educational programs, like Head Start, are often thought to boost both social-emotional and cognitive development. Capitalizing on the clustered design of the study, in which the same treatment was provided, we estimated intervention impacts for each grantee. We then applied meta-analytic techniques to test whether grantees that generated larger impacts on social-emotional or cognitive skills at posttest also produced larger impacts on the same, and opposite, skill domains measured consistently at follow-up.

This approach allowed us to employ straightforward cross-lagged-panel modeling in data that contained exogenously generated variation in child skills, rather than naturally occurring variation, to examine the development of social-emotional and cognitive skills. In doing so, we improved upon previous correlational designs by guarding against confounding that otherwise biases estimates of the correlations between child skills. Our approach allowed us to approximate the extent to which stronger earlier cognitive and social-emotional skills causally shaped stronger later cognitive and social-emotional skills by relying on variation in skills generated by randomization to Head Start. In addition to building on past nonexperimental work that has addressed similar questions but is likely susceptible to omitted variables bias, our model builds on experimental work that has examined intervention impacts on non-targeted skill domains without adjustments for interventions' impacts on targeted skills (e.g., to deduce whether SEL interventions improve academic achievement through improved social-emotional functioning or any number of other factors). Although we did not have strong *a priori* hypotheses, we believe that our approach provides important new estimates that can advance our understanding of causal early childhood developmental processes.

Methods

Data

The current study used data from the Head Start Impact Study (HSIS). The HSIS was an RCT evaluation of Head Start services provided in the 2002-2003 school year. The federal government launched Head Start in the 1960s with the ambition of reducing socioeconomic disparities in development with a particular focus on school readiness (Puma et al., 2012). To test the effects of the program, Congress mandated a government-commissioned study to evaluate the primary Head Start function at the time: providing preschool services, as well as medical, dental, mental health supports, and parenting resources, to racially diverse three- and four-year-old children from families with low incomes. Here, we provide a brief overview of the HSIS data relevant to our study with information drawn from the original HSIS reports published by the U.S. Department of Health and Human Services, Administration for Children and Families (Puma et al., 2010; 2012). The original reports provide additional details on all aspects of study design.

Children (n = 4,667) were randomly assigned to receive, or not receive, Head Start slots within 378 Head Start centers. The study investigators aimed to create a nationally representative sample of centers through an intensive multistep process. Only oversubscribed centers (i.e., centers where more families attempted to enroll than there were slots) were included. Within each oversubscribed center, a random sample of children was selected from the applicant pool. From this random sample, children were then randomized to the treatment condition (i.e., offer of a limited Head Start slot) or to the control condition (i.e., no Head Start slot). Random assignment occurred within two cohorts: (1) a three-year-old cohort, and (2) a four-year-old cohort. Both cohorts were randomly assigned to the offer of a single year of Head Start services, after which families with children in the three-year-old cohort were free to seek additional services (which could have included another year of Head Start). Randomization was successful, with few statistically significant differences in baseline characteristics between the treatment and control groups.

Relevant to the current study, centers were nested within 84 “grantee/delegate agencies” (henceforth referred to as “grantees”). Both public and private agencies (e.g., school districts, faith-based institutions, non- and for-profits) served as grantees. Grantees reported to the federal Office of Head Start and were responsible for ensuring that Head Start centers administered high-quality, comprehensive programming that was responsive to their community’s needs.

Thus far, the HSIS has collected six waves of data. In the current study, we used assessments collected on a consistent schedule for children in the three-year-old and four-year-old cohorts to maximize our analytic sample. The assessment waves included: (1) fall of 2002 pre-test assessment at baseline; (2) spring of 2003 posttest assessment at the end of the intervention; (3) spring of 2004 follow-up assessment collected a year after posttest; and (4) spring of 2005 follow-up assessment collected two years after posttest.

Participants

Table 1 presents baseline characteristics for children randomized to the treatment and control groups.[1] Recall that to be eligible for Head Start, a child’s family income was required to fall below the federal poverty line.[2] The sample was nearly equally split among Black, White, and Hispanic children. Approximately 25% of children spoke Spanish as their primary language, and close to 70% of families indicated Spanish as the primary language spoken in the home. More than 60% of mothers had a high school degree or higher. In about 50% of families, both parents lived in the home, and approximately 45% of mothers were married. Fewer than 20% of mothers gave birth as teenagers.

Measures

---

[1] We did not use the weights from the original HSIS reports as we were not interested in recovering nationally representative statistics.
[2] Some centers allowed up to 10% of enrollees to have incomes above the poverty line (see Puma et al., 2010).

Cognitive and Social-Emotional Composites

The primary variables used in our analyses were social-emotional and cognitive composite scores calculated at pre-test (for use as a covariate), posttest, 1-year follow-up, and 2-year follow-up assessment waves. To form these composites, we first restricted our dataset to only the measures that were collected consistently across the posttest and follow-up assessment waves.[3] Several parent- or teacher-report measures capturing behavioral problems, social skills, learning skills, and child-adult relationship quality composed the social-emotional skill measures, including the: (1) total problems score from the *Adapted Child Behavior Checklist*; (2) *Developing Skills Checklist*; (3) *Social Skills and Positive Approaches*; (4) *Adjustment Scales for Preschool Intervention;* (5) *Parent-child Relationship Scale*; and (6) *Teacher-child Relationship Scale*. Direct assessments of language and math performance, as well as parent reports of emerging literacy, composed the cognitive assessments including: (1) the *Peabody Picture Vocabulary Test* in both English and Spanish; (2) the oral comprehension, letter-word identification (in both English and Spanish)[4], and applied problems subscales from the *Woodcock-Johnson*; and (3) the *Emergent Literacy Scale*. The Puma et al. (2012) report contains additional details on each assessment.

To prepare to generate composite scores, we first standardized each measure using the control-group standard deviation. We multiplied standardized negatively valenced measures (e.g., behavioral problems) by -1 so that higher scores were consistently indicative of higher functioning performance (e.g., stronger vocabulary, fewer behavioral problems). We then averaged the valence-adjusted standardized cognitive and social-emotional measures to form the

[3] Other than *Adjustment Scales for Preschool Intervention, Parent-child Relationship Scale, and Teacher-child Relationship Scale* (only at posttest and follow-up) measures were collected at pre-test, posttest, and follow-up.
[4] Both English and Spanish assessments were included in composites for children who completed both.

composites. If children completed both the English and Spanish assessments, both scores were included in the composite. Across assessment waves, the Cronbach's Alpha ranged from .83 to .87 for the cognitive composites and from .73 to .77 for the social-emotional composites. Finally, we re-standardized the composite scores using the sample-wide control-group standard deviation. Composite scores were generated for participants if they had non-missing data for at least one measure collected at the assessment wave of focus.

(Quasi)experimental Grantee-level Treatment Impacts

Rationale. We relied on meta-analytic techniques to examine the extent to which intervention-induced changes to social-emotional and cognitive were stable over time and associated with subsequent intervention effects in the opposite domain. We were interested in examining the associations among intervention impacts on child skills because intervention impacts—generated through randomization to the treatment or control groups—should capture exogenous skill differences that are not confounded by the many observed and unobserved factors that otherwise lead child skills to be associated across time (Duncan et al., 2004). In a conventional correlational study of children's skills, we can examine whether children with stronger skills at time one tend to then have stronger skills in the same and other domains at time two. However, any number of factors that matter for development (e.g., environment, opportunities, relationships, genetics) could shape children's skills across time and confound estimates of the causal links among skills. For example, the observed association between skill one and two could be caused by socioeconomic pressures rather than skill building processes. In the current study, instead of examining variation in children's skills at the individual child level, we attempted to examine variation in intervention impacts.

In a traditional meta-analytic framework, one could test the associations among intervention impacts by compiling longitudinal intervention impacts from many studies with varied initial effects. In this paradigm, if auto-regressive and cross-domain skill building dynamics are at play, then the magnitude of effects on skill two should be a function of the magnitude of earlier effects on skill two and skill one; studies that generated larger effects on skill two and on skill one should produce larger effects on skill two. Although this kind of meta-analysis would provide useful insights regarding the extent to which intervention-induced changes in children's skills persist and transfer, it is challenging to accomplish in practice. Few interventions conduct follow-up assessments (Watts et al., 2019), and even fewer collect data on *both* social-emotional and cognitive outcomes. Among studies that do, longitudinal assessment schedules are rarely aligned across studies.

In the current study, we applied a similar meta-analytic framework using data from one intervention, the HSIS, instead of many. In the HSIS, children's social-emotional and cognitive skills were measured using the same instruments at the end of the program and several follow-up assessment waves. The cluster-based randomization of children (within centers and grantees) allowed us to compute grantee-level treatment effects that we then analyzed using meta-analytic techniques demonstrated in Watts et al. (2024). We essentially treated each grantee as a small-scale experiment. Given heterogeneity in the initial effects of Head Start across grantees (also observed by Bloom & Weiland, 2015 across centers), we were then able to longitudinally examine the extent to which larger benefits in one skill domain corresponded with larger subsequent benefits in the same and the other skill domain. Critically, variation in treatment impacts across grantees should be exogeneous and otherwise unrelated to the individual factors that might otherwise cause variation in child skills across time and, thus, our models should be guarded against omitted variables bias.

Estimation Method. First, we estimated treatment impacts on cognitive and social-emotional composites at the grantee level for each assessment wave as follows:

$$zComposite_{ct} = \beta_0 + \beta_1 TX_c + \alpha_g + \gamma_g TX_c + \delta' X_{ct} + \phi' X_c + \varepsilon_{ct} \tag{1}$$

where g indicated grantee, c indicated child, and t indicated assessment wave. In each model, we regressed the wave-specific composite of focus (social-emotional or cognitive) on a dummy variable indicator (TX_c) for the child's treatment status (0 = control; 1 = treatment), a fixed effect for the grantee the child was nested within (α_g), and the interaction between the grantee and the treatment status ($\gamma_g TX_c$). For each outcome wave, we also included time-specific controls for age at the assessment, and binary indicators for whether age was mean-imputed ($\delta' X_{ct}$). Additionally, we included baseline child controls ($\phi' X_c$). Baseline child controls included pre-test cognitive and social-emotional composites, as well as a host of additional covariates included in the Puma et al. (2012) HSIS report: child sex, child race/ethnicity, child's primary language, primary language spoken at home, primary caregiver's age, whether both biological parents live with the child, whether the biological mother is a recent immigrant, maternal educational attainment, maternal marital status, and whether the mother gave birth as a teenager.[5]

After fitting each linear model, we used a marginal effects function to identify the treatment effect, and associated standard error, for each grantee. Recall that the social-emotional and cognitive composites were standardized using concurrent assessment-specific control-group

[5] We included child age at assessment in our models, and not number of weeks elapsed from 9/1/02 and testing/the parent intervention which Puma et al., (2012) used.

standard deviations. As such, the grantee-specific impacts can be interpreted as effect sizes measuring the impact of receiving a randomly assigned slot to attend Head Start. Ultimately, this estimation process resulted in a treatment impact estimate for each grantee at each assessment wave for both the social-emotional and cognitive composites.

Assumptions. Importantly, this approach relies on the strong assumption that grantee-level treatment impacts are exogenously generated and not caused by pre-existing differences in achievement or social-emotional functioning between treatment and control children. If this exogeneity assumption were violated (i.e., random assignment did not produce groups with equal outcomes on expectation), associations between cognitive and social-emotional impacts could reflect underlying stability in child characteristics rather than causal transfer processes. We ran a number of sensitivity analyses (reported in the results section and further detailed in the supplement) to assess the degree to which this assumption held, and to reduce bias caused by any violations. Ultimately, we label our grantee-level impacts as "quasi-experimental" because we cannot rule out baseline differences between the two groups at the grantee level, and because many grantees included in our sample had small sample sizes (posttest $\bar{n}$ = 79), making it less likely that random assignment produced entirely balanced groups.

It is worth noting that concerns about exogeneity shaped our decision to compute effects at the grantee level. Indeed, the HSIS had two levels of clustering: centers (k = 378) and grantees (k = 84). In our initial conceptualization of these analyses, we had intended to compute center-level treatment impacts. However, we found that these impacts were highly imprecise (and standard errors were often uncorrelated with sample size) largely due to the fact that centers often contained very few children (posttest $\bar{n}$ = 13). For this reason, we did not proceed with running our analyses using the center-level estimates and, instead, we opted to estimate treatment impacts and run our analyses at the grantee level, for which we observed larger sample sizes and plausible standard error estimates.

The grantee-level impacts on social-emotional and cognitive composites largely reflected the child-level analyses reported in Puma et al. (2010; 2012). As shown in Table S1, we observed a meta-analytic average impact of ~.20 *SD* at posttest on the cognitive composite that faded to near zero at subsequent assessment waves. The impacts on social-emotional composites were near zero, on average, across all waves.

Importantly, we observed considerable variation in impacts across grantees and assessment waves (see Figure 1). Heterogeneity statistics suggested meaningful variation in posttest effects across grantees for both cognitive composites ($\tau = 0.12$) and social-emotional composites ($\tau = 0.11$). Q-tests indicated that overall between-grantee variation in the posttest impacts was statistically significant for the cognitive composites ($Q(79) = 108.79$, $p = .01$), but not the social-emotional composites ($Q(79) = 89.73$, $p = .19$). At 1- and 2-year follow-ups, estimable heterogeneity dropped to $\tau = 0.07$ ($Q(79) = 83.19$, $p = .35$) and then $\tau = 0.03$ ($Q(78) = 90.78$, $p = .15$) for cognitive composites. For the social-emotional composites, estimable between-grantee variation dropped to $\tau = 0.00$ at 1-year follow-up ($Q(80) = 80.84$, $p = .45$) and $\tau = 0.00$ at 2-year follow-up ($Q(80) = 74.86$, $p = .64$).

Data Coverage. Table 2 presents details on data coverage across assessment waves and outcomes. The HSIS originally randomized 4,667 children within 84 grantees. However, the publicly available data excludes children from Puerto Rico. As such, our base sample size was 4,442 children from 83 grantees. Within these 83 grantees, an average of 32 children were randomized to the treatment and 22 were randomized to control.

Of these 83 grantees, we were only able to compute impacts for 78 to 80 grantees (depending on the outcome and assessment wave) that had child-level data for at least one child in both the treatment and the control groups. Across cognitive and social-emotional impacts and assessment waves, 80% to 85% of participants contributed data to our grantee-level impacts (n = 3,580 to 3,820). On average, each grantee-level impact was estimated using non-missing data from 28 to 30 treatment children and 17 to 18 control children. It should be noted that some grantees had very few children in either the control or treatment group. Our meta-analytic approach (detailed in the "Analysis" section) addressed this issue by weighting grantee-level impacts according to their precision (a function of sample size) so that grantees with few children were down-weighted. We also ran supplemental models that dropped grantees with particularly low sample sizes that contained less than 10 treatment or control children, and models that included only children who consistently provided data at each assessment wave on both measures.

Analysis

Using the grantee-level, wave-specific treatment impacts on the cognitive and social-emotional composites, we ran a series of regressions to identify the auto-regressive and cross-lagged associations between social-emotional and cognitive functioning over time. We set out to test the theory that stronger earlier skills support the development of stronger later skills in the same, and opposite, domain. Table 3 outlines the six independent linear regression models we ran to test for auto-regressive and cross-lagged effects. Consider Figure 2, which depicts Model 1, as an example. In this model, we sought to examine the (1) auto-regressive association between posttest impacts on cognitive skills and 1-year follow-up effects on cognitive skills, and (2) cross-lagged association between posttest impacts on social-emotional skills and 1-year follow-up impacts cognitive skills. Thus, we estimated both the auto-regressive association between same-skill impacts measured over time (represented by "1" in the figure) and the cross-lagged association between different skills (represented by "2").

We ran the models in two steps. First, we fit a basic model in which we regressed grantee-level cognitive impacts at 1-year follow-up on grantee-level cognitive impacts at posttest. As explained in Hart et al., (2024), the slope from this model provides an estimate of the "conditional persistence rate" for an intervention impact over time. Next, we introduced social-emotional posttest impacts to the model, which enabled us to estimate the cross-lagged association between cognitive and social-emotional skills across time (depicted as "2" in Figure 2), while controlling for the initial impact on cognitive skills.

The auto-regressive association indicated what portion of the intervention-induced change in children's skills persisted at follow-up. A larger conditional persistence rate would be consistent with the expectation that stronger earlier skills beget stronger later skills. A smaller conditional persistence rate would indicate that intervention-induced changes to children's skills fade out with time. The cross-lagged association indicated the extent to which grantees that generated larger boosts to social-emotional functioning at posttest then observed corresponding benefits to cognitive functioning a year later, above and beyond the initial benefits to cognitive functioning. A larger slope coefficient for social-emotional posttest impacts would be consistent with the expectations of cross-skill transfer. Insofar as the grantee-level impacts were biased by exogeneity violations or were correlated with treatment impacts on other outcomes (other than measured cognitive impacts, which were accounted for), these confounds would bias any observed cross-lagged associations. In addition to the slope coefficient, the models also estimated an intercept term. In the context of our full cross-lagged model, the intercept captured the portion of the follow-up effect on cognitive skills that was unexplained by posttest effects on cognitive and social-emotional skills.

Model

We ran these full cross-lagged models with either posttest or 1-year follow-up impacts as the independent variable and 1- or 2-year follow-up effects as the dependent variable. Each model was executed in R using the metafor and clubSandwich packages as follows:

Level 1:

$$TX\_Impact_{kgt} = \beta_{0g} + \beta_1 TX\_Impact_{kg(t-x)} + \beta_2 TX\_Impact_{-kg(t-x)} + \varepsilon_{gt} \tag{2}$$

Level 2:

$$\beta_{0g} = \gamma_{00} + u_{0g} \tag{3}$$

where g indicated grantee, t represented assessment wave/time, x represented the time elapsed between assessment waves (one or two years), and k represented whether the outcome was social-emotional or cognitive (with $-k$ indicating the opposite of k). Thus, at Level 1 on the left-hand side of the equation, TX_Impact_{kgt} represented the effect size for grantee g at assessment t (1- or 2-year follow-up) on outcome k. On the right-hand side of the equation, TX_Impact_{-kgt-1} and TX_Impact_{kgt-1} indicated the effect size for the corresponding grantee at a prior assessment wave (posttest or 1-year follow-up) on the opposite skill domain ($-k$) and same skill domain (k), respectively. ε_{gt} represented error in the estimation of TX_Impact_{kgt}.

We included random intercepts in the model, represented in Level 2. γ_{00} represented the mean intercept across grantees and u_{0g} represented the site-specific deviation in follow-up effects (with an assumed variance of τ^2). Each model employed random-effects weighting for the inverse variances of the dependent variable (TX_Impact_{kgt}).

Considered as a whole, in the context of the hypothetical model depicted in Figure 2, β_1 represented the autoregressive association between cognitive skills at posttest and follow-up. A larger β_1 would indicate greater persistence of posttest cognitive impacts. β_2 represented the extent to which grantees that generated larger treatment benefits on social-emotional functioning at posttest then observed larger treatment effects on cognitive functioning at follow-up, holding constant the initial impacts of the intervention on cognitive skills. A positive β_2 would indicate cross-lagged transfer. Finally, the cross-grantee average intercept, γ_{00}, represented the portion of the follow-up effect on cognitive skills that was not explained by posttest impacts on either social-emotional or cognitive skills. A larger γ_{00} would indicate that a significant portion of the benefit of an intervention on cognitive skills at follow-up must have been generated through mechanisms other than initial intervention benefits on cognitive or social-emotional skills.

Results

Longitudinal Descriptives for Quasi-experimental Impacts

Figure 1 depicts the trajectories of grantee-level social-emotional and cognitive impacts across assessment waves. As shown in Table S1, the meta-analytic average of social-emotional impacts was consistently near 0 *SD* at all assessment waves, though estimates ranged from approximately –1 *SD* to 2 *SD* across waves (the middle 80% of the distribution ranged from about -.50 *SD* to .40 *SD*) suggesting variation across grantees in the extent to which randomization to Head Start affected social-emotional outcomes.[6] The meta-analytic average of cognitive impacts was .18 *SD* at posttest, but faded to approximately 0 *SD* at follow-up, with impacts ranging from about – 2 *SD* to 1.50 *SD* across assessment waves (middle 80% ranged from approximately -.40 *SD* to .50 *SD*).

Table 4 presents the raw, unweighted correlations between social-emotional and cognitive impacts across assessment waves. The associations among social-emotional composite impacts ranged from r = -.03 to .26 (p = .02 to .82) across waves. Associations among cognitive composite impacts were larger, ranging from r = .38 to .67 ($p < .001$). At posttest, we observed little correspondence between cognitive and social-emotional impacts (r = -.09, p = .42), suggesting that the grantees that generated larger end-of-treatment effects on cognitive skills were not necessarily the same grantees that generated larger posttest impacts on social-emotional skills. 1-year (r = .22, p = .05) and 2-year (r = .14, p = .21) social-emotional and cognitive follow-up impacts demonstrated greater correspondence. The larger association at follow-up is consistent with the possibility of transfer across skills; however, we cannot determine the direction of transfer between socioemotional and cognitive skills based on these associations, nor can we rule out the possibility of unmeasured mediators affecting both cognitive and social-emotional skill impacts in the post-treatment period.

Autoregressive Associations among Quasi-experimental Impacts

We first examined the auto-regressive associations among impacts over time, prior to entering cross-domain impacts to our model (see Table 5, columns 1 and 3). Overall, we observed that posttest impacts were predictive of 1- and 2-year follow-up effects, in line with past work (see Hart et al., 2024). At the 1-year follow-up wave, we observed that posttest

---

[6] Note that Figure 1 does not depict grantees with less than 10 participants and, hence, does not exactly align with the range of estimates for the full sample (including grantees with less than 10 participants) described here.

impacts persisted at a rate of 38% for cognitive skills (p = .001) and 20% for social-emotional skills (p = .11). At 2-year follow-up, conditional persistence further reduced to 27% for cognitive impacts (p = .04) and to 17% for social-emotional impacts (p = .14). 1-year follow-up impacts were very predictive of 2-year follow-up impacts for cognitive skills (67% persistence of remaining 1-year impact; p < .001) and, to a lesser extent, for social-emotional skills (31% of 1-year impact; p = .004). Across waves, estimates for cognitive outcomes were more precise and statistically significant than estimates for social-emotional outcomes.

We also observed near-zero, negative intercept effects (on average, β_0 = -.02) across the models that were statistically non-significant. The estimates indicated that no predictable portion of the observed variation in follow-up impacts was explained by factors other than earlier intervention impacts on cognitive or social-emotional skills; unmeasured mediators did not appear to drive the persistence of cognitive or socioemotional impacts in this study.

When we then introduced cross-domain posttest impacts to these models (Table 5, columns 2 and 4), the auto-regressive slope paths generally demonstrated a slight reduction (by .04 on average). The small decrease indicated that cognitive and social-emotional skill impacts did not share much common variance in predicting follow-up impacts. If anything, the intercepts became more negative, though remained small (on average, β_0 = -.04) with the inclusion of the cross-domain effects.

Skill-Building Dynamics using Quasi-experimental Impacts

Figure 3 and Table 5 present the results from our analyses examining whether grantees that generated larger initial impacts on social-emotional skills observed larger subsequent follow-up impacts on cognitive skills, and vice versa. Overall, cross-lagged associations ranged from .01 to .24. Cross-lagged estimates were consistently imprecise, and only one path was statistically significant (p < .05). Considered together, we found more support for cognitive-driven transfer effects (average β_1= .20) than social-emotional-driven transfer effects (average β_1= .10).

Larger cognitive impacts at posttest predicted larger social-emotional impacts at 1-year follow-up (β_1= .18, p = .23) and 2-year follow-up (β_1= .21, p = .15). Additionally, larger 1-year follow-up impacts predicted larger social-emotional impacts at 2-year follow-up (β_1= .21, p = .14). For every 1 *SD* increase in Head Start's impact on cognitive skills, the results indicate that one would expect a ~.20 *SD* increase in social-emotional performance a year or two later. In the context of the HSIS study, the magnitude of these effects is likely small. For example, compared with a grantee that observed a posttest cognitive impact at the meta-analytic average (.18 *SD*), a grantee that observed a posttest impact 1 standard deviation above the average (.30 *SD*, aligned with the observed τ =.12) would observe a .02 *SD* larger follow-up effect on social-emotional functioning.

Social-emotional impacts were less consistently predictive of cognitive impacts. From posttest to 1-year follow-up, and 1-year to 2-year follow-up, social-emotional effects at time 1 were minimally predictive of cognitive effects at time 2 (β_1= .01, p = .93 and β_1= .04, p = .70, respectively). Interestingly, in contrast, larger social-emotional posttest impacts were predictive of larger cognitive 2-year follow-up effects (β_1= .24, p = .03).

Additional Analyses

Exploratory Analysis of Skill-Building Dynamics using Observational Data

Next, we performed a within-study comparison of the estimates generated through our quasi-experimental grantee-level impacts and those generated through a more traditional correlational analysis. Within-study comparisons are a useful tool for examining the extent to which the patterns observed in observational data converge with those drawn from more causally relevant methods (see Wan et al., 2023). Past within-study analyses have found that correlational methods often produce auto-regressive associations that are larger than those generated using experimental data, highlighting the contribution of confounding bias (e.g., Bailey et al., 2018; Wan et al., 2023). However, this work has focused on cognitive outcomes and auto-regressive associations; the extent to which social-emotional estimates and cross-lagged associations are biased is less clear, though simulations suggest that bias is likely (Lüdtke & Robitzsch, 2022).

For the correlational analysis, we limited the sample to children from the control group and executed a series of models in which we regressed child-level composites in one domain at follow-up on child-level composites in the opposite skill domain at an earlier assessment wave. We included controls for functioning in the same-skill-domain at the earlier assessment wave, child and family characteristics, and pre-test social-emotional and cognitive composites.

Table S2 and Figure S1 depict the results. All of the auto-regressive and cross-lagged associations were statistically significant at $p < .05$. The auto-regressive paths in the child-level models were, on average, about 10% to 140% larger than those from the quasi-experimental models for both cognitive and social-emotional skills (average = 67% larger, 71% for social-emotional skills and 63% for cognitive skills).

In contrast, the cross-lagged paths generated using the correlational methods were not consistently larger than those estimated using our quasi-experimental approach. For cognitive skills, which showed more evidence of cross-lagged transfer in the quasi-experimental data, the cross-lagged paths were about 30% to 60% *smaller* when relying on non-experimental variation (.08 to .15 *SD* as opposed to ~.20 *SD*). For social-emotional skills, the non-experimental cross-lagged associations were consistently observed to be around .05 to .07 *SD,* marking small increases from the estimates observed in two of the three grantee-level estimates, and a decrease from the .24 *SD* cross-lagged coefficient observed at one wave. Consistent with the grantee-level estimates, cognitive skills were more strongly predictive of social-emotional skills in the child-level data than vice versa.

Sensitivity Analyses

We then ran a series of analyses to examine the extent to which the quasi-experimental grantee-level impacts reflected treatment-control differences due to randomization versus endogenous child characteristics. The findings from each sensitivity analysis are summarized here, with additional details provided in the supplemental text and Tables S3 – S8. In brief, we first examined grantee-level baseline balance, and we observed that several grantees showed treatment-control-group imbalance. To probe whether our estimates were biased by grantee-level imbalances, we first ran a model that substituted posttest impacts for pre-test "impacts," generated prior to random assignment. Had pre-test "impacts" shown similar patterns as the true posttest "impacts," it would have suggested that our grantee-level impacts may have captured endogenous differences between children in the skill composite variables, as opposed to variation generated through random assignment. However, in support of our assumption that the grantee-level effects captured variation due to random assignment, the results using pre-test "impacts" did not resemble the estimates from our primary models that used posttest impacts. Nonetheless, we ran additional models that further probed the sensitivity of our findings to the: (1) removal of particularly small-sample grantees (for which imbalance is most likely), (2) removal of pre-test demographic and child performance covariates, (3) reliance on the same analytic sample across assessment waves, and (4) inclusion of pre-test "impacts" as covariates in our second-stage meta-analytic models. The auto-regressive associations were largely aligned with our primary estimates. Although the cross-lagged estimates varied across analyses and remained imprecise and not statistically significant, we generally found more compelling support for cognitive to social-emotional transfer than vice versa. Finally, we ran a common-effects model, as a reasonable alternative to our random-effects approach. Our estimates were nearly identical under this specification.

Discussion

In the present study, we conducted a novel examination of cognitive and social-emotional skill development using grantee-level data from the HSIS. Changes to children's skills persisted at a rate of approximately 40% to 20% one to two years after program end, with greater stability for cognitive skills. We did not observe strong evidence of transfer effects from cognitive to social-emotional skills or vice versa, as estimates were generally imprecise and statistically non-significant across models and timepoints. However, of note, above and beyond the contribution of Head Start's initial impacts on social-emotional skills, a 1 *SD* increase in the impact of Head Start on cognitive skills consistently predicted an additional ~.20 *SD* magnitude increase in impact on social-emotional skills one and two years later. Traditional correlational methods appeared to produce upwardly biased auto-regressive, but not cross-lagged, estimates.

A key innovation of the present study was our examination of exogenously generated variation stemming from the *same treatment* on the *same outcomes* assessed longitudinally. We observed that changes in cognitive skills persisted at a rate very similar to that observed in Hart et al.'s (2024) meta-analysis of highly diverse educational programs, as well as Watts et al.'s (2024) analysis of an early math intervention. In comparison with Hart et al., social-emotional impact persistence was weaker at 1-year follow-up, but stronger at 2-year follow-up, in the present study. Across studies with different methodological strengths, we are beginning to see an accumulation of evidence suggesting approximately 40% stability of exogenous skill changes one year after intervention end and even less (~20-30%) at two years after intervention end, with more consistency for cognitive skills than social-emotional skills. In contrast with Hart et al. (2024), we observed consistently small intercept effects, suggesting no evidence of unmeasured mediational processes driving follow-up impacts.

Indeed, same-skill stability did not appear to be meaningfully confounded by, or mediated through, posttest impacts on the opposite skill domain. The auto-regressive paths minimally reduced with the inclusion of cross-domain impacts. In other words, we did not find evidence to suggest that, for example, social-emotional impacts mediated persistent follow-up impacts on cognitive outcomes, as has been theorized in the case of prominent early childhood programs (Heckman & Kautz, 2012). Of note, grantees that generated larger posttest impacts on social-emotional skills were generally not the same grantees that generated larger posttest impacts on cognitive skills (r = -.09; also see Jackson, 2018; Liu & Loeb, 2021).

We did observe that, above and beyond same-skill stability, delegates with larger earlier impacts on cognitive skills consistently observed larger follow-up impacts on social-emotional skills. Although estimates were not statistically significant and, as such, we do not interpret this finding as particularly strong evidence for transfer effects, the consistency of the estimates in the context of our method—which should avoid confounding by observed and unobserved factors—suggests the possibility of cognitive to social-emotional transfer processes that future research should further investigate. Researchers may be surprised that the opposite was not observed. Especially in recent years, there has been fervent enthusiasm for interventions targeting social-emotional skills (Cipriano et al., 2023). Our findings indicate the field should not dismiss the important role cognitive skills and learning opportunities may play in shaping development.

Insofar as the consistency in the cross-lagged estimates between cognitive and social-emotional skills is suggestive of transfer dynamics, the findings of the present study may align with the Large Interconnected Network Theory (LINT; Bailey et al., 2024), which articulates how programs might generate long-run effects despite short-term fadeout. In line with the cross-lagged associations we observed, LINT argues that initial intervention impacts ripple out to a network of correlated skills, contexts, and relationships that interact over time to form a network-level effect that ultimately shapes adult functioning. Future examinations of the co-development of skills can provide clarity as to the specific transfer dynamics underlying long-run intervention effects.

Our findings also speak to methodological concerns about the use of correlational approaches to examine skill building dynamics. We were surprised to find that while observational auto-regressive paths were predictably larger than our quasi-experimental estimates, cross-lagged estimates were not. The findings indicate that correlational methods may lead us astray in the expectation of same-domain skills-beget-skills processes, but are potentially less biased in the context of cross-domain transfer (also see Orth et al., 2024). Whether the estimates produced through our method were void of confounding hinges on whether the grantee-level impacts captured variation in child skills due to randomization versus endogenous variation. Our sample was small, and our cross-lagged estimates were noisy; replication is needed. Triangulating findings from the vast observational literature, growing longitudinal intervention impact literature, and within-study comparisons of the two methods, may help to move the field forward (Bailey et al., 2024).

It is worth considering how measurement differences may have influenced our results. Cognitive skills were measured using direct assessments, while social-emotional skills relied on parent- and teacher-reports, which could differ in reliability and validity. The social-emotional composites did have lower Cronbach's alpha's, less heterogeneity, and weaker auto-regressive paths. However, we find it hard to imagine that the social-emotional assessments did not capture *any* meaningful heterogeneity. Indeed, the auto-regressive impacts were consistently greater than zero. Additionally, the social-emotional measures were much like those that other interventions *have* impacted (e.g., Hart et al., 2024) and those that have shown strong predictive validity in forecasting adult functioning (e.g., Koepp et al., 2023). Nonetheless, larger questions in the field surrounding how to best conceptualize and operationalize social-emotional functioning with theoretical specificity are certainly relevant (Inzlicht et al., 2020; McCoy & Sabol, 2025; Morrison & Grammer, 2016). Poor explication of theoretical constructs may not only limit instruments' reliability and validity (Shadish et al., 2002), but also the ability of an intervention like Head Start to directly target social-emotional development (see Inzlicht & Roberts, 2024). Indeed, we observed that Head Start generated near-zero posttest impacts. The social-emotional measures may have tapped constructs that are hard to define and to change. Finally, if social-emotional to cognitive skill transfer processes are real, but much smaller in magnitude than cognitive to social-emotional transfer processes, it is possible that the current study was underpowered to detect such effects.

Conclusion

Longitudinal intervention data may provide a new inroad for triangulating the answers to causal questions about skill stability and transfer. The use of block-level RCT data affords unique opportunities toward these ends. We hope future studies will replicate our approach to test dynamics across a wide range of psychological constructs.

References

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*(2), 55–97. https://doi.org/10.1177/1529100620915848

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D. H., & Sarama, J. (2018). Risky business: Correlation and causation in longitudinal studies of skill development. *The American Psychologist*, *73*(1), 81–94. https://doi.org/10.1037/amp0000146

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. https://doi.org/10.1146/annurev-devpsych-010923-103044

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*(4), 1186–1206. https://doi.org/10.1111/cdev.12660

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, *66*(1), 711–731. https://doi.org/10.1146/annurev-psych-010814-015221

Bloom, H. S., & Weiland, C. (2015). *Quantifying Variation in Head Start Effects on Young Children's Cognitive and Socio-Emotional Skills Using Data from the National Head Start Impact Study* (SSRN Scholarly Paper No. 2594430). Social Science Research Network. https://doi.org/10.2139/ssrn.2594430

Breit, M., Scherrer, V., Tucker-Drob, E. M., & Preckel, F. (2024). The stability of cognitive abilities: A meta-analytic review of longitudinal studies. *Psychological Bulletin*, *150*(4), 399. https://doi.org/10.1037/bul0000425

Bronfenbrenner, U. (1992). Ecological systems theory. In *Six theories of child development: Revised formulations and current issues* (pp. 187–249). Jessica Kingsley Publishers.

Burchinal, M., & Vandell, D. L. (2025). School entry skills and young adult outcomes. *Early Childhood Research Quarterly*, *72*, 1–12. https://doi.org/10.1016/j.ecresq.2025.01.004

CASEL. (2025). *Making the case for SEL.* CASEL. https://casel.org/fundamentals-of-sel/how-can-you-make-the-case-for-sel/

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, *126*(4), 1593–1660. https://doi.org/10.1093/qje/qjr041

Chow, J. C., & Wehby, J. H. (2018). Associations between language and problem behavior: A systematic review and correlational meta-analysis. *Educational Psychology Review*, *30*(1), 61–82. https://doi.org/10.1007/s10648-016-9385-z

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development*, *94*(5), 1181–1204. https://doi.org/10.1111/cdev.13968

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

Denham, S. A., & Brown, C. (2010). “Plays nice with others”: Social-emotional learning and academic success. *Early Education and Development*, *21*(5), 652–680. https://doi.org/10.1080/10409289.2010.497450

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., & Duckworth, K. (2007). School readiness and later achievement. Developmental psychology, 43(6), 1428.

Eronen, M. I. (2020). Causal discovery and the problem of psychological interventions. *New Ideas in Psychology*, *59*, 100785. https://doi.org/10.1016/j.newideapsych.2020.100785

Gobet, F., & Sala, G. (2023). Cognitive training: A field in search of a phenomenon. *Perspectives on Psychological Science*, *18*(1), 125–141. https://doi.org/10.1177/17456916221091830

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. https://doi.org/10.1093/qje/qjac036

Hamaker, E., Kuiper, R., & Grasman, R. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. https://doi.org/10.1037/a0038889

Hamaker, E. L. (2023). The within-between dispute in cross-lagged panel research and how to move forward. *Psychological Methods*. https://doi.org/10.1037/met0000600

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

Hübner, N., Merrell, C., Cramman, H., Little, J., Bolden, D., & Nagengast, B. (2022). Reading to learn? The co-development of mathematics and reading during primary school. *Child Development*, *93*(6), 1760–1776. https://doi.org/10.1111/cdev.13817

Inzlicht, M., & Roberts, B. W. (2024). The fable of state self-control. *Current Opinion in Psychology*, *58*, 101848. https://doi.org/10.1016/j.copsyc.2024.101848

Inzlicht, M., Werner, K. M., Briskin, J., & Roberts, B. (2020). Integrating models of self-regulation. Annual review of psychology, 72(1), 319-345.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non–test score outcomes. *Journal of Political Economy*, *126*(5), 2072–2107. https://doi.org/10.1086/699018

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., & Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights*, *2*(4), 491–508. https://doi.org/10.1257/aeri.20200029

Koepp, A. E., Watts, T. W., Gershoff, E. T., Ahmed, S. F., Davis-Kean, P., Duncan, G. J., Kuhfeld, M., & 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), 1389–1406. https://doi.org/10.1037/dev0001533

Li, C., Hart, E. R., Duncan, R. J., & Watts, T. W. (2023). Bi-directional relations between behavioral problems and executive function: Assessing the longitudinal development of self-regulation. *Developmental Science*, *26*(3), e13331. https://doi.org/10.1111/desc.13331

Liou, G., Bailey, D. H., Baldwin, C. R., Duckworth, A. L., & Tay, L. (2023, July 14). Why life outcomes are hard to predict. https://doi.org/10.31234/osf.io/7q2rx

Liu, J., & Loeb, S. (2021). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, *56*(2), 343–379. https://doi.org/10.3368/jhr.56.2.1216-8430R3

Lüdtke, O., & Robitzsch, A. (2022). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(6), 888–907. https://doi.org/10.1080/10705511.2022.2065278

Lundberg, I., Brown-Weinstock, R., Clampet-Lundquist, S., Pachman, S., Nelson, T. J., Yang, V., Edin, K., & Salganik, M. J. (2024). The origins of unpredictability in life outcome prediction tasks. *Proceedings of the National Academy of Sciences*, *121*(24), e2322973121. https://doi.org/10.1073/pnas.2322973121

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491–495. https://doi.org/10.1017/S0954579410000222

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, *43*(4), 947–959. https://doi.org/10.1037/0012-1649.43.4.947

McCoy, D. C., & Sabol, T. J. (2025). Overcoming the streetlight effect: Shining light on the foundations of learning and development in early childhood. *American Psychologist*, *80*(2), 135–147. https://doi.org/10.1037/amp0001432

Miles, S. B., & Stipek, D. (2006). Contemporaneous and longitudinal associations between social behavior and literacy achievement in a sample of low-income elementary school children. *Child Development*, *77*(1), 103–117. https://doi.org/10.1111/j.1467-8624.2006.00859.x

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & 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. https://doi.org/10.1073/pnas.1010076108

Morrison, F. J., & Grammer, J. K. (2016). Conceptual clutter and measurement mayhem: Proposals for cross-disciplinary integration in conceptualizing and measuring executive function. In *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 327–348). American Psychological Association. https://doi.org/10.1037/14797-015

Napolitano, C. M., Soto, C. J., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2025). Changes in social, emotional, and behavioral skills are associated with changes in secondary school students' important outcomes. *European Journal of Personality*, 08902070251328046. https://doi.org/10.1177/08902070251328046

Orth, U., Meier, L. L., Bühler, J. L., Dapp, L. C., Krauss, S., Messerli, D., & Robins, R. W. (2024). Effect size guidelines for cross-lagged effects. *Psychological methods*, *29*(2), 421. https://doi.org/10.1037/met0000499

Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, *145*(2), 189–236. https://doi.org/10.1037/bul0000182

Perry, N. B., Calkins, S. D., Dollar, J. M., Keane, S. P., & Shanahan, L. (2018). Self-regulation as a predictor of patterns of change in externalizing behaviors from infancy to adolescence. *Development and Psychopathology*, *30*(2), 497–510. https://doi.org/10.1017/S0954579417000992

Protzko, J. (2017). Raising IQ among school-aged children: Five meta-analyses and a review of randomized controlled trials. *Developmental Review*, *46*, 81–101. https://doi.org/10.1016/j.dr.2017.05.001

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third Grade Follow-up to the Head Start Impact Study Final Report*, OPRE Report# 2012-45, Washington, DC: Office of Planning, Research and Evaluation, Administration for Childrenand Families, U.S. Department of Health and Human Services.

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.

Roemer, L., Lechner, C. M., & Rammstedt, B. (2022). Beyond competencies: Associations between personality and school grades are largely independent of subject-specific and general cognitive competencies. *Journal of Intelligence*, *10*(2), 26. https://doi.org/10.3390/ jintelligence10020026

Rohrer, J., & Lucas, R. (2020). *Causal Effects of Well-Being on Health: It's Complicated*. OSF. https://doi.org/10.31234/osf.io/wgbe4

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (pp. xxi, 623). Houghton, Mifflin and Company.

Shawn Green, C., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., Bingel, U., Chein, J. M., Colzato, L. S., Edwards, J. D., Facoetti, A., Gazzaley, A., Gathercole, S. E., Ghisletta, P., Gori, S., Granic, I., Hillman, C. H., Hommel, B., Jaeggi, S. M., … Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement. *Journal of Cognitive Enhancement*, *3*(1), 2–29. https://doi.org/10.1007/s41465-018-0115-y

Stansbury, K., & Zimmermann, L. K. (1999). Relations among child language skills, maternal socialization of emotion regulation, and child behavior problems. *Child Psychiatry and Human Development*, *30*(2), 121–142. https://doi.org/10.1023/A:1021954402840

Thorndike, E. L. (1924). Mental discipline in high school studies. *Journal of Educational Psychology*, 15(2), 83–98. https://doi.org/10.1037/h0071035

Trapp, S., Blömeke, S., & Ziegler, M. (2019). The openness-fluid-crystallized-intelligence (OFCI) model and the environmental enrichment hypothesis. *Intelligence*, *73*, 30-40. https://doi.org/10.1016/j.intell.2019.01.009

Ursache, A., Blair, C., & Raver, C. C. (2012). The promotion of self-regulation as a means of enhancing school readiness and early achievement in children at risk for school failure. *Child Development Perspectives*, *6*(2), 122–128. https://doi.org/10.1111/j.1750-8606.2011.00209.x

United States Department of Health and Human Services, Administration for Children and Families. (2018). *Head Start Impact Study (HSIS), 2002-2006 [United States]* (ICPSR 29462) [Dataset]. Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR29462.v7

van Geert, P. (1994). *Dynamic systems of development: Change between complexity and chaos* (pp. xii, 300). Harvester Wheatsheaf.

Vygotsky, L. S. (1986). *Thought and language* (8. print). Mit Press.

Wan, S., Brick, T. R., Alvarez-Vargas, D., & Bailey, D. H. (2023). Triangulating on developmental models with a combination of experimental and nonexperimental estimates. *Developmental Psychology*, *59*(2), 216–228. https://doi.org/10.1037/dev0001490

Watts, T. W., Bailey, D. H., & Li, C. (2019). Aiming further: Addressing the need for high-quality longitudinal research in education. *Journal of Research on Educational Effectiveness*, *12*(4), 648–658. https://doi.org/10.1080/19345747.2019.1644692

Watts, T. W., Botvin, C. M., Bailey, D. H., Hart, E. R., Mattera, S., Clements, D. H., Sarama, J., Farran, D., & Lipsey, M. W. (2024). *Predicting persistence and fadeout across multi-site RCTs of an early childhood mathematics curriculum intervention.* University of Chicago Voices. https://bpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/a/1176/files/2025/04/Watts-et-al.-under-review-BB-Multisite-Fadeout.pdf

Watts, T. W., Clements, D. H., Sarama, J., Wolfe, C. B., Spitler, M. E., & Bailey, D. H. (2017). Does early mathematics intervention change the processes underlying children's learning? *Journal of Research on Educational Effectiveness*, *10*(1), 96–115. https://doi.org/10.1080/19345747.2016.1204640

Zins, J. E. (2004). *Building Academic Success on Social and Emotional Learning: What Does the Research Say?* Teachers College Press.

Table 1
Baseline Equivalence on Child and Parent Characteristics

	Control Group		Treatment Group		
	M (*SD*) / proportion	*n*	*M* (*SD*) / proportion	*n*	*p*
Child Characteristics					
Pre-test cognitive composite	-0.01 (1.00)	1796	0.01 (1.04)	2646	0.67
Pre-test social-emotional composite	-0.00 (1.00)	1796	0.00 (1.00)	2646	0.82
Female child	0.49	1796	0.50	2646	0.68
Black child	0.30	1796	0.31	2646	0.69
White child	0.33	1796	0.31	2646	0.28
Hispanic child	0.37	1796	0.38	2646	0.52
Child's primary language is Spanish	0.25	1796	0.26	2646	0.47
Parent Characteristics					
Primary caregivers age	28.65 (7.06)	1796	29.08 (7.52)	2646	0.05
Primary home language is Spanish	0.71	1796	0.70	2646	0.34
Both parents live in the home	0.49	1796	0.50	2646	0.92
Mother is a recent immigrant	0.19	1796	0.20	2646	0.50
Mother has less than a HS degree	0.39	1796	0.37	2646	0.29
Mother has more than a HS degree	0.28	1796	0.29	2646	0.49
Mother has a HS degree	0.33	1796	0.34	2646	0.66
Mother was never married	0.39	1796	0.39	2646	0.74
Mother is currently married	0.45	1796	0.44	2646	0.59
Mother has other relationship status	0.16	1796	0.16	2646	0.82
Missing mother's relationship status	0.00	1796	0.00	2646	0.50
Mother gave birth as a teen	0.18	1796	0.16	2646	0.04
Joint Test					
$\chi^2(18) = 10.71$, $p = 0.83$, $n = 4{,}442$.					

Notes: To generate the *p* values, each characteristic was regressed on an indicator for treatment using robust standard errors. To provide a joint test of orthogonality for overall treatment-control group differences, a probit model with robust standard errors was executed (with reference groups dropped in the case of maternal education, maternal marital status, and child race/ethnicity). Note that no weights were used in generating these estimates. Also note that there were very few mothers with missing relationship status at baseline, hence the means of zero.

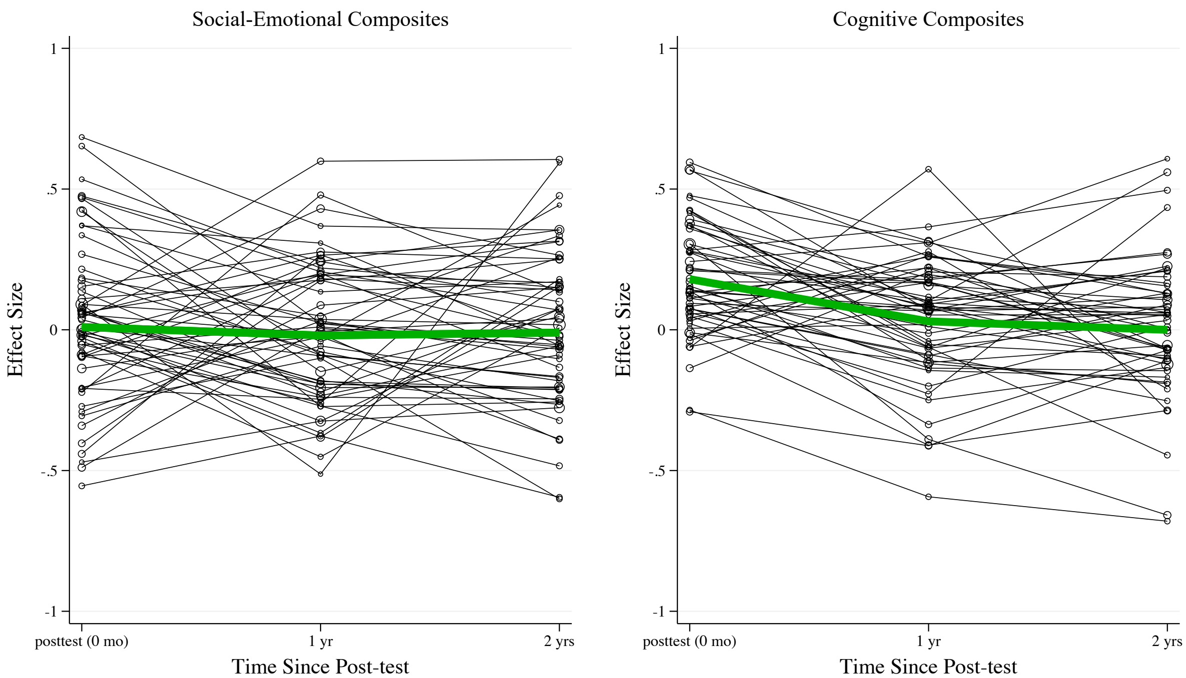
Table 2
Grantee and Child Level Sample Sizes

	Total		Treatment	Control
	n	*k*	Avg. n in k	Avg. n in k
Randomized Sample	4442	83	32	22
Analytic Sample				
Soc Posttest	3820	80	30	18
Soc 1-year Follow-up	3717	81	28	18
Soc 2-year Follow-up	3721	81	28	18
Cog Posttest	3783	80	29	18
Cog 1-year Follow-up	3668	80	28	18
Cog 2-year Follow-up	3580	79	28	17

Notes: “Soc” = social-emotional, “Cog” = cognitive. This table presents the child (*n*) and grantee (*k*) sample sizes in the original randomized sample and in the analytic sample of grantee-level effects.

Figure 1

Trajectory of Grantee-level Intervention Impacts for Social-emotional and Cognitive Measures

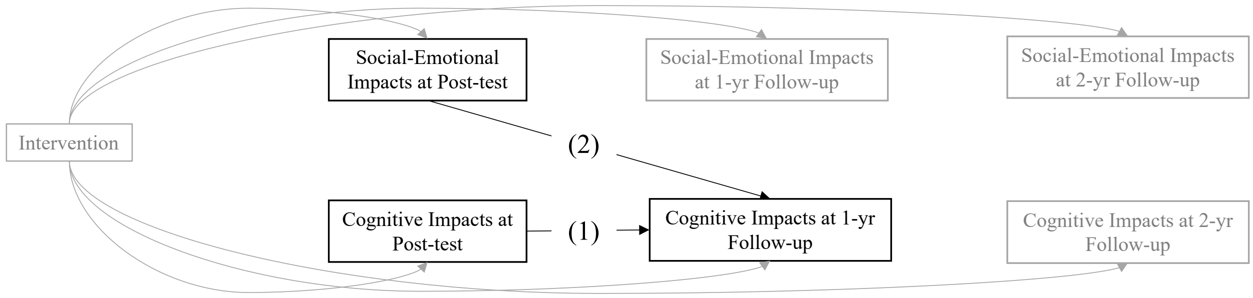


Notes: Each line indicates a grantee-specific trajectory of intervention impacts on the cognitive or social-emotional composite across posttest, 1-year follow-up, and 2-year follow-up. For each grantee, coordinates were weighted by the posttest inverse variances. Grantee-level intervention impacts were computed using a range of baseline child and family characteristics. Green lines indicate the meta-analytic average of grantee-level impacts, computed at each assessment wave in R using the metafor package. Models contained inverse variance weighting and a random intercept for grantee. Delegates with fewer than 10 participants were removed from this figure.

Table 3
Primary Linear Regression Models

Model	
1	$TX\_Impact_{cog\_d\_1yr} = \beta_{0d} + \beta_1 TX\_Impact_{cog\_d\_post} + \beta_2 TX\_Impact_{soc\_d\_post} + \varepsilon_{dt}$
2	$TX\_Impact_{cog\_d\_2yr} = \beta_{0d} + \beta_1 TX\_Impact_{cog\_d\_post} + \beta_2 TX\_Impact_{soc\_d\_post} + \varepsilon_{dt}$
3	$TX\_Impact_{cog\_d\_2yr} = \beta_{0d} + \beta_1 TX\_Impact_{cog\_d\_1yr} + \beta_2 TX\_Impact_{soc\_d\_1yr} + \varepsilon_{dt}$
4	$TX\_Impact_{soc\_d\_1yr} = \beta_{0d} + \beta_1 TX\_Impact_{soc\_d\_post} + \beta_2 TX\_Impact_{cog\_d\_post} + \varepsilon_{dt}$
5	$TX\_Impact_{soc\_d\_2yr} = \beta_{0d} + \beta_1 TX\_Impact_{soc\_d\_post} + \beta_2 TX\_Impact_{cog\_d\_post} + \varepsilon_{dt}$
6	$TX\_Impact_{soc\_d\_2yr} = \beta_{0d} + \beta_1 TX\_Impact_{soc\_d\_1yr} + \beta_2 TX\_Impact_{cog\_d\_1yr} + \varepsilon_{dt}$

Figure 2
Example Model



Notes: This model visualizes Model 1 in Table 3 which aimed to estimate auto-regressive associations among cognitive posttest impacts and cognitive 1-year follow-up impacts and cross-domain impacts between social-emotional posttest impacts and 1-year cognitive follow-up impacts.

Table 4
Correlation Matrix: Grantee-level Treatment Impacts

	Soc Post	Soc 1-year	Soc 2-year	Cog Post	Cog 1-year	Cog 2-year
Soc Post	1.00					
Soc 1-year	0.26*	1.00				
Soc 2-year	0.19+	-0.03	1.00			
Cog Post	-0.09	-0.15	0.06	1.00		
Cog 1-year	-0.09	0.22*	0.07	0.45***	1.00	
Cog 2-year	0.19	0.07	0.14	0.38***	0.67***	1.00

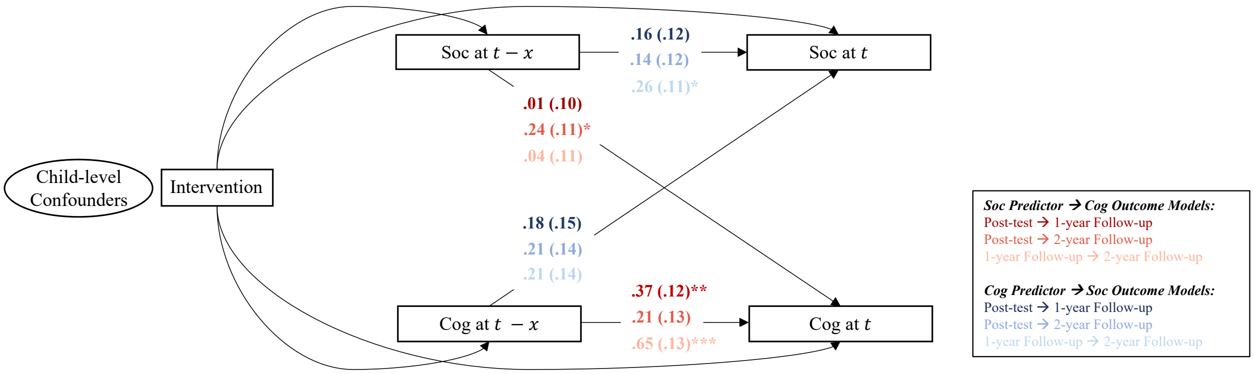
Notes: "Soc" = Social-emotional, "Cog" = Cognitive, "Post" = end of treatment impacts, "1-year" = 1-year follow-up impacts, "2-year" = 2-year follow-up impacts. This matrix reports the correlations between grantee-level treatment impacts.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 5
Auto-regressive and Cross-domain Dynamics in Grantee-Level Effects

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model (1)	Cross-Domain Model (2)	Same-Domain Model (3)	Cross-Domain Model (4)
Panel A: Posttest Predicting 1-year Follow-up Effects				
Intercept	-0.03 (0.03)	-0.06 (0.04)	-0.03 (0.03)	-0.03 (0.03)
Soc Posttest	0.20 (0.12)	0.16 (0.12)		0.01 (0.10)
Cog Posttest		0.18 (0.15)	0.38 (0.12)***	0.37 (0.12)***
$\tau_{intercept}$	0.04	0.02	0.00	0.00
Observations	80	80	80	80
Panel B: Posttest Predicting 2-year Follow-up Effects				
Intercept	-0.01 (0.03)	-0.05 (0.04)	-0.05 (0.04)	-0.04 (0.04)
Soc Posttest	0.17 (0.12)	0.14 (0.12)		0.24 (0.11)*
Cog Posttest		0.21 (0.14)	0.27 (0.13)*	0.21 (0.13)
$\tau_{intercept}$	0.02	0.00	0.00	0.00
Observations	80	80	79	79
Panel C: 1-year Follow-up Effects Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.03)	-0.01 (0.03)	-0.02 (0.03)	-0.02 (0.03)
Soc 1-year Follow-up	0.31 (0.11)**	0.26 (0.11)*		0.04 (0.11)
Cog 1-year Follow-up		0.21 (0.14)	0.67 (0.13)***	0.65 (0.13)***
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	81	80	79	79

Notes: "Soc" = Social-emotional. "Cog" = Cognitive. Treatment impacts were estimated at the grantee level with controls for baseline child and family characteristics. Same-domain and cross-domain transfer estimates were computed through a series of independent meta-regression models (12 presented in this table) in which grantee-level impacts from a later assessment wave (1- or 2-year follow-up) were regressed on grantee-level impacts from an earlier assessment wave (1-year follow-up or posttest). Models included weighting by the inverse variances of follow-up effects and a random intercept for grantee. Controls used for computing the grantee-level effects included: center-level fixed effects, pre-test cognitive and social-emotional composites, cohort, age at concurrent testing (e.g., in computing posttest impacts, we controlled for age at posttest), child gender, child race, child primary language, primary caregiver's age, primary language spoken at home, whether both biological parents live with the child, whether the child's mother is a recent immigrant, mother's educational attainment, mother's marital status, and whether the mother gave birth as a teenager.
* $p < .05$, ** $p < .01$, *** $p < .001$

Figure 3
Auto-regressive and Cross-domain Dynamics among Grantee-level Impacts



Notes: Model-based estimates come from Table 5.

Supplemental file for:

Using experimental variation to examine the (co-)development of cognitive and social-emotional skills in early childhood

Sensitivity Analyses

The exogeneity assumption is particularly tenuous in low-sample-size grantees. Fortunately, our primary meta-analytic approach inherently addressed this concern by down-weighting less-precise effect sizes from smaller samples via inverse-variance weighting. Additionally, we ran models in which we dropped grantees with treatment and/or control sample sizes smaller than 10 participants, and found substantive patterns that were generally similar to those observed from our primary models (see Table S3). Of note, the auto-regressive paths for social-emotional impacts were generally smaller, and the cross-lagged paths for cognitive impacts predicting social-emotional impacts were generally larger.

Importantly, there were some grantees that failed the joint test for baseline balance and did *not* have abnormally small sample sizes. To further address imbalance in baseline characteristics, we controlled for baseline child and family characteristics, including child social-emotional and cognitive functioning, when we computed our primary grantee-level treatment impacts. As expected, the inclusion of these controls attenuated the autoregressive associations among social-emotional and cognitive skills (see Table S4 for the models without controls). For cross-lagged associations, the inclusion of controls in some cases minimally affected the estimates and in other cases reduced the magnitude of the estimates.

As a further test, we ran models in which we substituted grantee-level posttest impacts for “impacts” estimated at pre-test. If it were the case that the grantee-level treatment impacts reflected nothing more than endogenous child-level differences in cognitive and social-emotional skills, then pre-test differences among the treatment and control groups should be just as predictive of follow-up effects as posttest differences. However, in support of the expectation that the quasi-experimental estimates captured exogenous differences due to the treatment itself, overall, the models using pre-test “impacts” produced auto-regressive and cross-lagged paths that did not resemble estimates from models using posttest impacts (see Table S5).

Given that the grantee-level pre-test “impacts” were not all zero, we then ran our primary cross-lagged models with additional controls for social-emotional and cognitive pre-test “impacts.” Building from our primary model that used grantee-level effects, estimated with controls for baseline characteristics, we entered both cognitive and social-emotional pre-test “impacts” to test for additional bias due to treatment-control-group differences in skills observed prior to randomization (see Table S6). The auto-regressive and cross-lagged estimates were minimally affected with the inclusion of the pre-test “impacts,” indicating that above and beyond controls for child-level baseline characteristics, treatment-control *differences* in baseline social-emotional and cognitive functioning did not appear to bias our estimates.

Next, we examined extent to which issues of missing data biased our estimates (see Table S7). In our primary grantee-level quasi-experimental impacts, about 80% to 85% of participants contributed data across assessment waves and outcome types. To probe how changes in the contributing sample affected estimates, we ran models in which we limited the sample to participants who contributed cognitive and social-emotional data at each assessment wave (n = 3,262 children from 79 grantees). The auto-regressive and social-emotional to cognitive transfer paths were minimally affected, but the cognitive to social-emotional transfer paths were significantly attenuated in two of the three models.

Finally, we tested the consistency of our results when using a common-effects model, a reasonable alternative to our primary random-effects model. The estimates from this model were very similar to that of our primary model (Table S8).

Table S1
Meta-analytic Averages for Cognitive and Social-Emotional Outcomes (B(SE))

	Posttest (1)	1-year Follow-up (2)	2-year Follow-up (3)
Social-Emotional Composite			
Meta-analytic Average	0.01 (0.03)	-0.02 (0.03)	-0.01 (0.03)
$\tau_{intercept}$	0.11	0.00	0.00
Q	Q(79) = 89.73 *p* = .19	Q(80) = 80.84 *p* = .45	Q(80) = 74.86 *p* = .64
Observations	80	81	81
Cognitive Composite			
Meta-analytic Average	0.18 (0.03)***	0.03 (0.03)	0.00 (0.03)
$\tau_{intercept}$	0.12	0.07	0.03
Q	Q(79) = 108.79* *p* = .01	Q(79) = 83.19 *p* = .35	Q(78) = 90.78 *p* = .15
Observations	80	80	79

Notes: This table presents the meta-analytic average of grantee-level treatment impacts on the social-emotional and cognitive composites. Grantee-level intervention impacts were computed using a range of baseline child and family characteristics. Meta-analytic averages were computed at each assessment wave for the social-emotional and cognitive composites, respectively. Models included weighting by the inverse variances of follow-up effects and a random intercept for grantee.

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

Table S2

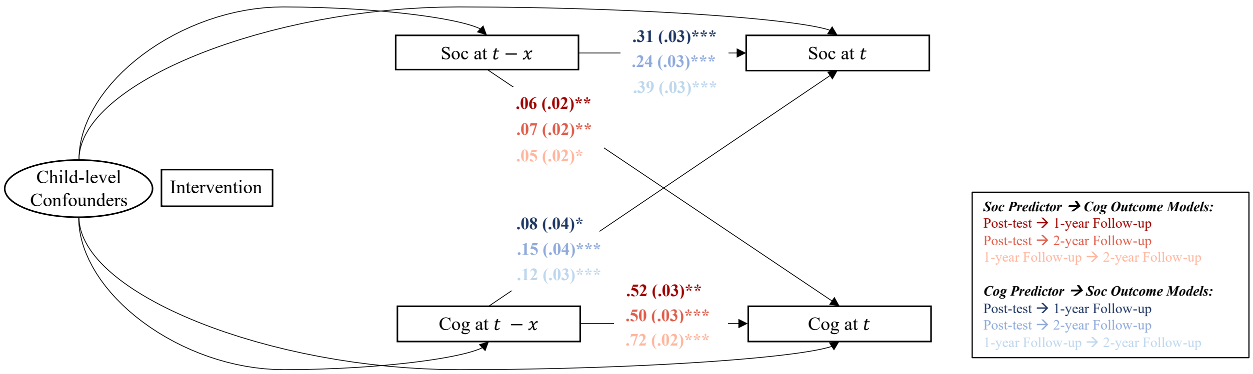
Auto-regressive and Cross-domain Associations in Child-level Observational Data

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model (1)	Cross-Domain Model (2)	Same-Domain Model (3)	Cross-Domain Model (4)
Panel A: Posttest Predicting 1-year Follow-up Effects				
Intercept	-0.06 (0.04)	-0.06 (0.05)	-0.05 (0.04)	-0.05 (0.04)
Soc Posttest	0.32 (0.03)***	0.31 (0.03)***		0.06 (0.02)**
Cog Posttest		0.08 (0.04)*	0.53 (0.03)***	0.52 (0.03)***
Observations	1329	1322	1308	1305
Panel B: Posttest Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.04)	0.03 (0.04)	0.04 (0.05)	0.03 (0.05)
Soc Posttest	0.25 (0.03)***	0.24 (0.03)***		0.07 (0.02)**
Cog Posttest		0.15 (0.04)***	0.52 (0.03)***	0.50 (0.03)***
Observations	1317	1310	1268	1268
Panel C: 1-year Follow-up Effects Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.03)	0.02 (0.04)	-0.07 (0.03)*	-0.08 (0.03)*
Soc 1-year Follow-up	0.41 (0.02)***	0.39 (0.03)***		0.05 (0.02)*
Cog 1-year Follow-up		0.12 (0.03)***	0.74 (0.02)***	0.72 (0.02)***
Observations	1352	1326	1287	1282

Notes: "Soc"= Social-emotional. "Cog"= Cognitive. Only children from the control group were included in these analyses. The analyses relied on child level data, not estimated treatment impacts. Same-domain and cross-domain transfer estimates were computed through a series of independent regression models (12 presented in this table) in which child-level functioning at a later assessment wave (1- or 2-year follow-up) was regressed on child level functioning at an earlier assessment wave (1-year follow up or posttest). Models included the following controls: pre-test cognitive and social-emotional composites, cohort, age at concurrent testing (for both the predictor and outcome), child gender, child race, child primary language, primary caregiver's age, primary language spoken at home, whether both biological parents live with the child, whether the child's mother is a recent immigrant, mother's educational attainment, mother's marital status, and whether the mother gave birth as a teenager.

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

Figure S1
Auto-regressive and Cross-domain Associations using Observational Data from the Control Group



Notes: Estimates are presented in Table S2.

Table S3
Auto-regressive and Cross-domain Dynamics in Grantee-level Effects with Small Grantees Dropped

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model	Cross-Domain Model	Same-Domain Model	Cross-Domain Model
	(1)	(2)	(3)	(4)
Panel A: Posttest Predicting 1-year Follow-up Effects				
Intercept	-0.02 (0.03)	-0.07 (0.05)	-0.03 (0.04)	-0.03 (0.04)
Soc Posttest	0.09 (0.13)	0.02 (0.14)		-0.04 (0.11)
Cog Posttest		0.29 (0.17)	0.37 (0.13)**	0.38 (0.14)**
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	57	57	57	57
Panel B: Posttest Predicting 2-year Follow-up Effects				
Intercept	-0.01 (0.03)	-0.07 (0.05)	-0.05 (0.04)	-0.04 (0.04)
Soc Posttest	0.12 (0.13)	0.04 (0.13)		0.15 (0.13)
Cog Posttest		0.33 (0.17)	0.33 (0.15)*	0.26 (0.16)
$\tau_{intercept}$	0.03	0.00	0.00	0.00
Observations	58	58	55	55
Panel C: 1-year Follow-up Effects Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.03)	-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
Soc 1-year Follow-up	0.53 (0.14)***	0.46 (0.14)***		0.10 (0.14)
Cog 1-year Follow-up		0.31 (0.16)	0.54 (0.15)***	0.51 (0.15)***
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	57	57	55	55

Notes: Treatment impacts were estimated at the grantee level with controls for baseline child and family characteristics. Grantees with fewer than 10 children in the control or treatment group were not included in these models. Same-domain and cross-domain transfer estimates were computed through a series of independent meta-regression models (12 presented in this table) in which grantee-level impacts from a later assessment wave (1- or 2-year follow-up) were regressed on grantee-level impacts from an earlier assessment wave (1-year follow-up or posttest). Models included weighting by the inverse variances of follow-up effects and a random intercept for grantee. Controls used for computing the grantee-level effects included: center-level fixed effects, pre-test cognitive and social-emotional composites, cohort, age at concurrent testing, child gender, child race, child primary language, primary caregiver's age, primary language spoken at home, whether both biological parents live with the child, whether the child's mother is a recent immigrant, mother's educational attainment, mother's marital status, and whether the mother gave birth as a teenager.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table S4
Auto-regressive and Cross-domain Dynamics in Grantee-level Effects with Minimal Controls

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model	Cross-Domain Model	Same-Domain Model	Cross-Domain Model
	(1)	(2)	(3)	(4)
Panel A: Posttest Predicting 1-year Follow-up Effects				
Intercept	-0.02 (0.04)	-0.03 (0.04)	-0.06 (0.04)	-0.05 (0.04)
Soc Posttest	0.30 (0.12)*	0.28 (0.12)*		0.08 (0.11)
Cog Posttest		0.07 (0.11)	0.51 (0.10)***	0.49 (0.10)***
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	80	80	80	80
Panel B: Posttest Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.03)	-0.02 (0.04)	-0.07 (0.04)	-0.07 (0.04)
Soc Posttest	0.29 (0.11)*	0.26 (0.12)*		0.22 (0.12)
Cog Posttest		0.11 (0.11)	0.39 (0.10)***	0.34 (0.11)**
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	80	80	79	79
Panel C: 1-year Follow-up Effects Predicting 2-year Follow-up Effects				
Intercept	0.01 (0.03)	0.00 (0.03)	-0.03 (0.03)	-0.03 (0.03)
Soc 1-year Follow-up	0.40 (0.11)***	0.33 (0.12)**		0.01 (0.12)
Cog 1-year Follow-up		0.24 (0.13)	0.73 (0.12)***	0.73 (0.12)***
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	81	80	79	79

Notes: Treatment impacts were estimated at the grantee level with a reduced set of controls (cohort, child age at concurrent testing, center-level fixed effects). Same-domain and cross-domain transfer estimates were computed through a series of independent meta-regression models (12 presented in this table) in which grantee-level impacts from a later assessment wave (1- or 2-year follow-up) were regressed on grantee-level impacts from an earlier assessment wave (1-year follow-up or posttest). Models included weighting by the inverse variances of follow-up effects and a random intercept for grantee.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table S5
Auto-regressive and Cross-domain Dynamics in Grantee-Level Effects using Pre-test Effects

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model	Cross-Domain Model	Same-Domain Model	Cross-Domain Model
	(1)	(2)	(3)	(4)
Panel A: Pre-test Effects Predicting 1-year Follow-up Effects				
Intercept	-0.02 (0.03)	-0.02 (0.03)	0.04 (0.03)	0.04 (0.03)
Soc Pre-test	0.03 (0.13)	0.08 (0.13)		0.14 (0.11)
Cog Pre-test		-0.21 (0.12)	-0.15 (0.09)	-0.18 (0.09)
$\tau_{intercept}$	0.02	0.00	0.05	0.04
Observations	81	81	80	80
Panel B: Pre-test Effects Predicting 2-year Follow-up Effects				
Intercept	-0.01 (0.03)	-0.01 (0.03)	0.01 (0.03)	0.01 (0.03)
Soc Pre-test	-0.11 (0.13)	-0.10 (0.13)		0.03 (0.12)
Cog Pre-test		-0.04 (0.11)	-0.10 (0.10)	-0.11 (0.11)
$\tau_{intercept}$	0.02	0.03	0.03	0.04
Observations	81	81	79	79

Notes: Treatment impacts were estimated at the grantee level with controls for baseline child and family characteristics. Same-domain and cross-domain transfer estimates were computed through a series of independent meta-regression models (12 presented in this table) in which grantee-level impacts from a later assessment wave (1- or 2-year follow-up) were regressed on grantee-level impacts at pre-test. Models included weighting by the inverse variances of follow-up effects and a random intercept for grantee. Controls used for computing the grantee-level effects included: center-level fixed effects, pre-test cognitive and social-emotional composites (in the case of 1- and 2-year follow-up effects), cohort, age at concurrent testing, child gender, child race, child primary language, primary caregiver's age, primary language spoken at home, whether both biological parents live with the child, whether the child's mother is a recent immigrant, mother's educational attainment, mother's marital status, and whether the mother gave birth as a teenager.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table S6
Auto-regressive and Cross-domain Dynamics in Grantee-level Effects controlling for Grantee-level Pre-test "Impacts"

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model	Cross-Domain Model	Same-Domain Model	Cross-Domain Model
	(1)	(2)	(3)	(4)
Panel A: Posttest Predicting 1-year Follow-up Effects				
Intercept	-0.02 (0.03)	-0.06 (0.04)	-0.04 (0.03)	-0.04 (0.03)
Soc Posttest	0.20 (0.12)	0.16 (0.12)		0.00 (0.10)
Cog Posttest		0.23 (0.15)	0.43 (0.12)***	0.43 (0.12)***
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	80	80	80	80
Panel B: Posttest Predicting 2-year Follow-up Effects				
Intercept	-0.01 (0.03)	-0.05 (0.04)	-0.05 (0.04)	-0.04 (0.04)
Soc Posttest	0.19 (0.12)	0.15 (0.12)		0.24 (0.11)*
Cog Posttest		0.23 (0.15)	0.30 (0.13)*	0.25 (0.14)
$\tau_{intercept}$	0.03	0.00	0.00	0.00
Observations	80	80	79	79
Panel C: 1-year Follow-up Effects Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.03)	-0.01 (0.03)	-0.02 (0.03)	-0.02 (0.03)
Soc 1-year Follow-up	0.32 (0.11)**	0.27 (0.11)*		0.04 (0.11)
Cog 1-year Follow-up		0.24 (0.14)	0.68 (0.13)***	0.66 (0.13)***
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	81	80	79	79

Notes: Treatment impacts were estimated at the grantee level with controls for baseline child and family characteristics (listed below). In these models, only data from participants that contributed data at each assessment wave on both cognitive and social-emotional composites were included when computing grantee-level treatment impacts. Same-domain and cross-domain transfer estimates were computed through a series of independent meta-regression models (12 presented in this table) in which grantee-level impacts from a later assessment wave (1- or 2-year follow-up) were regressed on grantee-level impacts from an earlier assessment wave (1-year follow-up or posttest) as well as pre-test "impacts" computed at the grantee level using the aforementioned approach. Models included weighting by the inverse variances of follow-up effects and a random intercept for grantee. Controls used for computing the grantee-level effects included: center-level fixed effects, pre-test cognitive and social-emotional composites (in the case of posttest, 1- and 2-year follow-up effects), cohort, age at concurrent testing, child gender, child race, child primary language, primary caregiver's age, primary language spoken at home, whether both biological parents live with the child, whether the child's mother is a recent immigrant, mother's educational attainment, mother's marital status, and whether the mother gave birth as a teenager.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table S7
Auto-regressive and Cross-domain Dynamics in Grantee-level Effects with Consistent Sample Across Waves

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model	Cross-Domain Model	Same-Domain Model	Cross-Domain Model
	(1)	(2)	(3)	(4)
Panel A: Posttest Predicting 1-year Follow-up Effects				
Intercept	-0.01 (0.03)	-0.01 (0.05)	-0.04 (0.04)	-0.03 (0.04)
Soc Posttest	0.17 (0.12)	0.17 (0.13)		0.06 (0.10)
Cog Posttest		0.00 (0.16)	0.41 (0.12)***	0.39 (0.13)**
$\tau_{intercept}$	0.00	0.00	0.00	0.01
Observations	79	79	79	79
Panel B: Posttest Predicting 2-year Follow-up Effects				
Intercept	-0.01 (0.03)	-0.02 (0.05)	-0.07 (0.04)	-0.06 (0.04)
Soc Posttest	0.18 (0.12)	0.18 (0.12)		0.19 (0.11)
Cog Posttest		0.04 (0.16)	0.40 (0.14)**	0.35 (0.14)*
$\tau_{intercept}$	0.00	0.00	0.05	0.06
Observations	79	79	79	79
Panel C: 1-year Follow-up Effects Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.03)	-0.02 (0.03)	-0.02 (0.03)	-0.02 (0.03)
Soc 1-year Follow-up	0.36 (0.11)**	0.31 (0.12)**		0.01 (0.10)
Cog 1-year Follow-up		0.25 (0.14)	0.76 (0.12)***	0.76 (0.13)***
$\tau_{intercept}$	0.00	0.00	0.00	0.00
Observations	79	79	79	79

Notes: Treatment impacts were estimated at the grantee level with controls for baseline child and family characteristics. In these models, only data from participants that contributed data at each assessment wave on both cognitive and social-emotional composites were included when computing grantee-level treatment impacts. Same-domain and cross-domain transfer estimates were computed through a series of independent meta-regression models (12 presented in this table) in which grantee-level impacts from a later assessment wave (1- or 2-year follow-up) were regressed on grantee-level impacts from an earlier assessment wave (1-year follow-up or posttest). Models included weighting by the inverse variances of follow-up effects and a random intercept for grantee. Controls used for computing the grantee-level effects included: center-level fixed effects, pre-test cognitive and social-emotional composites, cohort, age at concurrent testing, child gender, child race, child primary language, primary caregiver's age, primary language spoken at home, whether both biological parents live with the child, whether the child's mother is a recent immigrant, mother's educational attainment, mother's marital status, and whether the mother gave birth as a teenager.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table S8
Auto-regressive and Cross-domain Dynamics in Grantee-level Effects using a Common-Effects Approach (no Random Effect)

	Outcome: Social-Emotional Composite		Outcome: Cognitive Composite	
	Same-Domain Model (1)	Cross-Domain Model (2)	Same-Domain Model (3)	Cross-Domain Model (4)
Panel A: Posttest Predicting 1-year Follow-up Effects				
Intercept	-0.03 (0.03)	-0.06 (0.04)	-0.03 (0.03)	-0.03 (0.03)
Soc Posttest	0.19 (0.12)	0.16 (0.12)		0.01 (0.10)
Cog Posttest		0.18 (0.15)	0.38 (0.12)***	0.37 (0.12)***
Observations	80	80	80	80
Panel B: Posttest Predicting 2-year Follow-up Effects				
Intercept	-0.01 (0.03)	-0.05 (0.04)	-0.05 (0.04)	-0.04 (0.04)
Soc Posttest	0.17 (0.12)	0.14 (0.12)		0.24 (0.11)*
Cog Posttest		0.21 (0.14)	0.27 (0.13)*	0.21 (0.13)
Observations	80	80	79	79
Panel C: 1-year Follow-up Effects Predicting 2-year Follow-up Effects				
Intercept	0.00 (0.03)	-0.01 (0.03)	-0.02 (0.03)	-0.02 (0.03)
Soc 1-year Follow-up	0.31 (0.11)**	0.26 (0.11)*		0.04 (0.11)
Cog 1-year Follow-up		0.21 (0.14)	0.67 (0.13)***	0.65 (0.13)***
Observations	81	80	79	79

Notes: Treatment impacts were estimated at the grantee level with controls for baseline child and family characteristics (listed below). In these models, only data from participants that contributed data at each assessment wave on both cognitive and social-emotional composites were included when computing grantee-level treatment impacts. Same-domain and cross-domain transfer estimates were computed through a series of independent meta-regression models (12 presented in this table) in which grantee-level impacts from a later assessment wave (1- or 2-year follow-up) were regressed on grantee-level impacts from an earlier assessment wave (1-year follow-up or posttest) as well as pre-test "impacts" computed at the grantee level using the aforementioned approach. Models included weighting by the inverse variances of the follow-up effects. Controls used for computing the grantee-level effects included: center-level fixed effects, pre-test cognitive and social-emotional composites (in the case of posttest, 1- and 2-year follow-up effects), cohort, age at concurrent testing, child gender, child race, child primary language, primary caregiver's age, primary language spoken at home, whether both biological parents live with the child, whether the child's mother is a recent immigrant, mother's educational attainment, mother's marital status, and whether the mother gave birth as a teenager.
* $p < .05$, ** $p < .01$, *** $p < .001$