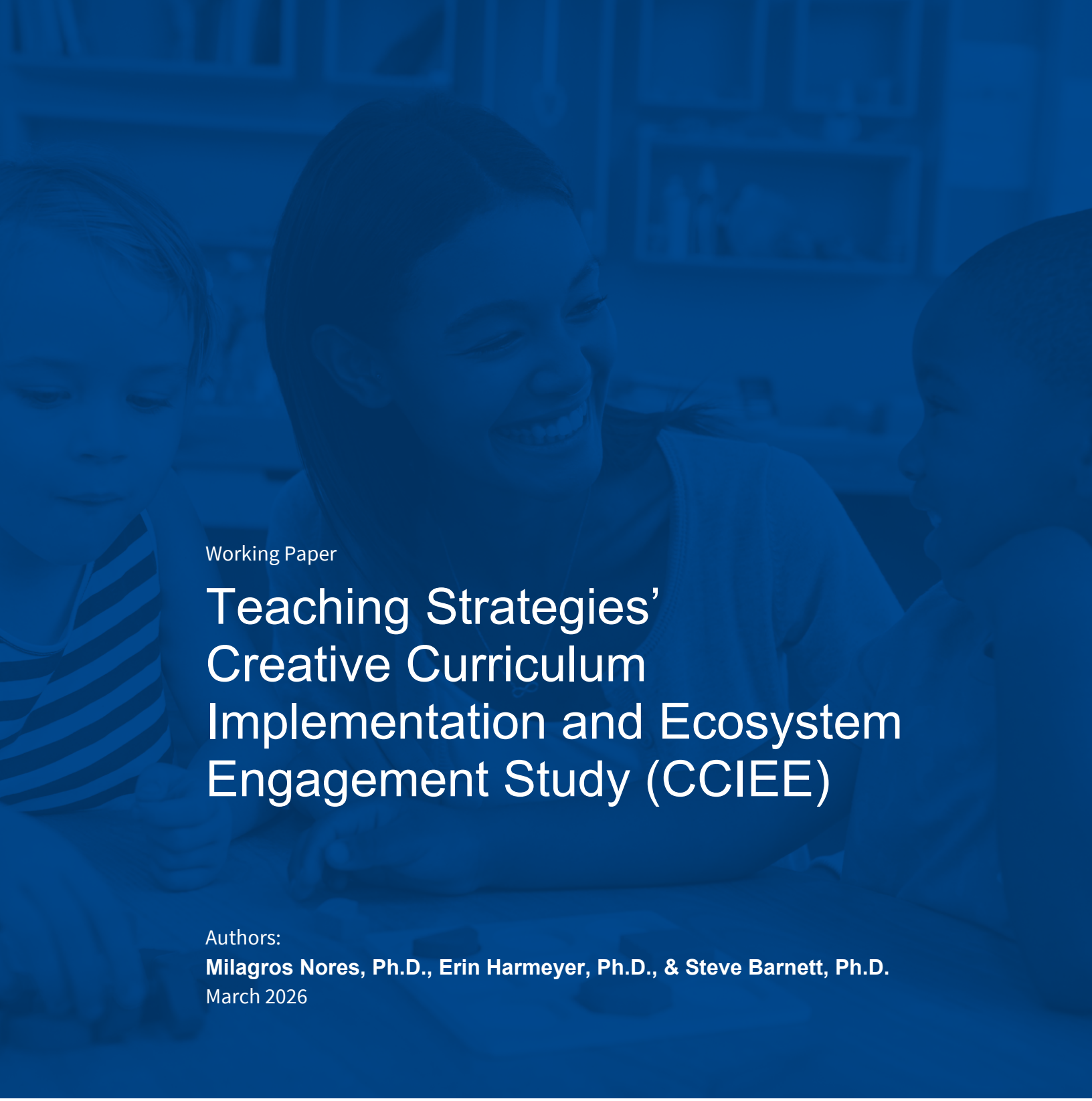


Working Paper

Teaching Strategies' Creative Curriculum Implementation and Ecosystem Engagement Study (CCIEE)

Authors:
Milagros Nores, Ph.D., Erin Harmeyer, Ph.D., & Steve Barnett, Ph.D.
March 2026

NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

Acknowledgments

Grateful acknowledgment is made to the districts, schools, and classrooms that participated in this study. In addition, the authors would like to thank Julie Macleod for her invaluable project coordination of data collection throughout the three years of the grant, as well as Facundo Luna Mallea for his invaluable research analysis support for this report. This work would not be possible without the partnership with Teaching Strategies, which provided a grant to support the work and worked closely with the research team to track all intervention and sample aspects relevant to the research.

About NIEER

The National Institute for Early Education Research (NIEER) at the Graduate School of Education, Rutgers University, New Brunswick, NJ, conducts and disseminates independent research and analysis to inform early childhood education policy.

About the Authors

Milagros Nores, Ph.D. Dr. Nores is a Co-Director for Research at The National Institute for Early Education Research (NIEER) at Rutgers University. Dr. Nores conducts research at NIEER on issues related to early childhood policy, programs, and evaluation, both nationally and globally.

Erin Harmeyer, Ph.D. Dr. Harmeyer is an Assistant Research Professor at the National Institute for Early Education Research (NIEER) at Rutgers University. Dr. Harmeyer conducts research at NIEER related to early childhood programs and evaluation, supporting projects in New Jersey, Rhode Island, and Pennsylvania.

W. Steve Barnett, Ph.D. Dr. Barnett is Founding Director of the National Institute for Early Education Research (NIEER) and Board of Governors Professor at Rutgers University. Dr. Barnett conducts research and evaluations on early childhood policy and programs, nationally as well as globally.

Correspondence regarding this report should be addressed to Milagros Nores at the National Institute for Early Education Research. Email: mnores@nieer.org.

Suggested Citation

Nores, M., E. Harmeyer & W.S. Barnett (2026). *Teaching Strategies' Creative Curriculum Implementation and Ecosystem Engagement Study (CCIEE).* Research Report. New Brunswick, NJ: National Institute for Early Education Research.

Abstract

This report presents findings from the *Creative Curriculum Implementation and Ecosystem Engagement Study* (CCIEE), a cluster randomized controlled trial conducted in New Jersey. The study assessed the impact of an intervention designed to strengthen curriculum implementation, effective teaching practices, and teachers' engagement with Teaching Strategies® fully digital ecosystem via additional professional development opportunities and aligned supports. This report summarizes the impact on teacher outcomes, classroom quality, curriculum fidelity, and child outcomes. Conducted across two low-income districts and including a "synthetic" comparison group, the study found that the intervention significantly improved teacher retention but had null or even negative effects on classroom quality and curriculum fidelity as measured in this study. Both control and treatment teachers improved across these measures, with greater improvement in the control group. There were no direct treatment (ITT) effects on children's externally assessed outcomes. However, findings suggest that teacher retention—significantly improved by the intervention—was significantly associated with positive peer play skills in the treatment group and marginally associated with gains in children's executive function in both groups. In contrast, there were positive treatment effects on Teaching Strategies *GOLD*® measures with statistically significant gains in language and mathematics and marginally significant gains in social-emotional development, and in *GOLD* literacy and cognitive measures for children in classrooms with retained teachers. There were also positive, statistically significant effects of treatment "dosage" on the *GOLD* measures for all domains and marginal dosage effects on the externally assessed literacy and math measures. In addition, greater *SmartTeach*™ platform engagement in the treatment group was associated with gains on the cognitive, literacy, and mathematics *GOLD* domains. The difference in outcomes could be explained by a combination of floor effects for some of the external assessment measures and the breadth of the constructs measured by the curriculum-aligned *GOLD* assessment. Teachers reported generally positive experiences with *The Creative Curriculum*®, with few differences emerging by treatment status. However, at the end of the study, teachers in the treatment group reported statistically significantly higher levels of personal accomplishment and marginally lower levels of emotional exhaustion than control group teachers. Survey findings also indicated that teachers found *The Creative Curriculum* easy to implement, engaging, and that it helped them teach academic skills. Overall, the findings highlight the importance of staffing stability and alignment between professional development, curriculum, and child measurement approaches in understanding how digital tools relate to classroom practice and improved child development.

Key Highlights

- As part of the intervention, teachers in the treatment group received on average 24 hours of virtual, synchronous professional development (PD) during the 3-year study period.
- Teachers in the control and treatment groups both had access to asynchronous PD (on-demand courses) and engaged with the *SmartTeach* online platform, with teachers in the treatment group averaging about 5 more hours of asynchronous PD and 94 more hours of platform engagement across the three years of the study.
- Both treatment and control groups showed improvements in classroom quality throughout the study period. There were no intervention effects observed on classroom quality nor curriculum fidelity, with results favoring the control group.
- Participation in the intervention significantly increased teacher retention by 23 percentage points.
- We found mostly null (and one marginally negative) treatment effects in children's externally assessed outcomes; however, treatment dosage was associated with statistically significant gains in the literacy measure and marginally significant gains in the math measure.
- We found positive treatment effects in children's outcomes measured by Teaching Strategies' *GOLD* with marginally significant gains in social-emotional development, and statistically significant gains in language and math.
- Retention emerged as important for executive function, peer play skills, and *GOLD* cognitive and literacy outcomes.
- Higher virtual, synchronous PD dosage was associated with significant gains across all *GOLD* domains, with statistically significant positive effects on cognitive, language, and literacy outcomes and marginally significant positive effects on social-emotional and math outcomes.
- Similarly, greater engagement with the digital platform in the treatment group was associated with statistically significant gains in the cognitive, literacy, and mathematics *GOLD* domains.
- Differences in outcomes between *GOLD* and the other measures may be explained by a combination of floor effects for some external measures, greater alignment between the curriculum and *GOLD*, and the breadth of the constructs measured by *GOLD*.
- On the external measures, children in the treatment and control group evidenced gains similar to children in a non-equivalent comparison group of 21 districts implementing the state preschool program across New Jersey, despite starting the school year lagging developmentally relative to peers in these districts.
- At follow-up, treatment group teachers reported higher levels of personal accomplishment and energy, plus lower levels of emotional exhaustion compared to control group teachers.
- Survey findings also show that teachers found *The Creative Curriculum* easy to implement, engaging, and helped teach academic skills.
- COVID-19 disruptions early into the study, including changes in district policies, (including changes in the rules for using the digital curriculum and assessment tools) and patterns in enrollment, constrained implementation and likely attenuated the impact of the ecosystem supports, likely making the study a conservative test of the intervention.
- The study highlights the importance of strong, aligned professional development and staffing continuity in achieving intended gains in classroom quality and child development through integrated digital curriculum within a connected early childhood ecosystem.

Introduction

Early childhood education (ECE) programs are widely recognized for fostering children's early learning and development, with long-term benefits extending into their academic, social, and economic trajectories (NASEM, 2023). Central to realizing long-term outcomes is achieving and sustaining high-quality implementation over time and at scale (Barnett & Frede, 2017; Nores et al., 2018). Although there is no universal agreement on what defines "quality," there is broad consensus that a well-designed curriculum that supports comprehensive child development across all domains is a key component (Barker et al., 2014; Goble & Pianta, 2017; Golinkoff & Hirsch-Pasek, 2006; NASEM, 2024; Pyle, Poliszczuk et al., 2018; Von Suchodoletz et al., 2023; Yoshikawa et al., 2013). Moreover, there is emerging consensus that a developmentally appropriate assessment system that supports individualized learning in culturally and linguistically responsive ways is equally essential, especially when well-aligned with curriculum (NAEYC, 2020). In many jurisdictions, state or local authorities designate approved curriculum and assessments (Friedman-Krauss, 2025); however, in the absence of such mandates, districts and providers are left to make their own independent choices. Unfortunately, research has offered limited practical guidance due to challenges in conducting rigorous curriculum evaluations, the inconsistent nature of past findings, and the rapid and continuous evolution of curricular models and their increasing integration of digital components. Further complicating the landscape is the substantial variability in contexts and quality among early childhood programs using the same curricula, raising concerns about consistency and equity (NASEM, 2024; Rege et al., 2024), as well as the limited evaluation base that exists on curriculum for specific subgroups of children such as dual language learners, children with disabilities, and children from historically marginalized communities (NASEM, 2024). In addition, there is little information on how specific curricula intersect with family engagement strategies and how they work in mixed-delivery settings (Engel et al., 2025).

Despite these challenges, evidence suggests that some curricular approaches can positively influence children's cognitive development, including language, literacy, and math skills, with some effects measured through third grade (Chambers et al., 2016; Lipsey et al., 2009). The IES-funded *Effects of Preschool Curriculum Programs on School Readiness* project examined 14 curricula but found limited positive effects overall (PCER, 2008). Subsequent analyses by Jenkins et al. (2018) of the PCER data indicated that classrooms using *The Creative Curriculum*® consistently outperformed those using locally developed curricula in classroom quality, although not in child outcomes. Jenkins and colleagues also found that implementation of a literacy-focused curriculum led to improved children's literacy outcomes but not in other domains, and that implementation of a structured math curriculum improved children's math outcomes but was associated with neutral or trending negative effects on other skills. Additional analyses by Jenkins et al. (2019) integrated data from other large-scale studies and echoed these mixed results: for instance, in the Head Start Family and Child Experiences Survey (FACES) 2009, children in *The Creative Curriculum* classrooms performed better in receptive vocabulary and math, while the absence of a formal curriculum in the National Center for Early Development and Learning (NCEDL) pre-K study was associated with increased behavioral issues. Lipsey et al. (2009) also found that two curricula, which included *The Creative Curriculum*, showed detectable effects in third-grade state tests, though these were not observable with standardized measures at the end of preschool. Engel et al. (2025) note that despite knowing the least about the effects of whole-child (i.e., "comprehensive") curricula on children, including *The Creative*

Curriculum, these curricula fare strongly in their supports for teachers across dimensions of ECE experiences that are important in many contexts, including professional development (PD) and materials that support implementation, linguistic responsiveness, individualization for children with special needs, individualization based on other child needs, and in family engagement—areas often absent or minimal in domain-specific curricula.

Taken together, these studies highlight the potential of *The Creative Curriculum* to contribute to high-quality preschool experiences but also underscore the complexity of effectively implementing and supporting curriculum across settings. Summarizing the evidence, a recent National Academies of Sciences, Engineering, and Medicine (NASEM, 2024) report states that "studies have found some evidence that comprehensive curricula sometimes have positive effects on the quality of classroom interactions and in some outcome domains (e.g., Fantuzzo et al., 2011; Jenkins et al., 2018). "In comparisons of domain-specific and comprehensive curricula, the evidence tends to show no difference in general interactions in the classroom but marked differences in instruction related to the skill area targeted by the domain-specific curriculum and positive impacts on child learning in that area" (p. 52). In addition, the report emphasizes that the variability in curriculum findings makes it difficult to draw definitive conclusions about causality (p. 372) and that implementation quality is a critical component. The report also notes that subgroup effects have been underexamined due to sample size limitations. Among the limited subgroup analyses conducted, findings are mixed for Black and Latinx children, multilingual learners, and children with lower baseline skills, underscoring the need for further research. The report further highlights the critical role of professional development, contextual program factors, and intensive training and ongoing coaching in supporting effective curriculum implementation (Davidson et al., 2009; Jenkins et al., 2019; Weiland et al., 2018).

Accordingly, *Teaching Strategies*® embarked on an effort to evaluate their newly revamped, fully digital Teaching Strategies ecosystem (including *The Creative Curriculum*) in combination with a set of aligned supports, relative to practices as usual. The ecosystem reflects a broader trend towards digitalization in the field, as digital tools are increasingly being developed and offered to preschool programs. These shifts have the potential to reshape teachers' day-to-day experiences and professional development, raising important questions about how digital integration supports high-quality implementation and professional learning, and further so, as artificial intelligence components may soon be critical to these trends. These tools have the potential to shift educators' in-service opportunities.

Against this backdrop of mixed research findings on curriculum and growing digital integration in early childhood education, this study evaluates the impact of implementing the ecosystem—supported by additional teacher trainings and enhanced coaching—compared to practices as usual. To reduce variation in context, the study was completed within the New Jersey preschool program in former "Abbott" districts. New Jersey has been providing high-quality, full-day pre-k to 3- and 4-year-olds in the poorest 31 districts since 1998, as mandated by the Abbott court decisions (see Frede & Barnett, 2011, and Barnett et al., 2013). Since 2017, NJ has expanded pre-k access to more districts and children throughout the state with a commitment to universal pre-k for 3- and 4-year-olds. Currently, the state serves 22% of 3-year-olds and 33% of 4-year-olds in state-funded pre-k (Friedman-Krauss et al., 2024), and the program is mixed delivery—offered in public schools, Head Start centers, and private child care centers. Doing this research within the NJ preschool program ensures minimal variation in structural aspects, such as teacher-child ratios (2:15), teacher qualifications (with a BA required for lead teachers and a CDA

required for assistant teachers), compensation parity, and high levels of per-child spending (averaging $16,302 per child in the 2022–2023 school year). Longitudinal research has shown that the NJ preschool program yields sustained benefits, with effects continuing through tenth grade (Barnett & Jung, 2021). Notably, districts participating in the program could select among three curriculum options at the start of the study, to be used across all programs in their district: *The Creative Curriculum*, *High Scope*, and *Tools of the Mind*. The two districts in the randomized arms of this study were using *The Creative Curriculum.*

In this context, teachers and coaches in both districts were randomized to treatment and control arms for this study. Treatment teachers and their coaches were offered a series of virtual, synchronous professional development sessions throughout the study. Control teachers and coaches represented "practices as usual" within the districts. Both arms had access to the ecosystem*,* and participating districts had longstanding requirements for using Teaching Strategies *GOLD,* the assessment component. This effort aimed to ensure that teachers and coaches were supported with professional development on their use of the ecosystem, understanding that there is a lot of variation in which components programs may be using at any given time and in the extent to which they engage in professional development efforts (and provide time and coverage for it) to support their teachers and coaches. This research report summarizes the study, procedures, and sample (more information is provided in Nores et al., 2026, and Harmeyer & Nores, 2026), as well as the analyses and results for teacher and child outcomes.

Teaching Strategies Creative Curriculum Implementation and Ecosystem Engagement Study (CCIEE)

The Teaching Strategies® ecosystem is a comprehensive, research-based digital solution designed to support high-quality early childhood education by combining curriculum, assessment, professional development, and family engagement into a cohesive, connected system. It is grounded in two overarching theories of change—that these resources must be integrated into an end-to-end ECE system and that the effectiveness of these components is mediated by the behavior of the adults in the system. The ecosystem aims to empower administrators, educators, and families to drive positive learning outcomes for young children. The five components of the Teaching Strategies ecosystem include:

- *Comprehensive and integrated approach*. The ecosystem provides an end-to-end solution, ensuring that all essential components of early learning—curriculum, formative and summative assessment, teacher support, and family engagement—are seamlessly connected.
- *Research-based, practice-proven*. Rooted in evidence-based practices, the ecosystem incorporates high-quality curriculum (*The Creative Curriculum*®), authentic assessment (*GOLD*®), professional development, and coaching to enhance instructional effectiveness.
- *Empowering educators*. Teachers have access to ongoing professional development and coaching, enabling them to implement developmentally appropriate and culturally responsive teaching practices with fidelity.
- *Family engagement*. The ecosystem fosters strong family-school partnerships through in-platform family sharing and add-on tools like *ReadyRosie*®, helping families actively participate in their child's learning journey.

- *Actionable data*. Built-in assessment, reports, and analytics provide educators and administrators with actionable insights to drive informed instructional decisions and measure program effectiveness.

The ecosystem integrates all these components to ensure that curriculum implementation is supported by strong assessment practices, professional learning is aligned with real classroom needs, and families are engaged as partners in their child's development.

With this in mind, this study implemented a cluster randomized controlled trial (RCT) within two low-income districts implementing New Jersey's state preschool program, with an additional non-equivalent comparison group of classrooms in districts implementing practices as usual across the state. The research assessed the effectiveness of engagement with the ecosystem, especially the aligned supports, which included professional development and coaching provided by Teaching Strategies. The study put forward five major goals:

(1) Measure teachers' engagement with the ecosystem, their teaching practices, and their well-being in contexts with relatively high levels of resources and structural quality (e.g., ratio, class size, wages, spending, among others);
(2) Describe children's experiences as measured via observational protocols;
(3) Create exogenous differences in ecosystem engagement through random assignment to the aligned supports to increase confidence in the causal inference relating engagement to outcomes;
(4) Assess the degree to which the variation in levels of support for engagement and implementation (induced by the random assignment) are associated with observed practices and child outcomes;
(5) Include measures of creativity and problem solving that ensure evaluation of curriculum effects is more relevant to *The Creative Curriculum* than past research.

Early care and education (ECE) programs are vital to children's development, and Bronfenbrenner's Ecological Systems Theory (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 2007) offers a comprehensive lens for understanding children's learning within ECE programs. Using the person–process–context–time framework, we focus on capturing induced variation in teaching practices and ecosystem engagement and assessing their relation to children's outcomes over time. Complementary theories further enrich this perspective: Vygotsky's Sociocultural Theory (Vygotsky & Cole, 1978) emphasizes learning as scaffolded through interactions with peers and adults, which is critical here for both the teachers receiving training and the children themselves; Piaget's Constructivist Theory (Piaget, 2013) centers on knowledge construction through active engagement with their environments; and Bandura's Social Learning Theory (Bandura, 1986) highlights learning through observation, modeling, and the role of self-efficacy, providing a lens to assess how families and teachers—key figures in children's microsystems—engage in learning. This theory suggests that individuals learn by observing and modeling behaviors and emphasizes self-efficacy and motivation as drivers of change, thereby informing the evaluation of professional development opportunities for teachers. Together, these theories offer a multidimensional foundation for exploring implementation and impact in the CCIEE study.

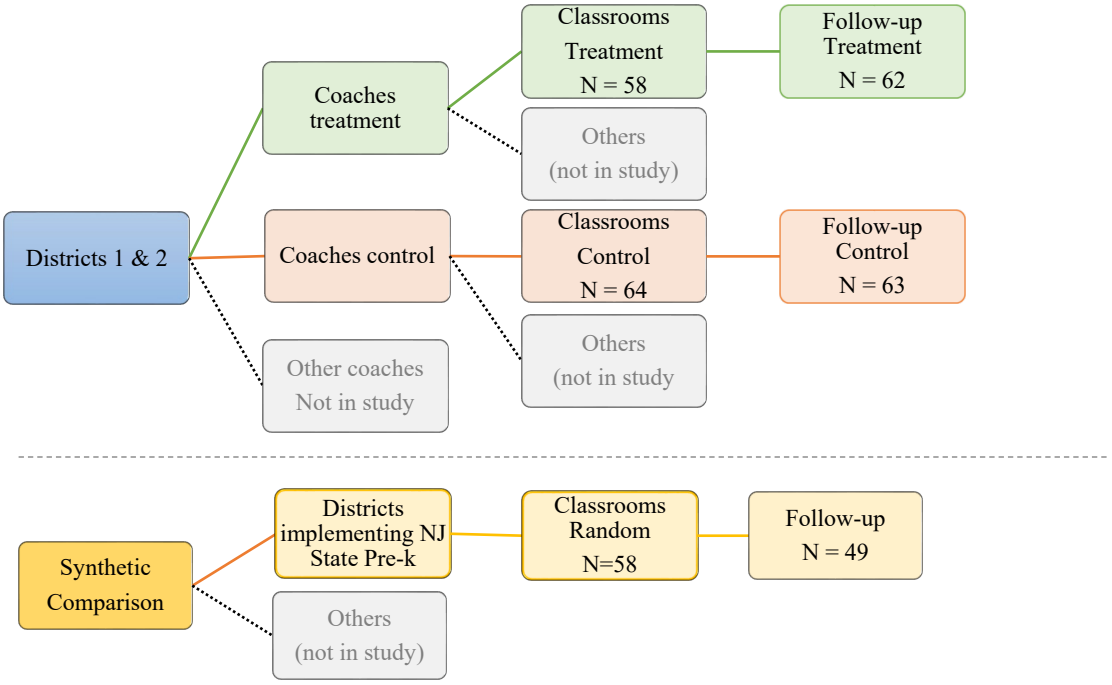
Methods

Randomization

We conducted a cluster randomized controlled trial (RCT) with coaches and their teachers assigned to treatment and control groups within two districts in New Jersey's state preschool program. In Figure 1, we show the study's sample selection. Specifically, one group of coaches and the teachers/classrooms they work with in each district were assigned to the primary control group (n = 64), while the other group of coaches and the teachers/classrooms they work with were assigned to the treatment group to receive the virtual, synchronous PD and coaching supports (n = 58). Randomization was completed in the Fall of 2021, before the commencement of study activities. Clustering at the coach level is designed to account for shared professional development activities within each cluster. The study includes a third "synthetic" arm consisting of a secondary control group of teachers/classrooms where outcomes were also assessed. This third group is a set of randomly selected preschool classrooms across districts in the same state (n = 58 classrooms across 25 districts),[1] regardless of which curriculum they were implementing. This allows the RCT sample to be situated within the larger context of the New Jersey preschool program. Post randomization, the primary units of analyses are the teachers, and when a teacher was replaced throughout the study, the replacement teacher followed the assignment status of the former teacher. At follow-up, inclusion of additional teachers added by a district to a treatment coach (assigned to treatment as per coach assignment), as well as closure of three classrooms (two treatment and one control) resulted in sample of 125 teachers (control n = 63 and treatment n = 62). Detailed information on the randomization flow-chart (as per CONSORT guidelines; Hopewell et al., 2025), data, timelines, roles and responsibilities, and processes is provided in Nores et al., 2026.[2]

[1] In year 3 only 21 out of the 25 "synthetic" districts participated in the follow-up observations and assessments.
[2] Three treatment classrooms had inconsistent teachers throughout the study and therefore ended up in the control group as treatment was not possible; we maintain these in the treatment group in intent-to-treat analyses, but we account for this shift in treatment-on-treated analyses. Forty-one classrooms had teacher churns, and in all these cases we followed the new teacher(s) who were given the same treatment/control assignment as the teacher that left. Seven were reassigned by the district to a control coach, but we continued to follow the treated teachers as such, and these remained in the treatment group.

Figure 1. CCIEE sample selection and randomization status for classrooms/teachers at baseline



Context: The study context was deliberate. We selected New Jersey districts given the state's relatively high levels of structural quality across sites/settings and selected districts that were already implementing the core curriculum (*The Creative Curriculum*) and the aligned assessment (*GOLD*) components. Consistency in structural quality mitigates contextual differences that could potentially confound internal validity. Preexisting familiarity with the core components bypasses the exploration and initial "ramping up" phase that occurs during the first several years of implementing any curriculum and assessment system, narrowing the focus to differences in use of the digital components of the ecosystem. This design choice entails a significant tradeoff. On the one hand, introducing additional supports into an ongoing implementation rather than starting with a brand-new curriculum reduces treatment contrast. On the other hand, adding onto a successful implementation obviates key barriers that often hamstring completely new implementations.

What was not a deliberate design choice, however, was the quick universal adoption of the digital ecosystem in both districts in the study, a consequence of the push for digital education technology during the COVID-19 pandemic. Due to the recent adoption, it was expected that the intervention supports would still induce higher engagement among treated teachers (relative to teachers in the primary control group). Given engagement with the ecosystem in the control group (which we describe below) and district changes in expectations for curriculum and assessment, the study is likely a conservative test of the ecosystem's intervention impact.[3]

[3] During the study's duration study trainers reported that one of the two districts implemented district-wide expectations for and monitoring of teachers' digital lesson plan activities and assessment practices in the platform across all teachers regardless of treatment status, thus reducing treatment contrast. This was reported by the trainer in an internal study TS e-mail: Education consultant, internal communication. February 17, 2023. This was confirmed by the study team upon review of the digital lesson plans.

Timeline: Randomization was conducted in the Fall of 2021, and baseline classroom quality and fidelity observations and teacher surveys were conducted between October of 2021 and February of 2022, as supports started to ramp up. Synchronous (virtual) professional development and coaching started in the Winter of the 2021–22 school year and continued throughout the 2022–23 and 2023–24 school years. In the Fall of 2023, classrooms were provided with consent forms (in English and Spanish) to distribute to all children in the classroom. Four children were randomly selected among those for whom a consent was received. Child assessments were collected in control, treatment, and synthetic classrooms between October and December of 2023[4] and again in the Spring during May and June of 2024.

Procedures: In years 2 and 3 of the study, teachers were provided with incentives of $125 for each of the three school-year quarters for participating in the study PD. For teachers who were not provided with school time to participate in PD sessions and had to attend these after work hours, an additional $20 was distributed per hour of training. All participating classrooms received Teaching Strategies curriculum *Essential Kits*. Synthetic sites received $100 per data collection wave (baseline observations, child assessments, and follow-up observations) for participating in the study. The research team tracked any changes in teachers and coaches throughout this period (see Nores et al., 2026). Post-test classroom observations and surveys were collected in the Spring of 2024. Classroom observers and child assessors were blind to the treatment status of participants (although teachers could have communicated their status, so there is the possibility that assessors learned this information when in the classrooms). Families received a gift card of $25 for completing the family survey, teachers received one of $50 for completing each teacher survey (pre and post), and coaches received one of $50 for completing the coach survey, regardless of treatment. In addition, teachers also received a gift card of $5 for completing a socio-emotional survey for each child at each assessment point (fall and spring), regardless of treatment. The RCT was registered (AEARCTR-0012262), and all study procedures, protocols, and surveys underwent Institutional Review Board approvals (Rutgers IRB #Pro2020000039).

Intervention

The professional development and coaching plan implemented by Teaching Strategies was designed to induce exogenous differences in how teachers used the ecosystem (which includes *The Creative Curriculum* and *GOLD*), with an emphasis on strengthening curriculum implementation and *SmartTeach* platform engagement to support children's learning. The live sessions were synchronous and delivered virtually to the treatment group by a Teaching Strategies trainer. The sessions included learning the ecosystem components, foundations of curriculum and assessment, learning how Teaching Strategies' solutions aid teachers in carrying out effective practices in the classroom and beyond, creating developmentally appropriate and meaningful experiences for children in their care, implementing formative assessment and reflecting on the teaching and assessment cycle, responsive planning, effective day-to-day use of digital and print resources, implementing the ecosystem to fidelity, and more. The early coaching sessions focused on the *Fidelity Tool for Administrators* and the *Coaching to Fidelity* tools to support teachers at all levels of fidelity with improving their practice. The intervention also delivered ongoing support through virtual side-by-side and group coaching sessions designed to deepen teachers' understanding of the curriculum and enhance instructional quality, while

[4] A small fraction of Spanish assessments took place through early January 2024.

centering educator expertise and promoting confidence and self-efficacy. Detailed information on the timeline of the different sessions and topics is reported in Nores et al. (2026).

Theory of Change

Our theory of change (TOC) connects the professional learning opportunities and aligned supports provided by *Teaching Strategies* to the coach and teacher dyads to outcomes in teachers and teaching practices (as measured via classroom quality) and, subsequently, to children's learning outcomes. We hypothesize that access to the ecosystem, and the aligned professional development, training, and supports for coaches and teachers, would bolster curriculum fidelity, *SmartTeach* platform engagement, and teacher retention, consequently increasing the quality of teaching practices, and in turn, impact children's learning outcomes across the various domains supported by *The Creative Curriculum* and *GOLD*. The TOC is further illustrated in Nores et al. (2026).

Measures

The measures selected for the study have been used previously in early childhood classrooms, demonstrate adequate psychometric properties and have detected program outcomes in other studies (e.g., Barnett et al., 2018; Macdonald et al., 2021; Siraj et al., 2023; Weiland et al., 2024). Data collectors for child assessment and classroom observation measures were trained by NIEER, with the exception of the fidelity measure, which was administered directly by Teaching Strategies observers. For classroom observation measures, separate trainings were provided on the ECERS-3 and SSTEW observation protocols. ECERS-3 observers were trained by an ECERS-3 certified trainer and met the ERSI reliability requirements for observer certification (85% reliability or above, measured by exact matches or one match away from the trainer score). ECERS-3 observers' reliability for this study ranged between 85% to 94% (mean 89%). SSTEW observers were trained by the authors in a three-day training (including online modules), followed by a live online discussion with the authors. Observers then underwent three reliability sessions with the NIEER anchor team and a final meeting with the developers for additional clarifications. SSTEW observers met reliability requirements with agreement percentages with anchors ranging between 80-93% (mean 89%). All observation score sheets were cleaned, entered, and analyzed at NIEER. For child assessments, data collectors received a two-day training on the measures, followed by two live reliabilities with children conducted by NIEER staff. Each assessor was shadow scored to ensure 100% agreement (with the exception of the TCAM, described further below). A refresher training took place prior to the beginning of spring data collection.

Measure on Retention. Retention was defined as a lead preschool teacher in a classroom at the start of the study supports who continued in that position in the same classroom throughout the study period (i.e., was still in the same classroom at the end of year 3). Churn was coded once that lead teacher was replaced by a new teacher, regardless of how many times a churn occurred after the first replacement. Further information is provided in Nores et al., 2026.

Measures on Classrooms. The study used three tools to assess classroom quality and curriculum fidelity. The **ECERS-3** (Harms et al., 2015) rates preschool environments across six domains using a 1–7 scale, with higher scores reflecting higher quality; the most recent scoring clarifications at the time of data collection were applied. The ECERS tool (this version or the previous one) is among the most widely used classroom observation tools in the early childhood

field (Early et al., 2018). Alphas for the ECERS at pre and post-test are all above 0.87 with factor analyses confirming a one-factor solution (Nores et al., 2026). The **SSTEW** (Howard et al., 2018) evaluates educator support for children's thinking, curiosity, questioning, confidence, risk-taking, and autonomy in children's learning processes and the display of different teaching and learning strategies and relevant content knowledge applied flexibly with contextual, individual, and socio-cultural sensitivity. Alphas for the SSTEW at pre and post-test are all above 0.77 with factor analyses confirming a one-factor solution as well (Nores et al., 2026). The **Fidelity Tool for The Creative Curriculum** assesses how well educators implement the curriculum, examining the classroom environment, use of materials, instructional planning, and individualized teaching based on observation and teacher interview or survey. Because the tool significantly overlaps with ECERS-3 and the SSTEW, we limited the measure to only include the curriculum "Use" (i.e., curriculum implementation) subscale.

Measures on Dosage and Engagement. Dosage and engagement are critical to measure as they capture variation in exposure to the intervention and variation in engagement with the *SmartTeach* platform. Dosage is captured in two ways: the number of hours of synchronous, virtual professional development that treatment teachers engaged in during the study and the number of asynchronous, on-demand course hours they completed via the ecosystem PD website (See Nores et al., 2026). Engagement is defined as the sum of the number of minutes (which we divide by 60 to calculate hours) spent on the *SmartTeach* platform in the following areas: Teach, Assess, Engage (Family), Report, and Library.

Measures on Children. We utilize a first battery of externally assessed measures that include socio-emotional development, executive function, creativity, and problem-solving assessments:

- Dimensional Change Card Sort task (DCCS; Zelazo, 2006) assesses attention-shifting.
- Peg Tapping test (PT; Diamond & Taylor, 1996) requires children to inhibit a natural tendency to mimic the experimenter while remembering the rule for the correct response.
- Strengths and Difficulties Questionnaire (SDQ: Goodman, 2001) is a brief behavioral questionnaire about 2–17 year-old children. The total difficulties score indicates overall behavioral and emotional problems (i.e., lower scores are preferred).
- Penn Interactive Peer Play Scale (PIPPS) is a behavioral rating instrument useful for measuring peer play behaviors for teachers (or parents) to respond (Fantuzzo & McWayne, 2002).
- Thinking Creatively in Action and Movement test (TCAM; Zachopoulou, Makri, & Pollatou, 2009) assesses the creativity of young children or others with limited verbal and drawing skills. This is an adaptation of the Torrance Tests of Creative Thinking (TTCT; Torrance, 1974) for younger children. The three subscales of the TCAM are Fluency, Imagination, and Originality.

A second battery includes cognitive standardized measures of receptive vocabulary, literacy, and math:

- The Peabody Picture Vocabulary Test–Fourth Edition (PPVT-IV; Dunn & Dunn, 2007) is a 204-item adaptive test of receptive vocabulary in standard English. For children with Spanish language proficiency, we also used the Test de Vocabulario en Imágenes Peabody (TVIP; Dunn et al., 1986).
- The Woodcock-Johnson Psycho-Educational Battery-Fourth Edition (WJ-IV; Schrank, Mather, & McGrew, 2014), including the Applied Problems and Letter-Word Identification

subtests, i.e., broad emerging math and reading skills. For children with Spanish language proficiency, we also used the equivalent subtests from the Woodcock-Muñoz Batería III (WM-III; Schrank et al., 2005).

For the PPVT/TVIP and WJ/WM measures, we use total raw score as our outcome measure, consistent with other pre-k studies that have used the PPVT (e.g., Barnett et al., 2018; Weiland et al., 2024). For the TCAM, the intraclass correlation (ICC) ranged from 0.81 to 0.97, and up to 4 reliability iterations were performed when necessary. Spanish-speaking children were also assessed in Spanish on the PPVT and Woodcock-Johnson (by Spanish-speaking bilingual data collectors) using the Peabody Picture Vocabulary Test in Spanish (*Test de Vocabulario en Imágenes Peabody*, TVIP; Dunn et al., 1986) and the Woodcock-Muñoz III (Schrank et al., 2005). Teachers were the primary informants on whether a child should be assessed in Spanish in addition to English. However, if a Spanish-bilingual child did not get past set 3 in the PPVT-IV in English, they were also assessed in Spanish. The DCCS and Peg Tap were administered only once, in the child's dominant language. The socio-emotional and play surveys for children (SDQ and PIPSS) were collected from teachers on paper or via an online platform (Qualtrics). We collected demographic information on the child from parents and from the districts' administrative data (with data sharing agreements in place).

In addition, we integrate the measures of language, cognitive, literacy, mathematics knowledge, skills, and abilities, and social-emotional development, assessed via teachers' regular practice of using the curriculum-aligned *GOLD* assessment (Lambert, 2020). The objectives measured by each domain (i.e., "area") of *GOLD* are listed below (additional information is provided in Nores et al., 2026).

- *Language:* listening and understanding of increasingly complex language, and expressive use of language.
- *Cognitive:* approaches to learning, connecting experiences, classification skills and use of symbols and images to represent abstract concepts.
- *Literacy:* phonological awareness, phonics skills, and word recognition, knowledge of the alphabet and print, comprehending and responding to texts, and emerging writing skills.
- *Mathematics:* number concepts and operations, spatial relationships and shapes, comparing and measurement, and knowledge of patterns.
- *Social-Emotional:* self-regulation of emotions and behaviors, establishing and sustaining positive relationships, and cooperative and constructive participation.

For the *GOLD* measures in the study, we assess growth for the children in our sample and also for all children in the treatment and control classrooms, since the measure is used to assess all children in these districts as part of their standard practices.

Teacher and Coach Surveys

We surveyed teachers on their experience and qualifications, some aspects of curriculum fidelity, parent/family engagement, the amount of training and professional development they received, and demographics. Teacher surveys were distributed in the fall of 2021 and collected throughout the first school year, then again in the fall of 2023 and collected throughout the third school year. A 74% (n=128) response rate was attained at baseline and 90% (n=156) at follow-up. Coaches also completed a survey requesting information on experience and qualifications, interaction and

coaching strategies typically used with teachers, practice and modeling strategies used with teachers, and coaching goals. Coaches also reflected on their success in areas like assessment and curriculum implementation.

Analytical strategy

For classrooms, we hypothesize that the intervention (with dosage varying across teachers) would impact teachers' engagement with the ecosystem and their curriculum implementation, and that this impact would be observable in the curriculum fidelity measure and the classroom quality measures. We also hypothesize impact on teacher well-being, which would translate into effects on teacher retention. We assess these three outcomes, as well as the role of retention, engagement, and dosage. We estimate intention-to-treat (ITT) effects on the outcome measures using the following ordinary least squares (OLS) specification (and a parallel Probit estimation for the estimates on teacher retention):

$$C_j^t = \beta_1 + \beta_2 ITT_j + \beta_4 X_j + \varepsilon_i \qquad (1)$$

where *ITTi* equals 1 if classroom *i* was randomly assigned to treatment and 0 otherwise. C_j^t is an outcome variable for classroom *j* in period *t* (follow-up), and X_j is a vector of baseline teacher controls including teacher gender, years of experience, race and ethnicity, education, and coach retention.[5] We also include teacher retention in all the estimations where retention is not the primary outcome, plus, district fixed effects with standard errors clustered at the coach level. As *ITT* was randomly assigned, we can be confident that $\widehat{\beta_2}$ estimated by OLS (and probit for retention) captures the causal impact of random assignment to treatment on the outcome. Variations of this model test moderation by dosage and engagement, by adding these as main effects, and as an interaction term with ITT_i. For teacher covariates, we use a missing data dummy strategy (as in Weiland et al., 2024) to retain all cases and produce estimates adjusted for missingness.

For children, we hypothesize that the impacts on teacher retention, classroom quality, and curriculum fidelity will moderate the effect on children's growth over the school year, controlling for their performance at the beginning of the school year. We hypothesize that the intervention would impact the cognitive and non-cognitive outcomes described above. We estimate intention-to-treat (ITT) effects on the outcome measures using the following ordinary least squares (OLS) specification:

$$A_{i,j}^t = \beta_1 + \beta_2 A_i^{Baseline} + \beta_3 ITT_j + \beta_4 C_i^t + \beta_5 C_j ITT_j + \beta_6 X_i + \varepsilon_{ij} \qquad (2)$$

where *ITTi* equals 1 if classroom *i* was randomly assigned to treatment and 0 otherwise. $A_{i,j}^t$ is the outcome for child *i* in classroom *j*, C_i^t is an outcome variable for classroom *i* in period *t* (follow-up), $A_i^{Baseline}$ is the same measure at baseline for child *i*, and X_i is a vector of baseline child controls including gender, race and ethnicity, and whether the child has an individualized education program (IEP) and whether the child is a dual-language/multilingual learner. We also include district fixed effects and cluster standard errors at the classroom level. Given the multiple

[5] Given that two classrooms were in fact not treated as they had continuous changes in teachers and substitutes that did not allow an opportunity to engage them we also estimated the impact of effective treatment (treatment-on-treated or TOT). Treatment-on-treated (TOT) specifications can be found in Nores, et al. (2025) and findings align. We were unable to match GOLD outcomes for these two classrooms and a third classroom, so three classrooms are excluded form GOLD estimations and ITT and TOT estimations for GOLD do not differ in these cases.

child outcomes, we adjust two-tailed tests' p-values for multiple hypotheses testing using the Romano and Wolf (2005) stepdown procedure, grouping cognitive outcomes as one set, and socio-emotional, executive function, and creativity as a three other groups. We report adjusted p-values in Appendix E. Variations in this model examine heterogeneous effects by age at baseline, gender, dual-language learner status, and baseline development level under the 25th percentile (following work by Lipsey et al., 2009; Harden et al., 2023, and McCormick et al., 2023). All coefficients are reported in terms of the standard deviation of the distribution of the control groups' scores at baseline for the corresponding measure.

Results

Baseline characteristics

Table 1 provides summary statistics at baseline by random assignment (ITT) for the teachers in the sample, while coach characteristics may be found in Harmeyer & Nores, 2026 and Nores et al., 2026. All coaches, regardless of randomization, had a master's degree and were female. They were also quite similar in their average years of experience as a coach (12 in the control group and 13 in the treatment group) and in years of experience in early childhood (22.2 in the control group and 24.2 in the treatment group). Teacher characteristics were also balanced between control and treatment groups in teacher education and gender distributions. The statistically significant difference across groups was for years of experience, with the teachers in the treatment group exhibiting higher average experience among those reporting it, and for the distribution across race and ethnicity. These variables are all included as controls in classroom-level estimations.

In addition, pretest and post-test external outcomes are included in Table 1. At pretest, there were no statistically significant differences in ECERS-3 nor SSTEW scores between the control and treatment group, while fidelity differences were trending in significance in favor of the treatment group. Without accounting for any baseline differences between groups, only retention is statistically significant in favor of the treatment group at post-test.[6] On average, both the treatment and control groups evidenced gains in quality and curriculum fidelity throughout the study (see Nores et al, 2026 for further details). Moreover, teachers in the treatment group had a retention rate of 80.64%; 30 percentage points higher than that of the control group (models reported further below estimate retention after adjusting for teacher characteristics and districts).

Table 1 also reports the study's implementation variables of (PD) dosage and platform engagement, which capture variation in the intervention. On average, treatment teachers participated in 14.18 (SD 8.46) synchronous sessions with an average total of 23.52 hours accumulated across the three years (SD 13.89). In the control group, as per the treatment, teachers did not participate in these sessions. Both treatment and control teachers also had access to *self-directed* trainings within the ecosystem. The vast majority (>97%) of self-directed hours were asynchronous on-demand course completions, with treatment teachers completing, on average, 9.91 hours of asynchronous courses (SD 14.07) and control teachers only completing 4.60 hours (SD 6.28), a meaningful difference (p=0.004).[7] Unlike the treatment PD, the

[6] Retention is defined as the absence of churn (as we followed the classroom). We exclude closed classrooms, and code two classrooms that started as vacant as non-retained. A third classroom that had been vacant was filled before finishing the pre-test, so that one case is coded as retained.
[7] See Nores, et al., 2025 for more details.

asynchronous courses were driven by teachers' own choices and interests. In terms of platform engagement, teachers in the treatment group spent on average 314 hours total over the three year period on the online *SmartTeach* platform (in a scale of weeks, this is equivalent to 8.97 weeks of 35 hours per week of total engagement across three years), while teachers in the control group spent on average statistically significantly less (about 94 hours less or 2.29 weeks less) during the three-year period ($p = 0.003$).[4] Teachers also chose how and how much to engage with the online *SmartTeach* platform.

Table 1: Teacher characteristics by Control vs. Treatment groups, n=125.

Variable	Control			Treatment			P-value
	N	Mean/ Percent	SD	N	Mean/ Percent	SD	
Descriptive Statistics							
Female	55	81.30%		57	91.94%		0.328
Male	2	3.17%		3	4.84%		
Missing	6	9.52%		2	3.23%		
Teaching experience (years)	56	6.21	6.10	61	11.18	8.89	0.001***
Missing	7			1			0.331
Race/Ethnicity[i]	63			62			0.012**
Black, African, or African American	23	36.51%		10	16.13%		
Hispanic/Latino	15	23.81%		22	35.48%		
White	10	15.87%		21	33.87%		
Asian, Multi-racial & Other	9	14.29%		6	9.68%		
Missing	6	9.52%		3	4.84%		
Education	63			62			0.427
Associate degree (2 year)	1	1.59%		1	1.61%		
Bachelor's degree (4 year)	31	49.21%		33	53.23%		
Master's degree	22	34.92%		25	40.32%		
Doctorate degree	1	1.59%		1	1.61%		
Missing	8	12.70%		2	3.23%		
Implementation Variables							
Dosage: Synchronous PD hours	63	0.00	0.00	62	23.52	13.89	0.000***
Dosage: Asynchronous PD hours	63	4.60	6.28	62	9.91	14.07	0.004***
Engagement: *SmartTeach* platform hours	63	219.42	156.40	62	313.90	188.75	0.003***
Quality Measures at Baseline							
ECERS-3	60	3.76	0.72	56	3.65	0.80	0.444
SSTEW	58	2.62	0.77	57	2.66	0.80	0.766
Fidelity (Curriculum Use)	55	0.47	0.19	53	0.52	0.18	0.099*
Quality and Teacher Measures at Post-Test							
Retention (%)	63	50.79		62	80.64		0.000***
ECERS-3	63	3.93	0.63	61	3.85	0.88	0.572
SSTEW	63	3.00	0.93	62	2.84	1.14	0.370
Fidelity (Curriculum Use)	62	0.56	0.21	61	0.54	0.20	0.630

Notes: ***p<0.01, **p<0.05, *p<0.1. P-values for chi-squared tests of independence, or two-tailed t-tests for differences between the distributions, respectively. The table includes demographics for all classrooms using data from pretest and post-test. For teacher covariates, we use a missing data dummy strategy to retain all cases.

Table 2 provides summary statistics for children's socio-demographic characteristics at pretest by random assignment (ITT) for children in the analytical sample. We define the analytical sample as all the children for whom we have any one assessment at post-test. Children in the analytical sample were, on average, 51 months old. About 6% of children had an IEP, 52% were

female, and about 43% were DLLs. There are statistically significant differences between treatment and control children in terms of the percentages that were Black (higher in the control group), Hispanic (higher in the treatment group) and DLL (higher in the treatment group). All subsequent analyses, therefore, control for these variables.

Table 2: Child characteristics by Control vs. Treatment groups at the beginning of school year 3.

Variable	N	Treatment			Control			P-value[a]
		N	Mean	SD	N	Mean	SD	
Age	477	240	51.02	6.67	237	51.27	7.13	0.689
IEP	475	238	0.05	0.23	237	0.06	0.24	0.835
Female	477	240	0.50	0.50	237	0.53	0.50	0.490
Black	477	240	0.31	0.46	237	0.47	0.50	0.000***
Hispanic	477	240	0.60	0.49	237	0.46	0.50	0.002***
White	477	240	0.02	0.14	237	0.02	0.13	0.751
Asian	477	240	0.03	0.16	237	0.01	0.11	0.323
Other	477	240	0.04	0.19	237	0.04	0.20	0.794
DLL	474	237	0.49	0.50	237	0.37	0.48	0.005***

Note: ***p<0.01, **p<0.05, *p<0.1 [a]Standard p-values for two-tailed t-tests.

Table 3 provides pretest data for children by random assignment (ITT) for children in the sample. At pretest, there were no statistically significant differences between treatment and control children for any of the outcomes. Table 3 also includes the subsample assessed in Spanish (39% of the sample), for which some baseline differences in favor of the treatment group emerge. For *GOLD* measures, we also include data on all children in each classroom in the study, not just the sample children who were followed as part of this study. Therefore, we also compared our study's sample to the full set of children in the treatment and control classrooms (Appendix A, Table A.1). We found few significant differences at baseline for any of the *GOLD* domains between the study's sample and the full set of children enrolled in the classrooms that made part of the study. There was only a significant difference at the trend level for the social-emotional and language domains in favor of the control group. Only one of these remains significant after multiple hypothesis testing checks.

To assess baseline equivalency, we also ran summary statistics by randomization status for only three-year-olds in the sample (Appendix A, Table A.2), as they would be new to preschool, given the age requirements for entering New Jersey's state-funded preschool program, and would not have been exposed to treatment the year prior. For 3 year-old children, we only found statistically significant differences in the PPVT standard and WJ Applied Problem scores favoring the control group, but no statistically significant differences between treatment and control children in adjusted p-values with the exception of for the WJ Applied Problems standard score (at a trend level, $p = 0.073$ after adjustments for multiple hypothesis testing) in favor of the control group (Table A.2). For *GOLD* domains we found no statistically significant differences in the 3-year-old study sample but found differences in the language and social-emotional domains for all 3-year-old children sustained after multiple hypothesis testing in favor of the control group.

Table 3. Child pretest summary statistics by randomization status.

Pretest	N	Treatment			Control			P-value[b]	RW p-value[c]
		N	Mean	SD	N	Mean	SD		
Executive Functions, creativity, and Socio-Emotional outcomes									
DCCS	438	216	8.89	5.55	222	8.32	4.76	0.242	0.457
Peg Tapping	434	215	3.67	4.80	219	4.01	4.83	0.451	0.878
TCAM Raw	426	208	14.50	4.59	218	15.14	4.78	0.161	0.736
TCAM Standardized	427	208	83.89	11.47	219	85.89	22.48	0.251	0.736
PIPPS	352	178	71.99	11.53	174	71.32	11.84	0.590	0.878
SDQ Behavioral Problems	352	178	10.62	5.95	174	10.08	5.24	0.365	0.777
SDQ Prosocial Behaviors	352	178	6.44	2.31	174	6.40	2.07	0.859	0.878
Cognitive Outcomes									
PPVT Raw	447	221	45.49	23.68	226	47.00	22.87	0.494	0.815
PPVT Standardized	447	221	81.21	18.18	226	82.65	18.02	0.401	0.749
WJ Letter Word Raw	446	220	5.75	6.50	226	5.69	5.03	0.914	0.983
WJ Letter Word Standardized	446	220	91.40	14.66	226	91.41	13.87	0.993	0.991
WJ Applied Problem Raw	444	220	5.14	4.02	224	5.54	3.77	0.276	0.639
WJ Applied Problem Standardized	444	220	76.58	16.36	224	78.86	16.87	0.148	0.434
Cognitive Outcomes – Spanish Assessments									
TVIP Raw[a]	173	104	13.84	11.26	69	11.71	12.55	0.247	0.466
TVIP Standardized[a]	173	104	85.03	14.45	69	79.51	17.24	0.024**	0.085*
WM Letter Word Raw[a]	173	104	5.69	3.13	69	5.90	3.39	0.682	0.701
WM Letter Word Standardized[a]	172	103	83.19	19.03	69	81.52	18.50	0.569	0.701
WM Applied Problem Raw[a]	172	103	7.23	4.75	69	5.97	4.64	0.087*	0.223
WM App. Probl. Standardized[a]	172	103	80.13	17.22	69	72.58	19.87	0.009***	0.048**
GOLD Measures									
Social Emotional	425	212	399.68	57.21	213	404.53	66.52	0.422	0.870
Language	421	208	427.51	72.82	213	434.30	79.30	0.361	0.593
Cognitive	412	206	409.54	61.97	206	413.17	62.03	0.553	0.476
Literacy	395	193	467.58	41.15	202	465.60	43.60	0.645	0.870
Math	395	187	327.07	53.19	208	324.79	52.23	0.668	0.870

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and control children. [a]Only for those also assessed with the Spanish measure. [b]Standard p-values for two-tailed t-tests. [c]Stepdown p-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications.

Standardized scores allow comparing children in the sample to children in the norming sample for each instrument, and for some measures, they allow comparing them to children in other studies. Children in the sample showed low standardized scores in receptive vocabulary, language, and math relative to the norming sample of peers their age, with overall PPVT standardized scores of 81.94, WJ Letter Word scores of 91.41, and WJ Applied Problem scores of 77.73.[8] This was also the case when summarizing only 3-year-olds. To put this into perspective, children in the fall of 2019, FACES study of Head Start programs scored on average 81.40 on the PPVT, 86.40 in the WJ Letter Word IV, and 80.30 in WJ Applied Problems. Therefore, children in our sample showed similar scores to Head Start children in the FACES study, with Head Start serving low-income children across the country.[9]

[8] Standardized scores for the PPVT, WJ, TVIP and WM are standardized with a mean of 100 and a standard deviation of 15. That is, scores under 100 represent underperformance relative to children of the same age.

[9] See Tables B4, B8 and B11 in: Kopack Klein, A., Aikens, N., Li, A., Bernstein, S., Reid, N., Dang, M., Blesson, E., Rakibullah, S., Scott, M., Cannon, J., Harrington, J., Larson, A., Malone, L., Tarullo, L. Descriptive Data on Head Start Children and Families from FACES 2019: Fall 2019 Data Tables and Study Design. OPRE Report #2021-77. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. https://www.acf.hhs.gov/opre/report/descriptive-data-head-start-children-and-families-faces-2019-fall-2019-data-tables-and

We also assess correlations between the external measure scores collected directly by the study team with children in the study and the *GOLD* domain scores (Tables A.5 and A.6). At pretest we find moderate to strong correlations between all *GOLD* domains and the PPVT ($r \sim 0.47–0.50$ for the *GOLD* literacy and language domains), and the WJ Applied Problems measures ($r=0.50$ for *GOLD* math). Correlations with the WJ Letter-Word measure are lower, though statistically significant, potentially reflecting differences in construct specificity and task demands relative to observational assessments. These patterns are also present at post-test, with slightly stronger correlations observed, relative to those at pretest. In contrast, correlations between fall-to-spring gains (Table A.7) are low, indicating that while the external measures and *GOLD* domains align in the developmental levels they capture at each time point, they diverge in their sensitivity to growth over time. These patterns suggest that the measures may differ in their sensitivity to the processes and contexts driving developmental change, including classroom experiences. Lagged correlations between pretest *GOLD* domains and post-test external assessments and vice versa (Table A.8) are stronger, reinforcing the presence of stability in children's relative developmental ordering across measures, even as growth trajectories differ. Interestingly, there are moderate to high cross-domain correlations across measures. At pre- and post-test, the WJ measures appear to have floor effects in the lower portion of the distributions (see Figures A.1 and A.2).

As the PIPPS measure on play and the SDQ on socio-emotional issues were teacher-reported, these measures have higher missing data rates (of 22% percent) than the external assessment measures. We assess whether missingness in PIPPS or SDQ was associated with either the treatment or the demographic characteristics of children, and find this was not the case (Table A.9). For all other external outcomes, missing data analyses show few significant differences across outcomes.

In addition, the study included a "synthetic" group of children from other NJ state preschool classrooms across the state.[10] We find the treatment group underperformed at pretest relative to children across the state (Table A.10) and this pattern was also observable at post-test (Table A.11). This finding illustrates how the sample in the study not only paralleled other low-income children across the country but was also developmentally disadvantaged relative to children enrolled in other districts in NJ state's preschool program. There were, however, no statistically significant differences in growth for children in the treatment versus synthetic group, with the exception of growth in DCCS, which marginally favored the synthetic group after adjustment for multiple hypothesis testing (Table A.12).

Estimations of program impact on primary outcomes: teacher retention, classroom quality, and curriculum fidelity

Table 4 reports treatment effects, standard errors, and standard P-values for the classroom-level outcomes. We first present ITT estimates along with their standard errors (TOT estimates are reported in Appendix Table A.13). We then report ITT estimates including main effects for, and interactions with, PD dosage and platform engagement (measured in a scale of weeks of engagement at 35 hours per week), correspondingly for all models with the exception of the model on retention, as total dosage and total engagement are both dependent on, and associated with, retention. When including the interaction terms, the ITT coefficient represents the impact of treatment when dosage or engagement is zero, which is why these coefficients shifted.

[10] In year three, this sample included 21 out of the 25 original districts recruited to the study.

Program effects are reported as fractions of a standard deviation relative to the control group's baseline quality or fidelity scores. For retention, program effects are reported as marginal probabilities.[11]

For retention, we found strong statistically significant positive effects, with the intervention increasing teacher retention consistently (model *a;* $\beta = 0.234$, $p < .01$, and this increases to $\beta = 0.270$, $p < .01$ in TOT models). For classroom quality and curriculum fidelity, we observed no treatment effects on the ECERS-3 scores and fidelity scores and a negative effect on SSTEW scores (see Nores et al., 2026 for additional analyses).

Results for the ECERS-3 quality measure (Models *b* through *e*) indicated that treatment assignment was not associated with significant improvements in classroom quality and neither were dosage of PD (whether synchronous or asynchronous) or platform engagement hours (*SmartTeach*). Models *f* through *i* in Table 4 show the association between treatment and the SSTEW measure of quality, while controlling for the different implementation variables of interest.[12] Synchronous PD did not have any effect on SSTEW scores, while asynchronous PD showed a positive and significant association with SSTEW scores ($\beta = 0.068$, $p < .05$), suggesting that more hours of self-directed online training are generally associated with higher classroom quality. However, interaction between ITT and online asynchronous dosage was negative and significant ($\beta = -0.060$, $p < .05$), indicating that this positive effect was reversed for teachers in the treatment group, although this tempers at higher dosages. This implies that while asynchronous dosage is generally supportive of higher quality, its alignment with the treatment content may have been mismatched. For platform engagement, the main effect on SSTEW is also significant. However, the interaction between engagement and ITT trends negative ($\beta = -0.086$, $p < .1$). In essence, high platform engagement may work against quality as measured by the SSTEW scale for the treatment group, suggesting potential over-engagement or misalignment with this observational measure.

Lastly, models *j* through *m* estimate the association between treatment and the curriculum fidelity measure and include the dosage and engagement variables. The intervention alone had no effect on curriculum fidelity nor was synchronous PD associated with fidelity. Asynchronous PD dosage was positive and significantly associated with fidelity ($\beta = 0.010$, $p < .05$), but the ITT*asynchronous dosage interaction was negative and significant ($\beta = -0.012$, $p < .05$), indicating that the intervention reduced the positive impact of asynchronous PD on fidelity. Platform engagement shows a similar pattern, with the main effect significantly associated with fidelity ($\beta = 0.009$, $p < .05$), but the ITT*asynchronous dosage interaction was negative and significant ($\beta = -0.008$, $p < .05$).

Analyses controlling for baseline quality , and analyses imputing the baseline quality for classrooms not observed at baseline are reported in Nores et al. (2026) and similarly align.

[11] Estimated as probits and reported in marginal probabilities. We do not estimate these with dosage or engagement as they are co-dependent variables.

[12] Estimations without the interactions are reported in Table A.14 for added reference.

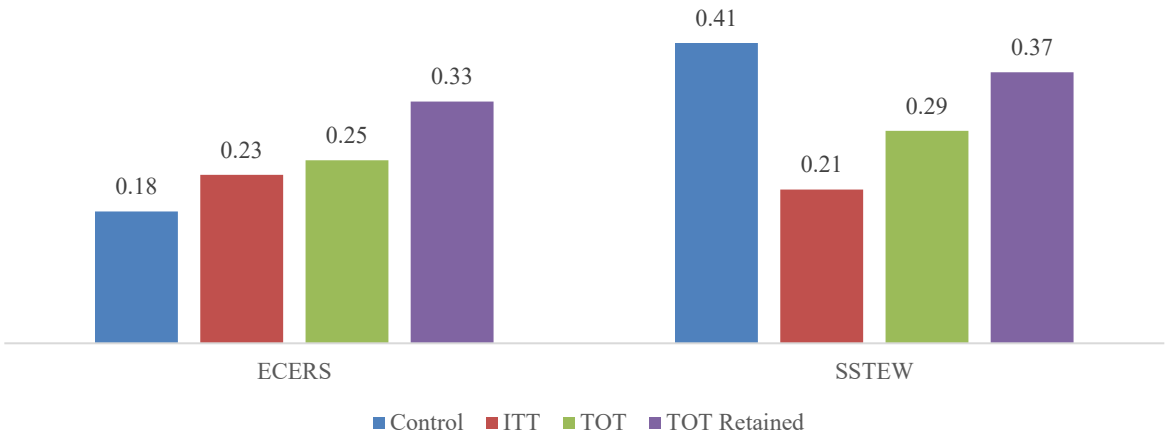
Table 4. ITT effects on primary outcomes (retention, classroom quality, and curriculum fidelity)

	Retention		ECERS-3			SSTEW				Fidelity Use			
Estimation	With ITT only	With ITT only	With Dosage Synch. PD	With Dosage Asynch. PD	With Platform Engage-ment	With ITT only	With Dosage Synch. PD	With Dosage Asynch. PD	With Platform Engage-ment	With ITT only	With Dosage Synch. PD	With Dosage Asynch. PD	With Platform Engage-ment
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)
ITT	0.234***	-0.127	0.164	0.143	-0.187	-0.392**	-0.667**	-0.119	0.280	-0.044	0.001	0.017	0.015
	(0.060)	(0.318)	(0.560)	(0.363)	(0.427)	(0.173)	(0.235)	(0.204)	(0.363)	(0.028)	(0.042)	(0.030)	(0.046)
Dosage[a]			-0.014	0.015			0.014	0.068***			-0.002	0.010***	
			(0.018)	(0.029)			(0.013)	(0.019)			(0.002)	(0.003)	
ITT*Dosage			-	-0.040			-	-0.060**				-0.012***	
				(0.030)				(0.023)				(0.003)	
Platform engagement[b]					-0.024				0.071***				0.009**
					(0.017)				(0.016)				(0.003)
ITT*Platform engagement					0.008				-0.086*				-0.008**
					(0.032)				(0.044)				(0.003)
Observations	125	124	124	124	124	125	125	125	124	123	123	123	123

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are retention (for all analyses except retention analyses), teacher demographics, years of experience and education, and district fixed effects; errors are clustered at the coach level. Robust standard errors in parentheses. [a]There are no dosage coefficients as dosage is zero for all the control group and is positive only for the treatment group, so the interaction captures this combined effect. [b]Engagement hours are included in a scale of weeks of engagement with 35 hours of engagement assumed in a week. Results are presented in standard deviations of the control group for all but retention.

Even though there are null or negative differences observed between treatment and control groups at post-test, analyses of growth in ECERS and SSTEW scores (Nores et al., 2026) show gains for all groups across both measures of quality (Figure 2). Moreover, growth in scores was higher for the treatment-on-treated group for teachers retained in classrooms.

Figure 2. Mean ECERS-3 and SSTEW growth in scores by Control vs. Treatment groups.



Estimations of effects on child outcomes

Table 5 reports treatment effects, standard errors, and statistical significance for the child estimations for the creativity and play measures, and for the executive function and socio-emotional measures. We report ITT estimates along with their standard errors (controlling for baseline outcomes for children). Program effects are reported as fractions of a standard deviation relative to the distribution of the control group's pretest outcomes for the corresponding measure.

Results in Table 5 show no statistically significant effects on any of the play, creativity, nor executive function outcomes externally assessed. For socio-emotional development, there are no statistically significant effects on the external measure used in the study (the SDQ). However, for the *GOLD* social-emotional domain there were positive and marginally significant effects for the study subsample ($\beta = 0.127$, $p < .1$) and for all children in the study classrooms ($\beta = 0.125$, $p < .1$). This significance is largely sustained after adjusting p-values for multiple hypothesis testing (Appendix E, Table E.2).[13]

[13] Estimations excluding the three classrooms for which we could not match any GOLD data (not-shown) are aligned in magnitude and significance for the external measures of children's creativity, play measures, executive functions, socio-emotional development, language, literacy and math.

Table 5. Standardized associations between ITT and children's creativity, play measures, executive functions, and socio-emotional development.

Variables	Creativity and Play		Executive Function		Socio-Emotional			
	TCAM	PIPPS	DCCS	PT	SDQ Behavioral Problems	SDQ Prosocial	GOLD Social Emotional	GOLD Social Emotional All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Pretest	0.068***	0.051***	0.054***	0.074***	0.115***	0.162***	0.008***	0.009***
	(0.012)	(0.003)	(0.010)	(0.013)	(0.006)	(0.028)	(0.001)	(0.001)
ITT	0.074	0.084	-0.107	0.163	-0.061	0.150	0.127*	0.125*
	(0.133)	(0.093)	(0.095)	(0.125)	(0.089)	(0.111)	(0.064)	(0.065)
Observations	389	302	418	410	302	302	397	1,331
R-squared	0.310	0.648	0.200	0.321	0.526	0.188	0.694	0.606

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, district fixed effects, and child demographics; errors are clustered at the school level. Robust standard errors in parentheses. Full estimations are included in Appendix Table B.1.

Table 6 shows the association between ITT and children's cognitive, language, literacy, and math outcomes. There were no statistically significant effects on *GOLD* for the cognitive domain. While there is a marginally significant negative effect on the external measure of receptive vocabulary (PPVT), the treatment showed a positive and significant effect on the *GOLD* language domain for the study's sub-sample (β =0.226, $p < .05$) and a marginal effect in estimations with all children the classrooms (β =0.192, $p < .1$). There was no statistically significant effect observed on children's *GOLD* literacy outcomes. For math there were no effects on the external measure (Woodcock-Johnson, Applied Problems subscale) but there was a positive and statistically significant effect on the *GOLD* math domain for the study's sub-sample (β =0.254, $p < .05$) and in estimations with all children in the study classrooms (β =0.211, $p < .05$). All of these effects remain aligned in significance after adjusting p-values for multiple hypothesis testing (Appendix E, Table E.3).[10]

There were no observed effects of the intervention on the external measures when assessed in Spanish for Spanish speakers (see Appendix Table B.5). We do not report *GOLD* Spanish language nor literacy outcomes because *GOLD* observations were predominantly collected only for the English language and literacy outcomes in all classrooms across the two districts.

Table 6. Standardized associations between ITT and children's cognitive outcomes.

Variables	Cognitive		Language			Literacy			Math		
	GOLD Cognitive	GOLD Cognitive All Children	PPVT	TSG Language	GOLD Language All Children	WJ-LW	GOLD Literacy	GOLD Literacy All Children	WJ-AP	GOLD Math	GOLD Math All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Pretest	0.010***	0.011***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.181***	0.013***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.002)	(0.001)	(0.017)	(0.001)	(0.001)
ITT	0.117	0.060	-0.143*	0.226**	0.192*	-0.052	0.053	0.052	-0.064	0.254**	0.211**
	(0.131)	(0.112)	(0.081)	(0.105)	(0.106)	(0.073)	(0.096)	(0.087)	(0.089)	(0.117)	(0.098)
Observations	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.635	0.610	0.718	0.730	0.663	0.726	0.734	0.693	0.651	0.727	0.651

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, district fixed effects, and child demographics; errors are clustered at the school level. Robust standard errors in parentheses. Full estimations are included in Appendix Table B.2.

We replicated these estimates for treatment-on-treated estimations (Tables B.3 and B.4), and results largely mirror the findings described above.

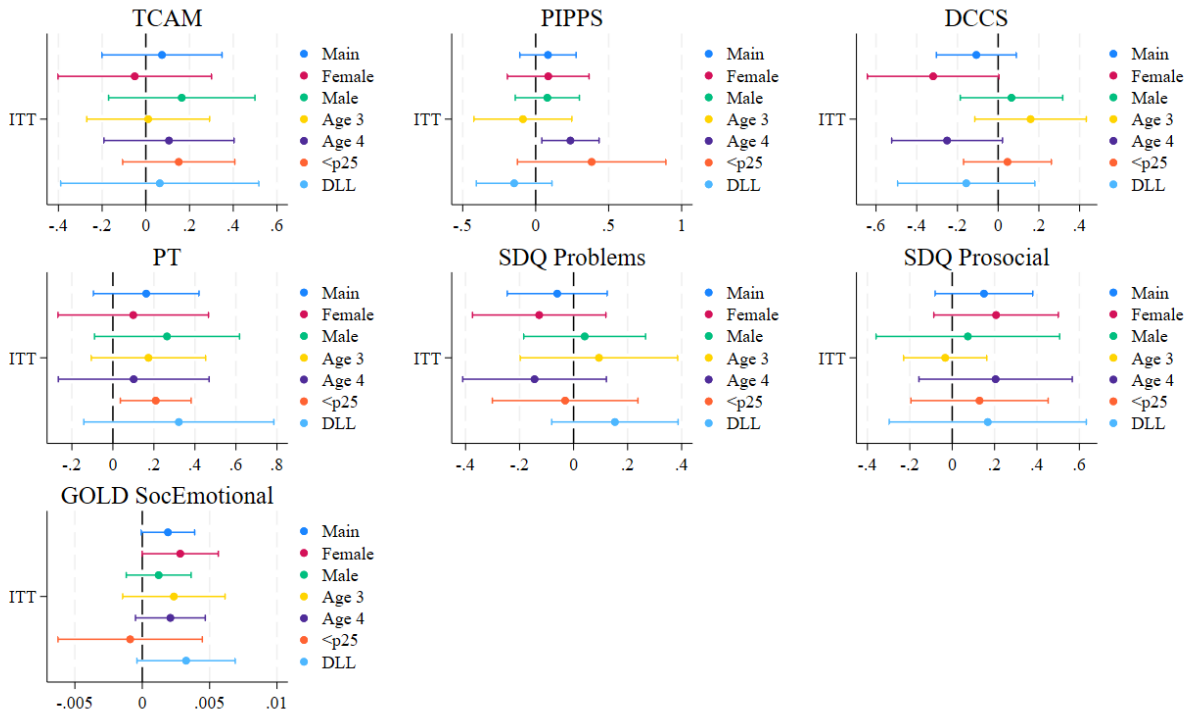
Heterogeneous effects on child outcomes

As noted earlier, we also estimated heterogeneous effects for the models in Tables 5 and 6 for key subgroups of children. More specifically, we estimate these models by gender, age at baseline (3 versus 4 years), dual-language learners, and low baseline development level (under the 25th percentile). Comparisons of ITT effects across outcomes and subgroups are depicted in Figures 3 and 4 (figures for TOT estimates are reported in Appendix C, Figures C.1 and C.2).

For children's creativity, play measures, executive functions, and socio-emotional development, some heterogeneous patterns emerged. For the play measure (PIPPS), positive effects were detected for older children (age 4) and children with low baseline levels on this measure. Findings for executive function differed across measures. For the DCCS negative significant estimates emerged for females and marginally for older children (age 4). For the PT effects were in the opposite direction; statistically significant positive effects emerge for children under the 25th percentile at baseline.

We also examined the heterogeneous effects for the SDQ and the *GOLD* social-emotional domain. No significant effects were observed on the SDQ Behavioral Problems and Prosocial scale. In contrast, for the *GOLD* social-emotional domain, positive and statistically significant effects emerged for females, and marginally significant effects emerged for older children (age 4), and dual-language learners. These effects were mostly sustained after adjusting for multiple hypothesis testing (Appendix E, Table E.6 for ITT and TOT estimations).

Figure 3. Standardized associations between ITT and children's creativity, play measures, executive functions, and socio-emotional development.

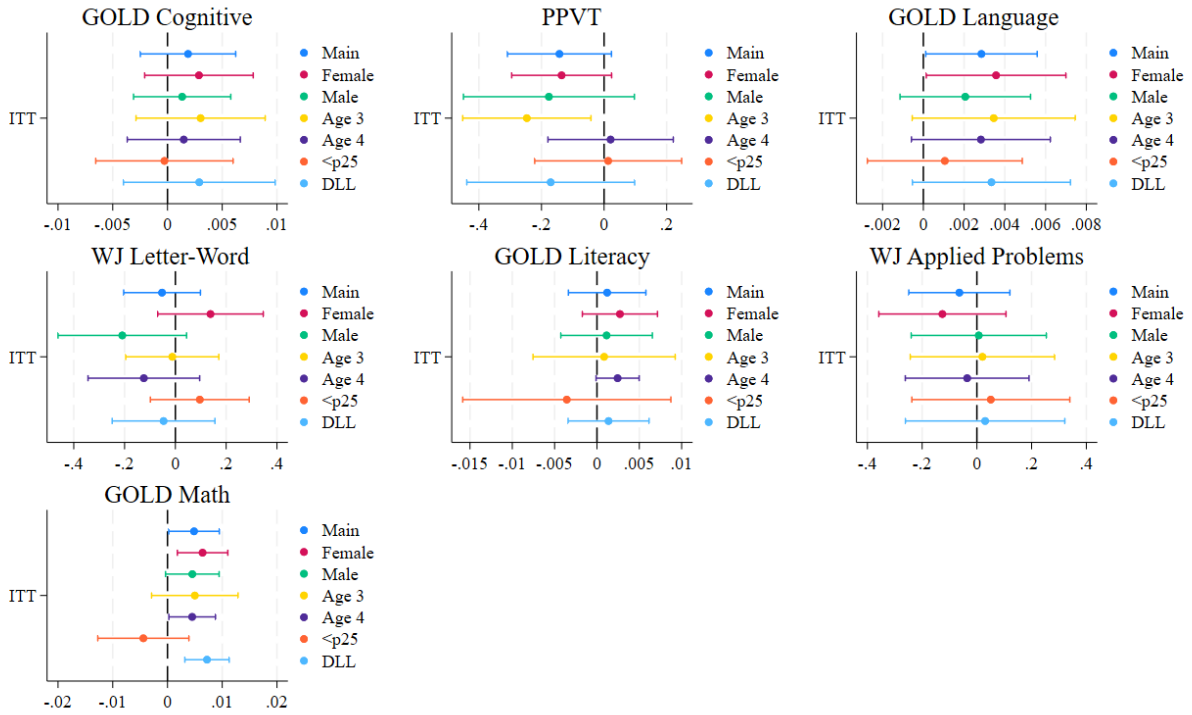


Note: Controls included are teacher demographics, experience and education, district fixed effects, and child demographics; errors are clustered at the School level. Figures present ITT mean effect sizes and the 95% confidence intervals. GOLD results are shown for the study sample. TOT estimations are included in Appendix C.1.

We also examined heterogeneous effects for children's cognitive outcomes, including cognition, receptive vocabulary, language, literacy, and mathematics, as shown in Figure 4. Overall, subgroup patterns for cognitive outcomes were mixed. Across most external measures (PPVT, WJ-LW, and WJ-AP), ITT estimates varied by subgroup, with generally limited evidence of systematic differential impacts across children's characteristics for external measures.

For the *GOLD* language measure, effects were positive and statistically significant for females and younger children (age 3) and marginally significant for males and older children. For the *GOLD* literacy measure, estimates were statistically significant and positive for older children. Lastly, for the *GOLD* math measure, estimates were statistically significant and positive for females, DLL children, and older children (age 4) and marginally significant for males. Most of these effects are sustained after adjusting for multiple hypothesis testing (adjusted p-values are reported in Appendix E, Table E.7) and stronger in TOT estimations.

Figure 4. Standardized associations between ITT and children's cognitive outcomes.



Note: Controls included are teacher demographics, experience and education, district fixed effects, and child demographics; errors are clustered at the School level. Figures present ITT mean effect sizes and the 95% confidence intervals. *GOLD* results are shown for the study sample. TOT estimations are included in Appendix C.2.

Pathways and Intervention dosage

To assess pathways that explain effects (or the lack of effects) on different domains, we explored the role of retention, intervention PD dosage, and *SmartTeach* platform engagement. To do this, we estimated the models above, but we now include in separate models, the teacher retention indicator, the two PD dosage variables and the platform engagement variable, independently.

Retention

Tables 7 and 8 show impacts on the different sets of outcomes, controlling for retention and its interaction with the ITT indicator. In Table 7, retention had a positive, marginally significant effect on the DCCS measure of executive function (estimation *e*; $\beta = 0.171$, $p < .1$), with children in classrooms with retained teachers more likely to evidence positive gains (not significant after accounting for multiple hypotheses, Appendix Table E.2). In addition, for treatment teachers, retention had a significant effect on the peer play scale (PIPPS; $\beta = 0.438$, $p < .05$)

With the inclusion of the retention variable, the ITT effects observed earlier for *GOLD* social-emotional development (Table 5) were strongly attenuated. Thus, teacher retention emerged as potentially associated with improvements in executive function (by one measure) and in social-emotional development. Retention did not significantly interact with ITT for the *GOLD* social-

emotional measure, indicating that retention does not moderate the treatment effect. Instead, these patterns are consistent with retention operating as a pathway (mediator) through which the intervention may influence children's socio-emotional development, particularly given the evidence of intervention impact on retention (satisfying a key precondition for mediation).

Table 7. Standardized associations between ITT and children's creativity, play measures, executive functions, and socio-emotional development, accounting for teacher retention.

	Creativity	Play Scale	Executive Functions		Socio-Emotional			
Variables	TCAM	PIPPS	DCCS	PT	SDQ Behavioral Problems	SDQ Prosocial	GOLD Social Emotional	GOLD Social Emotional Full Sample
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Pretest	0.069***	0.051***	0.055***	0.075***	0.115***	0.162***	0.008***	0.008***
	(0.012)	(0.003)	(0.010)	(0.013)	(0.007)	(0.028)	(0.001)	(0.001)
ITT	0.192	-0.265	0.062	0.122	0.039	-0.003	-0.040	0.043
	(0.215)	(0.196)	(0.171)	(0.197)	(0.221)	(0.431)	(0.198)	(0.165)
Retention	-0.139	-0.005	0.171*	0.174	-0.179	-0.207	0.103	0.106
	(0.159)	(0.113)	(0.098)	(0.122)	(0.118)	(0.176)	(0.113)	(0.100)
ITT * Retention	-0.121	0.438**	-0.259	0.022	-0.079	0.244	0.188	0.084
	(0.226)	(0.192)	(0.241)	(0.258)	(0.270)	(0.483)	(0.252)	(0.215)
Observations	389	302	418	410	302	302	397	1,331
R-squared	0.316	0.655	0.203	0.325	0.531	0.193	0.701	0.610

Note: ***p<0.01, **p<0.05, *p<0.1. Six models are reported. Controls included are teacher demographics, experience and education, district fixed effects, and child demographics; errors are clustered at the School level. Robust standard errors in parentheses. TOT estimations are included in Appendix Table D.1.

These same estimations are shown for cognitive outcomes in Table 8. In this set of estimations, outcomes showed varied patterns. With the introduction of retention, main effects for ITT attenuated to null, while main effects on the retention variable were positive but not significantly so. Importantly, ITT coefficients represent treatment effects in the absence of teacher retention, suggesting that the intervention alone would not be sufficient to produce cognitive gains when teachers were replaced (not retained).

In contrast, teacher retention showed sizable and marginally or statistically significant positive associations for the *GOLD* cognitive and literacy domains. More specifically, retention was positively associated with the *GOLD* cognitive domain for the full sample of children (estimation *b*; $\beta = 0.293$, $p < .05$), and marginally so with the *GOLD* literacy domain (estimation *h*; $\beta = 0.177$, $p < .1$) for the full sample of children in the study. In addition, the interaction term was positive and marginally significant for the *GOLD* literacy domain for the sub-study sample (estimation *g*; $\beta = 0.301$, $p < .1$), indicating that the intervention's effects on these outcomes were significantly stronger when teachers are retained. Together, these findings suggest that teacher retention plays an important moderating role at least for some domains, with intervention benefits evident in classrooms characterized by teacher stability.

These patterns were not observed for the external child assessments (e.g., PPVT, Woodcock-Johnson measures), a distinction we return to in the discussion section.

Table 8. Standardized associations between ITT and children's cognitive outcomes, accounting for teacher retention.

Variables	Cognitive			Language			Literacy			Emerging Math		
	GOLD Cognitive	GOLD Cognitive All Children	PPVT	GOLD Language	GOLD Language All Children	WJ-LW	GOLD Literacy	GOLD Literacy All Children	WJ-AP	GOLD Math	GOLD Math All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Pretest	0.010***	0.011***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.182***	0.014***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.001)	(0.001)	(0.017)	(0.001)	(0.001)
ITT	-0.107	-0.120	-0.163	-0.154	0.052	-0.202	-0.207	-0.124	-0.189	0.047	-0.020
	(0.251)	(0.207)	(0.103)	(0.213)	(0.163)	(0.132)	(0.186)	(0.105)	(0.145)	(0.186)	(0.144)
Retention	0.205	0.293**	-0.031	0.056	0.108	-0.084	0.075	0.177*	0.042	0.158	0.130
	(0.199)	(0.135)	(0.106)	(0.140)	(0.102)	(0.089)	(0.116)	(0.101)	(0.133)	(0.111)	(0.079)
ITT * Retention	0.223	0.162	0.035	0.457	0.153	0.218	0.301*	0.176	0.157	0.207	0.253
	(0.324)	(0.269)	(0.135)	(0.288)	(0.242)	(0.146)	(0.172)	(0.132)	(0.185)	(0.203)	(0.186)
Observations	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.646	0.626	0.718	0.739	0.666	0.727	0.741	0.702	0.652	0.733	0.657

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, district fixed effects, and child demographics; errors are clustered at the School level. Robust standard errors in parentheses. TOT estimations are included in Appendix Table D.2.

Dosage and Engagement

Next, we looked more closely at the process and output variables for teachers. Different patterns emerged for effects on children's outcomes relative to the amount of synchronous PD (only for treatment teachers), the amount of asynchronous PD (for all teachers), and the amount of *SmartTeach* platform engagement for treatment teachers (Tables 9 and 10).[14] All these estimations controlled for retention, which remained statistically significantly associated with various outcomes (see Appendix Tables D.3 and D.4 for estimations without retention).

For the estimations with synchronous dosage (top panel on Tables 9 and 10), it is important to note that since only the treatment teachers received any, the coefficient on synchronous dosage should be interpreted as the interaction effect. We found notable patterns for synchronous dosage on child outcomes. Synchronous PD hours were significantly associated with a higher externally assessed literacy outcomes ($\beta = 0.009$, $p < .01$) and marginally associated with a higher externally assessed math outcomes ($\beta = 0.008$, $p < .1$). In addition, for the *GOLD* measures, synchronous PD dosage was significantly positively associated with the social-emotional domain for the full sample ($\beta = 0.007$, $p < .05$), although marginally negative for the external measure of problem behaviors ($\beta = 0.008$, $p < .1$), with the later strongly reverting for retained teachers ($\beta = -0.204$, $p < .05$). We also found positive and statistically significant effects on the cognitive domain ($\beta = 0.012$, $p < .05$ for the sub-sample and $\beta = 0.010$, $p < .05$ for the full sample), for the language domain ($\beta = 0.016$, $p < .01$ for the sub-sample and $\beta = 0.012$, $p < .01$ for the full sample), the literacy domain ($\beta = 0.010$, $p < .01$ for the sub-sample and $\beta = 0.011$, $p < .01$ for the full sample), as well as marginally significant effects for the math domain for the full sample ($\beta = 0.009$, $p < .1$). These effects translate into trending increases of 0.21 standard deviations higher (.009* an average of 23.5 synchronous PD hours) for WJ literacy and 0.18 standard deviations

[14] Due to missing dosage and engagement data on the classrooms that shifted from the treatment to the control group, ITT and TOT estimations do not differ.

(.008* an average of 23.5 synchronous PD hours) for WJ math, and even larger effects across the *GOLD* measures.

Asynchronous PD (on-demand course) hours were positively associated with higher creativity scores on the TCAM ($\beta = 0.027$, $p < .01$). For the TCAM the ITT interaction term is significant and reverted in sign, indicating asynchronous hours show diminishing returns for the treatment group. Retention remains marginally associated with the external PT measure of executive function and with reductions in problem behaviors.

The third panel of results in Table 9 and Table 10 include Teachers' *SmartTeach* platform engagement. More platform engagement hours were marginally positively associated with DCCS scores ($\beta = 0.024$, $p < .1$). The engagement by ITT interaction was also positively associated with *GOLD* cognitive, literacy and math outcomes ($p < .1$ for math in the sub-sample, otherwise $p<.01$). These positive interaction effects suggest that platform engagement moderated the intervention effect consistently across this set of outcomes.

Table 9. Standardized associations between ITT and children's creativity, play measures, executive functions, and socio-emotional development, accounting for PD dosage and platform engagement, controlling for retention.

Variables	Creativity	Play Scale	Executive Function		Socio-Emotional			
	TCAM	PIPPS	DCCS	PT	SDQ Behavioral Problems	SDQ Prosocial	GOLD Socio-Emotional	GOLD Socio-Emotional All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Estimations with Synchronous PD dosage hours								
Pretest	0.068***	0.051***	0.054***	0.076***	0.114***	0.162***	0.008***	0.009***
	(0.012)	(0.003)	(0.010)	(0.013)	(0.007)	(0.028)	(0.001)	(0.001)
ITT	-0.079	0.107	-0.059	0.022	-0.198	0.243	-0.009	-0.034
	(0.180)	(0.154)	(0.139)	(0.193)	(0.119)	(0.204)	(0.070)	(0.069)
Dosage: Synch. PD hrs.	0.009	-0.002	-0.003	0.006	0.008*	-0.003	0.005	0.007**
	(0.006)	(0.004)	(0.005)	(0.008)	(0.004)	(0.006)	(0.003)	(0.003)
Retention	-0.190	0.099	0.093	0.172*	-0.204*	-0.148	0.140	0.120
	(0.129)	(0.106)	(0.082)	(0.100)	(0.112)	(0.195)	(0.089)	(0.078)
Obs.	389	302	418	410	302	302	397	1,318
R-squared	0.308	0.646	0.189	0.312	0.523	0.189	0.672	0.597
Estimations with Asynchronous PD dosage hours								
Pretest	0.072***	0.052***	0.054***	0.075***	0.116***	0.161***	0.008***	0.009***
	(0.013)	(0.003)	(0.010)	(0.013)	(0.007)	(0.027)	(0.001)	(0.001)
ITT	0.206	0.032	-0.174*	0.150	0.016	0.105	0.082	0.056
	(0.127)	(0.128)	(0.097)	(0.161)	(0.090)	(0.163)	(0.068)	(0.073)
Dosage: Asynch. PD hrs.	0.027***	-0.012	-0.013	0.007	0.005	-0.009	-0.005	-0.005
	(0.007)	(0.011)	(0.015)	(0.014)	(0.007)	(0.011)	(0.006)	(0.006)
ITT*Dosage Asynch. PD hrs.	-0.024***	0.010	0.012	-0.004	-0.007	0.012	0.004	0.008
	(0.006)	(0.013)	(0.014)	(0.014)	(0.007)	(0.011)	(0.006)	(0.006)
Retention	-0.169	0.098	0.085	0.183*	-0.203*	-0.136	0.141	0.129
	(0.127)	(0.114)	(0.081)	(0.103)	(0.117)	(0.204)	(0.085)	(0.081)
Obs.	389	302	418	410	302	302	397	1,331
R-squared	0.329	0.652	0.204	0.326	0.532	0.194	0.700	0.611
Estimations with SmartTeach Platform Engagement hours								
Pretest	0.071***	0.052***	0.054***	0.075***	0.114***	0.162***	0.008***	0.008***
	(0.013)	(0.003)	(0.010)	(0.013)	(0.007)	(0.028)	(0.001)	(0.001)
ITT	-0.095	0.220	-0.062	0.138	-0.168	0.288	0.019	-0.028
	(0.222)	(0.166)	(0.162)	(0.242)	(0.218)	(0.328)	(0.117)	(0.107)
Platform engagement	-0.016	0.004	0.024*	0.026	0.004	0.015	0.005	0.003
	(0.014)	(0.015)	(0.013)	(0.021)	(0.017)	(0.021)	(0.012)	(0.008)
ITT*Platform engagement	0.025	-0.016	-0.011	-0.005	0.015	-0.013	0.008	0.014
	(0.018)	(0.014)	(0.017)	(0.024)	(0.019)	(0.031)	(0.013)	(0.010)
Retention	-0.143	0.092	0.018	0.087	-0.230*	-0.214	0.126	0.107
	(0.121)	(0.119)	(0.099)	(0.116)	(0.124)	(0.204)	(0.086)	(0.072)
Obs.	386	302	414	406	302	302	397	1,331
R-squared	0.320	0.651	0.205	0.331	0.534	0.193	0.701	0.614

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in the GOLD estimations with all children. Engagement is presented in a scale of weeks of engagement throughout the three-year period at 35 hours of engagement per week.

Table 10. Standardized associations between ITT and children's cognitive outcomes, accounting for PD dosage and platform engagement, controlling for retention.

Variables	Cognitive		Language			Literacy			Emerging Math		
	GOLD Cognitive	GOLD Cognitive All Children	PPVT	GOLD Language	GOLD Language All Children	WJ-LW	GOLD Literacy	GOLD Literacy All Children	WJ-AP	GOLD Math	GOLD Math All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Estimations with Synchronous PD dosage hours											
Pretest	0.010***	0.011***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.180***	0.013***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.001)	(0.001)	(0.016)	(0.001)	(0.001)
ITT	-0.189*	-0.220**	-0.201**	-0.144*	-0.089	-0.227**	-0.194	-0.245***	-0.241*	0.060	-0.022
	(0.095)	(0.096)	(0.096)	(0.082)	(0.100)	(0.099)	(0.122)	(0.082)	(0.118)	(0.121)	(0.101)
Dosage: Synch. PD	0.012**	0.010**	0.003	0.016***	0.012***	0.009***	0.010***	0.011***	0.008*	0.007	0.009*
	(0.006)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)
Retention	0.247	0.325***	-0.025	0.148	0.140	-0.029	0.136	0.215**	0.079	0.191**	0.177**
	(0.153)	(0.109)	(0.080)	(0.114)	(0.089)	(0.079)	(0.109)	(0.094)	(0.091)	(0.089)	(0.077)
Obs.	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.653	0.632	0.719	0.747	0.674	0.729	0.745	0.710	0.655	0.734	0.660
Estimations with Asynchronous PD dosage hours											
Pretest	0.010***	0.010***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.183***	0.014***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.001)	(0.001)	(0.016)	(0.001)	(0.001)
ITT	0.042	-0.010	-0.149	0.140	0.097	-0.052	0.023	-0.009	-0.144	0.220*	0.162
	(0.129)	(0.112)	(0.089)	(0.119)	(0.126)	(0.075)	(0.109)	(0.099)	(0.097)	(0.117)	(0.111)
Dosage: Asynch. PD	0.001	0.007	-0.000	0.002	0.001	0.001	0.006	0.008	-0.007	0.012	0.015
	(0.010)	(0.009)	(0.008)	(0.007)	(0.009)	(0.005)	(0.008)	(0.008)	(0.013)	(0.008)	(0.009)
ITT*Dosage Asynch. PD	0.001	-0.003	0.001	0.006	0.009	-0.000	-0.003	-0.002	0.011	-0.006	-0.005
	(0.011)	(0.010)	(0.008)	(0.008)	(0.009)	(0.007)	(0.008)	(0.008)	(0.014)	(0.009)	(0.009)
Retention	0.265*	0.352***	-0.020	0.182	0.173*	-0.014	0.157	0.241**	0.093	0.226**	0.227**
	(0.153)	(0.110)	(0.080)	(0.125)	(0.097)	(0.084)	(0.106)	(0.095)	(0.088)	(0.098)	(0.089)
Obs.	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.645	0.628	0.718	0.736	0.672	0.727	0.738	0.704	0.653	0.737	0.666
Estimations with SmartTeach Platform Engagement hours											
Pretest	0.010***	0.010***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.184***	0.013***	0.012***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.016)	(0.002)	(0.001)	(0.017)	(0.001)	(0.001)
ITT	-0.256	-0.283**	-0.066	0.035	0.084	-0.146	-0.249	-0.222*	-0.226*	-0.041	-0.159
	(0.155)	(0.116)	(0.107)	(0.168)	(0.165)	(0.153)	(0.146)	(0.108)	(0.132)	(0.141)	(0.114)
Platform engagement	0.003	-0.006	0.009	-0.005	0.001	-0.013	-0.007	-0.003	-0.019	-0.010	-0.011
	(0.016)	(0.012)	(0.012)	(0.015)	(0.011)	(0.018)	(0.012)	(0.007)	(0.013)	(0.013)	(0.011)
ITT*Platform engagement	0.032**	0.031***	-0.010	0.018	0.009	0.013	0.029**	0.025**	0.020	0.028*	0.038**
	(0.014)	(0.010)	(0.013)	(0.016)	(0.015)	(0.015)	(0.012)	(0.009)	(0.014)	(0.014)	(0.014)
Retention	0.243	0.349***	-0.040	0.179	0.139	0.025	0.169	0.223**	0.145	0.234***	0.220**
	(0.150)	(0.114)	(0.079)	(0.127)	(0.104)	(0.140)	(0.117)	(0.103)	(0.105)	(0.083)	(0.083)
Obs.	378	1,289	438	395	1,340	436	356	1,194	429	367	1,252
R-squared	0.658	0.633	0.716	0.735	0.667	0.725	0.745	0.707	0.651	0.737	0.665

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in the GOLD estimations with all children. Engagement is presented in a scale of weeks of engagement throughout the three-year period at 35 hours of engagement per week.

Study Children in Context

As mentioned earlier, the study also incorporated a "synthetic" group that sampled classrooms and children across other districts in New Jersey, implementing the NJ state preschool program. Children in the study (treatment and control) started the school year underperforming relative to their peers across the state. Simple comparisons described on page 18 showed gains for children in the treatment group. They gained more than "synthetic" children on some measures, and gained less than them on others, but no systematic statistically significant differences emerged

across groups (Appendix Table A.12) with the exception of DCCS, for which children in the synthetic sample evidenced larger and statistically significant gains. In addition to assessing simple (uncontrolled) differences in gains, we run the same models above, including children in the treatment, control and synthetic samples (Table E.1). After controlling for pretest scores and other covariates, the estimations show no differences between the three groups for child post-test scores of creativity, play, behavioral problems, vocabulary, literacy, or math for all children. Likewise, no differences in externally assessed measures of vocabulary, literacy, or math emerged for Spanish-speakers.

Treatment children performed statistically significantly lower at post-test than their peers only on the DCCS measure of executive function ($\beta = -0.352$, $p < .05$) and marginally lower on the literacy measure for Spanish-speakers ($\beta = -0.519$, $p < .1$), while control children did so on the PT measure of executive function ($\beta = -0.324$, $p < .05$). In contrast, treatment and control children performed better than their "synthetic" peers on the prosocial measure (Treatment: $\beta = 0.322$, $p < .05$; Control: $\beta = 0.246$, $p < .1$). In essence, despite potential variation across all other districts in curriculum, leadership, supports and coaching, and quality, children randomized to either treatment or control performed at par at post-test with peers across the state after controlling for differences in child and classroom characteristics.

Teachers and Coaches Perspectives

We asked coaches to provide background information, coaching strategies, and perceived successes over the previous three years (Harmeyer & Nores, 2026). Ten of the 12 coaches who were working with teachers in the sample during year 3 completed the survey.[15] In general, coaches reported they were trained on observation tools to use with teachers, including having knowledge of and training in the ECERS-3. They reported engaging in different coaching practices frequently, with the most frequent strategy being sharing resources with teachers (100% of coaches identified doing this at least 3–4 times a month). Interestingly, 90% of coaches reported that they worked as an assistant in a classroom at least 3–4 times per month (and 40% reported doing this more than 4 times per month). It is feasible that spending significant amounts of time as substitute teachers across classrooms impacted the coaches' ability to provide direct and effective coaching support to teachers.

In asking the coaches to reflect on what they had done successfully as a coach, the majority of coaches in the treatment group (80%) mentioned that fostering social-emotional development and positive interactions in the classroom was most successful. In addition, when asked their perception on how the teachers they support had changed in their implementation of the program's curriculum, 80% mentioned that their teachers were using the curriculum more consistently and/or with fidelity. As one coach noted: "*Moderately to implementing and understand how to use it responsively, intentionally, and with fidelity*." This perception is aligned with the growth in fidelity over the years, although the analyses point to no differences between treatment teachers and their control peers.[16]

[15] In one district, two additional coaches were added (one in year two, one in year three) as co-coaches at three schools that included teachers in the sample. One of these coaches received incentives as part of the intervention; one coached teachers in the control group. Their surveys are included in the analysis. One preschool intervention and referral specialist (PIRS) also completed the survey. That data has been excluded from this analysis.

[16] It is important to note that baseline fidelity measurements took place after the intervention supports had already begun, and the initial coaching sessions focused exclusively on curriculum fidelity.

Teachers in our sample reported that they were receiving different amounts of professional development, depending on whether they were in the treatment group or not. As expected, teachers in the treatment group were significantly more likely to report receiving monthly trainings, underscoring the continuity of support associated with the intervention. Importantly, qualitative and quantitative feedback from the teachers indicated they found this professional development useful, with 88% of teachers who took place in a learning opportunity with Teaching Strategies (regardless of group) agreeing or strongly agreeing that they felt motivated to participate in these opportunities, and 93% of teachers agreeing or strongly agreeing that the PD provided knowledge they will continue to use in the future (Appendix F, Tables F.1 and F.2).

As with teachers in many contexts, the biggest challenge hindering participation in PD with Teaching Strategies reported by the teachers in our sample was time and the load of expectations they are required to fulfill in their role (Appendix F Tables F.3 and F.4). A total of 38% of teachers agreed or strongly agreed that insufficient time to plan and meet with coaches hindered their participation in learning experiences from Teaching Strategies, and 44% agreed or strongly agreed that the load of curricular and assessment requirements they are expected to fulfill was a challenge. One teacher noted they had *"no time during the school day to fully study and plan implementation of TS suggestions. I often take huge amounts of work home to complete."* A lack of time to participate was a greater hindrance to the treatment group (43% agreeing/strongly agreeing that lack of time was a challenge), compared to the control group (30% agreeing or strongly agreeing this was the case). It appears that the additional learning opportunities, while perceived as useful and motivating for teachers, also presented implementation challenges for teachers. At the same time, across all responses, 31% indicated no challenges in participating in the learning opportunities, with the difference between the treatment and control group emerging in the opposite direction (27% of teachers in the treatment group noted no challenges compared to 16% in the control group). The challenges seem to therefore emerge not in participating in PD, but rather in planning and implementing what they learn.

We also collected teachers' perceptions of positive and negative experiences in their professional environment using the Maslach Burnout Inventory for Education (MBI-ES; Maslach et al., 1997). Teachers reported on feelings of accomplishment in their work, along with emotional exhaustion and depersonalization (Appendix E, Tables E.5, E.6, and E.7). The full sample of teachers reported feeling competence and achievement often in their work, with most components (such as dealing effectively with the problems of students and dealing with emotional problems calmly) occurring on average at least a few times per week for the whole sample. The treatment group scored significantly higher on this subscale ($p < .05$) than the control group, more frequently reporting experiences like feeling energetic and feeling exhilarated after working closely with students than teachers in the control group. Teachers in the treatment group were also significantly less likely to report frequent experiences of feeling at the end of their rope or feeling fatigued than teachers in the control group. A perception of a lack of coaching support came up more frequently for the control group than for the treatment group in qualitative feedback, with one control teacher noting: "*lack of coaching/mentorship within my first 3 years and the increase of challenging behaviors in the classroom"* as a challenge.

Summary & Discussion

This clustered randomized controlled trial evaluated the impact of the fully digital Teaching Strategies ecosystem, supported by aligned professional development and coaching, within New Jersey's high-quality universal pre-k program. The professional development effort included opportunities to learn about curriculum, assessment, family engagement, and reporting supported by the ecosystem. Conducted across two low-income districts, including a third "synthetic" comparison group, the study focused on assessing impact on classroom quality and child outcomes, with a look at the role of retention, ecosystem engagement, and curriculum fidelity as the measurable pathways. In addition to assessing the impact of the intervention on externally assessed child development measures it also examined curriculum-aligned child outcomes. Finally, it examined teachers' and coaches' perceptions of the professional development and learning experiences.

Across pretests, there are no major differences between treatment and control groups, which suggests the randomization was balanced across children as well as across teachers and classrooms. Treatment teachers attended synchronous, virtually delivered live professional development sessions and completed asynchronous, on-demand professional development courses, as well as engaged with the *SmartTeach* online platform significantly more than control group teachers. Synchronous PD was offered to the teachers from the curriculum developers in a planned sequence, and attendance was strongly encouraged but not mandatory; whereas asynchronous PD course completion and platform engagement were driven by the teachers themselves (sometimes, prompted by district requirements), regardless of randomization. Overall, the study found significant effects on teacher retention but no overall positive impacts of the intervention on classroom quality or on curriculum fidelity. Specifically, the intervention increased the likelihood that teachers stayed in their jobs over the three-year study period by 23 percentage points.

Although both the treatment and control groups showed some improvements in classroom quality, as measured by SSTEW and ECERS-3, the intervention was not associated with quality improvements in the treatment group; rather, results for these measures showed null or negative patterns. PD dosage and platform engagement were the key implementation variables monitored by the study, with synchronous PD being the variable manipulated by the intervention effort. The results show differing patterns for synchronous PD, asynchronous PD, and platform engagement. These trends suggest possible misalignment between "synchronous" developer-planned training, "asynchronous" self-paced and self-selected training, and platform engagement, which may have overburdened teachers. Ultimately, improved alignment between these components may be critical, with synchronous training not only providing child development content but also carefully guiding the effective use of the ecosystem tools and with aligned asynchronous training seamlessly connecting components and content to curriculum fidelity and child outcomes.

The absence of ITT effects on externally assessed child measures aligns with prior large-scale curriculum evaluations that have shown limited short-term impacts on standardized or external measures. In this study, children in both treatment and control classrooms demonstrated developmental gains over the school year comparable to those observed across other districts statewide, but the intervention did not result in additional gains detectable by the external measures used. However, teacher retention, which was directly influenced by the intervention, was associated with better child outcomes on one measure of executive function. This pattern, as

well as outcomes across most domains measured with *GOLD*, suggests retention is functioning as a pathway through which the intervention influenced children's development. This aligns with research (Choi et al., 2018; Markowitz, 2024; Tran & Winsler, 2011) showing notable associations between teacher stability and children's development.

In contrast, children in the treatment group showed marginally or significantly positive gains in *GOLD* outcomes across three of five domains. This contrast between null (or marginally negative) effects on external assessments and positive effects on *GOLD* domains warrants careful consideration. *GOLD* is embedded within daily instructional practice and is closely aligned with *The Creative Curriculum*, which may increase its sensitivity to changes in teacher practices supported by the ecosystem. The moderate to strong correlations between *GOLD* and external assessments at each time point, coupled with weak correlations in growth, suggest that while the measures capture similar developmental levels at each point in time, they differ in their sensitivity to short-term change. These results align with those of Lipsey et al. (2009) as they considered that the lack of observed curriculum differences on traditionally used measures (some being earlier versions of the ones used in this study) may reflect limitations in the measures' sensitivity or construct coverage rather than an absence of meaningful curriculum effects. Specific to this study, some of the external assessments used exhibited floor effects for our sample of children, even at post-test. We agree with Lipsey et al.'s assessment that this aspect warrants further attention in curriculum research.

Findings emphasize that professional development—in this case, synchronous, virtual, live, and sustained—can significantly boost teacher retention, an important structural feature of quality, which ultimately sustains investments in PD and is supportive of children's development. However, curriculum fidelity and classroom quality outcomes did not show consistent improvements from the intervention as measured in this evaluation, possibly due to insufficient dosage or limited alignment between the PD modalities, teachers' self-guided engagement with the ecosystem, and the instructional practices in their classrooms. While child outcomes did not improve meaningfully across all domains, the role of teacher retention as a potential mediator for executive function and language and literacy outcomes offers a promising direction. Future studies should explore long-term effects of stable staffing across years for children, refine or provide guidance on engagement strategies, and explore alignment between synchronous and asynchronous PD to consistently focus on instructional strategies that can impact quality and support children's outcomes across all domains.

The investment in teachers through additional resources, including time and relationships, emerges as a significant contributor to teachers staying in the profession. The effects on retention, are aligned with what the research has shown on effective tutoring, coaching, and adult learning. Berry et al. (2021) highlight professional learning opportunities as a key factor influencing retention, and other research has shown that the presence of a curricular coach is associated with a substantial reduction in early-career teacher turnover (De Jong & Campioli, 2018). Positive professional relationships are an important reason teachers may stay engaged with the profession at a time when teacher retention remains a persistent problem (Waddell, 2010). Synchronous but virtual PD, therefore, may continue to be a key pathway towards teacher stability and stronger child outcomes in an increasingly digital world. Schaak et al. (2022) highlight that teachers who tend to stay in their job place greater value on professional development opportunities. Luesse et al. (2022) emphasize how such effects on retention from

PD may be aligned with similar impacts on excitement with the profession and perceived self-efficacy; aspects that are aligned with the impact coaches perceived on their teachers.

The null or negative intervention effects on quality and fidelity do not align with the impacts on children. This could be partially driven by the timing of the study and/or misalignment between the intervention and the observational tools used to assess quality. The study was planned for the year right after schools across New Jersey were back to in-person schooling, after almost a year of remote schooling for many. The pandemic's disruption of early childhood programs had significant impacts on enrollment, attendance, and staff stability, and NJ was no exception (Education Law Center, 2021; Hernandez & McElrath, 2023; Nores & Harmeyer, 2021; Weiland et al., 2024; Weisenfeld, 2021). As the 2021–2022 school year started, preschool districts in NJ continued to experience lags in enrollment, which mirrored national trends (Friedman-Krauss et al., 2023). In addition, in the fall and winter of that school year, the Omicron Covid-19 variant reduced attendance (for children and teachers) across education systems, including in New Jersey.[17] This aligned with the timing of our study, requiring extensive efforts from the professional development team to effectively engage teachers in the first disrupted year. While in years two and three, the Covid-19 pandemic subsided, the impacts on enrollments and attendance have remained (Education Law Center, 2025), with the districts in the study reporting issues through year three.

In addition, replacements and retirements of coaches and teachers in the districts had implications for the effective intervention dosage, as any new teacher was incorporated into the study but would have missed a portion of the synchronous PD (and have had less time to engage in the asynchronous PD and with the *SmartTeach* platform), which attenuates impact. Lastly, the two districts enrolled in the study show lower levels of child development at preschool entry compared to peers across the state, emphasizing the low-resource background of families and lack of developmental experiences and opportunities for children in the study, which may have been exacerbated by the pandemic. These contextual variables may have impeded effective uptake and/or implementation of some quality aspects. Despite this, the children in the two participating districts had similar child gains by end-line as their peers across 21 other districts in NJ implementing the state's preschool program.

The varied results on children's outcomes highlight the difficulty of improving programs in ways that have a consistent, positive impact across all children's developmental domains simultaneously. This may relate to specific curricular or PD interventions focusing on a subset of domains (e.g., executive functions or STEM) showing stronger child results in those specific domains, but not necessarily across all domains (Brunske et al., 2020; Jensen et al., 2017). To benefit children directly, teachers may need support that tightly connects to what they do in their classrooms, and child-level information across all areas of development. Additionally, as the field adopts more digital training and coaching tools, careful attention to their integration, alignment, and contextualization within early childhood settings will be essential to maximize their impact. Ultimately, the impact on effective teaching and learning strategies will be

[17] Bever, L. (2022, January 7). *Canceled buses and a superintendent in the classroom: How Omicron has thrown schools into crisis*. *The Washington Post*. https://www.washingtonpost.com/education/2022/01/07/student-teacher-absences-omicron/; Eyewitness News (2021, December 22) *NJ sets daily record with nearly 10,000 new COVID cases, some schools go remote.* Author. https://abc7ny.com/nj-covid-omicron-new-jersey-in-gov-phil-murphy/11370028/

determined by the opportunities teachers have to integrate their learning into their everyday classroom experiences (Luesse et al., 2022).

Overall, results indicate positive impacts on teacher retention and in some areas of children's learning and development, with the latter varying by domain, subgroup, and type of measure. Impacts on teacher retention appear to have been an important contributor to the improved child outcomes detected. Null or even negative effects were found on the measures of classroom quality included in this study. Findings should be interpreted within the context of a relatively high-quality, well-resourced, mixed-delivery preschool system serving children in communities with high concentrations of poverty, a modest treatment-control contrast, and pandemic-related disruptions early in the study period.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice Hall.

Barker, J. E., Semenov, A. D., Michaelson, L., Provan, L. S., Snyder, H. R., & Munakata, Y. (2014). Less-structured time in children's daily lives predicts self-directed executive functioning. *Frontiers in Psychology*, 5, 593.

Barnett, W. S., & Frede, E.C. (2017). Long-term effects of a system of high-quality universal preschool education in the United States. In H.-P. Blossfeld, N. Kulic, J. Skopek, & M. Triventi (Eds.), *Childcare, early education and social inequality: An international perspective*. Edward Elgar Publishing.

Barnett, W. S., & Jung, K. (2021). Effects of New Jersey's Abbott preschool program on children's achievement, grade retention, and special education through tenth grade. *Early Childhood Research Quarterly*, 56, 248–259.

Barnett, W. S., Jung, K., Friedman-Krauss, A., Frede, E. C., Nores, M., Hustedt, J. T., Howes, C., & Daniel-Echols, M. (2018). State prekindergarten effects on early learning at kindergarten entry: An analysis of eight state programs. *AERA open*, *4*(2).

Barnett, W. S., Jung, K., Youn, M., & Frede, E. C. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. National Institute for Early Education Research.

Berry, B., Bastian, K. C., Darling-Hammond, L., & Kini, T. (2021). The importance of teaching and learning conditions: influences on teacher retention and school performance in North Carolina. Research Brief. Learning Policy Institute.

Blair, C., & Raver, C. C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLOS ONE*, 9(11), e112393.

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American Psychologist*, *34*(10), 844.

Bronfenbrenner, U., & Morris, P. A. (2006). The Bioecological Model of Human Development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed., pp. 793–828). John Wiley & Sons, Inc.

Brunsek, A., Perlman, M., McMullen, E., Falenchuk, O., Fletcher, B., Nocita, G., Kamkar, N., & Shah, P. S. (2020). A meta-analysis and systematic review of the associations between professional development of early childhood educators and children's outcomes. *Early Childhood Research Quarterly*, *53*, 217–248.

Chambers, B., Cheung, A. C., & Slavin, R. E. (2016). Literacy and language outcomes of comprehensive and developmental-constructivist approaches to early childhood education: A systematic review. *Educational Research Review*, 18, 88–111.

Choi, J. Y., Horm, D., Jeon, S., & Ryu, D. (2019). Do stability of care and teacher–child interaction quality predict child outcomes in early head start?. *Early Education and Development*, *30*(3), 337–356.

Clarke, D., Romano, J. P., & Wolf, M. (2020). The Romano–Wolf multiple-hypothesis correction in Stata. *The Stata Journal*, *20*(4), 812-843.

Davidson, M. R., Fields, M. K., & Yang, J.(2009).A randomized trial study of a preschool literacy curriculum: The importance of implementation. *Journal of Research on Educational Effectiveness*, 2(3), 177–208.

De Jong, D., & Campoli, A. (2018). Curricular coaches' impact on retention for early-career elementary teachers in the USA: Implications for urban schools. *International Journal of Mentoring and Coaching in Education, 7(2),* 191–200.

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental Psychobiology*, 29(4), 315–334.

Dunn L, Lugo S., Padilla R., & Dunn, L. (1986). *Test de Vocabulario en Imágenes Peabody*. AGS.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test*. Pearson.

Early, D., Sideris, J., Neitzel, J., LaForett, D., & Nehler, C. (2018). Factor structure and validity of the Early Childhood Environment Rating Scale – Third Edition (ECERS-3). *Early Childhood Research Quarterly*, 44, 242–256.

Education Law Center. (2021, April 26). *ELC analysis shows a sharp decline in preschool enrollments caused by the COVID-19 pandemic*. Author. https://edlawcenter.org/elc-analysis-shows-sharp-decline-in-preschool-enrollments-caused-by-the-covid-19-pandemic/

Education Law Center. (2025, February 12). *Just Released ELC Report Uncovers Alarming Trends in Abbott Preschool Enrollment*. Author. https://edlawcenter.org/just-released-elc-report-uncovers-alarming-trends-in-abbott-preschool-enrollment/

Engel, M., Attaway, D. S., & Denker, H. (2025). Curricula in early childhood education preschool settings. In L. Cohen-Vogel, P. Youngs, & J. Scott (Eds.), *Handbook of education policy research* (2nd ed.). AERA.

Fantuzzo, J., & McWayne, C. (2002). The relationship between peer-play interactions in the family context and dimensions of school readiness for low-income preschool children. *Journal of Educational Psychology*, 94(1), 79.

File, N. (2012). The relationship between child development and early childhood curriculum. In J. J. Mueller & N. File (Eds.), *Curriculum in Early Childhood Education* (pp. 43-55). Routledge.

Frede, E., & Barnett, W. S. (2011). Why pre-k is critical to closing the achievement gap. *Principal*, *90*(5), 8–11.

Friedman-Krauss, A. H. (2025). State of curriculum in state-funded preschool. National Institute for Early Education Research. NASEM presentation: https://www.nationalacademies.org/documents/embed/link/LF2255DA3DD1C41C0A42D3BEF0989ACAECE3053A6A9B/file/DD05D67A460FCB7FACCAE45567E4FF934694CFB438A5?noSaveAs=1

Friedman-Krauss, A. H., Barnett, W. S., Hodges, K. S., Garver, K. A., Weisenfeld, G. G., Gardiner, B. A., & Jost, T. M. (2023). *The state of preschool 2022: state preschool yearbook.* National Institute for Early Education Research.

Friedman-Krauss, A., Barnett, W. S., Jost, T., & Garver, K. (2024). Help Wanted: Early Intervention and Early Childhood Special Education Workforce Needs. Findings from a National Survey. Research Report. National Institute for Early Education Research.

Goble, P., & Pianta, R. C. (2017). Teacher–child interactions in free choice and teacher-directed activity settings: Prediction to school readiness. *Early Education and Development*, 28(8), 1035–1051.

Golinkoff, D. G. S. R. M., & Hirsh-Pasek, K. (2006). *Play= Learning: How play motivates and enhances children's cognitive and social-emotional growth*. Oxford University Press.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.

Harden, B. J., Brett, B. E., Gross, J. T., Weiland, C., Berne, J., Klein, E. L., & Tirrell-Corbin, C. (2023). Benefits of pre-kindergarten for children in Baltimore, MD. *Early Childhood Research Quarterly*, 64, 1–12.

Harmeyer, E., & Nores, M. (2026). Teaching Strategies' Creative Curriculum Implementation and Ecosystem Engagement Study (CCIEE): Technical Report 2. National Institute for Early Education Research.

Harms, T., Clifford, D. M., & Cryer, D. (2015). *Early childhood environment rating scale, third edition*. Teachers College Press.

Hernandez, E. L., & McElrath, K. (2023, August 30). *Public and private preschool enrollment from 2019 to 2021 at its lowest since 2005*. U.S. Census Bureau. https://www.census.gov/library/stories/2023/08/preschool-enrollment.html

Hoffmann, J., & Russ, S. (2012). Pretend play, creativity, and emotion regulation in children. *Psychology of Aesthetics, Creativity, and the Arts*, 6(2), 175.

Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, Tunn R, Aggarwal R, Berkwits M, Berlin JA, Bhandari N, Butcher NJ, Campbell MK, Chidebe RCW, Elbourne D, Farmer A, Fergusson DA, Golub RM, Goodman SN, Hoffmann TC, Ioannidis JPA, Kahan BC, Knowles RL, Lamb SE, Lewis S, Loder E, Offringa M, Ravaud P, Richards DP, Rockhold FW, Schriger DL, Siegfried NL, Staniszewska S, Taylor RS, Thabane L, Torgerson D, Vohra S, White IR, Boutron I. CONSORT 2025 statement: updated guideline for reporting randomised trials. Lancet. 2025 Apr 14:S0140-6736(25)00672-5.

Howard, S. J., Siraj, I., Melhuish, E. C., Kingston, D., Neilsen-Hewett, C., de Rosnay, M., Duursma, E., & Luu, B. (2018). Measuring interactional quality in pre-school settings: introduction and validation of the Sustained Shared Thinking and Emotional Wellbeing (SSTEW) scale. *Early Child Development and Care*, 1–14.

Jenkins, J. M., Whitaker, A. A., Nguyen, T., & Yu, W. (2019). Distinctions without a difference? Preschool curricula and children's development. *Journal of Research on Educational Effectiveness*, 12(3), 514–549.

Jensen, B., Jensen, P., & Rasmussen, A. W. (2017). Does professional development of preschool teachers improve children's socio-emotional outcomes?. *Labour Economics*, *45*, 26–39.

Kopack Klein, A., Aikens, N., Li, A., Bernstein, S., Reid, N., Dang, M., Blesson, E., Rakibullah, S., Scott, M., Cannon, J., Harrington, J., Larson, A., Malone, L., Tarullo, L. Descriptive Data on Head Start Children and Families from FACES 2019: Fall 2019 Data Tables and Study Design. OPRE Report #2021-77. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation. https://www.acf.hhs.gov/opre/report/descriptive-data-head-start-children-and-families-faces-2019-fall-2019-data-tables-and

Lambert, R. (2020). Technical manual for the Teaching Strategies GOLD® assessment (second edition): Birth through third grade. Center for Educational Measurement and Evaluation, University of North Carolina Charlotte.

Lipsey, M. W., Farran, D. C., Hurley, S. M., Hofer, K. G., & Bilbrey, C. (2009). Effects of a literacy focused curriculum and a developmental curriculum on school readiness and subsequent state achievement test outcomes in rural prekindergarten classrooms. Society for Research on Educational Effectiveness.

Luesse, H. B., Luesse, J. E., Lawson, J., Camp, M. J., & Diaz, K. G. (2022). The academy for teachers' professional development program. A model to support teacher retention. *Cogent Education*, *9*(1).

Macdonald, N., Gealy, A. M., & Tinney, G. (2021). Exploring the effect of an attachment intervention in areas of multiple deprivation on adult–child interaction and the implications for children's social, emotional, and behavioural development. *Early Child Development and Care*, *191*(5), 670–684.

Markowitz, A. J. (2024). Within-year teacher turnover in Head Start and children's school readiness. *AERA Open*, *10*.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1997). *Maslach Burnout Inventory*. Scarecrow Education.

McCormick, M., Pralica, M., Hsueh, J., Weiland, C., Weissman, A. K., Shapiro, A., Xia, S., MacDowell, C., Maves, S., Taylor, A., & Sachs, J. (2023). Going the distance: disparities in pre-K enrollment in higher-quality schools by geographic proximity, race/ethnicity, family income, and home language. *AERA Open*, 9.

National Academies of Sciences, Engineering, and Medicine. 2023. *Closing the Opportunity Gap for Young Children*. The National Academies Press. https://doi.org/10.17226/26743

National Academies of Sciences, Engineering, and Medicine. 2024. *A New Vision for High-Quality Preschool Curriculum*. The National Academies Press.

National Association for the Education of Young Children (NAEYC). 2020. Developmentally Appropriate Practice (DAP). Position Statement. National Association for the Education of Young Children.

Nores, M., & Harmeyer, E. (2021). The impact of the pandemic on academic support for preschoolers: Key takeaways from a New Jersey district-level survey. National Institute for Early Education Research.

Nores, M., Figueras-Daniel, A., López, M. A., & Bernal, R. (2018). Implementing aeioTU: quality improvement alongside an efficacy study—learning while growing. *Annals of the New York Academy of Sciences,* 1419(1), 201–217.

Nores, M., Harmeyer, E., Barnett, W.S. & Macleod, J. (2026). Teaching Strategies' Creative Curriculum Implementation and Ecosystem Engagement Study (CCIEE): Technical Report 1. National Institute for Early Education Research.

Piaget, J. (2013). *The construction of reality in the child* (Vol. 82). Routledge.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2006). CLASS: Classroom assessment scoring system manual preschool (Pre-K) version. Teachstone.

Preschool Curriculum Evaluation Research Consortium (2008). Effects of Preschool Curriculum Programs on School Readiness. National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. U.S. Government Printing Office.

Pyle, A., Poliszczuk, D., & Danniels, E. (2018). The challenges of promoting literacy integration within a play-based learning kindergarten program: Teacher perspectives and implementation. *Journal of Research in Childhood Education*, 32(2), 219–233.

Rege, M., Størksen, I., Solli, I. F., Kalil, A., McClelland, M. M., Ten Braak, D., Lenes, R., Breive, S., Carlsen, M., Erfjord, I., & Hundeland, P. S. (2024). The effects of a structured curriculum on preschool effectiveness: A field experiment. *Journal of Human Resources*, *59*(2), 576–603.

Romano, J. P., & Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, *100*(469), 94-108.

Schaack, D. D., Donovan, C. V., Adejumo, T., & Ortega, M. (2022). To stay or to leave: Factors shaping early childhood teachers' turnover and retention decisions. *Journal of Research in Childhood Education*, *36*(2), 327–345.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV tests of achievement. Riverside Publishing.

Schrank, F. A., McGrew, K. S., Ruef, M. L., & Alvarado, C. G. (2005). Batería III Woodcock-Muñoz. *Assessment Service Bulletin*, (1).

Siraj, I., Kingston, D., & Melhuish, E. C. (2015). Assessing quality in early childhood education and care: sustained shared thinking and emotional well-being (SSTEW) scale for 2–5-year-olds provision. Trentham Books.

Siraj, I., Melhuish, E., Howard, S. J., Neilsen-Hewett, C., Kingston, D., De Rosnay, M., Huang, R., Gardiner, J., & Luu, B. (2023). Improving quality of teaching and child development: A randomised controlled trial of the leadership for learning intervention in preschools. *Frontiers in Psychology*, *13*, 1092284.

Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, *19*(3), 375–397.

Torrance, E. P. (1974). *Norms technical manual: Torrance tests of creative thinking*. Personal Press.

Tran, H., & Winsler, A. (2011). Teacher and center stability and school readiness among low-income, ethnically diverse children in subsidized, center-based child care. *Children and Youth Services Review*, *33*(11), 2241–2252.

Von Suchodoletz, A., Lee, D. S., Henry, J., Tamang, S., Premachandra, B., & Yoshikawa, H. (2023). Early childhood education and care quality and associations with child outcomes: A meta-analysis. *PLOS ONE*, *18*(5), e0285985.

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Waddell, J. H. (2010). Fostering relationships to increase teacher retention in urban schools. *Journal of Curriculum and Instruction, 4(1),* 70-85.

Weiland, C., Greenberg, E., Bassok, D., Markowitz, A., Rosada, P. G., Luetmer, G., Abenavoli, R., Gomez, C., Johnson, A., Jones-Harden, B., Maier, M., McCormick, M., Morris, P., Nores, M., Phillips, D., & Snow, C. (2021). Historic crisis, historic opportunities: Using evidence to mitigate the effects of the COVID-19 crisis on young children and early care and education programs. *Education Policy Initiative, University of Michigan. https://edpolicy. umich. edu/sites/epi/files/uploads/EPI-UI-Covid% 20Synthesis% 20Brief% 20 June*, *202021*.

Weiland, C., McCormick, M., Duer, J., Friedman-Krauss, A., Pralica, M., Xia, S., Nores, M., & Mattera, S. (2024). The mixed-delivery pre-k opportunity gap? Differences in demographics, quality, and children's gains in community-based versus public school programs across five large-scale systems. *Early Childhood Research Quarterly*, *68*, 247–259.

Weiland, C., McCormick, M., Mattera, S., Maier, M., & Morris, P. (2018). Preschool curricula and professional development features for getting to high-quality implementation at scale: A comparative review across five trials. *AERA Open*, *4*(1), 1–16. X`

Weisenfeld, G. G. (2021). Impacts of COVID-19 on preschool enrollment and spending. National Institute for Early Education Research. Published March.

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

Zachopoulou, E., Makri, A., & Pollatou, E. (2009). Evaluation of children's creativity: psychometric properties of Torrance's 'Thinking Creatively in Action and Movement' test. *Early Child Development and Care*, *179*(3), 317–328.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297.

Appendices

Appendix A

Table A.1. Pretest summary statistics by randomization status for *GOLD*, study sample vs. all children in the study classrooms.

Pretest	N	Treatment			Control			P-value T vs. C[a]	P-value Study vs. All Children: Treated	P-value Study vs. All Children: Control	RW p-values T vs. C[b]
		N	Mean	SD	N	Mean	SD				
Study Sample											
Social-emotional	425	212	399.68	57.21	213	404.53	66.52	0.422	0.106	0.195	0.870
Language	421	208	427.51	72.82	213	434.30	79.30	0.361	0.022	0.046	0.593
Cognitive	412	206	409.54	61.97	206	413.17	62.03	0.553	0.032	0.054	0.476
Literacy	395	193	467.58	41.15	202	465.60	43.60	0.645	0.027	0.043	0.870
Math	395	187	327.07	53.19	208	324.79	52.23	0.668	0.021	0.074	0.870
All Children											
Social-emotional	1522	754	392.40	58.14	768	397.92	65.61	0.083*	0.106	0.195	0.001***
Language	1512	743	413.94	75.75	769	421.63	82.59	0.060*	0.022	0.046	0.138
Cognitive	1473	731	399.08	61.72	742	403.31	65.45	0.202	0.032	0.054	0.347
Literacy	1400	678	460.06	41.75	722	458.36	45.14	0.466	0.027	0.043	0.604
Math	1415	667	317.01	52.17	748	317.31	53.63	0.917	0.021	0.074	0.914

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and control children. [a]Standard P-values. [b]Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications, for comparisons of T vs. C.

Table A.2. Pretest summary statistics by randomization status for 3-year-olds in the sample.

Pretest	N	Treatment			Control			P-value[b]	RW p-values[c]
		N	Mean	SD	N	Mean	SD		
Executive Functions, creativity and Socio-Emotional outcomes									
DCCS	197	97	7.70	4.79	100	7.23	4.02	0.455	0.924
Peg Tapping	195	96	1.96	3.63	99	2.45	3.84	0.355	0.924
TCAM Raw	187	91	13.02	4.66	96	13.27	4.77	0.719	0.924
PIPPS	153	77	72.45	10.99	76	73.21	14.29	0.714	0.970
TCAM Standardized	188	91	85.61	14.00	97	89.06	32.24	0.347	0.970
SDQ Behavioral Problems	153	77	11.31	5.76	76	11.36	6.46	0.965	0.970
SDQ Pro-social Behaviors	153	77	6.60	2.37	76	6.32	2.19	0.485	0.924
Cognitive Outcomes									
PPVT Raw	205	102	34.37	19.32	103	37.73	18.91	0.210	0.746
PPVT Standardized	205	102	78.25	18.56	103	83.65	16.43	0.028**	0.188
WJ Letter Word Raw	204	101	3.81	4.51	103	4.23	5.83	0.565	0.918
WJ Letter Word Standardized	204	101	92.43	13.89	103	93.59	15.23	0.569	0.915
WJ Applied Problem Raw	202	101	2.93	2.63	101	3.63	2.83	0.069*	0.418
WJ Applied Problem Standardized	202	101	74.34	14.76	101	79.99	15.70	0.009*	0.073*
Cognitive Outcomes – Spanish Assessments									
TVIP Raw[a]	81	50	11.32	9.26	31	8.39	10.72	0.196	0.746
TVIP Standardized[a]	81	50	87.44	11.33	31	83.39	15.89	0.185	0.671
WM Letter Word Raw[a]	81	50	5.02	3.10	31	5.19	3.91	0.825	0.918
WM Letter Word Standardized[a]	80	49	84.57	22.13	31	84.32	23.90	0.962	0.989
WM Applied Problem Raw[a]	80	49	5.41	4.46	31	5.03	4.34	0.712	0.918
WM App. Probl. Standardized[a]	80	49	78.51	18.92	31	77.97	20.16	0.903	0.989
GOLD Study Sample									
Social Emotional	174	89	364.26	55.61	85	366.12	54.59	0.824	0.870
Language	169	85	382.76	71.76	84	394.64	69.18	0.275	0.593
Cognitive	165	86	368.92	59.80	79	380.34	50.69	0.189	0.476
Literacy	152	76	441.53	44.22	76	437.74	41.66	0.587	0.870
Math	151	68	289.79	57.58	83	293.43	47.46	0.671	0.870
GOLD All Children									
Social Emotional	661	331	358.25	56.41	330	365.79	57.97	0.091*	0.041**
Language	644	319	371.01	72.36	325	386.50	76.52	0.009***	0.015**
Cognitive	630	317	364.28	59.94	313	373.12	59.13	0.063*	0.117
Literacy	565	272	432.64	42.32	293	434.24	45.87	0.668	0.672
Math	564	252	284.27	53.10	312	292.64	54.52	0.067*	0.117

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and control children. [a]Only for those also assessed with the Spanish measure. [b]Standard P-values. [c]Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications.

Table A.3. Post-test summary statistics by randomization status.

Post-Test Outcomes	N	Treatment			Control			P-value[b]	RW p-values[c]
		N	Mean	SD	N	Mean	SD		
Executive Functions, creativity and Socio-Emotional outcomes									
DCCS	453	221	9.28	5.28	232	9.77	5.38	0.323	0.852
Peg Tapping	451	222	5.46	5.56	229	5.13	5.49	0.523	0.927
TCAM Raw	429	211	15.02	4.17	218	15.16	4.58	0.755	0.991
TCAM Standardized	433	214	84.28	10.76	219	87.19	12.27	0.009***	0.074*
PIPPS	389	203	70.13	10.63	186	71.82	11.71	0.136	0.566
SDQ Behavioral Problems	389	203	10.34	4.92	186	10.40	5.35	0.903	0.991
SDQ Pro-social Behaviors	389	203	6.32	1.86	186	6.02	1.88	0.115	0.991
Cognitive Outcomes									
PPVT Raw	474	238	54.70	24.66	236	59.65	26.75	0.037**	0.138
PPVT Standardized	474	238	82.44	16.59	236	85.72	19.72	0.051*	0.179
WJ Letter Word Raw	473	237	8.06	7.22	236	8.28	6.15	0.716	0.844
WJ Letter Word Standardized	472	237	90.81	14.37	235	91.35	13.47	0.675	0.844
WJ Applied Problem Raw	468	236	7.45	4.46	232	7.82	4.54	0.375	0.679
WJ Applied Problem Standardized	467	236	79.28	16.95	231	80.85	17.63	0.326	0.669
Cognitive Outcomes – Spanish Assessments									
TVIP Raw[a]	167	95	18.49	16.19	72	15.31	14.88	0.194	0.464
TVIP Standardized[a]	167	95	82.65	19.24	72	76.85	18.75	0.053*	0.193
WM Letter Word Raw[a]	162	93	6.12	3.05	69	6.49	3.40	0.463	0.752
WM Letter Word Standardized[a]	161	92	80.07	17.40	69	79.61	17.82	0.871	0.870
WM Applied Problem Raw[a]	160	92	8.02	4.80	68	7.44	4.98	0.458	0.752
WM App. Probl. Standardized[a]	160	92	76.75	16.85	68	72.84	18.79	0.169	0.464
GOLD Study Sample									
Social Emotional	417	208	468.47	53.86	209	464.34	50.42	0.419	0.525
Language	421	216	503.23	89.89	205	495.58	73.05	0.340	0.525
Cognitive	407	205	482.09	61.17	202	473.54	58.86	0.152	0.310
Literacy	399	202	509.21	40.43	197	503.52	35.89	0.139	0.310
Math	399	198	389.10	53.49	201	376.73	46.46	0.014**	0.046**
GOLD All Children									
Social Emotional	1522	754	392.40	58.14	768	397.92	65.61	0.083*	0.022**
Language	1521	753	512.62	71.89	768	517.03	67.87	0.218	0.007***
Cognitive	1512	743	413.94	75.75	769	421.63	82.59	0.060*	0.025**
Literacy	1473	731	399.08	61.72	742	403.31	65.45	0.202	0.001***
Math	1400	678	460.06	41.75	722	458.36	45.14	0.466	0.017**

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and control children. [a]Only for those also assessed with the Spanish measure. [b]Standard P-values. [c]Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications.

Table A.4. Gains' summary statistics comparing the treatment group with the control group.

Gains	N	Treatment			Control			P-value[b]	RW p-values[c]
		N	Mean	SD	N	Mean	SD		
Executive Functions, creativity and Socio-Emotional outcomes									
DCCS	422	204	0.51	6.42	218	1.57	5.94	0.078*	0.327
Peg Tapping	414	202	1.92	5.65	212	1.08	5.59	0.131	0.418
TCAM Raw	393	189	0.56	5.03	204	-0.20	4.57	0.117	0.938
TCAM Standardized	397	192	0.69	15.04	205	0.89	25.41	0.926	0.053
PIPPS	306	163	-0.85	8.32	143	0.48	9.27	0.187	0.418
SDQ Behavioral Problems	306	163	0.20	4.67	143	-0.06	3.52	0.598	0.938
SDQ Pro-social Behaviors	306	163	-0.18	2.48	143	-0.31	1.99	0.634	0.926
Cognitive Outcomes									
PPVT Raw	446	221	9.26	12.69	225	12.96	16.83	0.009***	0.041**
PPVT Standardized	446	221	1.22	11.92	225	3.24	14.96	0.115	0.431
WJ Letter Word Raw	444	219	2.26	3.66	225	2.61	3.63	0.308	0.731
WJ Letter Word Standardized	444	219	-0.76	10.39	225	-0.15	10.63	0.540	0.897
WJ Applied Problem Raw	437	218	2.34	2.89	219	2.36	3.17	0.967	0.965
WJ Applied Problem Standardized	437	218	2.75	14.23	219	1.93	15.28	0.564	0.731
Cognitive Outcomes – Spanish Assessments									
TVIP Raw[a]	156	93	5.19	9.22	63	2.75	9.90	0.116	0.334
TVIP Standardized[a]	156	93	-1.61	12.34	63	-3.56	13.78	0.359	0.755
WM Letter Word Raw[a]	153	91	0.70	3.59	62	0.76	3.27	0.924	0.986
WM Letter Word Standardized[a]	151	89	-0.96	24.24	62	-0.65	21.40	0.936	0.986
WM Applied Problem Raw[a]	150	89	1.08	4.21	61	1.49	3.74	0.538	0.818
WM App. Probl. Standardized[a]	150	89	-2.30	16.92	61	0.75	15.91	0.267	0.643
GOLD Study Sample									
Social Emotional	403	196	69.92	34.87	207	58.17	43.17	0.003***	0.006***
Language	401	200	83.85	48.75	201	61.29	46.05	0.000***	0.001***
Cognitive	384	190	73.17	40.87	194	60.96	44.04	0.005***	0.013**
Literacy	362	175	43.76	21.54	187	39.07	26.52	0.067*	0.075*
Math	373	178	66.17	29.05	195	51.89	34.04	0.000***	0.001***
GOLD All Children									
Social Emotional	1345	656	83.02	42.53	689	69.59	48.19	0.000***	0.001***
Language	1353	671	100.62	56.97	682	79.79	56.74	0.000***	0.001***
Cognitive	1302	638	89.01	46.40	664	76.96	50.65	0.000***	0.001***
Literacy	1208	580	52.77	25.55	628	47.95	30.52	0.003***	0.003***
Math	1266	602	77.09	37.09	664	65.04	40.64	0.000***	0.000***

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and control children. [a]Only for those also assessed with the Spanish measure. [b]Standard P-values. [c]Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications.

Table A.5. Pretest correlations between external measures and GOLD.

Variables	External Assessment Outcomes								
	DCCS	PT	TCAM	PIPPS	SDQ Prob.	SDQ Prosocial	PPVT	WJ-LW	WJ-AP
DCCS	1								
PT	**0.39***	1							
TCAM	0.18*	0.20*	1						
PIPPS	0.19*	0.23*	0.07	1					
SDQ	-0.15*	-0.15*	-0.13*	**-0.72***	1				
Prosocial	-0.08	-0.06	0.02	-0.05	0.02	1			
PPVT	0.25*	0.29*	**0.40***	0.27*	-0.20*	-0.01	1		
WJ-LW	0.03	0.19*	0.03	0.02	0.01	-0.05	**0.35***	1	
WJ-AP	**0.32***	**0.34***	**0.36***	0.24*	-0.17*	-0.01	**0.73***	**0.37***	1
Social Emotional	0.21*	0.24*	**0.30***	**0.43***	-0.27*	-0.03	**0.38***	0.10	**0.41***
Language	0.23*	0.22*	**0.35***	**0.33***	-0.21*	0.01	**0.50***	0.16*	**0.52***
Cognitive	0.21*	0.20*	**0.35***	0.29*	-0.12*	-0.02	**0.43***	0.15*	**0.46***
Literacy	0.22*	0.22*	0.28*	0.28*	-0.16*	-0.04	**0.47***	0.27*	**0.47***
Math	0.19*	0.24*	**0.31***	**0.31***	-0.14*	0.01	**0.47***	**0.31***	**0.50***

Notes: Pairwise Pearson correlations reported. Statistical significance is based on two-tailed tests. * p<0.05. Moderate (>.3) to strong (>.50) correlations are bolded.

Table A.6. Post-test correlations between external measures and GOLD.

Variables	External Assessment Outcomes								
	DCCS	PT	TCAM	PIPPS	SDQ Prob.	SDQ Prosocial	PPVT	WJ-LW	WJ-AP
DCCS	1								
PT	**0.45***	1							
TCAM	0.16*	0.25*	1						
PIPPS	0.24*	**0.34***	0.14*	1					
SDQ Behav. Prob.	-0.20*	-0.29*	-0.11	**-0.69***	1				
SDQ Prosocial	-0.02	-0.1	-0.12*	-0.15*	0.13*	1			
PPVT	**0.40***	**0.46***	**0.38***	0.25*	-0.15*	0	1		
WJ-LW	0.15*	0.22*	0.08	0.00	0.01	0.05	0.26*	1	
WJ-AP	**0.51***	**0.61***	**0.33***	**0.35***	-0.28*	-0.05	**0.71***	**0.36***	1
Social-Emotional	0.23*	**0.43***	**0.37***	**0.40***	-0.28*	-0.06	**0.42***	0.09	**0.52***
Language	0.24*	**0.40***	**0.39***	**0.32***	-0.16*	-0.01	**0.50***	0.16*	**0.54***
Cognitive	0.25*	**0.44***	**0.39***	**0.32***	-0.17*	-0.03	**0.46***	0.17*	**0.56***
Literacy	0.27*	**0.47***	**0.39***	**0.30***	-0.17*	-0.05	**0.52***	**0.40***	**0.64***
Math	0.25*	**0.43***	**0.35***	0.27*	-0.13*	-0.09	**0.46***	**0.34***	**0.60***

Notes: Pairwise Pearson correlations reported. Statistical significance is based on two-tailed tests. * p<0.05. Moderate (>.3) to strong (>.50) correlations are bolded.

Table A.7. Gains' correlations between external measures and GOLD.

Variables	External Assessment Outcomes								
	DCCS	PT	TCAM	PIPPS	SDQ Prob.	SDQ Prosocial	PPVT	WJ-LW	WJ-AP
DCCS	1								
PT	0.24*	1							
TCAM	-0.06	0.08	1						
PIPPS	-0.05	-0.07	0.04	1					
SDQ Behav. Prob.	0.09	0.06	-0.10	-0.22*	1				
SDQ Prosocial	-0.06	-0.03	-0.03	0.06	-0.09	1			
PPVT	0.18*	0.15*	-0.04	-0.01	0.01	0.04	1		
WJ-LW	0.12	0.08	0.06	0.12	0.03	-0.10	0.03	1	
WJ-AP	0.09	0.20*	-0.03	0.00	0.11	-0.08	0.21*	0.23*	1
Social-Emotional	-0.07	-0.01	-0.04	**0.31***	-0.05	0.03	0.00	-0.11	-0.04
Language	-0.08	-0.05	-0.03	0.25*	-0.1	0.04	0.09	-0.02	0.05
Cognitive	-0.06	0.01	0.04	0.15*	0.01	-0.01	0.00	-0.07	0.02
Literacy	0.00	0.00	-0.03	0.14*	-0.01	-0.03	0.05	-0.08	-0.05
Math	-0.14*	0.01	0.05	0.16*	-0.03	-0.02	0.11	-0.07	0.04

Notes: Pairwise Pearson correlations reported. Statistical significance is based on two-tailed tests. * p<0.05. Moderate (>.3) to strong (>.50) correlations are bolded.

Table A.8. Predictive correlations between pretest and post-test GOLD and external outcome measures.

Variables	GOLD Outcomes: Pretest				
Post-Test	Social Emotional	Language	Cognitive	Literacy	Math
DCCS	0.19*	0.26*	0.22*	0.20*	0.24*
PT	**0.30***	**0.37***	**0.30***	0.33*	0.34*
TCAM	**0.33***	**0.40***	**0.34***	0.28*	0.29*
PIPPS	**0.30***	0.24*	0.23*	0.26*	0.25*
SDQ Behav. Prob.	-0.21*	-0.11	-0.13*	-0.19*	-0.14*
SDQ Prosocial	-0.05	0.00	0.01	0.03	0.01
PPVT	**0.34***	**0.44***	**0.38***	**0.39***	**0.39***
WJ-LW	0.17*	0.19*	0.20*	0.32*	**0.35***
WJ-AP	**0.43***	**0.50***	**0.46***	**0.54***	**0.53***
Pretest	GOLD Outcomes: Post-Test				
DCCS	0.23*	0.13*	0.21*	0.18*	0.21*
PT	0.29*	0.20*	0.27*	0.25*	0.26*
TCAM	**0.33***	**0.33***	**0.34***	**0.32***	**0.32***
PIPPS	**0.34***	0.23*	0.27*	0.29*	0.22*
SDQ Behav. Prob.	-0.24*	-0.17*	-0.14*	-0.17*	0.08
SDQ Prosocial	-0.12	-0.08	-0.11	-0.07	-0.11
PPVT	**0.41***	**0.50***	**0.47***	**0.51***	**0.46***
WJ-LW	0.08	0.18*	0.18*	**0.36***	**0.33***
WJ-AP	**0.47***	**0.53***	**0.50***	**0.57***	**0.53***

Notes: Pairwise Pearson correlations reported. Statistical significance is based on two-tailed tests. * p<0.05. Moderate (>.3) to strong (>.50) correlations are bolded.

Table A.9. Missing data analyses for post-test external outcome measures.

Variables	Missing PPVT (a)	Missing WJ-LW (b)	Missing WJ-AP (c)	Missing TCAM (d)	Missing PIPPS/SDQ (e)	Missing DCCS (f)	Missing Peg Tapping (g)
ITT	-0.000	0.004	-0.007	0.010	-0.065	0.041**	0.029
	(0.005)	(0.008)	(0.011)	(0.032)	(0.066)	(0.020)	(0.021)
Age	-0.001	-0.001*	-0.001	-0.005**	-0.005	-0.003**	-0.004***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.001)	(0.001)
IEP	-0.005	-0.010	0.022	0.152*	0.068	-0.018	-0.024
	(0.006)	(0.007)	(0.039)	(0.088)	(0.083)	(0.043)	(0.043)
Female	0.004	-0.013*	-0.010	-0.005	-0.025	0.019	0.022
	(0.008)	(0.007)	(0.013)	(0.026)	(0.032)	(0.019)	(0.021)
Black	-0.046	-0.034	-0.019	-0.031	0.074	-0.057	-0.046
	(0.040)	(0.040)	(0.043)	(0.053)	(0.087)	(0.055)	(0.055)
Hispanic	-0.049	-0.047	-0.034	0.050	0.045	-0.055	-0.052
	(0.048)	(0.047)	(0.050)	(0.066)	(0.083)	(0.065)	(0.065)
White	0.068	-0.047	-0.043	0.058	-0.009	-0.082	-0.079
	(0.106)	(0.044)	(0.046)	(0.112)	(0.137)	(0.060)	(0.060)
Asian	-0.049	-0.052	-0.044	0.019	0.008	-0.120*	-0.120*
	(0.048)	(0.048)	(0.050)	(0.124)	(0.134)	(0.068)	(0.070)
DLL	0.004	0.014	0.007	0.071	-0.047	0.054*	0.058*
	(0.012)	(0.011)	(0.017)	(0.045)	(0.059)	(0.031)	(0.031)
Observations	472	472	472	472	472	472	472
R-squared	0.088	0.038	0.010	0.093	0.028	0.051	0.043

Note: ***p<0.01, **p<0.05, *p<0.1. Controls include district fixed effects; errors are clustered at the school. Robust standard errors in parentheses.

Table A.10. Pretest summary statistics comparing the treatment group with for 'synthetic' sample.

Pretest	N	Treatment			Synthetic			P-value[b]	RW p-value[c]
		N	Mean	SD	N	Mean	SD		
Executive Functions, creativity and Socio-Emotional outcomes									
DCCS	353	216	8.89	5.55	137	9.79	5.6	0.142	0.236
Peg Tapping	352	215	3.67	4.8	137	5.16	5.58	0.008***	0.035**
TCAM Raw	338	208	14.50	4.59	130	15.52	4.20	0.042**	0.154
TCAM Standardized	335	208	83.89	11.47	127	85.84	8.95	0.103	0.18
PIPPS	268	178	75.23	14.01	90	80.39	10.65	0.002***	0.042**
SDQ Behavioral Problems	268	178	10.62	5.95	90	9.31	4.70	0.069*	0.236
SDQ Pro-social Behaviors	268	178	6.44	2.31	90	6.38	1.90	0.815	0.94
Cognitive Outcomes									
PPVT Raw	364	221	45.49	23.68	143	55.69	24.47	0.000***	0.001***
PPVT Standardized	364	221	81.21	18.18	143	88.9	18.89	0.000***	0.001***
WJ Letter Word Raw	363	220	5.75	6.50	143	7.13	5.58	0.037**	0.013**
WJ Letter Word Standardized	363	220	91.4	14.66	143	95.59	13.15	0.006***	0.003***
WJ Applied Problem Raw	363	220	5.14	4.02	143	7.31	4.35	0.000***	0.001***
WJ Applied Problem Standardized	363	220	76.58	16.36	143	85.49	18.63	0.000***	0.001***
Cognitive Outcomes – Spanish Assessments									
TVIP Raw[a]	123	104	13.84	11.26	19	13.32	10.28	0.851	0.988
TVIP Standardized[a]	123	104	85.03	14.45	19	85.42	10.86	0.911	0.988
WM Letter Word Raw[a]	123	104	5.69	3.13	19	8.16	3.55	0.002***	0.004***
WM Letter Word Standardized[a]	122	103	83.19	19.03	19	93.21	12.9	0.030**	0.041**
WM Applied Problem Raw[a]	122	103	7.23	4.75	19	9.37	5.52	0.082*	0.059*
WM App. Probl. Standardized[a]	122	103	80.13	17.22	19	87.26	15.43	0.095*	0.059*

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and 'synthetic' children. [a]Only for those also assessed with the Spanish measure. [b]Standard P-values. [c]Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications.

Table A.11. Post-test summary statistics comparing the treatment group with the 'synthetic' sample.

Post-Test Outcomes	N	Treatment			Synthetic			P-value[b]	RW p-value[c]
		N	Mean	SD	N	Mean	SD		
Executive Functions, creativity and Socio-Emotional outcomes									
DCCS	371	221	9.28	5.28	150	11.69	5.89	0.000***	0.001***
Peg Tapping	370	222	5.46	5.56	148	7.85	6.05	0.000***	0.002***
TCAM Raw	354	211	15.02	4.17	143	16.55	4.36	0.001***	0.007***
TCAM Standardized	358	214	84.28	10.76	144	87.12	10.36	0.014**	0.049**
PIPPS	304	203	5.22	0.92	101	5.54	0.82	0.003***	0.043**
SDQ Behavioral Problems	304	203	10.34	4.92	101	9.38	4.69	0.104	0.170
SDQ Pro-social Behaviors	304	203	6.32	1.86	101	6.04	1.67	0.209	0.194
Cognitive Outcomes									
PPVT Raw	391	238	54.70	24.66	153	66.52	24.21	0.000***	0.001***
PPVT Standardized	391	238	82.44	16.59	153	91.03	17.92	0.000***	0.001***
WJ Letter Word Raw	390	237	8.06	7.22	153	9.58	5.67	0.029**	0.023**
WJ Letter Word Standardized	389	237	90.81	14.37	152	94.74	13.30	0.007***	0.014**
WJ Applied Problem Raw	389	236	7.45	4.46	153	9.98	4.41	0.000***	0.001***
WJ Applied Problem Standardized	388	236	79.28	16.95	152	88.91	17.69	0.000***	0.001***
Cognitive Outcomes – Spanish Assessments									
TVIP Raw[a]	127	95	18.49	16.19	32	16.09	14.71	0.459	0.742
TVIP Standardized[a]	127	95	82.65	19.24	32	77.03	17.21	0.145	0.393
WM Letter Word Raw[a]	125	93	6.12	3.05	32	7.97	4.76	0.012**	0.067*
WM Letter Word Standardized[a]	123	92	80.07	17.40	31	83.77	15.86	0.296	0.646
WM Applied Problem Raw[a]	123	92	8.02	4.80	31	8.52	5.43	0.633	0.742
WM App. Probl. Standardized[a]	123	92	76.75	16.85	31	74.71	18.24	0.569	0.742

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and 'synthetic' children. [a]Only for those also assessed with the Spanish measure. [b]Standard P-values. [c]Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications.

Table A.12. Gains' summary statistics comparing the treatment group with the 'synthetic' sample.

Gains	N	Treatment			Synthetic			P-value[b]	RW p-value[c]
		N	Mean	SD	N	Mean	SD		
Executive Functions, creativity and Socio-Emotional outcomes									
DCCS	338	204	0.51	6.42	134	2.21	6.16	0.016**	0.077*
Peg Tapping	334	202	1.92	5.65	132	2.79	6.31	0.192	0.661
TCAM Raw	312	189	0.56	5.03	123	1.29	4.66	0.197	0.661
TCAM Standardized	312	192	0.69	15.04	120	2.15	10.28	0.352	0.759
PIPPS	226	163	-1.30	9.51	63	-2.57	7.49	0.342	0.342
SDQ Behavioral Problems	226	163	0.20	4.67	63	0.30	4.05	0.875	0.876
SDQ Pro-social Behaviors	226	163	-0.18	2.48	63	-0.44	1.87	0.451	0.783
Cognitive Outcomes									
PPVT Raw	364	221	9.26	12.69	143	11.34	13.88	0.143	0.480
PPVT Standardized	364	221	1.22	11.92	143	2.43	12.00	0.344	0.750
WJ Letter Word Raw	362	219	2.26	3.66	143	2.48	3.06	0.548	0.854
WJ Letter Word Standardized	361	219	-0.76	10.39	142	-0.94	8.67	0.869	0.881
WJ Applied Problem Raw	361	218	2.34	2.89	143	2.71	2.84	0.241	0.646
WJ Applied Problem Standardized	360	218	2.75	14.23	142	3.46	12.25	0.626	0.854
Cognitive Outcomes – Spanish Assessments									
TVIP Raw[a]	109	93	5.19	9.22	16	4.38	10.01	0.747	0.971
TVIP Standardized[a]	109	93	-1.61	12.34	16	-3.19	14.26	0.646	0.938
WM Letter Word Raw[a]	107	91	0.70	3.59	16	0.38	3.69	0.738	0.971
WM Letter Word Standardized[a]	104	89	-0.96	24.24	15	-1.87	17.25	0.889	0.971
WM Applied Problem Raw[a]	104	89	1.08	4.21	15	-0.07	5.27	0.350	0.795
WM App. Probl. Standardized[a]	104	89	-2.30	16.92	15	-5.07	19.71	0.569	0.938

Note: ***p<0.01, **p<0.05, *p<0.1 for differences in means between treatment and 'synthetic' children. [a]Only for those also assessed with the Spanish measure. [b]Standard P-values. [c]Stepdown P-values are for Romano and Wolf (2005) stepdown procedures applied by blocks of baseline variables, 1000 replications.

Table A.13: Proximal effects of intervention on Teacher Retention, Classroom Quality and Curriculum Fidelity.

Estimation	Retention		ECERS-3		SSTEW		Fidelity Use (%)	
ITT	0.234***		-0.127		-0.392**		-0.044	
	(0.060)		(0.318)		(0.173)		(0.028)	
TOT		0.270***		-0.044		-0.414**		-0.035
		(0.074)		(0.326)		(0.161)		(0.031)
Observations	125	125	124	124	125	125	123	123

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are retention (for all analyses except for retention analyses), teacher demographics, years of experience and education, district fixed effects and errors are clustered at the coach level. Robust standard errors in parentheses.

Table A.14. Estimations for PD dosage and platform engagement on instructional practice without the interaction terms.

Panel A. Asynchronous PD dosage hours

Variables	Retention	ECERS-3	SSTEW	Fidelity Use (%)
ITT	0.221***	-0.102	-0.359**	-0.032
	(0.065)	(0.312)	(0.159)	(0.030)
Dosage: Asynch. PD	0.004	-0.019	0.017	0.000
	(0.006)	(0.016)	(0.016)	(0.002)
Observations	125	124	125	123
R-squared	-	0.140	0.228	0.350

Panel B. SmartTeach Platform Engagement.

Variables	Retention	ECERS-3	SSTEW	Fidelity Use (%)
ITT	0.248***	-0.154	-0.294	-0.032
	(0.072)	(0.321)	(0.168)	(0.030)
Platform engagement	0.074***	0.056	0.144	0.022
	(0.027)	(0.082)	(0.087)	(0.013)
Observations	125	124	125	123
R-squared	-	0.115	0.230	0.371

Note: ***p<0.01, **p<0.05, *p<0.1. Controls include teacher demographics, experience and education, and district fixed effects; errors are clustered at the coach level. Robust standard errors in parentheses. Synchronous dosage main effects are not included as these are reported above as the interaction is not possible since dosage was zero for the control group. Estimations for retention are probit estimations and report marginal probabilities.

Figure A.1. Baseline (Fall) child assessment distributions for cognitive, literacy, language, math and socio-emotional development

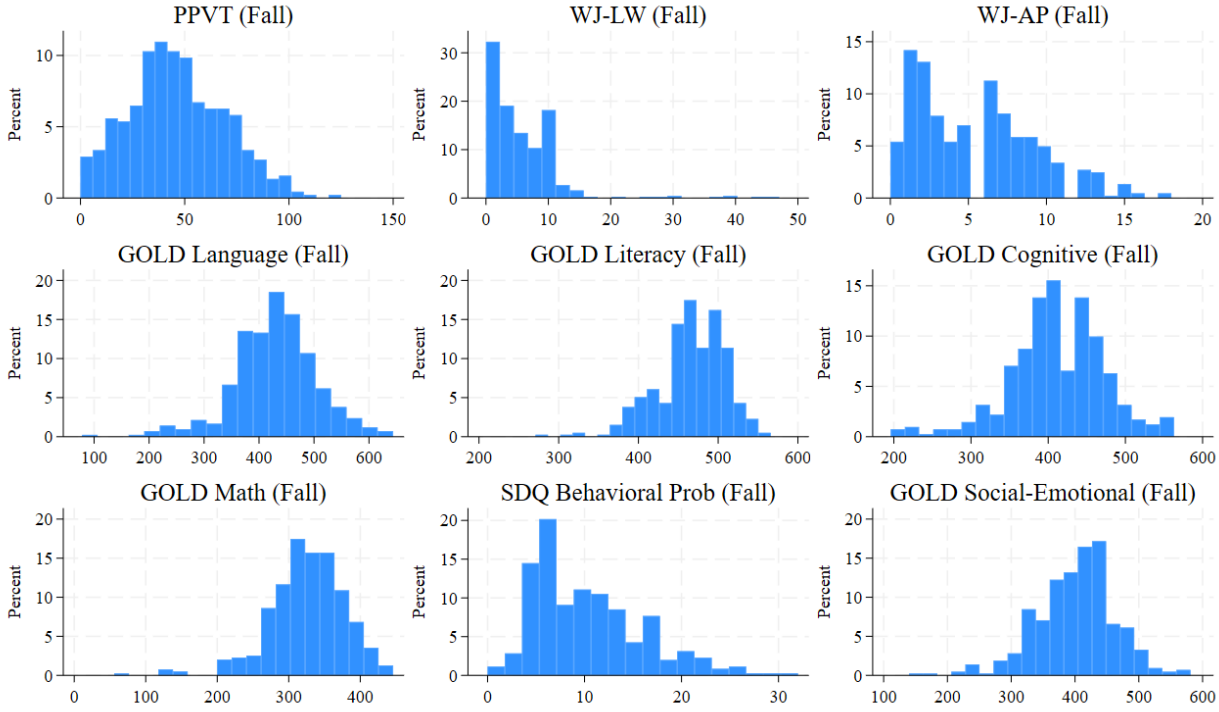
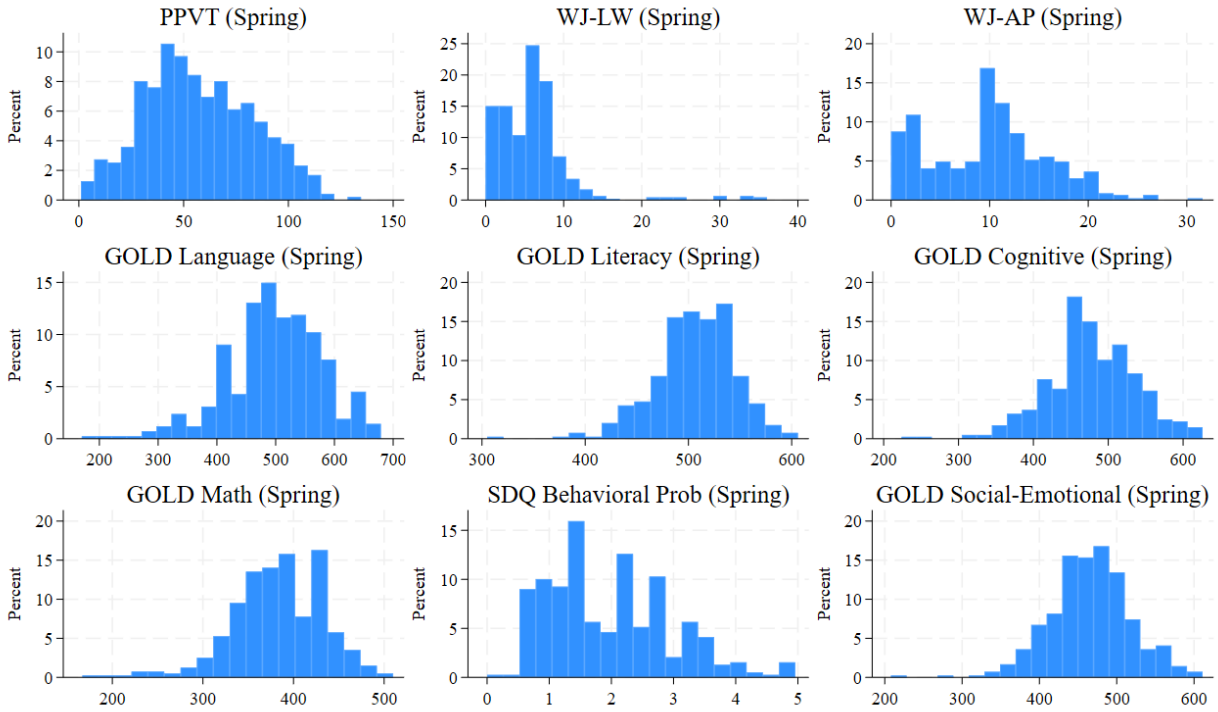


Figure A.2. Post-test (Spring) child assessment distributions for cognitive, literacy, language, math and socio-emotional development



Appendix B

Table B.1. Standardized associations between ITT and external children's creativity and play measures, executive functions and socio-emotional development.

VARIABLES	TCAM	PIPPS	DCCS	PT	SDQ Behav. Problems	SDQ Prosocial	Social Emotional	Social Emotional All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Pretest	0.068***	0.051***	0.054***	0.074***	0.115***	0.162***	0.008***	0.009***
	(0.012)	(0.003)	(0.010)	(0.013)	(0.006)	(0.028)	(0.001)	(0.001)
ITT	0.074	0.084	-0.107	0.163	-0.061	0.150	0.127*	0.125*
	(0.133)	(0.093)	(0.095)	(0.125)	(0.089)	(0.111)	(0.064)	(0.065)
ECERS	0.122*	-0.154***	0.086	0.096	0.032	-0.031	-0.017	-0.022
	(0.064)	(0.049)	(0.079)	(0.068)	(0.066)	(0.075)	(0.045)	(0.044)
SSTEW	0.044	0.048	0.004	0.072	-0.025	-0.016	0.031	0.032
	(0.050)	(0.038)	(0.053)	(0.073)	(0.052)	(0.064)	(0.031)	(0.025)
Age	0.034***	0.008	0.028***	0.049***	-0.008	-0.000	0.017***	0.024***
	(0.007)	(0.006)	(0.009)	(0.007)	(0.006)	(0.008)	(0.004)	(0.004)
IEP[a]	-0.283	-0.226*	-0.603***	-0.747***	0.261*	0.021	-0.493***	-
	(0.248)	(0.122)	(0.161)	(0.092)	(0.143)	(0.330)	(0.091)	-
Female	0.072	0.032	0.214*	0.174	-0.067	-0.064	0.095*	0.123***
	(0.078)	(0.071)	(0.116)	(0.129)	(0.079)	(0.116)	(0.049)	(0.022)
Black	0.125	-0.110	-0.207	0.045	-0.141	-0.043	-0.168**	0.043
	(0.225)	(0.190)	(0.259)	(0.244)	(0.219)	(0.217)	(0.076)	(0.074)
Hispanic	0.059	0.118	-0.180	0.153	-0.163	-0.147	-0.180**	0.086
	(0.270)	(0.219)	(0.311)	(0.244)	(0.251)	(0.214)	(0.086)	(0.106)
White	-0.356	0.516	0.135	0.236	0.054	0.086	-0.276*	0.181
	(0.389)	(0.329)	(0.593)	(0.468)	(0.469)	(0.324)	(0.147)	(0.139)
Asian	0.509*	0.448	0.325	0.589	0.074	0.377	-0.026	-
	(0.279)	(0.290)	(0.367)	(0.409)	(0.340)	(0.388)	(0.237)	-
DLL	-0.166	-0.121	-0.236	-0.317***	-0.017	0.011	-0.012	-0.025
	(0.135)	(0.094)	(0.185)	(0.110)	(0.110)	(0.106)	(0.041)	(0.032)
Observations	389	302	418	410	302	302	397	1,331
R-squared	0.310	0.648	0.200	0.321	0.526	0.188	0.694	0.606

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. [a]IEP status was only available for children in the study sample.

Table B.2. Standardized associations between ITT and children's cognitive outcomes.

VARIABLES	GOLD Cognitive	GOLD Cognitive All Children	PPVT	GOLD Language	GOLD Language All Children	WJ-LW	GOLD Literacy	GOLD Literacy All Children	WJ-AP	GOLD Math	GOLD Math All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Pretest	0.010***	0.011***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.181***	0.013***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.002)	(0.001)	(0.017)	(0.001)	(0.001)
ITT	0.117	0.060	-0.143*	0.226**	0.192*	-0.052	0.053	0.052	-0.064	0.254**	0.211**
	(0.131)	(0.112)	(0.081)	(0.105)	(0.106)	(0.073)	(0.096)	(0.087)	(0.089)	(0.117)	(0.098)
ECERS	0.046	0.013	-0.012	-0.096**	-0.115**	0.125***	0.055	0.029	0.219***	-0.110***	-0.078
	(0.070)	(0.066)	(0.043)	(0.042)	(0.053)	(0.035)	(0.038)	(0.039)	(0.049)	(0.039)	(0.059)
SSTEW	0.000	0.009	0.019	0.013	0.029	0.070	-0.018	-0.008	0.011	0.017	0.027
	(0.045)	(0.043)	(0.028)	(0.041)	(0.043)	(0.042)	(0.039)	(0.040)	(0.052)	(0.048)	(0.047)
Age	0.025***	0.026***	0.025***	0.021***	0.021***	0.014*	0.024***	0.022***	0.041***	0.015**	0.024***
	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)	(0.004)	(0.010)	(0.006)	(0.006)
IEP[a]	-0.639***	-	-0.541***	-0.461***	-	-0.082	-0.402***	-	-0.721***	-0.301***	-
	(0.145)	-	(0.179)	(0.118)	-	(0.244)	(0.135)	-	(0.113)	(0.107)	-
Female	0.084*	0.096***	0.065	0.086	0.102***	-0.020	0.091*	0.049**	0.032	0.073	0.046
	(0.046)	(0.028)	(0.058)	(0.053)	(0.022)	(0.076)	(0.046)	(0.021)	(0.076)	(0.050)	(0.030)
Black	-0.151	-0.065	-0.402**	-0.068	0.017	-0.091	-0.156	0.085*	-0.254	-0.024	0.115
	(0.146)	(0.110)	(0.192)	(0.097)	(0.084)	(0.208)	(0.092)	(0.048)	(0.228)	(0.146)	(0.075)
Hispanic	-0.141	0.006	-0.336*	-0.045	0.005	-0.207	-0.037	0.169***	-0.197	-0.028	0.117*
	(0.134)	(0.128)	(0.187)	(0.092)	(0.077)	(0.207)	(0.102)	(0.039)	(0.187)	(0.146)	(0.064)
White	-0.392	0.080	-0.289	-0.196	0.051	-0.451	-0.213	0.131	-0.365	-0.073	0.197
	(0.230)	(0.165)	(0.256)	(0.242)	(0.137)	(0.328)	(0.150)	(0.111)	(0.386)	(0.253)	(0.164)
Asian	0.237	-	-0.249	0.142	-	0.174	0.247	-	0.545**	0.035	-
	(0.392)	-	(0.221)	(0.258)	-	(0.242)	(0.323)	-	(0.259)	(0.271)	-
DLL	-0.030	-0.026	-0.233**	-0.196**	-0.080	0.086	-0.071	-0.035	-0.022	-0.078	-0.106*
	(0.060)	(0.042)	(0.106)	(0.072)	(0.050)	(0.082)	(0.066)	(0.039)	(0.081)	(0.076)	(0.053)
Other Eth.	-	0.062	-	-	-0.020	-	-	0.163*	-	-	0.072
	-	(0.147)	-	-	(0.144)	-	-	(0.080)	-	-	(0.106)
Observations	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.635	0.610	0.718	0.730	0.663	0.726	0.734	0.693	0.651	0.727	0.651

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. [a]IEP status was only available for children in the study sample.

Table B.3. Standardized associations between TOT and children's creativity and play measures, executive functions and socio-emotional development.

VARIABLES	TCAM	PIPPS	DCCS	PT	SDQ Behav. Problems	SDQ Prosocial	GOLD Social Emotional	GOLD Social Emotional All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Pretest	0.068***	0.051***	0.054***	0.075***	0.115***	0.162***	0.008***	0.009***
	(0.012)	(0.003)	(0.010)	(0.013)	(0.006)	(0.028)	(0.001)	(0.001)
TOT	0.048	0.084	-0.107	0.153	-0.061	0.150	0.127*	0.125*
	(0.138)	(0.093)	(0.094)	(0.127)	(0.089)	(0.111)	(0.064)	(0.065)
Observations	389	302	418	410	302	302	397	1,331
R-squared	0.310	0.648	0.200	0.321	0.526	0.188	0.694	0.606

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in GOLD estimations with all children.

Table B.4. Standardized associations between TOT and children's cognitive outcomes.

VARIABLES	GOLD Cognitive	GOLD Cognitive All Children	PPVT	GOLD Language	GOLD Language All Children	WJ-LW	GOLD Literacy	GOLD Literacy All Children	WJ-AP	GOLD Math	GOLD Math All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Pretest	0.010***	0.011***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.181***	0.013***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.002)	(0.001)	(0.017)	(0.001)	(0.001)
TOT	0.117	0.060	-0.154*	0.226**	0.192*	-0.041	0.053	0.052	-0.051	0.254**	0.211**
	(0.131)	(0.112)	(0.081)	(0.105)	(0.106)	(0.069)	(0.096)	(0.087)	(0.084)	(0.117)	(0.098)
Observations	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.635	0.610	0.719	0.730	0.663	0.726	0.734	0.693	0.651	0.727	0.651

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in GOLD estimations with all children.

Table B.5. Standardized associations between ITT and TOT and external children's cognitive outcomes, Spanish cognitive outcomes.

VARIABLES	TVIP	WM-LW	WM-AP	TVIP	WM-LW	WM-AP
	(a)	(b)	(c)	(d)	(e)	(f)
Pretest	0.078***	0.096**	0.113***	0.078***	0.096**	0.113***
	(0.007)	(0.042)	(0.025)	(0.007)	(0.042)	(0.025)
ITT	0.243	-0.099	0.017	0.243	-0.099	0.017
	(0.217)	(0.227)	(0.133)	(0.217)	(0.227)	(0.133)
TOT						
ECERS	-0.047	0.154	0.107	-0.047	0.154	0.107
	(0.103)	(0.120)	(0.098)	(0.103)	(0.120)	(0.098)
SSTEW	-0.064	0.128*	0.039	-0.064	0.128*	0.039
	(0.054)	(0.064)	(0.050)	(0.054)	(0.064)	(0.050)
Age	0.025*	0.010	0.032***	0.025*	0.010	0.032***
	(0.012)	(0.017)	(0.009)	(0.012)	(0.017)	(0.009)
IEP	-0.400**	0.215	-0.910***	-0.400**	0.215	-0.910***
	(0.187)	(0.246)	(0.139)	(0.187)	(0.246)	(0.139)
Female	0.079	-0.034	0.216	0.079	-0.034	0.216
	(0.148)	(0.162)	(0.151)	(0.148)	(0.162)	(0.151)
Observations	155	152	149	155	152	149
R-squared	0.683	0.281	0.566	0.683	0.281	0.566

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. The ITT and TOT estimations do not differ due to the classrooms where Spanish assessments were conducted.

Appendix C

Figure C.1. Standardized associations between TOT and children's creativity and play measures.

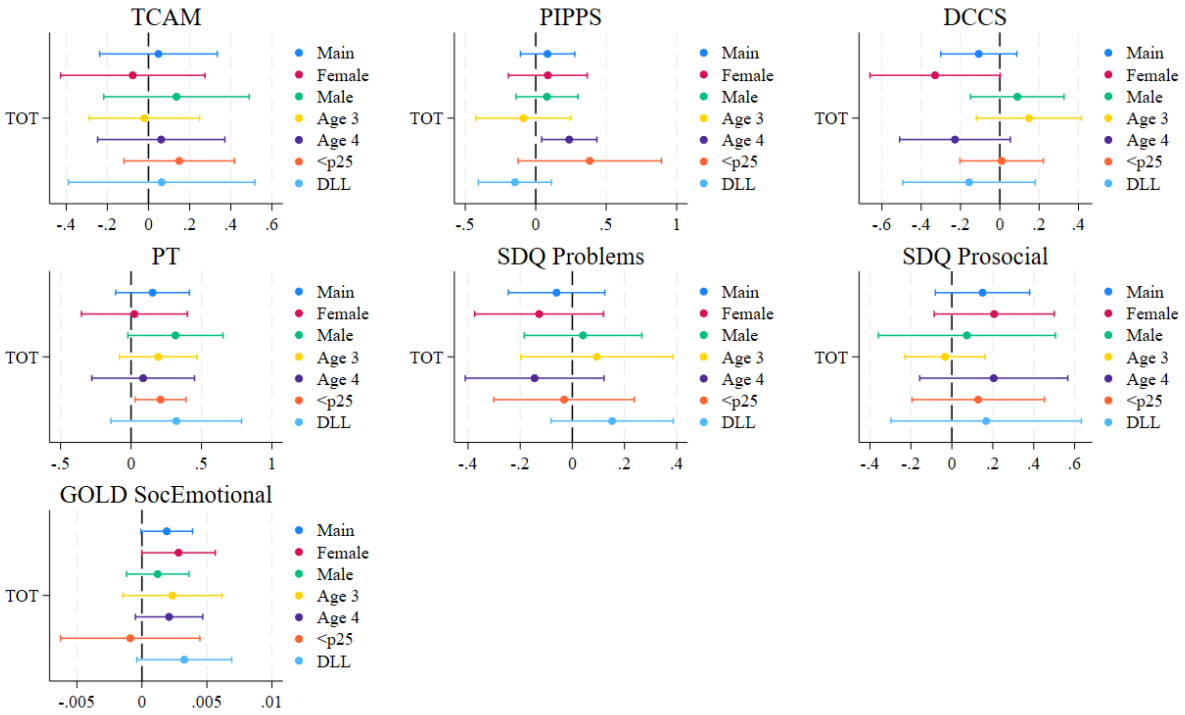
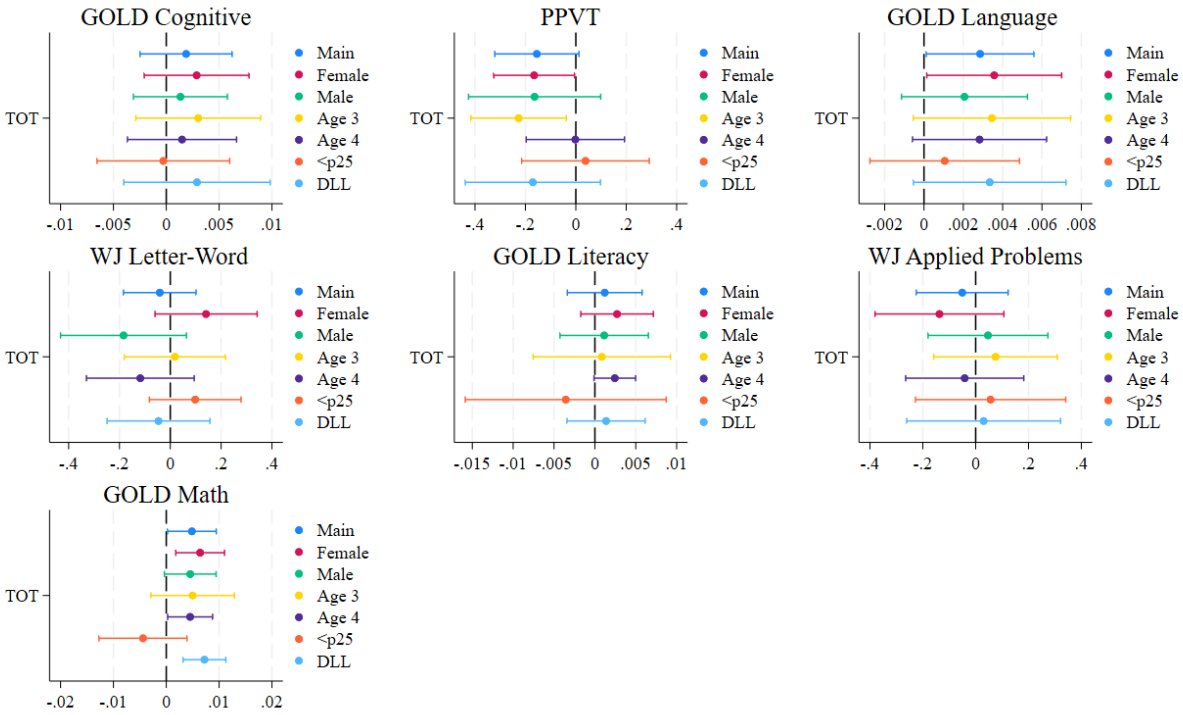


Figure C.2. Standardized associations between TOT and children's cognitive outcomes.



Appendix D

Table D.1. Standardized associations between TOT and children's creativity and play measures, executive functions and socio-emotional development, accounting for teacher retention.

Variables	Creativity	Play Scale	Executive Functions		Socio-Emotional			
	TCAM	PIPPS	DCCS	PT	SDQ Behav. Problems	SDQ Prosocial	GOLD Social Emotional	GOLD Social Emotional All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Pretest	0.069***	0.051***	0.055***	0.075***	0.115***	0.162***	0.008***	0.008***
	(0.012)	(0.003)	(0.010)	(0.013)	(0.007)	(0.028)	(0.001)	(0.001)
TOT	0.164	-0.265	0.048	0.037	0.039	-0.003	-0.040	0.043
	(0.224)	(0.196)	(0.184)	(0.209)	(0.221)	(0.431)	(0.198)	(0.165)
Retention	-0.159	-0.005	0.163*	0.146	-0.179	-0.207	0.103	0.106
	(0.148)	(0.113)	(0.093)	(0.115)	(0.118)	(0.176)	(0.113)	(0.100)
TOT * Retention	-0.096	0.438**	-0.245	0.107	-0.079	0.244	0.188	0.084
	(0.221)	(0.192)	(0.258)	(0.266)	(0.270)	(0.483)	(0.252)	(0.215)
Observations	389	302	418	410	302	302	397	1,331
R-squared	0.315	0.655	0.203	0.325	0.531	0.193	0.701	0.610

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in GOLD estimations with all children.

Table D.2. Standardized associations between TOT and children's cognitive outcomes, accounting for teacher retention.

Variables	Cognitive			Language			Literacy			Emerging Math		
	GOLD Cognitive	GOLD Cognitive All Children	PPVT	GOLD Language	GOLD Language All Children	WJ-LW	GOLD Literacy	GOLD Literacy All Children	WJ-AP	GOLD Math	GOLD Math All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Pretest	0.010***	0.011***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.181***	0.014***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.001)	(0.001)	(0.016)	(0.001)	(0.001)
TOT	-0.107	-0.120	-0.228**	-0.154	0.052	-0.186	-0.207	-0.124	-0.193	0.047	-0.020
	(0.251)	(0.207)	(0.109)	(0.213)	(0.163)	(0.148)	(0.186)	(0.105)	(0.134)	(0.186)	(0.144)
Retention	0.205	0.293**	-0.029	0.056	0.108	-0.064	0.075	0.177*	0.057	0.158	0.130
	(0.199)	(0.135)	(0.101)	(0.140)	(0.102)	(0.087)	(0.116)	(0.101)	(0.118)	(0.111)	(0.079)
TOT retention	0.223	0.162	0.102	0.457	0.153	0.204	0.301*	0.176	0.165	0.207	0.253
	(0.324)	(0.269)	(0.140)	(0.288)	(0.242)	(0.171)	(0.172)	(0.132)	(0.182)	(0.203)	(0.186)
Observations	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.646	0.626	0.719	0.739	0.666	0.727	0.741	0.702	0.652	0.733	0.657

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in GOLD estimations with all children.

Table D.3. Standardized associations between ITT and children's creativity, play measures, executive functions and socio-emotional development, accounting for PD dosage and platform engagement, without controlling for retention.

Variables	Creativity	Play Scale	Executive Function		Socio-Emotional			
	TCAM	PIPPS	DCCS	PT	SDQ Behav. Problems	SDQ Prosocial	GOLD Social Emotional	GOLD Social Emotional All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Estimations with Synchronous PD dosage hours								
Pretest	0.067***	0.051***	0.054***	0.075***	0.114***	0.161***	0.008***	0.009***
	(0.012)	(0.003)	(0.010)	(0.013)	(0.007)	(0.028)	(0.001)	(0.001)
ITT	-0.095	0.124	-0.053	0.033	-0.233	0.218	0.010	-0.017
	(0.182)	(0.156)	(0.144)	(0.213)	(0.141)	(0.198)	(0.068)	(0.073)
Dosage: Synch. PD	0.008	-0.002	-0.003	0.007	0.008	-0.003	0.006*	0.007**
	(0.005)	(0.004)	(0.005)	(0.008)	(0.005)	(0.007)	(0.003)	(0.003)
Obs.	389	302	418	410	302	302	397	1,331
R-squared	0.315	0.648	0.201	0.323	0.530	0.189	0.697	0.610
Estimations with Asynchronous PD dosage hours								
Pretest	0.071***	0.051***	0.054***	0.075***	0.115***	0.162***	0.008***	0.009***
	(0.013)	(0.003)	(0.010)	(0.013)	(0.007)	(0.026)	(0.001)	(0.001)
ITT	0.178	0.057	-0.162	0.175	-0.036	0.071	0.112	0.078
	(0.136)	(0.119)	(0.102)	(0.166)	(0.106)	(0.144)	(0.066)	(0.075)
Dosage: Asynch. PD	0.027***	-0.012	-0.013	0.007	0.005	-0.009	-0.006	-0.006
	(0.008)	(0.011)	(0.016)	(0.014)	(0.007)	(0.011)	(0.006)	(0.005)
ITT*Dosage Asynch. PD	-0.024***	0.009	0.012	-0.004	-0.005	0.013	0.004	0.009
	(0.008)	(0.012)	(0.014)	(0.015)	(0.007)	(0.011)	(0.006)	(0.006)
Obs.	389	302	418	410	302	302	397	1,331
R-squared	0.324	0.651	0.203	0.322	0.527	0.192	0.695	0.608
Estimations with SmartTeach Platform Engagement hours								
Pretest	0.070***	0.052***	0.054***	0.075***	0.114***	0.162***	0.008***	0.008***
	(0.013)	(0.003)	(0.010)	(0.013)	(0.007)	(0.028)	(0.001)	(0.001)
ITT	-0.158	0.260	-0.055	0.171	-0.271	0.194	0.083	0.026
	(0.220)	(0.156)	(0.150)	(0.240)	(0.226)	(0.278)	(0.114)	(0.112)
Platform engagement	-0.024	0.009	0.025**	0.031	-0.008	0.004	0.012	0.010
	(0.015)	(0.013)	(0.011)	(0.019)	(0.016)	(0.020)	(0.012)	(0.009)
ITT*Platform engagement	0.030	-0.019	-0.012	-0.008	0.023	-0.005	0.002	0.009
	(0.018)	(0.013)	(0.016)	(0.023)	(0.020)	(0.028)	(0.013)	(0.011)
Obs.	386	302	414	406	302	302	397	1,331
R-squared	0.318	0.650	0.205	0.330	0.530	0.188	0.699	0.612

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in GOLD estimations with all children. Due to missing dosage and engagement data on the classrooms that shifted from the treatment to the control group, ITT and TOT estimations do not differ. Engagement is presented in a scale of weeks of engagement throughout the three-year period at 35 hours of engagement per week.

Table D.4. Standardized associations between ITT and children's cognitive outcomes, accounting for PD dosage and platform engagement, without controlling for retention.

	Cognitive			Language			Literacy			Emerging Math	
Variables	GOLD Cognitive	GOLD Cognitive All Children	PPVT	GOLD Language	GOLD Language All Children	WJ-LW	GOLD Literacy	GOLD Literacy All Children	WJ-AP	GOLD Math	GOLD Math All Children
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
Estimations with Synchronous PD dosage hours											
Pretest	0.010***	0.011***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.180***	0.013***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.001)	(0.001)	(0.016)	(0.001)	(0.001)
ITT	-0.143	-0.169	-0.204**	-0.119	-0.067	-0.230**	-0.166	-0.207**	-0.235*	0.092	0.006
	(0.120)	(0.135)	(0.097)	(0.080)	(0.095)	(0.100)	(0.112)	(0.077)	(0.118)	(0.126)	(0.103)
Dosage:											
Synch. PD	0.012**	0.011*	0.003	0.016***	0.012***	0.009***	0.010***	0.011***	0.009**	0.007*	0.009*
	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.004)	(0.004)
Obs.	378	1,289	442	395	1,340	440	356	1,194	433	367	1,252
R-squared	0.645	0.617	0.719	0.744	0.671	0.729	0.741	0.702	0.654	0.729	0.656
Estimations with Asynchronous PD dosage hours											
Pretest	0.010***	0.010***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.182***	0.014***	0.012***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.016)	(0.002)	(0.001)	(0.016)	(0.001)	(0.001)
ITT	0.105	0.058	-0.152*	0.180	0.128	-0.054	0.061	0.041	-0.129	0.270**	0.203*
	(0.138)	(0.124)	(0.087)	(0.122)	(0.129)	(0.077)	(0.104)	(0.100)	(0.097)	(0.124)	(0.115)
Dosage:											
Asynch. PD	-0.001	0.004	0.000	0.000	-0.000	0.001	0.004	0.005	-0.007	0.010	0.013
	(0.010)	(0.009)	(0.008)	(0.007)	(0.008)	(0.005)	(0.008)	(0.008)	(0.014)	(0.008)	(0.009)
ITT*Dosage											
Asynch. PD	0.002	-0.002	0.001	0.006	0.009	-0.000	-0.003	-0.001	0.011	-0.006	-0.004
	(0.011)	(0.010)	(0.008)	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.014)	(0.008)	(0.008)
	378	1,289	414	395	1,340	412	356	1,194	405	367	1,252
R-squared	0.625	0.599	0.711	0.729	0.673	0.726	0.732	0.693	0.645	0.725	0.654
Estimations with SmartTeach Platform Engagement hours											
Pretest	0.010***	0.010***	0.034***	0.009***	0.009***	0.187***	0.015***	0.015***	0.183***	0.013***	0.012***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.016)	(0.002)	(0.001)	(0.017)	(0.001)	(0.001)
ITT	-0.123	-0.101	-0.083	0.130	0.156	-0.135	-0.155	-0.102	-0.165	0.083	-0.046
	(0.170)	(0.125)	(0.118)	(0.174)	(0.170)	(0.135)	(0.133)	(0.098)	(0.123)	(0.154)	(0.110)
Platform engagement	0.016	0.014	0.007	0.004	0.008	-0.012	0.002	0.009	-0.011	0.002	-0.000
	(0.016)	(0.011)	(0.012)	(0.014)	(0.010)	(0.012)	(0.011)	(0.007)	(0.011)	(0.013)	(0.011)
ITT*Platform engagement	0.020	0.014	-0.009	0.009	0.002	0.012	0.021	0.014	0.014	0.018	0.028**
	(0.016)	(0.010)	(0.014)	(0.016)	(0.015)	(0.012)	(0.012)	(0.010)	(0.013)	(0.014)	(0.013)
Obs.	378	1,289	438	395	1,340	436	356	1,194	429	367	1,252
R-squared	0.651	0.620	0.715	0.732	0.665	0.725	0.741	0.700	0.650	0.731	0.660

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics, experience and education, and district fixed effects; errors are clustered at the School level. Robust standard errors in parentheses. IEP status was only available for children in the study sample so it is not included in the GOLD estimations with all children. Due to missing dosage and engagement data on the classrooms that shifted from the treatment to the control group, ITT and TOT estimations do not differ. Engagement is presented in a scale of weeks of engagement throughout the three-year period at 35 hours of engagement per week.

Appendix E

Table E.1. Standardized post-test child outcomes comparing children in the Treatment and Control samples to children in the Synthetic sample.

Variables	Creativity	Play Scale	Executive Function		SDQ Socio-Emotional		Cognitive, English			Cognitive, Spanish		
	TCAM	PIPPS	DCCS	PT	Behav. Problems	Prosocial Behav.	PPVT	WJ-LW	WJ-AP	TVIP	WM-LW	WM-AP
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
Pretest	0.071***	0.059***	0.061***	0.072***	0.114***	0.161***	0.034***	0.185***	0.182***	0.077***	0.087**	0.104***
	(0.010)	(0.003)	(0.009)	(0.011)	(0.007)	(0.024)	(0.002)	(0.012)	(0.012)	(0.007)	(0.040)	(0.024)
Treated	-0.175	-0.048	-0.352**	-0.187	0.112	0.322**	-0.071	-0.009	-0.097	-0.012	-0.519*	-0.138
	(0.113)	(0.137)	(0.141)	(0.167)	(0.106)	(0.122)	(0.079)	(0.085)	(0.090)	(0.226)	(0.293)	(0.335)
Control	-0.223	0.063	-0.202	-0.324**	0.104	0.246*	0.073	0.057	-0.041	-0.289	-0.358	-0.117
	(0.137)	(0.159)	(0.123)	(0.136)	(0.098)	(0.141)	(0.070)	(0.067)	(0.100)	(0.248)	(0.262)	(0.331)
ECERS	0.084	0.063	0.088	0.108	-0.007	-0.009	0.012	0.108***	0.230***	-0.069	0.158	0.080
	(0.062)	(0.061)	(0.063)	(0.067)	(0.057)	(0.062)	(0.037)	(0.031)	(0.047)	(0.091)	(0.120)	(0.090)
SSTEW	0.042	-0.037	0.057	0.076	-0.008	-0.012	0.026	0.067**	0.004	-0.025	0.095*	0.039
	(0.048)	(0.043)	(0.050)	(0.057)	(0.039)	(0.054)	(0.026)	(0.032)	(0.038)	(0.053)	(0.047)	(0.054)
Age	0.029***	-0.006	0.025***	0.039***	-0.009*	-0.002	0.019***	0.010	0.032***	0.024**	0.020	0.044***
	(0.005)	(0.006)	(0.007)	(0.007)	(0.005)	(0.007)	(0.006)	(0.006)	(0.007)	(0.011)	(0.016)	(0.012)
IEP	-0.321*	0.030	-0.428***	-0.560***	0.189	0.125	-0.381**	-0.150	-0.596***	-0.416***	0.231	-0.756***
	(0.166)	(0.158)	(0.159)	(0.179)	(0.131)	(0.284)	(0.151)	(0.187)	(0.115)	(0.141)	(0.233)	(0.165)
Female	0.131*	-0.047	0.203**	0.220**	-0.124*	-0.089	0.061	0.020	0.059	0.106	-0.040	0.233
	(0.072)	(0.077)	(0.095)	(0.107)	(0.072)	(0.096)	(0.050)	(0.056)	(0.067)	(0.135)	(0.149)	(0.146)
Black	0.120	0.014	-0.162	0.066	-0.084	-0.131	-0.372*	-0.157	-0.211			
	(0.188)	(0.197)	(0.248)	(0.222)	(0.197)	(0.181)	(0.187)	(0.181)	(0.204)			
Hispanic	0.109	0.002	-0.117	0.227	-0.140	-0.218	-0.296	-0.191	-0.164			
	(0.217)	(0.187)	(0.275)	(0.230)	(0.222)	(0.186)	(0.183)	(0.182)	(0.176)			
White	0.371	-0.359	0.285	0.516*	-0.207	-0.118	-0.237	-0.080	0.076			
	(0.268)	(0.270)	(0.333)	(0.260)	(0.229)	(0.207)	(0.195)	(0.187)	(0.239)			
Asian	0.392	-0.215	0.585*	0.619*	-0.085	0.037	-0.308	0.045	0.381			
	(0.269)	(0.229)	(0.302)	(0.341)	(0.280)	(0.288)	(0.210)	(0.232)	(0.253)			
DLL	-0.312***	-0.016	-0.279*	-0.419***	0.023	-0.018	-0.225***	0.037	-0.041			
	(0.099)	(0.078)	(0.142)	(0.116)	(0.088)	(0.093)	(0.074)	(0.060)	(0.073)			
Observations	512	365	552	542	365	365	584	582	575	171	168	164
R-squared	0.311	0.540	0.238	0.312	0.498	0.189	0.715	0.727	0.659	0.654	0.299	0.533

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are teacher demographics and education; errors are clustered at the classroom level. Robust standard errors in parentheses.

Table E.2. Romano-Wolf adjusted p-values for children's creativity, play measures, executive functions and socio-emotional development.

Variables	Creativity	Play Scale	Executive Functions		Socio Emotional		
	TCAM	PIPPS	DCCS	PT	SDQ Behav. Problems	SDQ Prosocial	GOLD Social Emotional
Panel A ITT							
ITT	0.370	0.370	0.200	0.200	0.378	0.150	**0.033**
ITT	0.571	0.330	0.864	0.755	0.982	0.995	0.982
Retention	0.571	0.943	0.222	0.309	0.217	0.372	0.590
ITT*Retention	0.770	0.050	0.478	0.923	0.968	0.891	0.698
Panel B TOT							
TOT	0.561	0.363	0.234	0.234	0.378	0.150	**0.033**
TOT	0.663	0.350	0.921	0.921	0.982	0.995	0.982
Retention	0.437	0.943	0.229	0.427	0.217	0.372	0.590
TOT*Retention	0.848	0.050	0.559	0.905	0.968	0.891	0.698

Note: We use Romano-Wolf multiple-hypothesis corrections (Clarke, Romano & Wolf, 2020; Romano & Wolf, 2005).

Table E.3. Romano-Wolf adjusted p-values children's cognitive outcomes.

Variables	Cognitive	Language		Literacy		Emerging Math	
	GOLD Cognitive	PPVT	GOLD Language	WJ-LW	GOLD Literacy	WJ-AP	GOLD Math
Panel A ITT							
ITT	0.554	**0.069**	**0.025**	0.691	0.691	0.691	**0.025**
ITT	0.991	0.441	0.889	0.455	0.740	0.591	0.991
Retention	0.772	0.991	0.991	0.815	0.923	0.991	0.500
ITT*Retention	0.906	0.991	0.441	0.468	0.355	0.817	0.775
Panel B TOT							
TOT	0.566	**0.047**	**0.021**	0.786	0.786	0.786	**0.021**
TOT	0.955	0.155	0.938	0.676	0.740	0.552	0.955
Retention	0.782	0.955	0.955	0.938	0.938	0.955	0.552
TOT*Retention	0.938	0.938	0.418	0.711	0.322	0.866	0.788

Note: We use Romano-Wolf multiple-hypothesis corrections (Clarke, Romano & Wolf, 2020; Romano & Wolf, 2005).

Table E.4. Romano-Wolf adjusted p-values for children's creativity, play measures, executive functions and socio-emotional development, accounting for PD dosage and platform engagement.

Variables	Creativity	Play Scale	Executive Functions		Socio Emotional		
	TCAM	PIPPS	DCCS	PT	SDQ Behav. Problems	SDQ Prosocial	GOLD Social-Emotional
Panel A ITT							
ITT	0.707	0.587	0.858	0.858	0.189	0.363	0.857
Dosage: Synch. PD	0.179	0.707	0.856	0.717	0.189	0.816	0.169
ITT	0.228	0.494	0.194	0.438	0.856	0.856	0.268
Dosage: Asynch. PD	**0.005**	0.276	0.539	0.629	0.856	0.856	0.689
ITT*Dosage Asynch. PD	**0.007**	0.494	0.539	0.713	0.856	0.575	0.856
ITT	0.521	0.175	0.892	0.752	0.442	0.795	0.795
Platform engagement	0.182	0.521	**0.091**	0.223	0.887	0.969	0.508
ITT*Platform engagement	0.175	0.182	0.752	0.892	0.458	0.969	0.969
Panel B TOT							
TOT	0.576	0.576	0.840	0.912	0.189	0.363	0.857
Dosage: Synch. PD	**0.094**	0.590	0.840	0.630	0.189	0.816	0.169
TOT	0.416	0.494	0.200	0.539	0.856	0.856	0.268
Dosage: Asynch. PD	**0.005**	0.349	0.539	0.663	0.856	0.856	0.689
TOT*Dosage Asynch. PD	**0.009**	0.494	0.539	0.709	0.856	0.575	0.856
TOT	0.341	0.162	0.883	0.785	0.442	0.795	0.795
Platform engagement	0.162	0.341	0.084	0.248	0.887	0.969	0.508
TOT*Platform engagement	0.133	0.168	0.761	0.883	0.458	0.969	0.969

Table E.5. Romano-Wolf adjusted p-values children's cognitive outcomes, accounting for PD dosage and platform engagement.

Variables	Cognitive	Language		Literacy		Emerging Math	
	GOLD Cognitive	PPVT	GOLD Language	WJ-LW	GOLD Literacy	WJ-AP	GOLD Math
Panel A ITT							
ITT	0.321	0.087	0.269	**0.068**	0.269	**0.096**	0.371
Dosage: Synch. PD	0.087	0.371	**0.001**	**0.020**	**0.005**	**0.088**	0.106
ITT	0.971	0.307	0.486	0.975	0.987	0.607	**0.099**
Dosage: Asynch. PD	1.000	1.000	1.000	0.997	0.987	0.987	0.607
ITT*Dosage Asynch. PD	1.000	1.000	0.971	1.000	0.997	0.970	0.975
ITT	0.880	0.880	0.873	0.873	0.823	0.717	0.914
Platform engagement	0.851	0.914	0.914	0.873	0.958	0.873	0.958
ITT*Platform engagement	0.735	0.914	0.914	0.873	0.462	0.846	0.738
Panel B TOT							
TOT	0.305	**0.089**	0.260	**0.089**	0.260	0.153	0.311
Dosage: Synch. PD	**0.089**	0.311	**0.001**	**0.017**	**0.006**	0.153	0.138
TOT	0.972	0.201	0.483	0.989	0.989	0.712	**0.098**
Dosage: Asynch. PD	1.000	1.000	1.000	0.998	0.989	0.989	0.613
TOT*Dosage Asynch. PD	1.000	1.000	0.972	1.000	0.998	0.971	0.983
TOT	0.909	0.909	0.909	0.909	0.825	0.746	0.932
Platform engagement	0.854	0.932	0.932	0.909	0.958	0.909	0.958
TOT*Platform engagement	0.746	0.932	0.932	0.909	0.469	0.854	0.746

Table E.6. Romano-Wolf adjusted p-values for children's creativity, play measures, executive functions and socio-emotional development for heterogeneous effects.

Variables	Creativity	Play Scale	Executive Functions		Socio-Emotional		
	TCAM	PIPPS	DCCS	PT	SDQ Behav. Problems	SDQ Prosocial	GOLD Social Emotional
Panel A ITT							
Female	0.673	0.628	**0.035**	0.490	0.227	0.154	**0.044**
Male	0.380	0.394	0.566	0.130	0.863	0.863	0.496
3 years old	0.928	0.766	0.269	0.269	0.709	0.718	0.337
4 years old	0.342	0.015	**0.061**	0.457	0.249	0.249	0.108
DLL	0.666	0.291	0.283	0.193	0.215	0.349	**0.093**
Below p25	0.241	0.241	0.612	0.025	0.896	0.734	0.896
Panel B TOT							
Female	0.632	0.632	**0.037**	0.857	0.227	0.154	**0.044**
Male	0.529	0.529	0.398	**0.069**	0.863	0.863	0.496
3 years old	0.866	0.760	0.238	0.198	0.709	0.718	0.337
4 years old	0.588	**0.015**	**0.092**	0.531	0.249	0.249	0.108
DLL	0.666	0.291	0.283	0.193	0.215	0.349	**0.093**
Below p25	0.254	0.230	0.911	**0.037**	0.896	0.734	0.896

Table E.7. Romano-Wolf adjusted p-values children's cognitive outcomes for heterogeneous effects.

Variables	Cognitive	Language		Literacy		Emerging Math	
	GOLD Cognitive	PPVT	GOLD Language	WJ-LW	GOLD Literacy	WJ-AP	GOLD Math
Panel A ITT							
Female	0.268	0.135	0.048	0.268	0.268	0.268	**0.006**
Male	0.765	0.348	0.348	0.177	0.808	0.939	**0.090**
3 years old	0.528	**0.023**	0.146	0.992	0.992	0.992	0.390
4 years old	0.782	0.889	0.148	0.445	0.097	0.889	**0.050**
DLL	0.670	0.404	0.151	0.845	0.845	0.845	**0.002**
Below p25	0.985	0.985	0.954	0.824	0.954	0.969	0.812
Panel B TOT							
Female	0.211	**0.055**	**0.054**	0.211	0.211	0.211	**0.007**
Male	0.776	0.358	0.355	0.253	0.811	0.811	**0.088**
3 years old	0.546	**0.025**	0.148	0.956	0.956	0.821	0.402
4 years old	0.789	0.976	0.154	0.431	**0.085**	0.862	**0.046**
DLL	0.670	0.404	0.151	0.845	0.845	0.845	**0.002**
Below p25	0.953	0.953	0.953	0.793	0.953	0.953	0.793

Appendix F

Table F.1 How accurately do the following statements describe your learning experience with Teaching Strategies this year? Full sample.

	Obs.	Mean	Std. dev	% Agree & Strongly Agree
I felt motivated to participate in these learning opportunities.	134	4.28	0.69	88.05
I felt motivated to implement the strategies I learned in my classroom.	135	4.46	0.60	94.81
The strategies I learned were easy to begin and I could sustain them over time.	135	4.21	0.71	88.14
The trainings were engaging.	132	4.18	0.78	84.09
The trainings helped me improve my practice.	135	4.35	0.64	91.11
Is benefitting the children in my classroom.	134	4.32	0.64	91.79
Has influenced my teaching practice and classroom management.	135	4.30	0.73	88.89
Has been a valuable professional development opportunity for me.	134	4.31	0.73	88.06
Is providing me with knowledge I will likely continue to use in the future.	135	4.37	0.68	92.60
I would recommend these trainings to other teachers.	135	4.36	0.70	90.37

Table F.2. How accurately do the following statements describe your learning experiences with Teaching Strategies this year? Control vs. Treatment.

	Control			Treatment			P-value
	Obs.	Mean	Std. dev	Obs.	Mean	Std. dev	
I felt motivated to participate in these learning opportunities.	51	4.20	0.78	55	4.40	0.66	0.146
I felt motivated to implement the strategies I learned in my classroom.	51	4.39	0.60	55	4.56	0.60	0.146
The strategies I learned were easy to begin and I could sustain them over time.	51	4.08	0.72	55	4.31	0.74	0.107
The trainings were engaging.	50	4.04	0.83	54	4.41	0.71	0.017**
The trainings helped me improve my practice.	51	4.22	0.64	55	4.51	0.66	0.023**
Is benefitting the children in my classroom.	50	4.26	0.69	55	4.42	0.64	0.234
Has influenced my teaching practice and classroom management.	51	4.22	0.81	55	4.40	0.74	0.222
Has been a valuable professional development opportunity for me.	50	4.14	0.83	55	4.47	0.69	0.028**
Is providing me with knowledge I will likely continue to use in the future.	51	4.24	0.79	55	4.51	0.64	0.051*
I would recommend these trainings to other teachers.	51	4.25	0.72	55	4.45	0.77	0.170

Note: Synthetic is excluded. Mean values are calculated based on the following Likert Scale: Strongly disagree=1, Disagree=2, Neither agree nor disagree=3, Agree=4, Strongly Agree=5. Excluded 'Don't know'. ***p<0.01, **p<0.05, *p<0.1. N=106.

Table F.3. How accurately do the following statements describe your learning experiences with Teaching Strategies this year? Control vs. Treatment.

	Control			Treatment			P-value
	Obs.	Mean	Std. dev	Obs.	Mean	Std. dev	
Lack of resources and materials.	52	2.65	1.03	55	2.40	1.13	0.228
Insufficient time to plan and meet with coaches.	52	3.00	1.18	54	2.81	1.17	0.420
A mismatch in language with children and/or coaches.	52	2.00	0.84	53	2.17	1.07	0.367
The absence of support from my coach.	52	1.96	0.89	54	1.94	0.98	0.925
The absence of support from leadership at my school/center.	51	2.16	0.95	54	2.06	1.11	0.606
The load of curricular and assessment requirements I am expected to fulfill.	52	3.35	1.14	55	3.04	1.22	0.177

Note: Synthetic is excluded. Mean values are calculated based on the following Likert Scale: Strongly disagree=1, Disagree=2, Neither agree nor disagree=3, Agree=4, Strongly Agree=5. Excluded 'Don't know'. ***p<0.01, **p<0.05, *p<0.1. N=106.

Table F.4. Please rate the following statements with how often they occur. Control vs. Treatment.

	Control			Treatment			P-value
	Obs.	Mean	Std. dev	Obs.	Mean	Std. dev	
I can easily understand how my students feel about things	53	5.30	1.40	54	5.26	1.42	0.876
I deal very effectively with the problems of my students.	53	5.34	1.31	56	5.70	0.81	0.093*
I feel I'm positively influencing other people's lives through my work.	52	4.94	1.73	56	5.27	1.20	0.256
I feel very energetic.	53	4.21	1.70	56	4.80	1.29	0.043**
I can easily create a relaxed atmosphere with my students.	54	5.48	1.06	56	5.63	0.65	0.391
I feel exhilarated after working closely with my students.	54	4.02	2.22	56	5.16	1.27	0.001***
I have accomplished many worthwhile things in this job.	53	4.96	1.64	57	5.35	1.01	0.142
In my work, I deal with emotional problems very calmly.	53	5.17	1.47	56	5.48	0.76	0.173
Personal accomplishment subscale	54	4.92	0.91	57	5.34	0.61	0.005***

Note: Synthetic is excluded. Mean values are calculated based on the following Likert Scale: Never=0, A few times a year or less=1, Once a month or less=2, A few times a month=3, Once a week=4, A few times a week=5, Everyday=6. N=150.***p<0.01, **p<0.05, *p<0.1. N=106.

Table F.5. Please rate the following statements with how often they occur. Control vs. Treatment.

	Control			Treatment			P-value
	Obs.	Mean	Std. dev	Obs.	Mean	Std. dev	
I feel emotionally drained from my work.	54	3.04	2.08	56	2.73	1.91	0.425
I feel used up at the end of the workday.	54	3.26	2.26	56	2.77	1.90	0.220
I feel fatigued when I get up in the morning and have to face another day on the job.	54	2.72	2.26	56	2.00	1.87	0.074*
Working with people all day is really a strain for me.	53	0.85	1.46	56	0.82	1.54	0.924
I feel burned out from my work.	53	2.72	2.11	56	2.13	1.85	0.122
I feel frustrated by my job.	53	2.04	1.76	55	1.95	1.69	0.782
I feel I'm working too hard on my job.	52	2.65	2.06	55	2.45	2.05	0.617
Working with people directly puts too much stress on me.	53	0.75	1.41	56	0.54	1.09	0.366
I feel like I'm at the end of my rope.	54	1.06	1.72	56	0.57	1.31	0.098*
Emotional exhaustion subscale	54	2.10	1.30	56	1.77	1.52	0.222

Note: Synthetic is excluded. Mean values are calculated based on the following Likert Scale: Never=0, A few times a year or less=1, Once a month or less=2, A few times a month=3, Once a week=4, A few times a week=5, Everyday=6. N=150.***p<0.01, **p<0.05, *p<0.1. N=106.

Table F.6. Please rate the following statements with how often they occur. Control vs. Treatment.

	Control			Treatment			P-value
	Obs.	Mean	Std. dev	Obs.	Mean	Std. dev	
I feel I treat some students as if they were impersonal objects.	54	0.26	0.87	56	0.16	0.60	0.489
I've become more callous toward people since I took this job.	52	0.67	1.40	55	0.53	1.37	0.587
I worry that this job is hardening me emotionally.	53	0.92	1.40	56	0.71	1.34	0.425
I don't really care what happens to some students.	53	0.09	0.49	56	0.07	0.53	0.816
I feel students blame me for some of their problems.	54	0.07	0.43	56	0.29	0.93	0.127
Depersonalization subscale	54	0.40	0.57	56	0.35	0.66	0.667

Note: Synthetic is excluded. Mean values are calculated based on the following Likert Scale: Never=0, A few times a year or less=1, Once a month or less=2, A few times a month=3, Once a week=4, A few times a week=5, Everyday=6. N=150. ***p<0.01, **p<0.05, *p<0.1. N=106.

Table F.7 How accurately do the following statements describe the curriculum you currently use? Full sample.

	Obs.	Mean	Std. dev	% Agree & Strongly Agree
Is easy for me to implement.	146	4.33	0.95	89.04
Is at the right level for the majority of my students.	146	4.02	1.09	78.08
Supports my student in developing skills to control emotions and behaviors.	146	4.10	1.02	78.77
Supports my students in developing academic skills.	145	4.26	0.91	87.59
Is engaging for my students.	147	4.27	0.80	90.48
Provides differentiated materials to meet the needs of all my students.	146	4.09	1.03	78.08
Meets the needs of students with special needs.	139	3.69	1.15	62.59
Meets the needs of English Language Learners (ELLs).	144	3.95	1.01	75.00

Note: Mean values are calculated based on the following Likert Scale: Strongly disagree=1, Disagree=2, Neither agree nor disagree=3, Agree=4, Strongly Agree=5. Excluded 'Don't know'. N=147.

Table F.8 How accurately do the following statements describe the curriculum you currently use? Treatment and Control group combined.

	Obs.	Mean	Std. dev	% Agree & Strongly Agree
Is easy for me to implement.	107	4.30	0.92	89.72
Is at the right level for the majority of my students.	107	3.96	1.07	76.64
Supports my student in developing skills to control emotions and behaviors.	107	4.10	1.01	79.44
Supports my students in developing academic skills.	106	4.28	0.85	87.74
Is engaging for my students.	108	4.25	0.77	90.74
Provides differentiated materials to meet the needs of all my students.	107	4.02	1.01	75.70
Meets the needs of students with special needs.	101	3.70	1.14	62.38
Meets the needs of English Language Learners (ELLs).	107	4.01	0.99	79.44

Note: Mean values are calculated based on the following Likert Scale: Strongly disagree=1, Disagree=2, Neither agree nor disagree=3, Agree=4, Strongly Agree=5. Excluded 'Don't know'. N=109.