NIEER
NATIONAL INSTITUTE FOR EARLY EDUCATION RESEARCH

Teaching Strategies' Creative Curriculum Implementation and Ecosystem Engagement Study (CCIEE): Technical Report 1

March 2026

Milagros Nores, Ph.D., Erin Harmeyer, Ph.D., Steve Barnett, Ph.D., & Julie Macleod, M.S. The National Institute for Early Education Research

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 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.

Julie MacLeod, **M.S.** Macleod is Project Coordinator I at the National Institute for Early Education Research (NIEER). She leads NIEER's field work on New Jersey evaluations. She has contributed to NIEER's research in curriculum, inclusion and classroom quality.

Grateful acknowledgment is made to the districts, schools and classrooms that participated in this study. In addition, the authors would like to thank Facundo Luna Mallea for his invaluable research analyses support for this report. This work would not be possible without the partnership with Teaching Strategies®, which provided a grant to support the work but also worked closely with the research team in order to track all intervention and sample aspects relevant to the research.

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

Table of Contents

Introduction

Early childhood programs have been lauded for their short-term contributions to children's learning and development and their long-term contributions to children's educational, social, and economic success and wellbeing (e.g., Britto et al., 2017). The foundation of effective programs is quality and mechanisms to sustain quality over time and at scale are essential to success (Barnett & Frede, 2017; Nores et al., 2018;). While there is no exact agreement on what constitutes quality, there is consensus that an element of high quality is a strong curriculum that supports children's learning and development in all domains (Barker et al., 2014; Goble & Pianta, 2017; Golinkoff & Hirsch-Pasek, 2006; Pyle et al., 2018; Yoshikawa et al., 2018). Moreover, there is emerging consensus that a developmentally appropriate assessment system that supports individualized learning in ways that are culturally and linguistically responsive is equally essential, especially when well-aligned with curriculum (NAEYC, 2020). Many state and city programs have stipulated approved curriculum/a and assessment(s), and where this is not the case, providers are responsible for making informed choices. Unfortunately, research provides limited guidance to support these decisions, in part because sound research on curriculum is difficult. Much of the past research is equivocal or has null findings, and curricula have evolved over the past five to 10 years and now include substantial digital components (with the pandemic accelerating change), making much of the past research outdated. In addition, large variation in observed center quality within early childhood programs has contributed to quality concerns across contexts and curricula (Rege et al., 2021; NASEM 2024).

Overall, some curricular approaches in early childhood programs have been found to improve cognitive outcomes, including language, literacy, and mathematical skills (Chambers et al.2016) even through third grade (Lipsey et al., 2009). A total of 14 curricula were studied in the *Effects of Preschool Curriculum Programs on School Readiness* IES funded effort (PCER, 2008), with mostly null results. Jenkins et al.'s (2018) separate analyses of the PCER data showed *The Creative Curriculum*® performing uniformly better than locally developed ones in various classroom quality measures but not on child outcome measures. The authors also found that implementation of a literacy curriculum led to better literacy outcomes for children but not to improved math nor cognitive or socio-emotional outcomes; and implementation of a structured math curriculum showed stronger effects in math but not in literacy nor for socio-emotional skills, the latter effect trending negative. In follow up analyses that incorporated additional sets of studies and samples, Jenkins et al. (2019) found that in the FACES 2003 study, children in programs implementing "other" curriculum scored lower on socio-emotional skills, in the FACES 2009 study children in *The Creative Curriculum* classrooms performed better in receptive vocabulary and math, and in the IES-funded National Center for Early Development & Learning study (NCEDL) children in programs without a published curriculum had higher behavioral problems. Lipsey, et al. (2009) found positive effects on children's state test (reading/language arts, mathematics, science, and social studies) for two different preschool curricula, including *The Creative Curriculum,* by third grade (even in the absence of end of pre-K effects on typically used measures).

Summarizing the evidence, a recent National Academies of Sciences, Engineering, and Medicine (NASEM) 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). Overall, the report recognizes significant variation in the findings on curriculum studies (NASEM 2024). It also highlights that "[s]ubgroup effects for preschool curricula have been less well studied than average effects, in part because of sample size requirements" (p.54), and among the few that have studied specific subgroups some found no effects and others some effects for Black and Latinx children, multi-lingual children and children starting the school year with lower baseline scores, indicating this is an area where further research is needed. This landmark NASEM report also recognizes, among many important factors, the importance of professional development in supporting curriculum implementation, as well as understanding the contextual program factors in which curricula are implemented. Other related factors of importance are equipping classrooms with intensive training and ongoing coaching (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* (the "ecosystem") in combination with an initial set of aligned supports, relative to practices as usual. To reduce variation in context and structural quality, the study was completed within the New Jersey preschool program in former "Abbott" (now NJSDA) 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 due to the Abbott court decisions (see Frede & Barnett, 2011). 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 (2:20), 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 $17,911 per child in the 2023–2024 school year; Friedman-Krauss et al., 2025). Evaluations of the NJ state pre-k program to date have shown long-term sustained effects through tenth grade (Barnett & Jung, 2021), regardless of variation across the three state-approved curricula options that districts could choose from, with *The Creative Curriculum* among the three.

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 virtually delivered 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 preexisting requirements to use the *GOLD* assessment that comprises part of it. 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 districts 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.[1]

This technical report starts by summarizing the goals of the study, as well as the theoretical framework and theory of change behind it. It then delves into the study design, describing the sample and data sources. The report also summarizes the study's timeline,

[1] This randomized trial was registered in the American Economic Association's registry for randomized trials. RCT ID: AEARCTR-0012262.

methods and procedures, plus, provides analyses on baseline and post-test classroom data for the sample, including the randomized classrooms as well as a group of classrooms representing "practices as usual" across New Jersey districts implementing the state's preschool program. Appendices provide additional information on the tools used in the study, internal consistency at baseline, the factor structure of the tools, and additional baseline comparisons.

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

Teaching Strategies® Ecosystem

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 is built to empower administrators, educators, and families to drive positive learning outcomes for young children.

The five components of the ecosystem include:

1. 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.
2. 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.
3. Empowering educators. Teachers have access to ongoing professional development and coaching, enabling them to implement developmentally appropriate and culturally responsive teaching practices with fidelity.
4. Family engagement at the core. 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.
5. Actionable data reporting. Built-in assessment, reports, and analytics provide educators and administrators with actionable insights to drive informed instructional decisions and measure program effectiveness.

Rather than treating early childhood education as separate, disconnected components, Teaching Strategies integrates all components into one unified system. This approach aims 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. Through this ecosystem, Teaching Strategies aims to support early learning programs with building sustainable, high-quality education systems that prepare children for success in school and beyond.

Study Goals

This study implements a cluster Randomized Controlled Trial (RCT) within two low-income districts in New Jersey's state preschool program, with an additional non-equivalent comparison group of practices as usual across the state. Relative to practice-as-usual in two districts already using *The Creative Curriculum* and *GOLD*, the research aims to assess the effectiveness of engagement with the ecosystem and the aligned supports, particularly, professional development and coaching provided by Teaching Strategies. The study has 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.

Theoretical Framework

Early care and education (ECE) programs play a crucial role in shaping children's developmental trajectories. Bronfenbrenner's Ecological Systems Theory provides a comprehensive framework for understanding children's learning within ECE programs. The theory emphasizes the interplay between children and their environments across five interconnected systems. The microsystem refers to the child's immediate environments, such as family, teachers, peers, and ECE settings (including curriculum), where direct interactions significantly impact development. The mesosystem encompasses the connections between these microsystems, such as the relationship between home and school, highlighting the importance of cohesive support across contexts. Other systems represent the broader cultural, economic, and political contexts, encompassing societal values, norms, and policies that shape opportunities for development. This theory underscores the complex, interdependent nature of ECE programs and provides valuable insights for both research and practice (Bronfenbrenner, 1979). We therefore use this person-process-context-time framework (Bronfenbrenner & Morris, 2007) with a focus on capturing induced variation in ecosystem engagement and teaching practices and assessing their relation to children's outcomes. Time addresses progress and changes in the environment over time, and relatedly, in children.

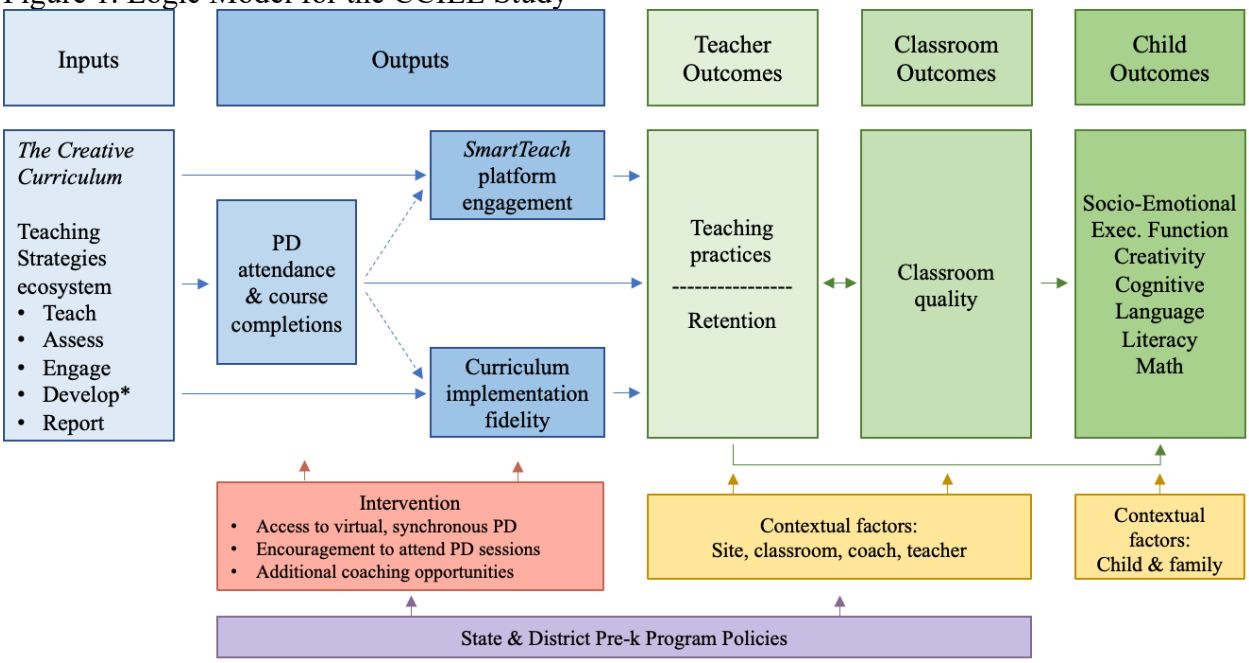
Additional theoretical frameworks complement Bronfenbrenner's model in understanding children's development. Vygotsky's Sociocultural Theory (Vygotsky & Cole, 1978) highlights the role of social and cultural contexts, emphasizing that children's learning is scaffolded through interactions with peers and adults. Piaget's Constructivist Theory (Piaget, 2013) focuses on how children actively construct knowledge through environmental interactions. Additionally, Social Learning Theory (Bandura, 1986) provides a lens to understand how caregivers 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, which then informs the evaluation of professional development opportunities for teachers. Together, these theories offer a rich, multidimensional foundation for exploring the implementation and impact of the CCIEE study.

Theory of Change in the CCIEE Study

Our theory of change connects the support strategies that increase the workforce proficiency within the macro (district and state) and micro (school and classroom) systems that focus on children's development. It connects the professional learning opportunities and other aligned supports provided by Teaching Strategies to the coach and teacher to the processes and outputs that emerge from the Teaching Strategies ecosystem. These processes and outputs are expected to result in improved teacher outcomes (retention and well-being) and teaching practices (as measured via fidelity and quality) and, subsequently, enhanced learning opportunities that lead to positive changes in children's outcomes. Figure 1 depicts the logic model. All components of the ecosystem, including *The Creative Curriculum,* are encompassed under the inputs, with the aligned professional development training (PD) and supports for coaches and teachers intended to bolster the ecosystem engagement and teacher well-being and knowledge, which are hypothesized to increase teacher practices and in turn, impact children's outcomes across the various domains supported by *The Creative Curriculum* and *GOLD*.

Figure 1. Logic Model for the CCIEE Study



*Teachers in the treatment group could access synchronous PD and asynchronous courses; teachers in the control group could only access asynchronous courses.

The Intervention

The intervention was designed to create exogenous differences in how teachers use the Teaching Strategies ecosystem, with an emphasis on strengthening curriculum implementation and

ecosystem engagement in ways expected to support children's learning. To this end, the treatment group received additional professional development, coaching, and updated curriculum materials. Together, these elements were intended to deepen teachers' understanding of the curriculum, improve their day-to-day use of digital and print resources, and, ultimately, enhance instructional quality. Teaching Strategies delivered the intervention through virtually delivered synchronous PD sessions for teachers and ongoing (virtual) side-by-side and group sessions with treatment group coaches. This model centers respect for educator expertise; therefore, the intervention also aimed to enhance teacher well-being, especially confidence and self-efficacy.

In Year 1, PD reinforced foundational understanding of the curriculum and assessment and trained teachers in navigating the digital ecosystem. In Years 2 and 3, PD and coaching focused more directly on using the digital resources—lesson planning, documentation tools, reports, and the other components—in ways designed to improve children's learning. Both districts had access to the full ecosystem throughout the study; however, only teachers and coaches in the treatment group received the PD/coaching, updated curriculum boxes, and preschool "Essentials" kits—play-based and instructionally supportive classroom materials. Figure 4 (further below) provides detailed information about the timing and content of all PD/coaching sessions.

The CCIEE Study Design

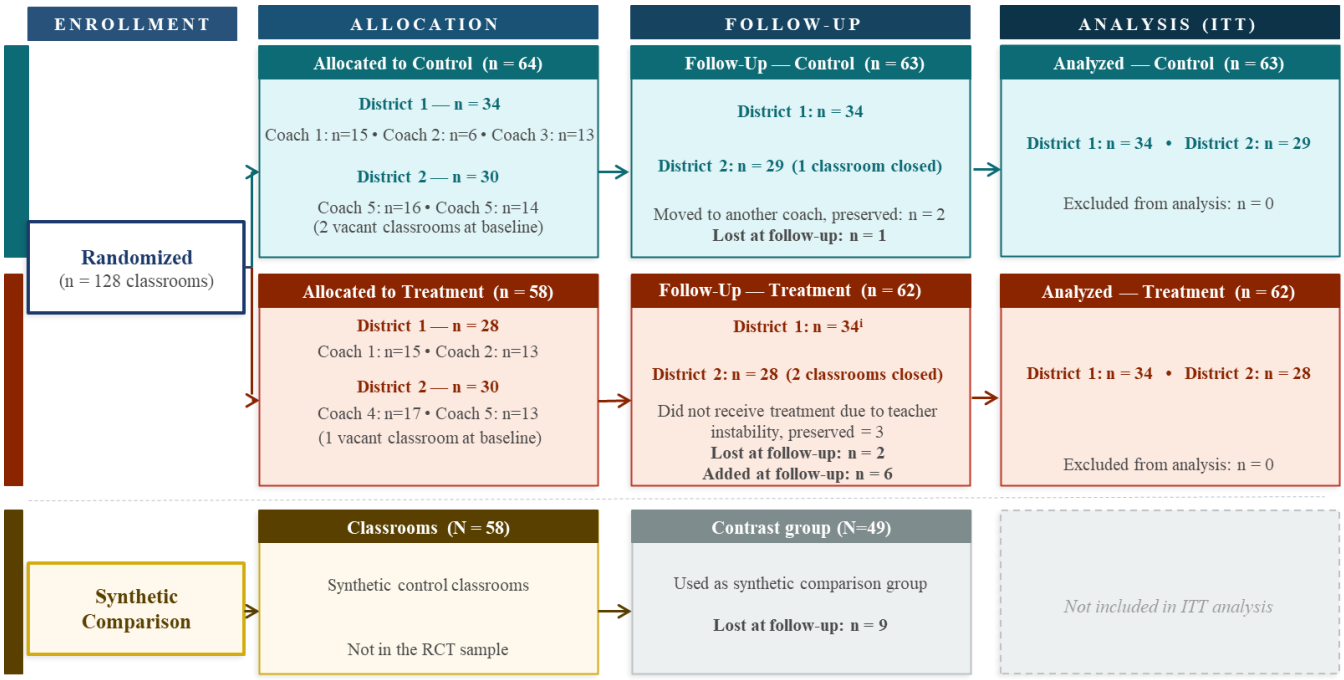
This study implements a cluster randomized controlled trial (RCT) across two districts in New Jersey's state preschool program. The research aims to assess the effectiveness of using the full ecosystem with exogenous differences induced by the additional professional development and coaching provided by Teaching Strategies to the treatment group, relative to practice-as-usual in these two districts (with a control group in each) and relative to other districts across the state using any of the four state-approved curricula at the time of the study (an additional synthetic control arm).

Specifically, one group of coaches and the teachers/classrooms they work with in two districts serve as the primary control group (n = 64) receiving no intervention, while the other group of coaches and the teachers/classrooms they work with form the treatment group receiving the 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 arm consisting of a secondary control group of teachers/classrooms where the outcomes are assessed. This third group is a set of randomly selected preschools classrooms across districts in the same state, regardless of which curriculum (of the four state approved options) they were implementing (n = 58).

Figure 2 illustrates the CCIEE Multi-Arm Design, with the sample at baseline (fall of year 1) and follow-up (spring of year 3) and tracks changes in the sample, as per CONSORT guidelines (Hopewell et al., 2025). Two treatment classrooms and one control classroom closed; plus, lead teacher positions in one treatment classroom and two control classrooms were vacant at baseline but filled at follow-up (one filled shortly after baseline). Three treatment classrooms had inconsistent teachers and therefore ended up in the control group; 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 who remained in the treatment group.

Figure 2. Randomization tree for the CCIEE Multi-Arm Design



Note: [i] Teachers added to coach 1 and considered treatment despite late entry: n = 6. Teachers that were shifted from coach 1 to a control coach but remained in the treatment group: n = 7.

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 aligned assessment (*GOLD*) components. Consistency in structural quality mitigates contextual differences that could potentially confound internal validity. In addition, preexisting familiarity with the core components bypasses the exploration and initial "ramping up" phase that occurs during the first several years in the implementation of any curriculum, narrowing the focus to differences in the 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 implementation, 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 universal adoption of the digital ecosystem in both districts in the study, a consequence of the push for digital education technology during and after the COVID-19 pandemic. Since the ecosystem was quite new, it was expected that the intervention supports would still induce higher engagement among treated teachers (relative to the controls). However, during the study, one of the two districts implemented district-wide expectations for and monitoring of teacher's digital curriculum

activities and assessment practices, reducing treatment contrast.[2] As a result, the study should be interpreted as a conservative test of the ecosystem's impact.

Implementation processes and outputs are captured via intervention "dosage," curriculum fidelity of implementation, and engagement. Intervention dosage is measured by attendance to synchronous, virtual professional development sessions, and completion of supplementary on-demand asynchronous PD courses. Curriculum fidelity is captured with Teaching Strategies' fidelity tool that was specifically designed to measure implementation of *The Creative Curriculum*. *SmartTeach*™ platform engagement is measured as total time spent using the core digital components of the ecosystem—the Teach, Assess, Family, Reports, and Library areas—throughout the 3-year study period.

Primary outcomes include teacher retention and classroom quality outcomes. Teacher retention is operationalized as whether a teacher remained at the same site throughout the three-year study period. Classroom outcome measures used to evaluate the impact of the intervention on teaching practices include two measures of classroom quality with baseline observations conducted in the fall/winter of the 2021-22 school year (year 1) and follow-up observations conducted in the winter/spring of the 2023-24 school year (year 3). Among other constructs, teacher well-being and burnout were also measured via survey at the end of years 1 and 3.

Child outcomes were measured via direct assessment of children's cognitive development (language, literacy, and mathematics), children's executive functions, socio-emotional development (teacher report), and creativity. These direct measures were complemented with ongoing authentic, observational assessments of children's social-emotional and cognitive development, as well as language, literacy, and mathematics skills using *GOLD*. Child assessments were conducted in year 3, during the fall of 2023 (baseline) and again the spring of 2024 (post-test) to assess children's growth over the school year.

All elements were measured in the treatment arm, control arm, and synthetic comparison group, except platform engagement, fidelity of curriculum use, and children's formative assessment (*GOLD*), which were limited to the main arms of the study.

[2] This was noted in a coaching session on February 17, 2023, in which lesson plan requirements by the district for assessment and curriculum were discussed and documentation was shared.

Table 1. Research Data Sources

Domain	Attendance	Ecosystem Engagement	Classroom Observations	Surveys	External Child Assessment	*GOLD* Child Assessment	Focus Groups /Feedback*
Coaches							
Training	✓			✓			
Dosage	✓						
Characteristics				✓			
Teachers							
Training	✓			✓			
Dosage	✓						
Characteristics				✓			
Context							
Teacher Burnout				✓			
Barriers/Facilitators				✓			✓
Engagement		✓					
Fidelity			✓				
Quality			✓				
Child Outcomes							
Language					✓	✓	
Literacy					✓	✓	
Mathematics					✓	✓	
Ex. Function					✓		
Creativity					✓		
Socio-Emotional				✓		✓	
Cognitive						✓	

*Focus group with treatment coaches. Teacher feedback elicited through surveys.

Table 2. Roles and responsibilities of the CCIEE Study

Organization	Roles & Responsibilities
Teaching Strategies	Provide the materials for coaches and teachers, grant access to the Teaching Strategies digital platform, coordinate and provide professional development sessions, track PD session attendance, track use of the digital platform, conduct curriculum fidelity observations, and deliver *GOLD* assessment data for children in the sample classrooms.
Trainer/facilitator	Provide virtual PD to teachers and coaches in the treatment group.
Coaches	Attend virtual PD (treatment), utilize the tools and coaching supports in their work with teachers, attend remotely facilitated individual "side-by-side" and group coaching sessions. Complete a survey.
Lead Teachers	Attend virtual PD (treatment), use the online tools, implement the curriculum, track child progress in *GOLD*. Complete a baseline and post-test survey.
NIEER Research Team	Conduct classroom quality observations & direct child assessments, disburse incentives, and track the sample.
District Leadership	Provide information about the coaches and teachers at baseline to randomize the sample; support communication with teachers and coaches; support efforts to collect data by providing information to the schools/centers, coaches, and teachers; and provide information on changes in the sample over time, provide child demographics for children in the sample.

Evaluation Processes

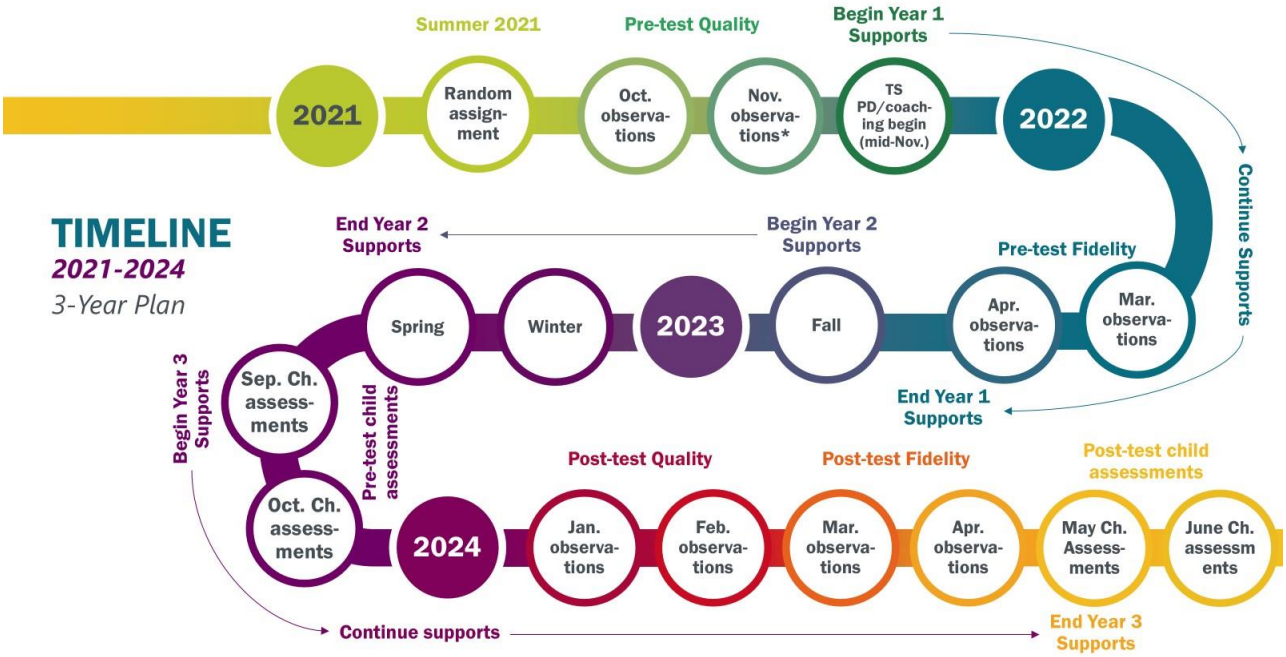
This section delves into the study's timeline, measures, and procedures. It is a 3-year study with professional development and supports starting in year 1, continuing throughout years 2 and 3, and with data collection in years 1 and 3. Two districts were recruited among those implementing the New Jersey preschool program for some decades. Both were already utilizing *The Creative Curriculum* and *GOLD* for many years prior to the study. One district is located in the northern region of New Jersey and the other one in the southern region of the state. Both districts are highly diverse although the one in the south is almost universally composed of Hispanic/Latinx (about 50% of the student population) and African American (about 40%) children. We discuss this district throughout as district one. The district in the north has about two-thirds Hispanic/Latinx children, with an additional large proportion (about 20%) of African American children. We discuss this district as district two.

Timeline

The study timeline is depicted in Figure 3. Random assignment was conducted in fall 2021, once the participating districts had finalized their teacher and coach lists. Teaching Strategies then started defining supports and planning for treatment coaches and teachers with each of the two districts. The randomly selected treatment and control teachers in the study were observed starting in the fall of 2021 through February of 2022 (baseline data collection), as supports were just beginning.[3] This was also the case in the "synthetic" group recruited across the state. Supports continued through spring 2022, throughout the 2022–23 school year and finally ended in the spring of the 2023–24 school year. The research team tracked any changes in teachers and coaches throughout this period. Classroom observations were again collected in the spring of 2024. Child assessments were also collected in control, treatment and "synthetic" classrooms in fall 2023 and again in spring 2024.

[3] As the U.S. appeared to be coming out of the COVID-19 pandemic, the study was launched. However, the Delta-Omicron waves of the COVID-19 pandemic occurred in the fall 2021 through spring 2022, which created delays in the implementation of the ecosystem, PD and baseline data collection on classrooms throughout this first school year. A few observations at baseline, therefore, occurred in May and June 2022.

Figure 3. CCIEE Study Timeline



Note: Includes teacher, coach, and family surveys layered in, plus, treatment coach focus groups (spring 2024). *Extended into Feb. 2022 due to the COVID-19 pandemic. Supports were provided between Nov. and June of the first school year, Oct. and June of the second school year, and Sept. and May of the third school year.

The professional development and coaching plan implemented by Teaching Strategies is depicted in Figure 4. The plan involved providing teachers with PD on the ecosystem, including *The Creative Curriculum* and *GOLD*, as well as helping coaches refine their coaching skills to support teachers with use of the ecosystem. The sessions included learning about the components of the new ecosystem solution, examining the 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, reflecting on the teaching and assessment cycle and responsive planning, implementing the ecosystem to fidelity, and more. The early coaching sessions focused on understanding the *Fidelity Tool for Administrators* and *Coaching to Fidelity*—tools that enable programs to measure teachers' fidelity of implementation—and use of that information to help teachers at all levels of fidelity improve their practice. After these early level-setting sessions, coaching involved ongoing remote support in the form of individual "side-by-side" and group follow-up sessions with Teaching Strategies' trainer.

Figure 4. Professional Development and Coaching Plan

Study Year	Quarter	*District 1*		*District 2*		*Coaches*	
		Topic(s)	**Hours**	**Topic(s)**	**Hours**	**Topic(s)**	**Hours**
1	Fall	Getting Started with *The Creative Curriculum*	3	Getting Started with *The Creative Curriculum*	3	Coaching to Fidelity	6
1	Winter	Getting Started with *The Creative Curriculum* (3 hours), Introduction to *GOLD* (6 hours)	9	Getting Started with *The Creative Curriculum* (3 hours), Supporting Distance Learning (1.5 hours), Introduction to *GOLD* (4.5 hours)	9	Individual side-by-side (6 hours) & group follow-up coaching sessions (4 hours)	10
1	Spring	Introduction to the Digital Curriculum (3 hours), Supporting Meaningful Experiences (6 hours)	9	Introduction to *GOLD* (1.5 hours), Responsive Planning (3 hours)	4.5	Individual side-by-side (3 hours) & group follow-up coaching sessions (4 hours)	7
2	Fall	Digital Ecosystem Implementation	1.5	Assessment: Observe and Collect Facts (2 hours), Digital Ecosystem Implementation (1.5 hours)	3.5	Individual side-by-side (3 hours) & group follow-up coaching sessions (2 hours)	5
2	Winter	Curriculum Implementation Fidelity (1.5 hours), Connecting Curriculum & Assessment (1.5 hours), Supporting Learning & Development with the Digital Curriculum (1.5 hours)	4.5	Curriculum Implementation Fidelity (1.5 hours), Connecting Curriculum & Assessment (1.5 hours), Supporting Learning & Development with the Digital Curriculum (1.5 hours)	4.5	Individual side-by-side (6 hours) & group follow-up coaching sessions (4 hours)	10
2	Spring	Supporting Learning & Development with the Digital Curriculum (1.5 hours), Language Development & Multilingual Learners (1.5 hours), Building a Bridge to Kindergarten (1.5 hours)	4.5	Supporting Learning & Development with the Digital Curriculum (1.5 hours), Language Development & Multilingual Learners (1.5 hours), Building a Bridge to Kindergarten (1.5 hours)	4.5	Individual side-by-side (3 hours) & group follow-up coaching sessions (2 hours)	5
3	Fall	*The Creative Curriculum* for Preschool: The First Six Weeks (1.5 hours), Reinforcing All Areas of Learning & Development with Curriculum and Assessment (3 hours), Supporting Language/Literacy Development and Learning (1.5 hours)	6	Connecting All Areas of Learning & Development with Curriculum and Assessment (1.5 hours), Supporting Language/Literacy Development and Learning (1.5 hours)	3	Individual side-by-side (2 hours) & group follow-up coaching sessions (2 hours)	4
3	Winter	Unlocking the Power of Formative Assessment (1.5 hours), Using Intentional Teaching Experiences to Scaffold Learning & Development (3 hours)	4.5	Unlocking the Power of Formative Assessment (1.5 hours), Using Intentional Teaching Experiences to Scaffold Learning & Development (3 hours)	4.5	Individual side-by-side (2 hours) & group follow-up coaching sessions (2 hours)	4
3	Spring	Using Intentional Teaching Experiences to Scaffold Learning & Development (1.5 hours), Spotlight on End of Year Expectations (1.5 hours)	3	Spotlight on End of Year Expectations (1.5 hours)	1.5	Individual side-by-side (2 hours) & group follow-up coaching sessions (2 hours)	4

Study Measures

Measures on Classrooms

The study included three measures assessing classroom quality and fidelity:

- The *Early Childhood Environment Rating Scale—Third Ed.* (ECERS-3; Harms, Clifford & Cryer, 2015) is an observation and rating instrument for preschool classrooms serving children ages three to five. The total ECERS-3 score represents an average of the scores on the 35 items under 6 domains. A rating scale between 1 and 7 is used, where a rating of 1 indicates inadequate quality, a rating of 3 indicates minimal quality, a rating of 5 indicates good quality, and a rating of 7 indicates excellent quality. The most updated Notes for Clarification[4] were utilized when scoring all classrooms in this sample.
- The *Sustained Shared Thinking and Emotional Wellbeing (SSTEW)* departs from other measures in that it assesses the presence of a culture in the setting that supports children's curiosity, thinking, and questioning and that children are seen engaged in appropriate, cognitively challenging activities and discussions. It also looks at educator support for 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 (Howard et al., 2018).
- The *Fidelity Tool: The Creative Curriculum* is an observational assessment designed to evaluate how effectively educators implement the curriculum in early childhood classrooms. The tool assesses physical classroom environment; the use of curricular materials; structure and balance of daily plans inclusive of large group, small group, and transitions, and the quality of interactions; while also reviewing how teachers observe, document, and use data to individualize instruction and engage with families. The tool supports teachers, coaches, and administrators by providing actionable insights related to the adherence to developmentally appropriate practices and the curriculum's framework. Because the tool significantly overlaps with ECERS-3 and the SSTEW, we limited the measure to only include the curriculum "Use" subscale for *The Creative Curriculum*.

Measures on Children

A first battery of child measures includes socio-emotional skills, executive functioning, creativity and problem solving, with all but the socio-emotional skills measured via standardized assessment of children in their classrooms:

- 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.
- The Strengths and Difficulties Questionnaire (SDQ: Goodman, 2001) is a brief behavioral questionnaire for 2- to 17-year-old children.
- The Penn Interactive Peer Play Scale (PIPPS) is a behavioral rating instrument used by teachers (or caregivers) to observe and measure peer play behaviors (Fantuzzo & McWayne, 2002).

[4] As published in https://www.ersi.info/additionalnotes at the time of the observations.

- Given the focus of the curriculum, we also included a measure of creativity. We use the Thinking Creatively in Action and Movement test (TCAM; Zachopoulou, Makri, & Pollatou, 2009) which 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.

A second battery includes cognitive measures of receptive vocabulary, literacy and math, also measured through standardized assessment of children.

- The Peabody Picture Vocabulary Test–Fourth Edition (PPVT-IV; Dunn & Dunn, 2007) a 204-item 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 Letter-Word Identification and Applied Problems subtests, broad reading and math measures.
- For children with Spanish language proficiency, we also used the equivalent subtests from the Woodcock-Muñoz Batería III (Schrank et al., 2005).

In addition, we integrated measures of language, literacy, and mathematics knowledge, skills and abilities as well as social-emotional and cognitive development measured via teachers' regular practice of using the curriculum-aligned *GOLD* assessment. The objectives measured by each domain (i.e., "area") of *GOLD* are listed below.[5]

- The social-emotional domain measures the following objectives:
  - Regulates own emotions and behaviors
  - Establishes and sustains positive relationships
  - Participates cooperatively and constructively in group situations
- The cognitive domain measures the following objectives:
  - Demonstrates positive approaches to learning
  - Remembers and connects experiences
  - Uses classification skills
  - Uses symbols and images to represent something not present
- The language domain measures the following objectives:
  - Listens to and understands increasingly complex language
  - Uses language to express thoughts and needs
  - Uses appropriate conversational and other communication skills
- The literacy domain measures the following objectives:
  - Demonstrates phonological awareness, phonics skills, and word recognition
  - Demonstrates knowledge of the alphabet
  - Demonstrates knowledge of print and its uses
  - Comprehends and responds to books and other texts
  - Demonstrates writing skills
- The mathematics domain measures the following objectives:
  - Uses number concepts and operations
  - Explores and describes spatial relationships and shapes

[5] We excluded the *Physical* domain from analyses, as this developmental domain was not in the study protocol.

- Compares and measures
- Demonstrates knowledge of patterns

Survey Measures and Administrative Data

The study also collected child demographic information (from administrative data and family surveys), as well as surveyed teachers on their experience and qualifications, some curriculum fidelity components, family engagement, the amount of training and professional development they received, and demographics. Families were surveyed on the amount of time they engage in language and math practices with children in a typical week. Lastly, the work also included coach surveys with information requested on experience and qualifications, typical interaction and coaching strategies 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.

Study Procedures

NIEER and Teaching Strategies worked collaboratively to select the measures for the study. All study procedures, protocols and surveys underwent Institutional Review Board approvals. Data collectors for child assessment and classroom observation measures were trained by NIEER. For child assessments, data collectors received a two-day training on the measures. Following the two-day training, data collectors were successfully shadowed by expert staff on two iterations of the assessments for reliability. Each assessor was shadow scored to ensure 100% accuracy in assessment. For the TCAM, intraclass correlation (ICC) varied between 0.81 and 0.97 with up to four reliability iterations done when necessary. A refresher training took place prior to the beginning of spring data collection. Classrooms were provided with consent forms (in English and Spanish) to distribute to all children in the classroom. Children were randomly selected among those for whom a consent was received, and in most classrooms, four children were included in the sample. Children were assessed in the fall between October and December of 2023 (with some Spanish assessments taking place in January/February 2024) and again in the spring between May and June of 2024. Spanish-speaking children were assessed in English and in Spanish on the PPVT and Woodcock Johnson (by Spanish-speaking bilingual data collectors). Teachers were the primary informants of whether a child should be assessed in Spanish in addition to English. However, if a Spanish-English bilingual child did not get past set 3 in the PPVT-IV in English, they were also assessed in Spanish. The DCCS and Pencil Tap tasks were administered only once, in the child's dominant language. Each assessor was shadow scored to ensure assessment accuracy. The socio-emotional and play surveys on children (SDQ and PIPPS) were collected from the teachers in paper or online. Teachers received a gift card of $5 per child completed at each assessment point (fall and spring).

For classroom observation measures, initial training was provided in administering the observation protocol that includes the ECERS-3 and the SSTEW for pre-k classrooms. Training took place in separate full-day workshops. ECERS-3 observers were trained by an ECERS-3 certified trainer and met the Environment Rating Scales Institute (ERSI) reliability requirements for observer certification. The trainee must complete three observations with the trainer with 85% or above exact matches or one-away from the true score. All data collectors met the ECERS-3 reliability requirements with agreement percentages ranging between 85-94% (with a

mean of 89%). SSTEW observers were trained by the authors in a three-day training (which included the developers' online modules), followed by an online discussion with the authors. This was then followed by three reliability sessions with the NIEER anchor team, and another meeting with the developers for additional clarifications based on reliability. All data collectors met reliability requirements for the SSTEW with agreement percentages ranging between 80-93% (with a mean of 89%). All observation score sheets were cleaned, entered and analyzed at NIEER. Appendices provide additional information on internal consistency and the factor structure of the tools. These analyses confirm a one factor solution for both measures. We therefore use the average SSTEW and ECERS-3 scores across our analyses.

Family surveys were initially distributed in the fall of 2023 and could be completed either online or on paper in English and Spanish. Families received a $20 gift card for completing the survey. NIEER staff followed up with families and collaborated with schools throughout the school year to obtain the survey information. Only a 60% response rate was attained with families (see Harmeyer & Nores, 2026). Demographics for children are therefore primarily incorporated from district data.

Teacher surveys were distributed in the fall of 2021 and collected throughout the school year, and again in the fall of 2023 (also collected throughout the school year). A 70% (n=123) response rate was attained at baseline and 90% (n=156) at follow-up. Teachers received a gift card of $50 for completing these. In years 2 and 3, teachers were also provided with incentives of $125 per each of the three school-year quarters for completing the PD sessions with Teaching Strategies. For teachers who were not provided with school time to engage in the trainings and had to do these after hours, an additional $20 was distributed per hour of training. All participating classrooms received *The Essentials Kit* from Teaching Strategies. Synthetic sites received $100 in each data collection wave (baseline observations, child assessments, and follow-up observations) per site for participating in the study.

Baseline Analyses

Coach and Teacher characteristics

All coaches regardless of control versus treatment randomization had a master's degree. In addition, they were all female. Average experience as a coach was 12.4 years (12 in the control group and 13 in the treatment group). Their experience in early childhood was on average 23.3 years (22.2 in the control group and 24.2 in the treatment group). The coach groups were, on average, quite similar at baseline.

Teacher characteristics were also balanced between control and treatment groups in terms of teacher education and gender. The statistically significant differences across groups were the years of experience, with the teachers in classrooms randomized to treatment exhibiting on average higher experience among those reporting it, and for the distribution of race and ethnicity.

Table 3: Teacher self-reported characteristics by Control vs. Treatment groups at post-test, n=125.

Variable	Control				Treatment				P-value
	N	Mean/ Percent	SD	95% CI	N	Mean/ Percent	SD	95% CI	
Female	55	87.30%			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	4.58-7.85	61	11.18	8.89	8.90-13.45	0.001***
Missing	7				1				0.331
Race/ethnicity	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%			

Notes: P-values for chi-squared tests of independence. The table includes demographics for all classrooms, regardless of the data collection time-point, as the baseline response rate was lower. For teachers without post-test data, but retained throughout the study, pre-test survey demographic data is used. Closed classrooms are not included. ***p<0.01, **p<0.05, *p<0.1 Descriptives for all 128 classroom included in the study as per Figure 2 similarly align.

Classroom Quality

Table 4 compares mean scores for various ECERS and SSTEW subscales at baseline between control and treatment groups. Most differences are not statistically significant, with the exception of Personal Care Routines, where the control group scored significantly higher than the treatment group (p=0.033). Other dimensions, such as Language and Literacy and Program Structure, showed negligible differences, suggesting similar practices at baseline across groups. Total ECERS-3 scores did not differ. For the SSTEW subscales, the mean scores also reflect no statistically significant differences across groups at baseline. The total SSTEW score is nearly identical across groups, reinforcing limited systematic differences overall.

Table 4: Difference in ECERS and SSTEW scores by Control vs. Treatment groups at baseline.

Variables	Control				Treatment				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	60	3.74	0.82	3.53-3.95	56	3.60	0.82	3.38-3.82	0.353
Personal Care Routines	60	4.00	1.07	3.72-4.28	56	3.56	1.11	3.26-3.86	0.033**
Language and Literacy	60	3.99	1.00	3.73-4.25	56	4.01	1.06	3.73-4.3	0.901
Learning Activities	60	3.17	0.79	2.96-3.37	56	3.12	0.76	2.92-3.33	0.756
Interaction	60	4.35	1.12	4.06-4.64	56	4.26	1.35	3.89-4.62	0.695
Program Structure	59	4.23	1.17	3.93-4.54	56	4.20	1.23	3.87-4.53	0.875
Total ECERS	*60*	*3.76*	*0.72*	*3.58-3.95*	*56*	*3.65*	*0.80*	*3.44-3.87*	*0.444*
SSTEW									
Building trust, confidence, & indep.	58	3.16	1.08	2.87-3.44	57	3.22	1.14	2.91-3.52	0.769
Social & emotional wellbeing	58	2.19	1.03	1.92-2.46	57	2.44	1.18	2.13-2.75	0.231
Supporting & extending language	58	3.36	1.30	3.02-3.7	57	3.33	1.30	2.99-3.68	0.920
Supporting learning & critical th.	58	1.60	0.35	1.51-1.69	57	1.70	0.39	1.59-1.8	0.156
Assessing learning & language	58	2.59	0.89	2.36-2.83	56	2.52	0.81	2.3-2.73	0.630
Total SSTEW	*58*	*2.62*	*0.77*	*2.42-2.82*	*57*	*2.66*	*0.80*	*2.45-2.87*	*0.766*

Note: ***p<0.01, **p<0.05, *p<0.1. ECERS-3 was completed in 116 classrooms, SSTEW was completed in 115 classrooms.

In contrast, the synthetic and treatment groups evidenced important differences (Table 5). The synthetic group generally outperformed the treatment group at baseline, with significant differences observed in Language and Literacy (p=0.047), Learning Activities (p=0.006), Program Structure (p=0.020) and Total ECERS score (p=0.011). For SSTEW, the subscales for Building Trust, Confidence, and Independence (p=0.025), Social and Emotional Wellbeing (p=0.033) and Supporting and Extending Language (p=0.002) also favor the synthetic group. The total SSTEW score is therefore higher for the synthetic group (p=0.011). These results illustrate the study's focus on two low-income districts in need of additional resources.

Table 5: Difference in ECERS and SSTEW scores Synthetic vs. Treatment groups at baseline

Variables	Synthetic				Treatment				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	58	3.70	0.64	3.54-3.87	56	3.60	0.82	3.38-3.82	0.434
Personal Care Routines	58	3.92	1.04	3.65-4.2	56	3.56	1.11	3.26-3.86	0.077*
Language and Literacy	58	4.38	0.89	4.14-4.61	56	4.01	1.06	3.73-4.30	0.047**
Learning Activities	58	3.58	0.96	3.33-3.84	56	3.12	0.76	2.92-3.33	0.006***
Interaction	58	4.68	1.11	4.39-4.98	56	4.26	1.35	3.89-4.62	0.069*
Program Structure	58	4.74	1.20	4.42-5.05	56	4.20	1.23	3.87-4.53	0.020**
Total ECERS	*58*	*4.02*	*0.72*	*3.83-4.21*	*56*	*3.65*	*0.80*	*3.44-3.87*	*0.011***
SSTEW									
Building trust, confidence, & indep.	55	3.70	1.09	3.4-3.99	57	3.22	1.14	2.91-3.52	0.025**
Social & emotional wellbeing	55	2.95	1.30	2.59-3.3	57	2.44	1.18	2.13-2.75	0.033**
Supporting & extending language	55	4.12	1.32	3.76-4.48	57	3.33	1.30	2.99-3.68	0.002***
Supporting learning & critical th.	55	1.86	0.52	1.72-2.01	57	1.70	0.39	1.59-1.80	0.058*
Assessing learning & language	54	2.44	0.90	2.19-2.68	56	2.52	0.81	2.30-2.73	0.612
Total SSTEW	*55*	*3.06*	*0.83*	*2.83-3.28*	*57*	*2.66*	*0.80*	*2.45-2.87*	*0.011***

Note: ***p<0.01, **p<0.05, *p<0.1.

Classroom Fidelity

Table 6 reports baseline scores for the fidelity measure on curriculum use, for treatment and control classrooms (synthetic classrooms were not observed using the fidelity tool) that were observed. The measure reports the percentage of indicators present (rated as 1) in the classroom at the time the classrooms were observed by Teaching Strategies' raters. At baseline fidelity scores across the treatment and control classrooms were likely higher in the treatment group (p=.099).

Table 6: Difference in Fidelity "Use" scores by Control vs. Treatment groups at baseline.

Variables	Control				Treatment				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
Fidelity Use Items (%)	55	0.465	0.191	0.413-0.516	53	0.524	0.179	0.480-0.574	0.099*

Note: ***p<0.01, **p<0.05, *p<0.1. The fidelity measure included only indicators on the use of the curriculum. Excludes classrooms that closed.

Outcome Analyses

Descriptives

Teacher Retention

Retention differed significantly between treatment and control groups, with higher retention in the treatment group for both districts, but this is driven mostly by higher retention rates in District 1. In contrast, a larger portion of the control teachers retained their coach in district 2.[6]

Table 7: Retention Rates by Control vs. Treatment Status.

Variables	**Control**		**Treatment**		P-value (T-test
	N	%	N	%	% Retention)
Teachers					
Total	63	50.79	62	80.64	0.000***
District 1	34	44.12	34	82.35	0.001***
District 2	29	58.62	28	78.57	0.110
Teachers retained their coach					
Total	63	44.44	62	53.22	0.330
District 1	34	35.29	34	64.70	0.015**
District 2	29	55.17	28	39.28	0.237

Note: ***p<0.01, **p<0.05, *p<0.1. Does not count closed classrooms. Considers two vacant classrooms filled early on as retained, and only considers a third one that was filled in the last year of the study as a churn. Similarly, a classroom with a disability leave at the beginning of the study that was then filled with a teacher, is considered churn as well. When not counting any of the vacant classrooms that got filled as churn, the retention rate for control classrooms is 50.82% and the retention rate for treatment classrooms is 80.32%. The statistically significant difference remains.

[6] This is driven by one retirement of a coach in the treatment group for which the new coach was brought into the study and received the same supports; also, a set of teachers moved to a control coach (but this set of teachers continued the supports throughout).

Classroom Quality

To start, we report ECERS-3 and SSTEW scores at follow-up for a set of comparison groups. Table 8 reports these by intent-to-treat (ITT). This includes all the classrooms randomized at baseline, regardless of whether they were treated. As reported earlier, three classrooms were in fact not treated as they had continuous changes in teachers and substitutes that did not allow an opportunity to engage them. As a consequence, we also report in Table 9 by Treatment-on-Treated (TOT), where these three classrooms cross into the control group.[7] Lastly, we also report quality (Table 10) comparisons for classrooms for which there were no teacher changes. Average effects and confidence intervals for control, ITT, TOT and the TOT retained group are shown in Figure 4.

As reported in Table 8, compared to average scores in the control group, average scores in the ITT group at follow-up were slightly lower across most subscales, with the exception of ECERS Program Structure and SSTEW Supporting Learning & Critical Thinking. The differences between the control and ITT group were, however, not statistically significant for ECERS subscales nor the total score. They were trending statistically significant ($p < 0.10$) for two of the SSTEW subscales but not for the total score.

Table 8: Mean ECERS-3 and SSTEW scores by Control vs. Intent-to-Treat at Follow-up.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	63	3.91	0.59	3.77-4.06	61	3.87	0.69	3.69-4.05	0.711
Personal Care Routines	63	3.57	0.97	3.33-3.82	61	3.39	1.19	3.09-3.70	0.356
Language and Literacy	63	4.29	0.91	4.06-4.51	61	4.26	1.02	4.00-4.52	0.893
Learning Activities	63	3.66	0.63	3.5-3.82	61	3.56	0.91	3.32-3.79	0.451
Interaction	63	4.33	1.17	4.03-4.62	61	4.21	1.42	3.85-4.58	0.627
Program Structure	63	4.16	1.13	3.87-4.44	61	4.23	1.37	3.88-4.58	0.763
Total ECERS	*63*	*3.93*	*0.63*	*3.77-4.09*	*61*	*3.85*	*0.88*	*3.63-4.08*	*0.572*
SSTEW									
Building trust, confidence, & indep.	63	3.37	1.34	3.03-3.71	62	2.90	1.54	2.51-3.29	0.070*
Social & emotional wellbeing	63	4.27	1.99	3.77-4.77	62	4.10	2.06	3.57-4.62	0.633
Supporting & extending language	63	3.27	1.23	2.96-3.58	62	3.22	1.27	2.89-3.54	0.816
Supporting learning & critical th.	63	2.06	0.62	1.90-2.22	62	2.15	0.93	1.91-2.38	0.544
Assessing learning & language	63	3.17	1.41	2.81-3.52	62	2.73	1.44	2.36-3.09	0.085*
Total SSTEW	*63*	*3.00*	*0.93*	*2.77-3.24*	*62*	*2.84*	*1.14*	*2.55-3.12*	*0.370*

Note: ***p<0.01, **p<0.05, *p<0.1. ECERS-3 was completed in 124 classrooms. SSTEW was completed in 125 classrooms.

Table 9 reports average quality scores for the ECERS and SSTEW by TOT status. There are no statistically significant differences between groups at follow-up, with average scores being quite closer across groups for all subscales and total scores. Table 10 reports mean quality scores constraining the TOT group to classrooms for which there were no teacher changes, and while scores are a bit higher for some subscales, these are not statistically significantly different from the control group. Analyses of the change in scores between pre and post-test in control and treatment classrooms provide similar patterns (Appendix C).

[7] This crossover occurred only in district 1.

Table 9: Mean ECERS-3 and SSTEW scores by Control vs. Treatment-on-Treated at Follow-up.

Variables	Control				Treatment-on-Treated				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	66	3.88	0.59	3.74-4.03	58	3.91	0.69	3.73-4.09	0.825
Personal Care Routines	66	3.55	0.97	3.31-3.79	58	3.41	1.21	3.09-3.72	0.468
Language and Literacy	66	4.24	0.95	4.01-4.47	58	4.31	0.98	4.06-4.57	0.669
Learning Activities	66	3.61	0.68	3.44-3.78	58	3.61	0.88	3.38-3.84	1.000
Interaction	66	4.27	1.19	3.98-4.56	58	4.27	1.42	3.90-4.65	0.991
Program Structure	66	4.13	1.13	3.85-4.41	58	4.26	1.38	3.90-4.62	0.565
Total ECERS	*66*	*3.89*	*0.66*	*3.73-4.05*	*58*	*3.90*	*0.87*	*3.67-4.13*	*0.944*
SSTEW									
Building trust, confidence, & indep.	66	3.29	1.36	2.95-3.62	59	2.97	1.55	2.56-3.37	0.219
Social & emotional wellbeing	66	4.20	1.99	3.71-4.69	59	4.17	2.07	3.63-4.71	0.940
Supporting & extending language	66	3.22	1.24	2.92-3.53	59	3.27	1.27	2.94-3.60	0.847
Supporting learning & critical th.	66	2.03	0.62	1.88-2.19	59	2.18	0.94	1.93-2.42	0.309
Assessing learning & language	66	3.09	1.42	2.74-3.44	59	2.79	1.44	2.41-3.16	0.240
Total SSTEW	*66*	*2.95*	*0.94*	*2.72-3.18*	*59*	*2.89*	*1.14*	*2.59-3.18*	*0.746*

Note: ***p<0.01, **p<0.05, *p<0.1. Three classrooms assigned to the treatment group where never treated due to teacher instability. ECERS-3 was completed in 124 classrooms. SSTEW was completed in 125 classrooms.

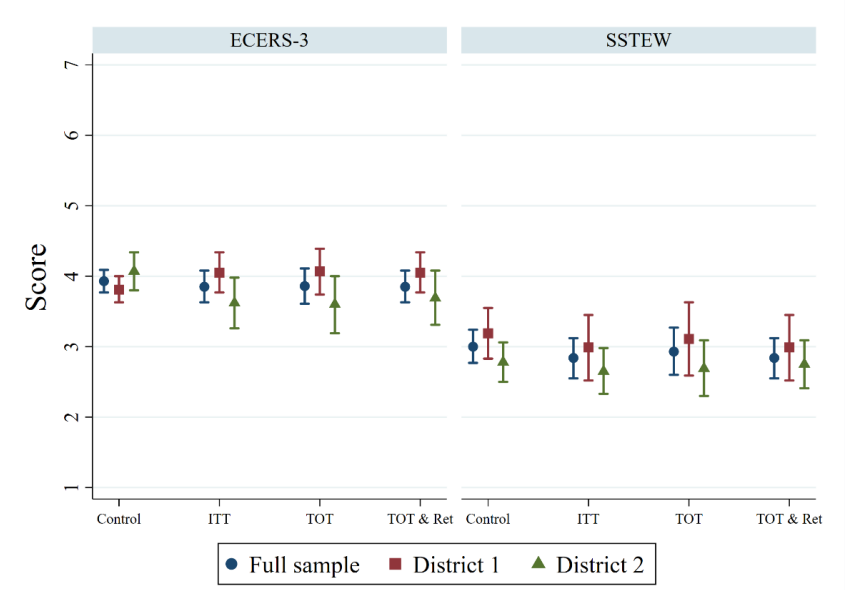
Table 10: Mean ECERS-3 and SSTEW scores by Control vs. Treatment-on-Treated & Retained at Follow-up.

Variables	Control				Treatment-on-Treated Retained				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	66	3.88	0.59	3.74-4.03	49	3.9	0.73	3.69-4.11	0.872
Personal Care Routines	66	3.55	0.97	3.31-3.79	49	3.34	1.19	3.00-3.68	0.298
Language and Literacy	66	4.24	0.95	4.01-4.47	49	4.26	0.95	3.99-4.53	0.903
Learning Activities	66	3.61	0.68	3.44-3.78	49	3.51	0.85	3.27-3.76	0.508
Interaction	66	4.27	1.19	3.98-4.56	49	4.2	1.36	3.81-4.59	0.770
Program Structure	66	4.13	1.13	3.85-4.41	49	4.25	1.41	3.84-4.65	0.622
Total ECERS	*66*	*3.89*	*0.66*	*3.73-4.05*	49	3.91	0.95	3.64-4.18	0.817
SSTEW									
Building trust, confidence, & indep.	66	3.29	1.36	2.95-3.62	50	2.99	1.6	2.53-3.44	0.276
Social & emotional wellbeing	66	4.20	1.99	3.71-4.69	50	4.26	2.09	3.67-4.85	0.869
Supporting & extending language	66	3.22	1.24	2.92-3.53	50	3.32	1.26	2.96-3.68	0.680
Supporting learning & critical th.	66	2.03	0.62	1.88-2.19	50	2.21	0.95	1.94-2.48	0.231
Assessing learning & language	66	3.09	1.42	2.74-3.44	50	2.89	1.45	2.48-3.30	0.457
Total SSTEW	*66*	*2.95*	*0.94*	*2.72-3.18*	50	3.13	1.26	2.78-3.49	0.878

Note: ***p<0.01, **p<0.05, *p<0.1.

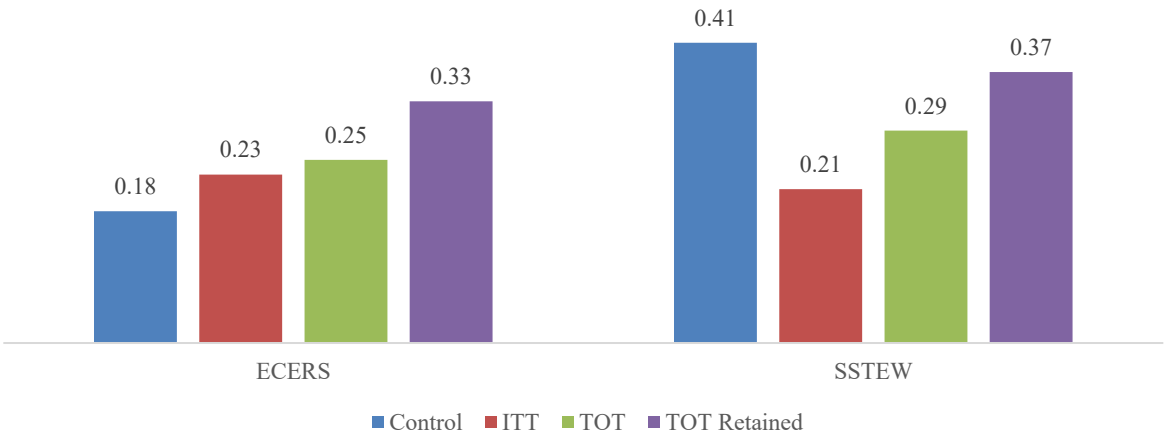
Analyses by district show a slightly different pattern (Figure 4). On average, for the ECERS, District 1's scores are above those of District 2 in all groups except the control group. Similarly, for the SSTEW, District 1's scores are above those of District 2 across all groups. These differences were however not statistically significant.

Figure 4. Mean ECERS-3 and SSTEW scores by Control vs. Treatment groups and district at Follow-up.



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 (Appendix C) show gains for all groups across both measures of quality (Figure 5). Moreover, growth in scores was higher for the treatment-on-treated group for teachers retained in classrooms.

Figure 5. Mean ECERS-3 and SSTEW growth in scores by Control vs. Treatment groups.



Classroom Fidelity

For fidelity, we report the percentage of indicators present in the treatment and control classrooms at the time these classrooms were observed by Teaching Strategies' raters in the spring of 2024. Table 11 reports these as a percentage of the indicators present in the classroom

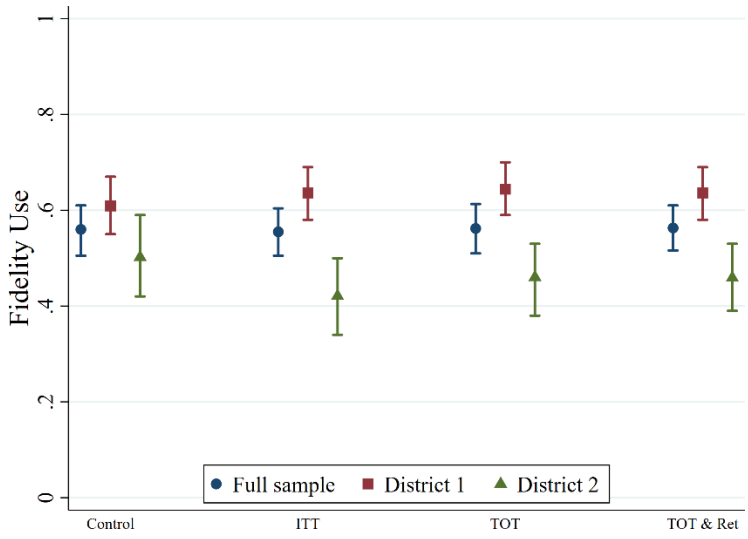
(rated as 1) for the curriculum use indicators only. There are no statistical differences for curriculum use at post-test. District differences are illustrated in Figure 6. District 1 outperformed District 2 at the post-test. In District 1, there are no statistically significant differences at post-test as captured by the fidelity measure. Any differences in favor of the control group are driven by District 2. It is important to note that baseline fidelity observations took place in the spring of year 1, after the intervention supports had already begun, and the initial coaching sessions in year 1 focused exclusively on curriculum fidelity. The timing of fidelity measurements may contribute to the higher baseline fidelity scores observed in the treatment group and, therefore, the smaller change in that group.

Table 11: Mean Fidelity Use scores by Control vs. Treatment status at Follow-up.

Variables	Control				Treatment status				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ITT Estimations	62	0.560	0.21	0.505-0.610	61	0.541	0.20	0.500-0.586	0.629
TOT Estimations	65	0.539	0.222	0.484-0.594	58	0.563	0.179	0.516-0.610	0.508
TOT Retained Estimations	65	0.539	0.222	0.484-0.594	50	0.562	0.18	0.510-0.613	0.553

Note: ***p<0.01, **p<0.05, *p<0.1. Only 123 classrooms out of the 125 classrooms had fidelity observed at the follow-up.

Figure 6. Mean Fidelity scores by Control vs. Treatment groups and district at Follow-up.



Analyses

Overall Results

We ran separate analyses by Intent-to-Treat (the randomization condition) and for Treatment-on-Treated (instrumented by the randomization condition). We include covariates in the analyses in relation to teacher background and demographics as well as cluster the standard errors by

coaches. The covariate that showed differences at baseline was years of experience and race/ethnicity. We control for these accordingly. Results are reported in Table 12.

Other than retention, which is reported as percentage point difference in the probability of retention, impacts are reported in terms of standard deviation units as measured at baseline for the control group. Estimates show a positive and statistically significant impact on retention for ITT and TOT estimations. No statistically significant impacts on ECERS-3 and large negative effects on the SSTEW are observed. In Appendix D we show additional analyses for those classrooms for which we have pre- and post-test data, controlling for the respective ECERS-3 and SSTEW baseline score, as well as estimations imputing the pre-test for those classrooms for which we do not have a pre-test. Results are overall consistent.

Table 12: Proximal effects of intervention on Teacher Retention, Classroom Quality and Curriculum Fidelity.

Estimation	Retention		ECERS-3		SSTEW		Fidelity Curr. 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.

Effect of Implementation Process and Outputs: PD Dosage & Platform Engagement

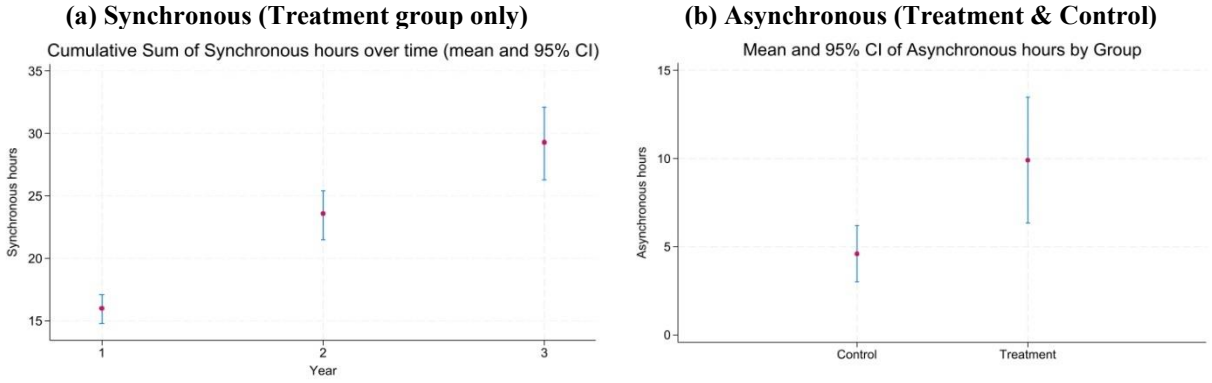
We next explore whether program impacts are moderated by (1) dosage, defined as the number of hours teachers participated in PD (either synchronous virtual sessions or asynchronous courses), and (2) *SmartTeach* platform engagement, measured as hours spent interacting with the Teach, Assess, Family, Library, and Report areas of the ecosystem.

Dosage

We measure ecosystem PD dosage in two ways, the number synchronous, virtual professional development hours that treatment teachers participated in during the study period (Figure 7, panel a), as well as the number of asynchronous, on-demand course hours any teacher completed in the ecosystem (Figure 7, panel b & Table D.4). Synchronous PD was only available to the treatment group, while asynchronous courses were accessible to both groups, though compared to treatment teachers, control teachers completed them statistically significantly less. For each year, we have individual level information on the number of hours of synchronous PD for treatment teachers. On average, treatment teachers participated in a total of 14.77 (SD 8.46) virtual sessions in which they received an average of 23.52 hours accumulated across these sessions (SD 13.89). In addition, both treatment and control teachers had access to on-demand courses within the ecosystem (reported as "asynchronous hours"). Treatment teachers completed on average 9.91 hours of asynchronous courses (SD 14.07) while control teachers only completed 4.60 hours (SD 6.28), a meaningful difference ($p = 0.007$). Both of these dosages together result in an average of 33.43 (SD 23.63) hours of total PD for the treatment group and

an average of 4.60 hours for the control group; this difference is statistically significant ($p < 0.001$). Figure 8 below shows the average number of synchronous professional development hours across the years (for the treatment group; the number of sessions are reported in Appendix Figure D.1), as well as the number of asynchronous hours (for the control and treatment group).

Figure 7. Total hours of synchronous (virtual) PD and on-demand (asynchronous) courses.



(a)

To estimate the association between dosage and quality as well as dosage and fidelity we rerun the previous set of estimations but now include the effect of dosage as the number of synchronous and asynchronous hours (see Appendix Tables D.5 and D.6 for full set of estimations). We do not estimate these models for retention, as dosage and retention are dependent on each other, i.e., an impact on retention is associated with more opportunities to participate in PD.[8] When including the interaction, the ITT coefficient now represents the impact of treatment when dosage is zero, which is why this coefficient shifts.

Synchronous PD dosage does not significantly predict classroom quality measures (ECERS-3, SSTEW, nor fidelity). In contrast, asynchronous course hours positively predict SSTEW scores and fidelity, although less so for the treatment group (negative interaction), and the inclusion of asynchronous hours tapers the negative association reported in Table 12. That is, more asynchronous hours are associated with higher quality, as per the estimations for the SSTEW and fidelity, but less so for the treatment group (and even reverting for fidelity).

Overall, dosage is a key component of PD, with asynchronous and synchronous dosages operating differently in this study, and with asynchronous PD operating more strongly for the control group (as shown by the negative interaction). Since control teachers were likely to be aware of the synchronous PD offered to their peers in their district, engagement with synchronous PD in the control group may have feasibly been driven by the absence of synchronous PD and/or a drive to stay up to par with their colleagues.

[8] There is a positive association between dosage and retention, with a positive and significant Pearson correlation of 0.339 for the measure on PD sessions and 0.371 for the measure on PD hours; estimations of dosage on retention has a significant positive association, with synchronous PD predicting higher retention.

Table 13. Proximal effects of intervention on Classroom Outcomes, accounting for dosage.

Estimation	ECERS-3		SSTEW		Fidelity Use	
	Synch. hrs.	Asynch. hrs.	Synch. hrs.	Asynch. hrs.	Synch. hrs.	Asynch. hrs.
ITT	0.164	0.143	-0.667**	-0.119	0.001	0.017
	(0.560)	(0.363)	(0.235)	(0.204)	(0.042)	(0.030)
Dosage	-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)
Observations	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. 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 effect.

SmartTeach Platform Engagement

To estimate whether results are moderated by *SmartTeach* platform engagement, we first created an engagement variable that is the sum of the number of minutes (which we divide by 60 to calculate hours) spent on the digital ecosystem in the following areas:[9] Teach, Assess, Family, Library and Report. Most of the engagement occurred across the Teach and Assess areas. Table 15 below summarizes the mean number of hours of platform engagement across the different areas. Teachers in the treatment group spent on average 314 hours on the online ecosystem (approximately 35 minutes per day, or about the equivalent of 8.97 "working" weeks total across the 3 years estimated at 35 hours of engagement in a week), while teachers in the control group spent on average about 94 hours less during the three-year period.

Table 15. Mean and standard deviations for platform engagement variables, by ITT.

	All			Control			Treatment		P-Value
Variable	n	Mean	SD	n	Mean	SD	n	Mean	SD	
Teach	125	105.67	84.9	63	92.87	71.96	62	118.48	94.99	0.093*
Assess	125	132.59	110.98	63	118.28	90.15	62	146.89	127.63	0.152
Family	125	8.2	8.84	63	7.11	7.67	62	9.3	9.81	0.17
Library	125	8.52	9.33	63	8.8	11.14	62	8.25	7.16	0.744
Report	125	12.32	11.77	63	9.51	8.83	62	15.13	13.62	0.007***
Total	125	265.90	178.76	63	219.42	156.40	62	313.90	188.7	0.003*

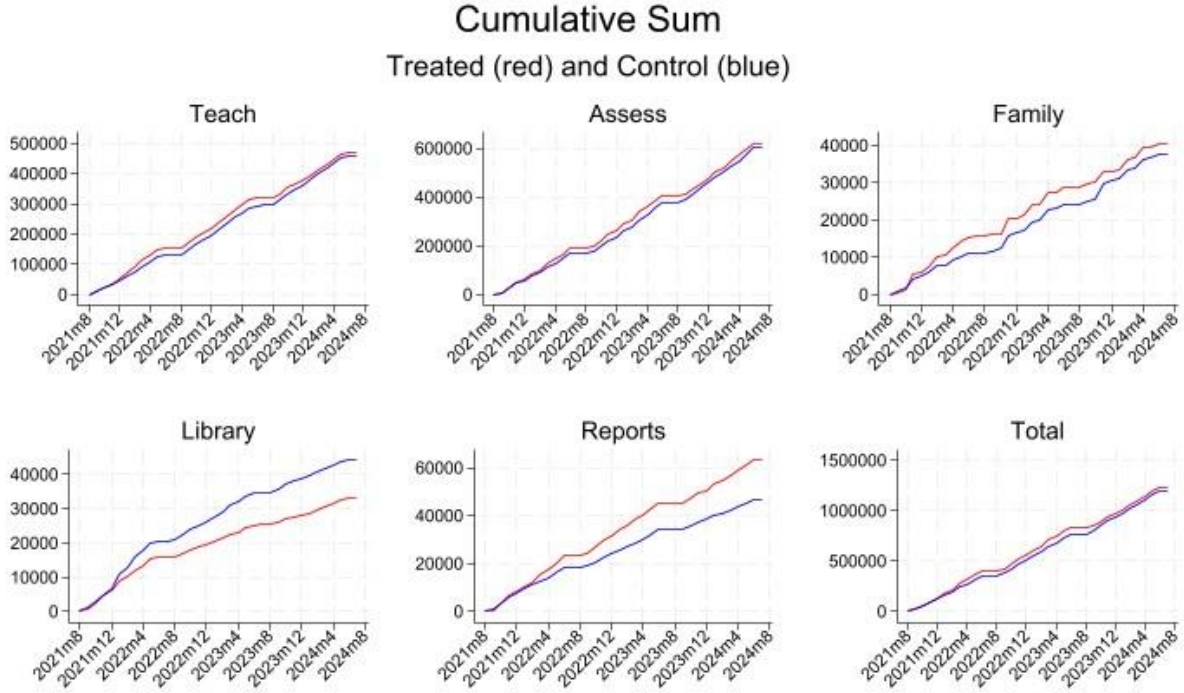
Note: ***p<0.01, **p<0.05, *p<0.1.

Figure 8 below shows the cumulative time spent in each area throughout the 3-year period. In line with the table above, the figure shows cumulative platform engagement for the treatment group started to increase towards the spring of year 1 and was higher for all types of activities for the treatment group, except for use of the Library, which was minimally used by all teachers. It does appear that teachers in the control group used the library more over time, but this difference is only about a half hour, which is negligible. Similar to dosage, platform

[9] Areas defined: *Teach* is where teachers conduct planning and teaching, *Assess* is where users upload and rate documentation of children's development and learning, *Family* is where teachers can participate in two-way communication with families, *Report* is where teachers can generate reports, and *Library* is where they can access their digital Creative Curriculum content.

engagement is dependent on retention and vice-versa, with a positive and significant correlation; we therefore do not estimate effects of engagement on retention.[10]

Figure 8. Number of minutes for which teachers were engaged in the *SmartTeach* platform.



As for platform engagement, we then replicate the estimations including this variable, as well as an interaction with ITT status (Table 16). When including the interaction, the ITT coefficient now represents the impact of treatment when engagement is zero, which is why this coefficient shifts. Platform engagement results are significantly and positively associated with the SSTEW scores, but this effect reverts for the treatment group (although only trends significant, $p < 0.10$). With platform engagement included, TOT estimations do not show any significant effects on the SSTEW scores (Appendix Table D.7), although the engagement by TOT interaction trends negative. This suggests that the combination of platform engagement and the intervention did not operate in complementary ways, even though engagement itself showed a positive effect on the SSTEW scores. No significant moderation by platform engagement was observed for ECERS-3 but a similar pattern of effects is observed for fidelity ($p < 0.05$ for the main effect and the negative interaction).

[10] There is positive association between platform engagement and retention, with a positive and significant Pearson correlation of 0.350.

Table 16: Proximal effects of intervention on Classroom Outcomes, accounting for *SmartTeach* platform engagement, ITT.

Estimation	ECERS-3	SSTEW	Fidelity Use
ITT	-0.304	0.319	-0.018
	(0.443)	(0.404)	(0.064)
Engagement	-0.043	0.050*	0.007
	(0.028)	(0.026)	(0.006)
ITT*Engagement	0.026	-0.090**	-0.004
	(0.037)	(0.040)	(0.006)
Observations	124	125	123

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are retention, teacher demographics, years of experience and education, district fixed effects and errors are clustered at the coach level. Robust standard errors in parentheses. We also replicate these estimations with pre-test and while the number of observations decreases, the results are aligned. This is also the case for the TOT estimations.

Summary & Discussion

This report summarizes Teaching Strategies' Creative Curriculum Implementation and Ecosystem Engagement Study (CCIEE) and provides baseline analyses as well as information on study design and procedures. In addition, it provides a first set of analyses for retention and classroom quality, the proximal outcomes for the study. The study emerges as a result of expansions to Teaching Strategies' digital ecosystem, which was previously focused on assessment and reporting, using random assignment to additional professional supports to drive exogeneous differences in engagement with the ecosystem components.

At baseline, the randomization of classrooms within districts shows equivalence in the treatment and control groups across the classroom quality and fidelity measures. This speaks to a successful randomization at baseline. By follow-up, there was substantial teacher churn, with significant differences in retention in favor of the treatment group.

In addition, we looked at classroom quality for the two measures in the study at follow-up by initial randomization status (ITT), as well as by effective treatment (TOT), as three classrooms were not treated given continuous changes in teachers; plus, three classrooms closed and six joined (as these teachers were assigned to a treatment coach). Classroom quality increased on average across all groups between baseline and follow-up. At follow-up, we find null (ECERS-3) or negative (SSTEW) effects on the classroom quality measures by ITT and TOT when controlling for baseline quality and teacher demographics. The analyses presented also look at retained teachers only, who have higher quality scores at follow-up, but still found no statistically significant differences for this group. Overall, both groups, treatment and control, improved across the quality measures, and negative effects appear driven by greater improvements in the control group.

We also looked at fidelity by ITT and TOT and find that the indicators in the fidelity measure that relate to use of the curriculum do not statistically differ between treatment and control classrooms at follow-up, having increased for both groups between baseline and post-test. Additional analyses controlling for baseline fidelity and teacher demographics show null or marginally negative effects on the fidelity measure due to the increase in curriculum fidelity being larger for the control group. As previously mentioned, this may be attributed, in part, to the timing of the baseline fidelity observations.

We repeated the analyses above including dosage and platform engagement as our key implementation variables. When we account for how much teachers participated in PD—either synchronous or asynchronous—we find that how much they engaged plays a key role in whether the program worked. Asynchronous PD was accessible to treatment and control groups, and it emerges as related to improvements in classroom quality (SSTEW) and curriculum fidelity across all classrooms. Interestingly, including asynchronous PD attenuates the negative associations with quality for the treatment group. In terms of platform engagement, teachers who used the digital tools more had better quality outcomes (SSTEW), although being assigned to the intervention did not translate into these effects on quality.

The mixed findings seem to suggest that it is not just whether the synchronous professional development program was offered, but rather whether teachers actually participated meaningfully or in complementary ways in relation to the PD. Asynchronous courses seemed to support aspects of quality and fidelity, but this association was reduced for the treatment group, possibly pointing to different aspects in both types of PD that work well for teachers, but that are not currently aligned in ways that complementarily support both retention and quality.

In summary, all classrooms showed improvements in observed quality, and the exogenous variation induced through the availability of PD and coaching for the treatment group did not generate differences by overall treatment status. Some implementation aspects could explain the lack of intervention impacts on classroom fidelity and observed quality. First, control teachers had access to the ecosystem, and the platform engagement variables show that they did in fact engage with it, and with the asynchronous PD accessible to both groups in the system. In addition, as mentioned earlier, one of the districts had considerable expectations for engagement.

Separate analyses assess whether differences in treatment, dosage or platform engagement relate to children's growth in year 3 and the overall results for the study. These will also include analyses on teachers' perceptions of their practice and the intervention.

References

Bandura, A. (1986). Social foundations of thought and action. Englewood Cliffs, NJ, 1986(23-28), 2.

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 Childcare, Early Education and Social Inequality. 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., Youn, M., & Frede, E. C. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. *New Brunswick, NJ: National Institute for Early Education Research*, *10*, 2001-2004.

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.

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., ... & MacMillan, H. (2017). Nurturing care: promoting early childhood development. The Lancet, 389(10064), 91-102.

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

Bronfenbrenner, U. & Morris, P. (2006). The Bioecological Model of Human Development. Volume I. Theoretical Models of Human Development, Handbook of child psychology.

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.

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.

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 Imagenes Peabody*. *American Guidance Service*.

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

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.

Fantuzzo, J. W., Gadsden, V. L., & McDermott, P. A. (2011). An integrated curriculum to improve mathematics, language, and literacy for Head Start children. American Educational Research Journal, 48(3), 763–793.

File, N. (2012). The relationship between child development and early childhood curriculum. In 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., Barnett, W. S., Hodges, K. S., Garver, K. A., Duer, J., Weisenfeld, G., & Siegel, J. (2025). *The State of Preschool 2024: State Preschool Yearbook.* National Institute for Early Education Research

Friedman-Krauss, A., Barnett, W. S., Jost, T., & Garver, K. (2024). *Help eanted: 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.

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

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

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 A.W., Collins, G.S., Hróbjartsson A., Moher, D., Schulz, K.F., Tunn, R., Aggarwal, R., Berkwits, M., Berlin, J.A., Bhandari, N., Butcher, N.J., Campbell, M.K., Chidebe, R.C.W., Elbourne, D., Farmer, A., Fergusson, D.A., Golub, R.M., Goodman S.N., Hoffmann, T.C., Ioannidis, J.P.A., Kahan, B.C., Knowles, R.L., Lamb, S.E., Lewis, S., Loder, E., Offringa, M., Ravaud, P., Richards, D.P., Rockhold, F.W., Schriger, D.L., Siegfried, N.L., Staniszewska, S., Taylor, R.S., Thabane, L., Torgerson, D., Vohra, S., White, I.R., & Boutron, I. (2025). CONSORT 2025 statement: updated guideline for reporting randomised trials. *Lancet.* :S0140-6736(25)00672-5.

Howard, S. J., Siraj, I., Melhuish, E. C., Kingston, D., Neilsen-Hewett, C., de Rosnay, M., ... & 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., Duncan, G. J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education Review, 65*, 107–125.

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.

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*.

National Academies of Sciences, Engineering, and Medicine. 2024. A New Vision for High-Quality Preschool Curriculum. Washington, DC: The National Academies Press. https://doi.org/10.17226/27429.

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

Nores, M., Figueras-Daniel, A., Lopez, 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.

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.

Preschool Curriculum Evaluation Research Consortium (2008). Effects of Preschool Curriculum Programs on School Readiness (NCER 2008-2009). Washington, DC: 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., ... & Hundeland, P. S. (2024). The effects of a structured curriculum on preschool effectiveness: A field experiment. *Journal of Human Resources*, *59*(2), 576-603.

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*. Riverside Publishing.

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.

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.

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

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. doi:10.1177/2332858418757735

Weiland, C., Rosada, P. G., Taylor, A., Penfold, L., Kushner, R., Snow, C., Xia, Y., & McCormick, M. (2026). Scaling high quality: An implementation study of Boston's Universal Pre-K expansion to community-based programs. *Early Childhood Research Quarterly*, *74*, 11-25.

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. Classroom quality supplement.

Table A.1. Alpha analysis for ECERS subscales.

	All		Treatment		Control		Synthetic	
Subscale	Obs	Alpha	Obs	Alpha	Obs	Alpha	Obs	Alpha
Baseline								
Space and Furnishings	174	0.909	56	0.924	60	0.900	58	0.894
Personal Care Routines	174	0.921	56	0.929	60	0.921	58	0.912
Language and Literacy	174	0.895	56	0.921	60	0.875	58	0.884
Learning Activities	174	0.900	56	0.920	60	0.888	58	0.883
Interaction	174	0.886	56	0.903	60	0.876	58	0.876
Program Structure	173	0.892	56	0.918	59	0.877	58	0.875
Total ECERS	174	0.866	56	0.892	60	0.850	58	0.848
Post-Test								
Space and Furnishings	173	0.931	61	0.952	63	0.912	49	0.908
Personal Care Routines	173	0.929	61	0.941	63	0.902	49	0.927
Language and Literacy	173	0.919	61	0.942	63	0.882	49	0.909
Learning Activities	173	0.917	61	0.937	63	0.891	49	0.904
Interaction	173	0.915	61	0.938	63	0.879	49	0.900
Program Structure	173	0.915	61	0.938	63	0.878	49	0.904
Total ECERS	173	0.895	61	0.924	63	0.851	49	0.878

Note: The alpha reliability analysis demonstrates strong internal consistency across ECERS subscales and for control, treatment and synthetic groups.

Table A.2. Alpha analysis for SSTEW subscales.

	All		Treatment		Control		Synthetic	
Subscale	Obs	Alpha	Obs	Alpha	Obs	Alpha	Obs	Alpha
Baseline								
Building trust, confidence, & independence	170	0.780	57	0.788	58	0.783	55	0.784
Social & emotional wellbeing	170	0.841	57	0.847	58	0.827	55	0.869
Supporting & extending language	170	0.771	57	0.781	58	0.774	55	0.771
Supporting learning & critical thinking	170	0.840	57	0.868	58	0.844	55	0.829
Assessing learning & language	168	0.864	56	0.856	58	0.880	54	0.830
Total SSTEW	170	0.852	57	0.860	58	0.855	55	0.850
Post-Test								
Building trust, confidence, & independence	174	0.840	62	0.873	63	0.807	49	0.849
Social & emotional wellbeing	174	0.859	62	0.901	63	0.808	49	0.874
Supporting & extending language	174	0.829	62	0.871	63	0.777	49	0.841
Supporting learning & critical thinking	174	0.868	62	0.904	63	0.833	49	0.851
Assessing learning & language	174	0.865	62	0.892	63	0.838	49	0.876
Total SSTEW	174	0.879	62	0.909	63	0.845	49	0.883

Note: The alpha reliability analysis demonstrates strong internal consistency across SSTEW subscales and for control, treatment and synthetic groups.

Table A.3: Factor analysis for ECERS subscales.

Factor	Eigenvalue	Difference	Proportion	Cumulative
Baseline				
Factor1	3.642	2.840	0.607	0.607
Factor2	0.801	0.179	0.134	0.741
Factor3	0.622	0.165	0.104	0.844
Factor4	0.457	0.183	0.076	0.920
Factor5	0.274	0.070	0.046	0.966
Factor6	0.204	.	0.034	1.000
Post-Test				
Factor1	3.956	3.324	0.659	0.659
Factor2	0.631	0.128	0.105	0.765
Factor3	0.504	0.035	0.084	0.849
Factor4	0.469	0.214	0.078	0.927
Factor5	0.255	0.069	0.042	0.969
Factor6	0.185	.	0.031	1.000

Note: N=173. LR test: independent vs. saturated: chi2(15) = 610.95, Prob>chi2 = 0.000. The results support a one factor solution.

Table A.4: Factor loadings and scoring coefficients (ECERS subscales).

Subscale	Factor loadings	Scoring coefficients
Baseline		
Space and Furnishings	0.702	0.193
Personal Care Routines	0.592	0.163
Language and Literacy	0.822	0.226
Learning Activities	0.778	0.214
Interaction	0.888	0.244
Program Structure	0.854	0.234
Post-Test		
Space and Furnishings	0.719	0.182
Personal Care Routines	0.745	0.188
Language and Literacy	0.828	0.209
Learning Activities	0.835	0.211
Interaction	0.867	0.219
Program Structure	0.867	0.219

Note: Factor loadings for factor 1.

Table A.5: Factor analysis for SSTEW subscales

Factor	Eigenvalue	Difference	Proportion	Cumulative
Baseline				
Factor1	3.183	2.408	0.637	0.637
Factor2	0.775	0.232	0.155	0.791
Factor3	0.543	0.208	0.109	0.900
Factor4	0.335	0.170	0.067	0.967
Factor5	0.165	.	0.033	1.000
Post-Test				
Factor1	3.376	2.803	0.675	0.675
Factor2	0.573	0.082	0.115	0.790
Factor3	0.491	0.160	0.098	0.888
Factor4	0.331	0.103	0.066	0.954
Factor5	0.228	.	0.046	1.000

Note: N=174. LR test: independent vs. saturated: chi2(10) = 451.52, Prob>chi2 = 0.0000. The results support a one factor solution.

Table A.6: Factor loadings and scoring coefficients (SSTEW subscales).

Variable	Factor loadings	Scoring coefficients
Baseline		
Building trust, confidence, & independence	0.906	0.285
Social & emotional wellbeing	0.738	0.232
Supporting & extending language	0.922	0.290
Supporting learning & critical thinking	0.735	0.231
Assessing learning & language	0.654	0.205
Post-Test		
Building trust, confidence, & independence	0.859	0.254
Social & emotional wellbeing	0.804	0.238
Supporting & extending language	0.890	0.264
Supporting learning & critical thinking	0.770	0.228
Assessing learning & language	0.780	0.231

Note: Factor loadings for factor 1.

Table A.7: Pairwise correlations among ECERS subscales (Full sample).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Baseline						
(1) Space and Furnishings	1.000					
(2) Personal Care Routines	0.3408***	1.000				
(3) Language and Literacy	0.4164***	0.3795***	1.000			
(4) Learning Activities	0.4973***	0.2529***	0.6887 ***	1.000		
(5) Interaction	0.5824***	0.4752***	0.6489***	0.5891***	1.000	
(6) Program Structure	0.4740***	0.4657***	0.6377***	0.5505***	0.7824***	1.000
Post-Test						
(1) Space and Furnishings	1.000					
(2) Personal Care Routines	0.4136***	1.000				
(3) Language and Literacy	0.4443***	0.5932***	1.000			
(4) Learning Activities	0.5786***	0.5127***	0.7118***	1.000		
(5) Interaction	0.5815***	0.5764***	0.6037***	0.6213***	1.000	
(6) Program Structure	0.5318***	0.5598***	0.6558***	0.6262***	0.8038***	1.000

Note: Pairwise correlation coefficients. *** p<0.01, ** p<0.05, * p<0.1. Baseline N=174. Post-test N=173.

Table A.8: Pairwise correlations among ECERS subscales (Control and Treatment sample).

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Baseline						
(1) Space and Furnishings	1.000					
(2) Personal Care Routines	0.368***	1.000				
(3) Language and Literacy	0.4399***	0.398***	1.000			
(4) Learning Activities	0.5247***	0.2569***	0.6707***	1.000		
(5) Interaction	0.5801***	0.5223***	0.6889***	0.6216***	1.000	
(6) Program Structure	0.4895***	0.4528***	0.6458***	0.5518***	0.7734***	1.000
Post-Test						
(1) Space and Furnishings	1.000					
(2) Personal Care Routines	0.3887***	1.000				
(3) Language and Literacy	0.3962 ***	0.6478***	1.000			
(4) Learning Activities	0.5624***	0.5514***	0.729***	1.000		
(5) Interaction	0.5609***	0.6357***	0.6114***	0.6071***	1.000	
(6) Program Structure	0.4025***	0.6548***	0.6683***	0.6272***	0.7790***	1.000

Note: Pairwise correlation coefficients. *** p<0.01, ** p<0.05, * p<0.1. Baseline N=116. Post-test N=124.

Table A.9: Pairwise correlations among SSTEW subscales (Full sample).

Variables	(1)	(2)	(3)	(4)	(5)
Baseline					
(1) Building trust, confidence, & independence	1				
(2) Social & emotional wellbeing	0.5949***	1			
(3) Supporting & extending language	0.8308***	0.6129***	1		
(4) Supporting learning & critical thinking	0.5737***	0.4642***	0.6025***	1	
(5) Assessing learning & language	0.5356***	0.2638***	0.5676***	0.3041***	1
Post-Test					
(1) Building trust, confidence, & independence	1				
(2) Social & emotional wellbeing	0.6629***	1			
(3) Supporting & extending language	0.6708***	0.6983***	1		
(4) Supporting learning & critical thinking	0.5609***	0.4803***	0.6261***	1	
(5) Assessing learning & language	0.6143***	0.4494***	0.6252***	0.5173***	1

Note: Pairwise correlation coefficients. *** p<0.01, ** p<0.05, * p<0.1. Correlations among SSTEW subscales highlight strong interrelations, particularly between Building trust, confidence, and independence and Supporting and extending language. These findings underscore interconnectedness of these two. Baseline N=170. Post-test N=174.

Table A.10: Pairwise correlations among SSTEW subscales (Control and Treatment sample).

Variables	(1)	(2)	(3)	(4)	(5)
Baseline					
(1) Building trust, confidence, & independence	1				
(2) Social & emotional wellbeing	0.6615***	1			
(3) Supporting & extending language	0.7947***	0.666***	1		
(4) Supporting learning & critical thinking	0.5734***	0.3259***	0.559***	1	
(5) Assessing learning & language	0.5854***	0.3661***	0.6578***	0.3125***	1
Post-Test					
(1) Building trust, confidence, & independence	1				
(2) Social & emotional wellbeing	0.6661***	1			
(3) Supporting & extending language	0.6623***	0.7254***	1		
(4) Supporting learning & critical thinking	0.5631***	0.4689***	0.6172***	1	
(5) Assessing learning & language	0.6325***	0.4596***	0.6461***	0.5065***	1

Note: Pairwise correlation coefficients. *** p<0.01, ** p<0.05, * p<0.1. These findings underscore interconnectedness of these two. Baseline N=115. Post-test N=125.

Appendix B. Baseline comparisons supplement.

Table B.1: Mean ECERS and SSTEW scores at baseline (Control vs. Treatment), District 1.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	32	3.89	0.84	3.58-4.19	27	3.89	0.70	3.62-4.17	0.977
Personal Care Routines	32	4.14	1.05	3.76-4.52	27	4.01	1.07	3.59-4.43	0.637
Language and Literacy	32	4.01	0.93	3.67-4.34	27	4.13	1.06	3.71-4.55	0.646
Learning Activities	32	3.07	0.71	2.82-3.33	27	3.24	0.67	2.97-3.50	0.360
Interaction	32	4.58	0.93	4.25-4.92	27	4.71	1.20	4.24-5.18	0.645
Program Structure	31	4.33	1.08	3.94-4.73	27	4.57	1.15	4.11-5.02	0.426
Total ECERS score	*32*	*3.83*	*0.59*	*3.61-4.04*	*27*	*3.92*	*0.63*	*3.67-4.17*	*0.572*
SSTEW	31	2.60	0.98	2.24-2.96	28	2.67	1.11	2.24-3.10	0.813
Building trust, confidence, & indep.	31	1.90	1.08	1.51-2.30	28	1.96	0.84	1.64-2.29	0.810
Social & emotional wellbeing	31	2.56	1.06	2.17-2.95	28	2.83	1.29	2.33-3.33	0.376
Supporting & extending language	31	1.50	0.35	1.37-1.63	28	1.60	0.43	1.43-1.76	0.338
Supporting learning & critical th.	31	2.34	0.91	2.01-2.67	28	2.18	0.83	1.86-2.50	0.484
Assessing learning & language	31	2.19	0.71	1.93-2.45	28	2.29	0.77	1.99-2.59	0.602
Total SSTEW score	*31*	*2.60*	*0.98*	*2.24-2.96*	*28*	*2.67*	*1.11*	*2.24-3.1*	*0.813*

Note: ***p<0.01, **p<0.05, *p<0.1

Table B.2: Mean ECERS and SSTEW scores at baseline (Control vs. Treatment), District 2.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	28	3.57	0.76	3.27-3.86	29	3.32	0.83	3.01-3.63	0.247
Personal Care Routines	28	3.84	1.08	3.42-4.26	29	3.15	1.00	2.76-3.53	0.015**
Language and Literacy	28	3.96	1.10	3.54-4.39	29	3.90	1.08	3.49-4.31	0.834
Learning Activities	28	3.28	0.88	2.93-3.62	29	3.02	0.84	2.70-3.33	0.255
Interaction	28	4.08	1.27	3.59-4.57	29	3.83	1.37	3.31-4.36	0.490
Program Structure	28	4.12	1.26	3.63-4.61	29	3.85	1.23	3.38-4.32	0.419
Total ECERS score	*28*	*3.69*	*0.85*	*3.36-4.02*	*29*	*3.41*	*0.87*	*3.08-3.74*	*0.226*
SSTEW									
Building trust, confidence, & indep.	27	3.79	0.83	3.46-4.12	29	3.75	0.91	3.40-4.09	0.854
Social & emotional wellbeing	27	2.52	0.89	2.17-2.87	29	2.90	1.29	2.41-3.39	0.211
Supporting & extending language	27	4.28	0.85	3.94-4.62	29	3.82	1.12	3.39-4.25	0.093*
Supporting learning & critical th.	27	1.71	0.32	1.59-1.84	29	1.79	0.32	1.67-1.92	0.351
Assessing learning & language	27	2.89	0.79	2.58-3.20	28	2.86	0.64	2.61-3.10	0.870
Total SSTEW score	*27*	*3.12*	*0.50*	*2.92-3.31*	*29*	*3.03*	*0.65*	*2.78-3.27*	*0.563*

Note: ***p<0.01, **p<0.05, *p<0.1

Appendix C. Follow-up comparisons supplement.

Table C.1: Change in ECERS and SSTEW change in scores by Control vs. Intent-to-Treat between Baseline and Follow-up.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	59	0.18	0.90	-0.06-0.41	54	0.25	0.75	0.04-0.45	0.657
Personal Care Routines	59	-0.39	1.46	-0.78--0.01	54	-0.08	1.39	-0.46-0.30	0.241
Language and Literacy	59	0.28	1.32	-0.06-0.63	54	0.28	1.17	-0.04-0.60	0.976
Learning Activities	59	0.50	0.93	0.25-0.74	54	0.43	0.90	0.19-0.68	0.725
Interaction	59	0.01	1.68	-0.43-0.44	54	-0.01	1.39	-0.39-0.37	0.964
Program Structure	58	-0.03	1.52	-0.44-0.37	54	0.16	1.63	-0.28-0.61	0.508
Total ECERS	*59*	*0.18*	*0.91*	*-0.06-0.42*	*54*	*0.23*	*0.82*	*0.00-0.45*	*0.780*
SSTEW									
Building trust, confidence, & indep.	58	0.24	1.77	-0.22-0.71	55	-0.28	1.79	-0.76-0.21	0.124
Social & emotional wellbeing	58	2.12	2.29	1.52-2.72	55	1.71	2.22	1.11-2.31	0.334
Supporting & extending language	58	-0.04	1.98	-0.56-0.48	55	-0.08	1.71	-0.54-0.38	0.912
Supporting learning & critical th.	58	0.47	0.77	0.27-0.68	55	0.50	0.98	0.23-0.76	0.898
Assessing learning & language	58	0.59	1.78	0.13-1.06	54	0.26	1.76	-0.22-0.74	0.318
Total SSTEW	*58*	*0.41*	*1.35*	*0.06-0.77*	*55*	*0.21*	*1.36*	*-0.15-0.58*	*0.445*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.2: Change in ECERS and SSTEW change in scores by Control vs. Treatment-on-Treated between Baseline and Follow-up.

Variables	Control				Treatment-on-Treated				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	62	0.18	0.91	-0.05-0.41	51	0.25	0.72	0.05-0.46	0.622
Personal Care Routines	62	-0.40	1.44	-0.76--0.03	51	-0.06	1.41	-0.45-0.34	0.213
Language and Literacy	62	0.25	1.36	-0.09-0.60	51	0.32	1.11	0.01-0.63	0.780
Learning Activities	62	0.47	0.97	0.22-0.71	51	0.47	0.84	0.23-0.70	0.997
Interaction	62	-0.01	1.67	-0.44-0.41	51	0.01	1.38	-0.38-0.40	0.940
Program Structure	61	-0.03	1.52	-0.42-0.36	51	0.17	1.64	-0.29-0.63	0.492
Total ECERS	*62*	*0.16*	*0.94*	*-0.07-0.4*	*51*	*0.25*	*0.78*	*0.03-0.47*	*0.603*
SSTEW									
Building trust, confidence, & indep.	61	0.12	1.82	-0.35-0.59	52	-0.17	1.76	-0.66-0.32	0.399
Social & emotional wellbeing	61	1.98	2.31	1.39-2.58	52	1.85	2.20	1.23-2.46	0.748
Supporting & extending language	61	-0.11	1.95	-0.61-0.39	52	-0.01	1.73	-0.49-0.47	0.782
Supporting learning & critical th.	61	0.45	0.78	0.25-0.65	52	0.53	0.99	0.25-0.80	0.621
Assessing learning & language	61	0.51	1.78	0.05-0.96	51	0.34	1.77	-0.16-0.84	0.625
Total SSTEW	*61*	*0.34*	*1.36*	*-0.01-0.68*	*52*	*0.29*	*1.36*	*-0.09-0.67*	*0.858*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.3: Change in ECERS and SSTEW scores by Control vs. Treatment-on-Treated & Retained between Baseline and Follow-up.

Variables	Control				TOT Retained				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	62	0.18	0.91	-0.05-0.41	43	0.27	0.73	0.04-0.49	0.583
Personal Care Routines	62	-0.40	1.44	-0.76--0.03	43	-0.06	1.39	-0.48-0.37	0.232
Language and Literacy	62	0.25	1.36	-0.09-0.60	43	0.44	1.02	0.13-0.76	0.438
Learning Activities	62	0.47	0.97	0.22-0.71	43	0.46	0.85	0.20-0.72	0.987
Interaction	62	-0.01	1.67	-0.44-0.41	43	0.10	1.26	-0.28-0.49	0.709
Program Structure	61	-0.03	1.52	-0.42-0.36	43	0.37	1.57	-0.12-0.85	0.194
Total ECERS	*62*	*0.16*	*0.94*	*-0.07-0.4*	*43*	*0.30*	*0.76*	*0.07-0.53*	*0.429*
SSTEW									
Building trust, confidence, & indep.	61	0.12	1.82	-0.35-0.59	43	-0.11	1.87	-0.68-0.47	0.534
Social & emotional wellbeing	61	1.98	2.31	1.39-2.58	43	1.91	2.24	1.22-2.60	0.867
Supporting & extending language	61	-0.11	1.95	-0.61-0.39	43	0.08	1.83	-0.49-0.64	0.632
Supporting learning & critical th.	61	0.45	0.78	0.25-0.65	43	0.61	0.99	0.31-0.91	0.346
Assessing learning & language	61	0.51	1.78	0.05-0.96	42	0.49	1.80	-0.07-1.05	0.955
Total SSTEW	*61*	*0.34*	*1.36*	*-0.01-0.68*	*43*	*0.37*	*1.43*	*-0.06-0.81*	*0.891*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.4: Mean ECERS and SSTEW scores by Synthetic vs. Intent-to-Treat at Follow-up.

Variables	Synthetic				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	49	4.07	0.73	3.86-4.28	61	3.87	0.69	3.69-4.05	0.151
Personal Care Routines	49	3.66	1.11	3.34-3.98	61	3.39	1.19	3.09-3.70	0.230
Language and Literacy	49	4.67	0.95	4.4-4.95	61	4.26	1.02	4.00-4.52	0.033**
Learning Activities	49	3.90	0.90	3.64-4.16	61	3.56	0.91	3.32-3.79	0.048**
Interaction	49	5.00	1.43	4.59-5.41	61	4.21	1.42	3.85-4.58	0.005***
Program Structure	49	4.78	1.24	4.43-5.14	61	4.23	1.37	3.88-4.58	0.030**
Total ECERS	*49*	*4.26*	*0.80*	*4.03-4.49*	*61*	*3.85*	*0.88*	*3.63-4.08*	*0.015***
SSTEW									
Building trust, confidence, & indep.	49	3.23	1.38	2.84-3.63	62	2.90	1.54	2.51-3.29	0.238
Social & emotional wellbeing	49	4.08	1.62	3.62-4.55	62	4.10	2.06	3.57-4.62	0.966
Supporting & extending language	49	3.13	1.00	2.84-3.42	62	3.22	1.27	2.89-3.54	0.685
Supporting learning & critical th.	49	2.32	0.96	2.05-2.6	62	2.15	0.93	1.91-2.38	0.329
Assessing learning & language	49	2.82	1.04	2.52-3.12	62	2.73	1.44	2.36-3.09	0.712
Total SSTEW	*49*	*2.94*	*0.94*	*2.67-3.21*	*62*	*2.84*	*1.14*	*2.55-3.12*	*0.594*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.5: Mean ECERS and SSTEW scores by Synthetic vs. Treatment-on-Treated at Follow-up.

Variables	Synthetic				Treatment-on-Treated				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	49	4.07	0.73	3.86-4.28	58	3.91	0.69	3.73-4.09	0.244
Personal Care Routines	49	3.66	1.11	3.34-3.98	58	3.41	1.21	3.09-3.72	0.268
Language and Literacy	49	4.67	0.95	4.40-4.95	58	4.31	0.98	4.06-4.57	0.058*
Learning Activities	49	3.90	0.90	3.64-4.16	58	3.61	0.88	3.38-3.84	0.092*
Interaction	49	5.00	1.43	4.59-5.41	58	4.27	1.42	3.90-4.65	0.009***
Program Structure	49	4.78	1.24	4.43-5.14	58	4.26	1.38	3.90-4.62	0.045**
Total ECERS	*49*	*4.26*	*0.80*	*4.03-4.49*	*58*	*3.96*	*0.94*	*3.71-4.21*	*0.028***
SSTEW									
Building trust, confidence, & indep.	49	3.23	1.38	2.84-3.63	59	2.97	1.55	2.56-3.37	0.354
Social & emotional wellbeing	49	4.08	1.62	3.62-4.55	59	4.17	2.07	3.63-4.71	0.809
Supporting & extending language	49	3.13	1.00	2.84-3.42	59	3.27	1.27	2.94-3.60	0.534
Supporting learning & critical th.	49	2.32	0.96	2.05-2.6	59	2.18	0.94	1.93-2.42	0.434
Assessing learning & language	49	2.82	1.04	2.52-3.12	59	2.79	1.44	2.41-3.16	0.909
Total SSTEW	*49*	*2.94*	*0.94*	*2.67-3.21*	*59*	*3.07*	*1.25*	*2.75-3.40*	*0.849*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.6: Mean ECERS and SSTEW scores by Synthetic vs. Treatment-on-Treated & Retained at Follow-up.

Variables	Synthetic				Treatment				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	49	4.07	0.73	3.86-4.28	49	3.90	0.73	3.69-4.11	0.263
Personal Care Routines	49	3.66	1.11	3.34-3.98	49	3.34	1.19	3.00-3.68	0.172
Language and Literacy	49	4.67	0.95	4.40-4.95	49	4.26	0.95	3.99-4.53	0.034**
Learning Activities	49	3.90	0.90	3.64-4.16	49	3.51	0.85	3.27-3.76	0.031**
Interaction	49	5.00	1.43	4.59-5.41	49	4.20	1.36	3.81-4.59	0.005***
Program Structure	49	4.78	1.24	4.43-5.14	49	4.25	1.41	3.84-4.65	0.049**
Total ECERS	*49*	*4.26*	*0.80*	*4.03-4.49*	*49*	*3.91*	*0.95*	*3.64-4.18*	*0.018***
SSTEW									
Building trust, confidence, & indep.	49	3.23	1.38	2.84-3.63	50	2.99	1.60	2.53-3.44	0.417
Social & emotional wellbeing	49	4.08	1.62	3.62-4.55	50	4.26	2.09	3.67-4.85	0.636
Supporting & extending language	49	3.13	1.00	2.84-3.42	50	3.32	1.26	2.96-3.68	0.403
Supporting learning & critical th.	49	2.32	0.96	2.05-2.60	50	2.21	0.95	1.94-2.48	0.562
Assessing learning & language	49	2.82	1.04	2.52-3.12	50	2.89	1.45	2.48-3.30	0.773
Total SSTEW	*49*	*2.94*	*0.94*	*2.67-3.21*	*50*	*3.13*	*1.26*	*2.78-3.49*	*0.938*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.7: Mean ECERS and SSTEW scores by Control vs. Intent-to-Treat (and TOT) at Follow-up, District 1.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	34	3.99	0.55	3.80-4.18	33	4.19	0.57	3.98-4.39	0.162
Personal Care Routines	34	3.20	0.62	2.98-3.42	33	3.65	1.22	3.22-4.09	0.057*
Language and Literacy	34	3.98	0.79	3.71-4.26	33	4.32	0.93	3.98-4.65	0.119
Learning Activities	34	3.66	0.62	3.44-3.88	33	3.73	0.73	3.47-3.99	0.673
Interaction	34	4.10	1.02	3.74-4.46	33	4.48	1.44	3.97-5.00	0.211
Program Structure	34	4.01	0.99	3.66-4.36	33	4.31	1.39	3.82-4.8	0.314
Total ECERS	*34*	*3.81*	*0.53*	*3.63-4.00*	*33*	*4.05*	*0.80*	*3.77-4.34*	*0.151*
SSTEW									
Building trust, confidence, & indep.	34	3.39	1.52	2.86-3.92	34	3.06	1.69	2.47-3.65	0.395
Social & emotional wellbeing	34	4.59	2.18	3.83-5.35	34	4.32	2.17	3.57-5.08	0.617
Supporting & extending language	34	3.59	1.47	3.08-4.10	34	3.35	1.50	2.83-3.88	0.516
Supporting learning & critical th.	34	2.03	0.54	1.84-2.22	34	2.14	1.08	1.76-2.52	0.596
Assessing learning & language	34	3.72	1.51	3.19-4.25	34	3.18	1.59	2.62-3.73	0.152
Total SSTEW	*34*	*3.19*	*1.04*	*2.83-3.55*	*34*	*2.99*	*1.33*	*2.52-3.45*	*0.484*

Note: ***p<0.01, **p<0.05, *p<0.1.
^Since this district did not experience any attrition from the treatment group to the control group, ITT and TOT analyses do not differ.

Table C.8: Mean ECERS and SSTEW scores by Control vs. Treatment-on-Treated & Retained at Follow-up, District 1.

Variables	Control				Treatment-on-Treated Retained				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	34	3.99	0.55	3.80-4.18	27	4.23	0.59	4.00-4.47	0.107
Personal Care Routines	34	3.20	0.62	2.98-3.42	27	3.67	1.18	3.20-4.14	0.049**
Language and Literacy	34	3.98	0.79	3.71-4.26	27	4.27	0.92	3.91-4.64	0.188
Learning Activities	34	3.66	0.62	3.44-3.88	27	3.68	0.74	3.38-3.97	0.919
Interaction	34	4.10	1.02	3.74-4.46	27	4.58	1.37	4.04-5.12	0.123
Program Structure	34	4.01	0.99	3.66-4.36	27	4.41	1.43	3.85-4.98	0.198
Total ECERS	*34*	*3.81*	*0.53*	*3.63-4.00*	*27*	*4.14*	*0.92*	*3.77-4.50*	*0.151*
SSTEW									
Building trust, confidence, & indep.	34	3.39	1.52	2.86-3.92	28	3.18	1.74	2.50-3.85	0.608
Social & emotional wellbeing	34	4.59	2.18	3.83-5.35	28	4.61	2.11	3.79-5.43	0.973
Supporting & extending language	34	3.59	1.47	3.08-4.10	28	3.49	1.48	2.92-4.07	0.797
Supporting learning & critical th.	34	2.03	0.54	1.84-2.22	28	2.21	1.11	1.78-2.63	0.416
Assessing learning & language	34	3.72	1.51	3.19-4.25	28	3.30	1.61	2.68-3.93	0.298
Total SSTEW	*34*	*3.19*	*1.04*	*2.83-3.55*	*28*	*3.30*	*1.42*	*2.80-3.90*	*0.788*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.9: Mean ECERS and SSTEW scores by Control vs. Intent-to-Treat at Follow-up, District 2.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	29	3.82	0.62	3.59-4.06	28	3.50	0.63	3.25-3.75	0.057*
Personal Care Routines	29	4.01	1.13	3.58-4.44	28	3.08	1.10	2.65-3.51	0.003***
Language and Literacy	29	4.64	0.92	4.29-4.99	28	4.20	1.13	3.76-4.64	0.111
Learning Activities	29	3.67	0.65	3.42-3.91	28	3.35	1.06	2.94-3.76	0.182
Interaction	29	4.59	1.29	4.10-5.09	28	3.89	1.35	3.37-4.42	0.051*
Program Structure	29	4.33	1.27	3.85-4.81	28	4.13	1.37	3.60-4.66	0.565
Total ECERS	*29*	*4.07*	*0.71*	*3.80-4.34*	*28*	*3.62*	*0.93*	*3.26-3.98*	*0.045***
SSTEW									
Building trust, confidence, & indep.	29	3.34	1.12	2.92-3.77	28	2.70	1.35	2.18-3.22	0.055*
Social & emotional wellbeing	29	3.90	1.70	3.25-4.54	28	3.82	1.93	3.07-4.57	0.876
Supporting & extending language	29	2.90	0.73	2.62-3.17	28	3.05	0.93	2.69-3.41	0.479
Supporting learning & critical th.	29	2.09	0.71	1.83-2.36	28	2.15	0.72	1.87-2.43	0.765
Assessing learning & language	29	2.52	0.94	2.16-2.87	28	2.18	1.00	1.79-2.57	0.193
Total SSTEW	*29*	*2.78*	*0.73*	*2.50-3.06*	*28*	*2.65*	*0.84*	*2.33-2.98*	*0.534*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.10: Mean ECERS and SSTEW scores by Control vs. Treatment-on-Treated at Follow-up, District 2.

Variables	Control				Treatment-on-Treated				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	32	3.76	0.62	3.54-3.99	25	3.54	0.66	3.27-3.81	0.189
Personal Care Routines	32	3.92	1.13	3.51-4.33	25	3.08	1.14	2.61-3.55	0.008***
Language and Literacy	32	4.51	1.04	4.14-4.89	25	4.31	1.06	3.87-4.75	0.477
Learning Activities	32	3.56	0.75	3.29-3.83	25	3.45	1.04	3.02-3.88	0.659
Interaction	32	4.45	1.33	3.97-4.93	25	3.99	1.37	3.43-4.56	0.209
Program Structure	32	4.26	1.26	3.81-4.71	25	4.20	1.40	3.62-4.78	0.865
Total ECERS	*32*	*3.97*	*0.78*	*3.69-4.25*	*25*	*3.69*	*0.93*	*3.31-4.08*	*0.231*
SSTEW									
Building trust, confidence, & indep.	32	3.18	1.19	2.75-3.61	25	2.84	1.36	2.28-3.4	0.324
Social & emotional wellbeing	32	3.78	1.70	3.17-4.39	25	3.96	1.95	3.16-4.76	0.713
Supporting & extending language	32	2.84	0.77	2.56-3.11	25	3.15	0.88	2.79-3.51	0.158
Supporting learning & critical th.	32	2.04	0.70	1.79-2.29	25	2.23	0.71	1.94-2.52	0.317
Assessing learning & language	32	2.42	0.95	2.08-2.76	25	2.26	1.02	1.84-2.68	0.540
Total SSTEW	*32*	*2.69*	*0.76*	*2.42-2.96*	*25*	*2.75*	*0.82*	*2.41-3.09*	*0.770*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.11: Mean ECERS and SSTEW scores by Control vs. Treatment-on-Treated & Retained at Follow-up, District 2.

Variables	Control				Treatment-on-Treated Retained				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ECERS-3									
Space and Furnishings	32	3.76	0.62	3.54-3.99	22	3.49	0.69	3.19-3.80	0.139
Personal Care Routines	32	3.92	1.13	3.51-4.33	22	2.93	1.10	2.44-3.42	0.002***
Language and Literacy	32	4.51	1.04	4.14-4.89	22	4.25	1.01	3.80-4.69	0.352
Learning Activities	32	3.56	0.75	3.29-3.83	22	3.32	0.95	2.90-3.74	0.302
Interaction	32	4.45	1.33	3.97-4.93	22	3.74	1.23	3.19-4.28	0.052*
Program Structure	32	4.26	1.26	3.81-4.71	22	4.05	1.39	3.43-4.66	0.557
Total ECERS	*32*	*4.08*	*0.85*	*3.77-4.39*	*22*	*3.63*	*0.93*	*3.22-4.04*	*0.072**
SSTEW									
Building trust, confidence, & indep.	32	3.18	1.19	2.75-3.61	22	2.74	1.40	2.12-3.36	0.225
Social & emotional wellbeing	32	3.78	1.70	3.17-4.39	22	3.82	2.02	2.92-4.71	0.942
Supporting & extending language	32	2.84	0.77	2.56-3.11	22	3.10	0.89	2.71-3.50	0.248
Supporting learning & critical th.	32	2.04	0.70	1.79-2.29	22	2.22	0.73	1.89-2.54	0.376
Assessing learning & language	32	2.42	0.95	2.08-2.76	22	2.36	1.04	1.90-2.82	0.832
Total SSTEW	*32*	*2.85*	*0.84*	*2.55-3.15*	*22*	*2.85*	*0.97*	*2.42-3.28*	*0.992*

Note: ***p<0.01, **p<0.05, *p<0.1.

Table C.12: Change in Fidelity scores by Control vs. Intent-to-Treat and Treatment-on-Treated between Baseline and Follow-up.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ITT	55	0.104	0.238	0.039-0.168	52	0.01	0.176	-0.098	0.023**
TOT[i]	57	0.095	0.239	0.031-0.158	50	0.016	0.176	-0.11	0.060*
TOT Retained[i]	57	0.008	0.239	0.031-0.158	42	0.008	0.18	0.047-0.063	0.051*

Note: ***p<0.01, **p<0.05, *p<0.1. [i] Control includes the classroom that dropped out of treatment.

Table C.13: Mean Fidelity scores by Control vs. Intent-to-Treat and Treatment-on-Treated at Follow-up, District 1.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ITT	33	0.609	0.180	0.55-0.67	34	0.636	0.149	0.58-0.69	0.500
TOT[i]	33	0.609	0.180	0.54-0.67	34	0.636	0.149	0.58-0.69	0.500
TOT Retained[i]	33	0.609	0.180	0.54-0.67	28	0.644	0.143	0.59-0.70	0.370

Note: ***p<0.01, **p<0.05, *p<0.1. [i] Control includes the classroom that dropped out of treatment.

Table C.14: Mean Fidelity scores by Control vs. Intent-to-Treat and Treatment-on-Treated at Follow-up, District 2.

Variables	Control				Intent-to-Treat				P-value
	N	Mean	SD	95% CI	N	Mean	SD	95% CI	
ITT	29	0.502	0.221	0.42-0.59	27	0.421	0.193	0.34-0.5	0.150
TOT[i]	32	0.466	0.240	0.41-0.58	24	0.459	0.168	0.39-0.53	0.900
TOT Retained[i]	32	0.466	0.24	0.38-0.55	21	0.462	0.212	0.38-0.53	0.868

Note: ***p<0.01, **p<0.05, *p<0.1. [i] Control includes the classroom that dropped out of treatment.

Appendix D. Additional Estimations.

Table D.1: Proximal effects on Classroom Outcomes, controlling for pre-test.

Estimation	ECERS-3		SSTEW		Fidelity Use	
ITT	-0.036		-0.394		-0.066*	
	(0.321)		(0.245)		(0.036)	
TOT		0.056		-0.438*		-0.070*
		(0.311)		(0.219)		(0.035)
Observations	113	113	113	113	107	107

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

Table D.2: Proximal effects on Classroom Outcomes, controlling for pre-test and estimating pre-test with multiple imputation.

Estimation	ECERS-3		SSTEW		Fidelity Use	
ITT	-0.052		-0.401**		-0.058*	
	(0.296)		(0.178)		(0.027)	
TOT		0.024		-0.425**		-0.052
		(0.290)		(0.169)		(0.029)
Obs.	124	124	124	124	122	122

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are retention, coach retention, teacher demographics and education, district fixed effects and errors are clustered at the coach level. Robust standard errors in parentheses. Baseline ECERS and SSTEW scores imputed through multiple imputation analyses.

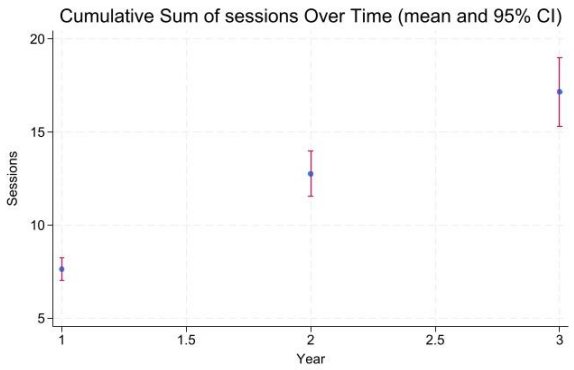
Table D.3: Proximal effects on Classroom Outcomes, by district.

Estimation	District 1				District 2			
	Retention	ECERS-3	SSTEW	Fidelity Use	Retention	ECERS-3	SSTEW	Fidelity Use
ITT (No pretest)	0.265***	0.304	0.015	0.003	0.182	-0.537*	0.019	-0.084
	(0.081)	(0.463)	(0.040)	(0.048)	(0.118)	(0.214)	(0.386)	(0.054)
Observations	68	67	67	67	57	57	57	56
ITT (With pretest)	n/a	0.249	-0.803*	-0.021	n/a	-0.260	-0.077	-0.106*
	n/a	(0.540)	(0.366)	(0.043)	n/a	(0.269)	(0.402)	(0.050)
Observations	n/a	59	59	56	n/a	54	54	51

Note: ***p<0.01, **p<0.05, *p<0.1. Controls included are retention (for all analyses except for retention analyses), coach retention, teacher demographics and education, district fixed effects and errors are clustered at the coach level. Robust standard errors in parentheses. There is no applicable pre-test for the retention models.

Figure D.1 Number of synchronous PD sessions and total (synchronous sessions and asynchronous hours) training.

(a) Number of Synchronous PD sessions



(b) Total number of PD hours

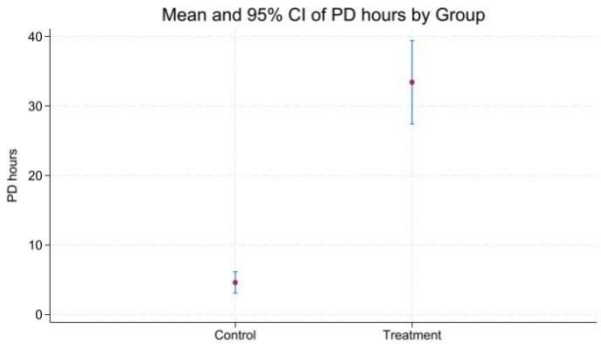


Table D.3 Sessions and hours of synchronous PD, asynchronous hours, and total training hours.

Variable	All	Control			Treatment			P-value
	N	N	Mean	SD	N	Mean	SD	
Synch. sessions	125	63	0	0	62	14.18	8.86	0.000***
Synch. hours	125	63	0	0	62	23.52	13.89	0.000***
Asynch. hours	125	63	4.6	6.28	62	9.91	14.07	0.007***
Total hours	125	63	4.6	6.28	62	33.43	23.63	0.000***

Note: ***p<0.01, **p<0.05, *p<0.1.

Table D.5. Proximal effects of intervention on Classroom Outcomes, accounting for dosage, ITT.

Estimation	ECERS-3				SSTEW				Fidelity Use			
	Synch. Sessions	Synch. Hours	Asynch. Hours	Synch. + Asynch. Hours	Synch. Sessions	Synch. Hours	Asynch. Hours	Synch. + Asynch. Hours	Synch. Sessions	Synch. Hours	Asynch. Hours	Synch. + Asynch. Hours
ITT	0.378	0.164	0.143	0.371	-0.587*	-0.667**	-0.119	-0.299	0.011	0.001	0.017	0.036
	(0.601)	(0.560)	(0.363)	(0.485)	(0.269)	(0.235)	(0.204)	(0.224)	(0.049)	(0.042)	(0.030)	(0.039)
ITT*Dosage	-0.041	-0.014	-0.040	-0.028	0.017	0.014	-0.060**	-0.061***	-0.005	-0.002	-0.012***	-0.011***
	(0.036)	(0.018)	(0.030)	(0.028)	(0.022)	(0.013)	(0.023)	(0.019)	(0.003)	(0.002)	(0.003)	(0.003)
Dosage			0.015	0.013			0.068***	0.070***			0.010***	0.010***
			(0.029)	(0.028)			(0.019)	(0.019)			(0.003)	(0.003)
Observations	124	124	124	124	125	125	125	125	123	123	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. There are no dosage coefficients for synchronous PD as dosage is zero for all the control group and is positive only for the treatment group so the interaction captures this effect.

Table D.6. Proximal effects of intervention on Classroom Outcomes, accounting for dosage, TOT.

Estimation	ECERS-3				SSTEW				Fidelity Use			
	Synch. Sessions	Synch. Hours	Asynch. Hours	Synch. + Asynch. Hours	Synch. Sessions	Synch. Hours	Asynch. Hours	Synch. + Asynch. Hours	Synch. Sessions	Synch. Hours	Asynch. Hours	Synch. + Asynch. Hours
ITT	0.572	0.346	0.287	0.561	-0.636***	-0.707***	-0.162	-0.367*	0.034	0.021	0.031	0.056
	(0.565)	(0.541)	(0.375)	(0.484)	(0.196)	(0.151)	(0.193)	(0.183)	(0.051)	(0.046)	(0.034)	(0.043)
ITT*Dosage	-0.050	-0.019	-0.047	-0.034	0.018	0.015	-0.058**	-0.059***	-0.006	-0.003	-0.012***	-0.012***
	(0.033)	(0.017)	(0.032)	(0.029)	(0.021)	(0.012)	(0.023)	(0.019)	(0.003)	(0.002)	(0.003)	(0.003)
Dosage			0.019	0.016			0.067***	0.070***			0.011***	0.010***
			(0.030)	(0.029)			(0.019)	(0.019)			(0.003)	(0.003)
Observations	124	124	124	124	124	124	124	124	123	123	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. There are no dosage coefficients for synchronous PD as dosage is zero for all the control group and is positive only for the treatment group so the interaction captures this effect.

Table D.7: Proximal effects of intervention on Classroom Outcomes, accounting for *SmartTeach* platform engagement, TOT.

Estimation	-0.079	0.266	-0.002
	(0.519)	(0.353)	(0.068)
TOT	-0.033	0.045	0.007
	(0.030)	(0.029)	(0.006)
Engagement	0.009	-0.083*	-0.005
	(0.042)	(0.039)	(0.006)
TOT*Engagement	3.315***	2.298***	0.128**
	(0.518)	(0.520)	(0.042)
Observations	124	125	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. Engagement is missing for one classroom. We also replicate these estimations with pre-test and while the number of observations decreases, the results are aligned. This is also the case for the TOT estimations.