ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-776

Instructional Alignment is Associated with Sustained Benefits of PreK

Meghan P. McCormick
MDRC

Cullen MacDowell
MDRC

Christina Weiland
University of Michigan

JoAnn Hsueh
MDRC

Michelle Maier
MDRC

Mirjana Pralica
MDRC

Samuel Maves
MDRC

Catherine Snow
Harvard University

Jason Sachs
Boston Public Schools

This study uses implementation fidelity data from PreK to 1st grade in the Boston Public Schools (BPS) to measure instructional alignment and examine whether stronger alignment is associated with sustained benefits of BPS PreK on children's language, literacy, and math skills through first grade. The study includes N = 498 students (mean age = 5.47, SD = 0.30 in K fall). Children who experienced strong instructional alignment across grades had faster gains in literacy (SD = .47) and math (SD = .28) skills through the spring of first grade compared with non-BPS PreK attenders. Mis-alignment predicted faster convergence in literacy skills. Results highlight that instructional alignment may help to sustain the initial benefits of PreK programs through first grade in a subset of outcome domains. Implications for further research measuring alignment in a broader range of settings and implications for practice are discussed.

VERSION: May 2023

Suggested citation: McCormick, Cullen, MacDowell, Christina Weiland, JoAnn Hsueh, Michelle Maier, Mirjana Pralica, Samuel Maves, Catherine Snow, and Jason Sachs. (2023). Instructional Alignment is Associated with Sustained Benefits of PreK. (EdWorkingPaper: 23-776). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/8s3g-tz12

Instructional Alignment is Associated with Sustained Benefits of PreK

Meghan P. McCormick
MDRC

Cullen MacDowell
MDRC

Christina Weiland
University of Michigan

JoAnn Hsueh
MDRC

Michelle Maier
MDRC

Mirjana Pralica
MDRC

Samuel Maves
MDRC

Catherine Snow
Harvard Graduate School of Education

Jason Sachs
Boston Public Schools

WORKING DRAFT

The research reported here was conducted as a part of a study funded by grant R305N160018 – 17 from the Institute of Education Sciences to MDRC with subcontracts to the University of Michigan and the Boston Public Schools. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Thanks to the Boston Public Schools, Josette Williams, Sonia Gomez-Banrey, Brian Gold, the BPS Department of Early Childhood coaches and staff, the BPS Department of Research, and the MDRC team (Jennifer Yeaton, Kelly Terlizzi, Desiree Alderson, Marissa Strassberger and Sharon Huang). The project name is ExCEL P-3 and the IRB approval number is 860661-15. Correspondence should be addressed to Meghan McCormick, Meghan.mccormick@mdrc.org.

Abstract

This study uses implementation fidelity data from PreK to 1st grade in the Boston Public Schools (BPS) to measure instructional alignment and examine whether stronger alignment is associated with sustained benefits of BPS PreK on children's language, literacy, and math skills through first grade. The study includes N = 498 students (mean age = 5.47, SD = 0.30 in K fall). Children who experienced strong instructional alignment across grades had faster gains in literacy (SD = .47) and math (SD = .28) skills through the spring of first grade compared with non-BPS PreK attenders. Mis-alignment predicted faster convergence in literacy skills. Results highlight that instructional alignment may help to sustain the initial benefits of PreK programs through first grade in a subset of outcome domains. Implications for further research measuring alignment in a broader range of settings and implications for practice are discussed.

Keywords: *PreK, instructional alignment, transition, PreK quality*

Instructional Alignment is Associated with Sustained Benefits of PreK

Instructional alignment across PreK and early elementary school – defined here as an approach wherein standards, curriculum and assessments from each grade level serve as a foundation on which to build the next grade (Stipek et al., 2017) – has been proposed as a leading strategy for maintaining the PreK boost (McCormick et al., 2022; Cohen-Vogel et al., 2020). However, outside a few randomized controlled trials of aligned math curricula (Mattera et al., 2021; Clements et al., 2012), there is very little empirical research on the efficacy and scalability of this approach. There are two key issues explaining this lack of evidence – limited take-up of models designed to align instruction and few tools to *measure* alignment. Localities respond to what is measured (Weiland & Guerrero-Rosada, 2022); without actionable data on *mis-alignment,* there is unlikely to be increased movement towards building curricula that promote continuity across PreK and elementary school systems.

In part to address this need, our team developed grade-specific measures of intervention fidelity to the Boston Public Schools' (BPS) aligned PreK to 2nd grade curriculum. In this study, fidelity is defined as the extent to which real-world implementation of the curriculum reflected the original conceptualization, design, and planning of that curriculum (Hulleman & Cordray, 2009). Using data from 163 classrooms across PreK, kindergarten, and first grade, we describe within- and across-grade fidelity to Boston's aligned curricula and use it to measure instructional alignment, compare the language, literacy, and math outcomes of children who experience varying levels of alignment and those who do not attend the BPS PreK program. This is just one approach for measuring instructional alignment but can help provide new information to the field to add to the existing evidence base and the strategies that have already been used to assess alignment (and mis-alignment) (e.g., Cohen-Vogel et al., 2021; Vitiello et al., 2022).

This paper makes three key contributions to the literature. First, this study uses rich observational data to develop a measure of instructional alignment across PreK and elementary school. Previous efforts to measure alignment have examined differences across PreK and kindergarten in teachers' self-reported instructional practices (Cohen-Vogel et al., 2021; Engel et al., 2013), more general observed teacher-child interactions (Vitiello et al., 2022), and time spent in various content areas and learning formats (Early et al., 2010; Justice et al., 2018). These papers have developed important strategies for measuring alignment but have yet to consider the extent to which the curriculum supports continuity and non-redundance in instruction across grades (Stipek et al., 2017). Moreover, there is limited existing evidence on whether stronger alignment sustains the PreK boost. Second, we examine an approach where kindergarten and first grade map onto developmentally appropriate practices from PreK, with a focus on play-based instruction, exposure to rich, thematic content, and significant time in learning centers and small groups (Weiland et al., 2018). Previous research has described how play was removed from many kindergartens as standards in kindergarten increased (Bassok et al., 2016), yet there have few studies of changes meant to counter this trend. And finally, this is the first attempt to our knowledge to examine whether there is empirical evidence that the initial positive effects of a *district-wide* PreK program on children's academic skills are more likely to be sustained when children experience high levels of instructional alignment from PreK to first grade.

Instructional Alignment to Sustain the PreK Boost

Children who attend a formal center-based PreK program as four-year-olds generally begin kindergarten scoring higher on assessments of academic skills than children who do not (Yoshikawa et al., 2013). Yet, these impacts on cognitive and academic skills tend to diminish in early elementary school—a phenomenon commonly known as *fadeout* (wherein impacts

of PreK programs diminish over time) or *convergence* (wherein the children who do not attend PreK program catch up to their PreK-attending peers over time). Although there are a number of theories to explain this pattern (e.g., Bailey et al., 2020; Bailey et al., 2017), lack of *vertical instructional alignment* across early childhood and elementary school settings has been proposed as a key explanation (Stipek et al., 2017). In line with frameworks put forth by Kauerz (2010) and Stipek et al. (2017), when schools are vertically aligned, standards, curriculum and assessments from each grade level serve as a foundation on which to build the next grade. Vygotsky's sociocultural theory suggests that strong alignment will provide children with regular exposure to content that is beyond their current skill level and within their range of abilities – both critical conditions for children's cognitive development (Vygotsky, 1978). When children enter elementary school with the foundational early math and literacy skills learned during PreK (e.g., letter recognition, cardinality), they should be exposed to sequentially more challenging tasks and concepts as they progress in order to maintain the same developmental trajectory.

In reality, children's early learning experiences are likely to be characterized by *mis-alignment* (Abry et al., 2015; Cohen-Vogel et al., 2020). Data from both the U.S. and Australia have shown that teachers in the early elementary school grades report spending substantial instructional time focused on skills already mastered by most kindergarten students, including those who did and did not attend PreK (e.g., Engel et al., 2013). Further work using observational data to measure alignment via the quality of teacher-child interactions (Vitiello et al., 2022), teacher reports of their instructional practices (Cohen-Vogel et al., 2021) and time spent in content areas and learning formats (Early et al., 2010; Justice et al., 2021) has found further evidence for this misaligned pattern.

There is a small but growing empirical literature highlighting the importance of

strengthening instructional alignment via implementation of aligned curricula across the early grades to sustain the PreK boost (Stipek et al., 2017). In the context of two separate randomized controlled trials, cross-grade content-specific curricula paired with training and coaching had positive impacts on student outcomes at the end of kindergarten compared to business-as-usual practice across grades or experiencing the curricula in PreK followed by typical instruction in elementary school (Clements et al., 2013) with further follow-up finding lasting effects through third grade (Mattera et al., 2021). These impacts, however, represent the difference in outcomes for PreK attenders who experienced instructional alignment compared to those who did not. It is still unclear how these children's outcomes compare to those who *did not attend PreK at all*. Finally, these studies were conducted in the context of demonstration studies with significant research team and curriculum developer involvement (Weiland et al., 2018). To date, there are few (if any) examples of school districts that have implemented content-rich, evidence-based curricula to align the early grades in real world, scaled systems (Cohen-Vogel et al., 2020).

Domain-specific Curricula across Grades and Instructional Alignment

To address this need, in 2012 the Boston Public Schools Department of Early Childhood (BPS DEC) began rolling out a program called *Focus on Early Learning* to align the content and mode of instruction, as well as teachers' professional development, within and across PreK through 2nd grade (Bornfreund & Loewenberg, 2018). The public PreK program in Boston is full-day, takes place in public school settings, and is staffed by master's-level teachers (or teachers working to attain a master's degree within five years). The basic iteration of the instructional model has been operating since 2007 and consists of two evidence-based curricula: 1) a language and literacy curriculum incorporating social-emotional components that was adapted from Opening the World of Learning (Schickedanz & Dickinson, 2004) to be culturally

relevant to the BPS population and to integrate principles from Universal Design for Learning; and 2) Building Blocks (Clements et al., 2013), an early mathematics curriculum that also promotes language development by requiring children to explain their mathematical reasoning verbally. Implementation is supported through initial and on-going training and coaching. The design of this approach is rooted in rigorous research showing that content-specific curricula – or those explicitly focusing on a key learning domain (e.g., language/literacy, math) with a specific scope and sequence – have the strongest evidence base for boosting children's learning outcomes (Yoshikawa et al., 2013; Nguyen et al., 2018). Weiland and Yoshikawa (2013) found that the Boston PreK program had moderate-to-large effects on students' language, literacy, math, and executive functioning skills at the start of kindergarten.

The *Focus on Early Learning* approach adjusts elementary school grades so they more closely mirror child-directed early learning practices such as small groups and play-based learning (Bornfreund & Loewenberg, 2018). In line with recommendations from Kauerz (2010) and Stipek et al. (2017), *Focus on Early Learning* content builds in complexity across grades but follows a similar set of thematic units. Because the PreK themes are reintroduced in the higher grades, the content can be expanded in depth and complexity across time. In theory, the instruction should then be vertically aligned across grades. In practice, the field continues to lack tools that allow for measurement of instructional alignment. It is challenging to determine whether this model has been successful in achieving its goals of promoting greater continuity across PreK and early elementary school and sustaining the initial benefits of high-quality PreK.

Fidelity of Implementation as a Tool to Measure Instructional Alignment

Assessing fidelity to the *Focus on Early Learning* model across grades may be one key strategy for measuring instructional alignment in this context. Hulleman and Cordray (2009)

have conceptualized intervention fidelity as having multiple indicators, including dosage (or the amount of the intervention received), adherence (or the extent to which the intervention was delivered as designed), and quality (or the extent to which the intervention the intervention was delivered using high-quality practices). In their own work Hulleman and Cordray (2009) found that although an educational program they tested in the lab with high levels of intervention fidelity yielded a moderate effect size on students' motivation (g = .45), the magnitude of this impact was significantly reduced (g = .05) when implemented in real-world conditions where intervention fidelity was much lower.

Studies examining implementation of programs in early childhood settings have found a similar pattern, highlighting the potential importance of quality and dosage in particular for explaining intervention effects (e.g., Davidson et al., 2009; Hamre et al., 2010; Mendive et al., 2016; Piasta et al., 2015). For example, Hamre and colleagues (2010) examined the use of a literacy and language curriculum supplement in 154 state-funded PreK classrooms serving at-risk populations in the mid-Atlantic and found descriptive evidence that larger impacts on children's emergent literacy and print awareness were associated with higher levels of dosage and better quality of implementation. In a study of a different PreK literacy curriculum implemented in a low-income school district, researchers found that students in PreK classrooms with high levels of fidelity significantly outperformed students in classrooms with low implementation fidelity on measures of phonological awareness (Davidson et al., 2009).

These findings suggest that it may be possible to assess intervention fidelity in order to measure instructional alignment, with greater fidelity within and across years being indicative of stronger instructional alignment across PreK and elementary school. Key to this work, however, is being able to *measure* intervention fidelity in a reliable and valid way for programs

implemented at-scale. To this end, recent work aimed to develop within-grade measures of intervention fidelity to the *Focus on Early Learning* curriculum. Psychometric analyses showed that measures of fidelity – specifically those capturing content-rich instruction and cognitive demand – demonstrated reliable and concurrent validity (Maier et al., 2022). These measures also predicted gains in children's math skills during the academic year, even after controlling for the three domains of the CLASS and other student and classroom-level features, such as the amount of time spent on math instruction. Findings provide preliminary evidence that measures of intervention fidelity to content-rich curricula – like those implemented as part of the *Focus on Early Learning* curriculum – can be helpful tools for describing the extent to which students' experiences are aligned or mis-aligned, and using that information to understand whether alignment is a factor that can sustain the PreK boost for a diverse group of students.

It is important to note that using this approach – focused in assessing fidelity of implementation to an aligned curriculum to measure the broader construct of instructional alignment – is limited in generalizability. At this point, it is unclear if the structural conditions in other settings would allow both implementation and testing of such a model across grades. However, the general pattern of PreK fadeout/convergence in the early elementary grades found in the field is also consistent in Boston (Weiland et al., 2020), suggesting that the same challenges that are happening there align with findings in other localities (e.g., Ansari et al., 2020; Burchinal et al., 2023; Yoshikawa et al., 2013). Learning from what is happening in Boston can serve as a proof of concept exercise to being building needed evidence on whether instructional alignment – when implemented at-scale across typical school-based settings – can sustain the PreK boost. Further work replicating this approach in other settings can then build more evidence that extends beyond this one city and can inform a broader set of districts working

to address these same challenges.

The Current Study

The current study aims to answer two overarching research questions:

1. Do PreK to first grade teachers implement the *Focus on Early Learning* curriculum with high intervention fidelity? And to what extent do children from different backgrounds experience aligned instruction from PreK to first grade?
2. Does instructional alignment predict sustained associations between PreK attendance and language, literacy, and math skills through 1st grade?

Taken together, results will shed important light on measuring features of early learning classrooms that may stand to promote better outcomes for young children within and across time.

Method

Participants

The study sample (N = 498) consists of two groups of students (see supplemental online Appendix E). The first includes 256 students (nested within 41 classrooms and 20 schools in PreK) who enrolled in the public school BPS PreK program during the 2016 – 2017 year and continued to attend a BPS elementary school for kindergarten (2017 – 2018) and first grade (2018 – 2019) and has valid fidelity and outcome data for either of these later grades. These students are a subsample of the 292 who enrolled in the study in 2016 – 2017 and the team followed into elementary school, with 12% attriting from our sample. The second group includes 242 students who did not attend BPS PreK and were recruited in the fall of 2017 from the 49 kindergarten classrooms in 21 public schools where the BPS PreK attenders enrolled. These students had experienced a range of settings in the year prior to kindergarten – 30% of them stayed home with a parent or other adult, or attended a home-based daycare and the remaining 70% enrolled in a non-public PreK program. These students make up the comparison group.

See Table 1 for demographics. As illustrated there, the majority of children in the sample

across grades were eligible for free- or reduced-price lunch and the students were diverse with respect to racial/ethnic background and parental education, among other characteristics. The public schools in the sample are representative of the broader population of BPS elementary schools. The students who enrolled in the study during PreK and kindergarten were also representative of the broader population of students who enrolled in kindergarten in the district.

Procedure

School and classroom recruitment. In 2016, before the start of the PreK, we randomly selected 25 public schools to participate in the study from the 76 schools in the district offering the public school BPS PreK program. Twenty-one agreed. We used one school as a pilot school for developing new measures and the remaining 20 schools made up the public school sample in the first year of the study. All PreK classrooms in the current study implemented the *Focus on K1* curriculum (the PreK component of the broader *Focus on Early Learning* curriculum described above). Teachers received training and coaching as needed. The BPS PreK program is implemented in public schools and co-located with elementary grades, a structural feature theorized to support program quality (Choi et al., 2016). All public school BPS PreK teachers must have a master's degree or be working to attain a master's degree within five years. They are paid on the same scale as other elementary school teachers.

We asked all PreK teachers assigned to general education or inclusion classrooms in each of the 20 public schools to participate in the study in the fall of 2016. Ninety-six percent ($N = 41$) of teachers across public schools agreed to participate. We followed sample children into kindergarten (2017 – 2018) – where we also recruited the non-BPS PreK attenders for comparison purposes -- and then into first grade (2018 – 2019). We were able to collect intervention fidelity data in 49 kindergarten classrooms and 53 first grade classrooms.

Student recruitment. We attempted to collect active consent for all PreK students enrolled in participating classrooms to participate in research activities across multiple years. Eighty-one percent of all children in participating classrooms had parent consent to participate in the study. Of the children with parent consent, we randomly selected 50% (~6 – 10 per classroom) to participate in student-level data collection. For the current study sample, we followed these children across kindergarten and first grade. We repeated this process in the kindergarten classrooms participating in the study in the fall of 2017 and enrolled 78% of students in the participating classrooms who had not attended the public school BPS PreK program in the 2016 – 2017 academic year.

Direct assessments. We trained research staff to reliability and then collected direct assessments of academic skills in the fall and spring of kindergarten (2017-2018), and the spring of first grade (2019). We used the Pre-language Assessment Scale (preLAS; Duncan & DeAvila, 1998) Simon Says and Art Show tests in PreK and kindergarten to determine the administration language for a subset of assessments (Barrueco et al., 2012). In kindergarten, 23 students did not pass the preLAS and completed a subset of assessments in Spanish in the fall and 6 in the spring. We did not administer preLAS in first grade due to high English proficiency.

Classroom observations. We provide a description of procedures for classroom observations in the main text and further details in supplemental online Appendix B.

Live classroom observations. In the winter and early spring (February – early May) in each of the study years, trained instructional coaches from the BPS district observed each participating classroom for two 2-hour blocks of academic instruction (see Table 2 for total length of observation time each year). Coaches only completed fidelity observations in classrooms they did not normally coach. Observations focused on the full classroom including

the lead teacher, the children, and any assistant teachers and other adults. BPS instructional coaches trained as observers participated in a 3-day training in January of each study year to learn how to rate classrooms and teachers on different indicators of intervention fidelity for each component of the *Focus on Early Learning* curriculum within that particular year. Senior members of the research team conducted the training. Coaches (N = 17) were majority female (80%) and diverse with respect to race/ethnicity (33% White, 33% Black, 20% Hispanic, 13% Asian). All were former early childhood or elementary school teachers and about 85% of coaches conducted observations across more than one year. Before collecting data in the field, all coaches had to pass a reliability test with a master coded video. Specifically, they had to agree with 80% of all quality codes across the full tool "within 1" point and needed to demonstrate 80% agreement with all adherence codes. The same procedure for reliability was used for each grade and observational period. During each of the fidelity data collection periods, coaches double-coded 20% of observations to assess interrater reliability, showing high levels of agreement, with 89% of quality double codes being reliable "within 1" (aligned with reliability standards on observational measures like the CLASS). There was 82% agreement on average across grades on adherence codes. To create the overall fidelity scores described in the measures section below, we took the average score across both observations. The team also coded all classrooms using the Classroom Assessment Scoring Systems (Pianta et al., 2008) and used those data for robustness checks. See supplemental online Appendix B for more information.

Teacher surveys. In the spring of each year, we asked teachers to report on their demographic characteristics and teaching experience. They also reported on their own implementation of the *Focus on Early Learning* curriculum and their receipt of training and coaching on the curriculum.

Parent surveys. In the fall of 2016 and 2017, using text messages, emails, and sending forms home, we contacted the consenting parents of all students selected for the study sample to complete a twenty-minute survey. We translated the surveys into Spanish, Vietnamese, and Mandarin. Across both survey waves, eighty-six percent of respondents were mothers and 12% were fathers. All parents received a $25 gift card for completing the survey. 94% of students in the current study had a parent who completed the survey in 2016 or 2017.

Administrative data from the school district. We accessed administrative records from the BPS district on students' demographic characteristics, history of enrollment in the BPS PreK program, classroom and school membership at the end of each year, and teacher-collected assessments of student literacy skills in the fall and spring of kindergarten.

Measures

Intervention fidelity to the BPS *Focus on Early Learning* model within grades: dosage, adherence, and quality. The research team worked closely with the BPS DEC to create an observational tool to assess fidelity to the *Focus on Early Learning* model within each grade. The fidelity measures had consistent items and components across years (see summary of components in supplemental online Appendices C and D) but also had some year-specific components. For the purposes of the current study, we focus on the components and fidelity domains that were consistent across years. Existing studies have already reported on the creation of this tool (McCormick et al., 2020) and its reliability and validity (Maier et al., 2022). In line with recommendations from Hulleman and Cordray (2009), we first used these tools to measure 3 dimensions of intervention fidelity – dosage, adherence, and quality – within each year.

Implementation dosage. Dosage captured the number of curricular components observed and the amount of time spent doing them.

Adherence to the curriculum. Adherence items captured the extent to which each curricular component was implemented as intended. We coded adherence to the curriculum as the proportion of adherence items that the teacher was observed to implement (maximum score of 1) within each component. We then averaged across all of the observed components to calculate a total adherence score (which was a percentage ranging from 0 to 100).

Quality of curricular implementation. Quality items captured the manner by which the curricular components were delivered and whether teachers used particular instructional practices. These items included a detailed set of anchors and descriptors using a 5-point Likert scale where 1 = low quality and 5 = high quality. We calculated the average quality score for each teacher across all quality items in curricular components that were observed.

Instructional alignment using cross-component constructs. Because almost all classrooms were implementing the *Focus on Early Learning* intervention, we decided to use measures of intervention fidelity across study years to measure instructional alignment. Specifically, we used a face validity approach to create four conceptually meaningful constructs that mapped onto the key domains that the *Focus on Early Learning* curriculum aimed to support across years – rich vocabulary (or the extent to which students are exposed to advanced vocabulary aligned with the curriculum across different activities), scaffolding and differentiation (the extent to which teachers provide individualized supports to students), extending/building on knowledge (the extent to which teachers used the curriculum as a tool to extend what children were learning and helping them build on their knowledge), and summarizing/making connections across thematic units (or succinctly summarizing key themes and goals of each lesson and connecting them to other activities happening inside and outside of the classroom). Items are listed in supplemental online Appendix D. By measuring these key domains we were able to maintain consistency in

the observational items that were used to capture intervention fidelity across time. We then mapped these domains across years in order to operationalize instructional alignment. We based the decision to average across core components on work examining implementation of the Responsive Classroom intervention (Abry et al., 2015) and work by Nelson et al. (2010).

Language skills. We used the Peabody Picture Vocabulary Test IV (PPVT IV; Dunn & Dunn, 2007) to directly assess children's receptive language skills in the fall and spring of kindergarten and the spring of first grade. The PPVT IV is a nationally normed measure that has been used widely in diverse samples of children. The test has excellent split-half and test– retest reliability estimates, as well as strong qualitative and quantitative validity properties (Dunn & Dunn, 2007). It requires children to choose (verbally or nonverbally) which of four pictures best represents a stimulus word. We used the PPVT raw score as our outcome. We assessed all children on the PPVT– regardless of whether they passed the PreLAS language screener – in order to describe an equivalent measure of language skills in English across the sample.

Literacy skills. We used the Letter Naming Fluency (LNF) subtest from the teacher-reported *Dynamic Indicators of Basic Literacy Skills - Next* (DIBELS; Cummings et al., 2011) to measure children's literacy skills in the fall and spring of kindergarten. This subtest has high levels of reliability, good concurrent, predictive, and discriminant validity properties, is widely used, and is sensitive to intervention (e.g., Biancarosa et al., 2010). Although there were other subtests available, we chose to use only the LNF subtest because it was the only one collected in both the fall and the spring of kindergarten. This decision restricted the number of statistical tests and reduced concerns about multiple comparisons. It was only available in English.

The district stopped collecting the DIBELs during the 2018 – 2019 school year. As such, we used the *Woodcock Johnson Letter-Word ID* (WJLWID) subtest to assess students' literacy

skills in the spring of first grade. The Letter-Word ID subtest assesses letter naming and word decoding skills by asking children to identify a series of letters and words presented in isolation. It is a nationally normed and widely used achievement test with demonstrated internal consistency and validity (e.g., Wechsler, 1989). In models predicting the WJLWID, we used the DIBELS LNF subtest as a conceptually similar covariate.

Math skills. We first used the Woodcock Johnson Applied Problems III (Woodcock, McGrew, & Mather, 2001) subtest to assess math skills in the fall and spring of kindergarten and the spring of first grade. We assessed Spanish-speaking children who did not pass the PreLAS language screener using the Spanish language version from the Batería III Woodcock Muñoz (Schrank et al., 2005). The WJ/WM Applied Problems assessment is a numeracy and early mathematics measure that requires children to perform calculations to analyze and solve arithmetic problems. Its estimated test–retest reliability for 2- to 7- year-old children is 0.90 and it has been used with diverse populations (Woodcock et al., 2001). We present results using the raw score, with English and Spanish combined.

The Woodcock-Johnson Applied Problems subtest has been criticized by some math experts because it is not particularly sensitive in the early childhood years, skips quickly to difficult items, and does not include geometry. As such, we also used the Research-based Early Mathematics Assessment (REMA; Clements et al., 2008) to assess math skills in the fall and spring of kindergarten, and the spring of first grade (we were unable to collect this assessment in the fall of PreK). The REMA is a hands-on, one-on-one assessment that measures core mathematical abilities of children ages 3 – 8. As described further in Sarama and colleagues (2012), children's abilities are assessed on the REMA according to theoretically and empirically based developmental progressions that underlie learning trajectories. The assessment captures

verbal counting, object counting, subitizing, number comparison, number sequencing, connection of numerals to quantities, number composition and decomposition, adding and subtracting, place value, shape recognition, congruence, construction of shapes, and spatial imagery, as well as measurement and patterning. Assessors code not only whether the child answered each item correctly, but also the difficulty of the strategy used to respond. The alpha reliabilities for the total scores (referred to as t scores) range from .92 to .94 (Sarama et al., 2012) with prior evidence of construct and concurrent validity (Clements et al., 2008).

Child characteristics from administrative data. Using administrative data, we created a series of indicators to describe children's race/ethnicity (Black, Hispanic, Asian, or Other Race/Ethnicity), coding 1 if the child fell into the indicated category and 0 otherwise (reference group White). We used similar indicators to describe eligibility for free or reduced price lunch (FRPL; 1 if eligible; 0 if not) and gender (1 = female; 0 = not female). We set a dummy variable for Dual Language Learner (DLL) equal to 1 if the parent reported that there was a language other than English spoken at home and 0 otherwise. We used the child's birthdate to calculate age on September 1st, 2017 (start of kindergarten). These covariates – and those below -- have been shown to predict outcomes in prior work (Choi et al., 2018).

Family characteristics from parent survey. Parents reported on demographic characteristics in the fall of the PreK year and we used these indicators as covariates. These variables indicated whether there was at least one parent in the home working full-time (35 hours/week or more) and whether the parent was married or lived with a partner. We used continuous variables to describe the age of the child's mother at her first birth, household size, and parental age in the fall of the prekindergarten year. We then included three dummy variables to describe the reporting parent's level of education – high school diploma or GED or less, some

college/2-year degree, and 4-year degree (with graduate study + as the reference group).

Analytic approach

Missing data. Overall, there was a relatively low amount of missing data. All students had complete data on child-level information provided by the school district. About 13% of students were missing parent survey covariate data, 10% were missing some outcome data on one of the assessments our research team collected, and 24% of students were missing information on the DIBELS outcome that was collected by the district. We used complete case analysis to deal with missing data but compared results to multiple imputation with $N = 100$ datasets (Enders, 2017; see summary of findings in supplemental online Appendix A).

Creation of alignment profiles. As noted above, we created four cross-component constructs to describe overall indicators of intervention fidelity to the *Focus on Early Learning* curriculum -- rich vocabulary, scaffolding/differentiation, extending/building knowledge, and summarizing/making connections. We held a number of research meetings involving content experts, early childhood researchers, and BPS instructional coaches to finalize the items included in the domains. Through this process, we reached greater than 80% agreement for including each item within its target domain. We recoded all adherence items (originally coded 1/0) so that a score of 1 was coded as 5 and a score of 0 was coded as 1. We then created an average score for each domain within year by dividing the sum of the items by the total number of non-missing items. Overall scores for each cross-component construct ranged from 1 (low fidelity) to 5 (high fidelity). Inter-item reliability was high ($\alpha = .84 – .96$ across domains and years) and constructs had low to moderate correlations, demonstrating that they were not redundant ($r = .19$ to $.74$).

We then used Latent Profile Analysis (LPA) to group students who attended BPS PreK (N = 256) into profiles based on their within-grade fidelity scores from PreK to first grade. In

line with recommendations from Masyn (2013) and used most recently by McCoy et al. (2022), we used five statistical and substantive criteria to determine the number of profiles (Masyn, 2013): (a) Akaike's information criterion (AIC), Bayesian information criterion (BIC), and sample-adjusted BIC, with lower values representing better fit; (b) the Vuong–Lo–Mendell–Rubin likelihood ratio test (VLMR-LRT), with a significant VLMR-LRT, suggesting better fit relative to the model with one fewer profile and a nonsignificant VLMR-LRT, suggesting no difference in overall fit; (c) entropy, with values closer to one indicating higher levels of separation between profiles; (d) the number of students in each profile and what would be appropriate for statistical power in predictive analyses; and (e) the extent to which the number of profiles mapped onto what was conceptually and practically meaningful for the district.

Four of our five statistical and substantive criteria for selecting an LPA solution suggested that a four-profile solution best fit the data from our 256 BPS public PreK attenders. In particular, reductions in the AIC, BIC, and sample-adjusted BIC diminished in models with greater than four profiles; the four-profile solution showed an adequately high entropy value of .82; each of the four profiles was adequately sized (smallest $N = 40$); and the four-profile solution demonstrated theoretical relevance. The only criterion not optimized was the VLMR-LRT, which showed that the four-profile solution did not differ in model fit from the three-profile solution ($p = .69$). As illustrated in Figure 1, we named the profiles: 1) High alignment (high values across time: $N = 69$); 2) Mixed: high-low-high (high fidelity in PreK followed by lower fidelity in kindergarten and higher in first grade; $N = 42$; also called Mixed: HLH); 3) Mixed: gradual decline (decline across grades; $N = 40$, also called Mixed: GD); and 4) Low alignment (low fidelity in all years; $N = 105$).

Research question 1: Evidence of instructional alignment from PreK to first grade

and variation across different groups of students. We used descriptive statistics to examine the means and SDs for dosage, adherence, and quality of intervention fidelity in PreK, kindergarten, and first grade. We then examined descriptive statistics for the cross-component constructs within and across grades as well. We used ANOVA with post-hoc tests to examine whether these measures were significantly different from one another across grades. We examined how profile membership varied by children's demographic characteristics.

Research question 2: Associations between alignment profiles and child outcomes. We used multi-level modeling to examine associations between fidelity profiles and gains in children's skills across kindergarten and first grade, relative to children who did not attend public PreK. Examination of unconditional means models and intraclass correlations (ICCs) revealed that two-level models with students nested within kindergarten public schools was the best fit to the data. This decision also mapped onto the measurement of alignment, with students changing classrooms across time but being nested in the same public schools for the full period.

We then regressed each of the outcomes – measured separately at the end of kindergarten and the end of first grade – on the alignment profile dummies. Non-PreK attenders were the reference group. The base equation for these models is included in supplemental online Appendix F. The coefficients on the alignment indicators and their associated *p*-values are the parameters of interest. These represent the difference in gains made in the outcome between that group and the PreK non-attenders. We calculated standardized associations for significant associations by dividing parameter estimates by the standard deviation of the outcome.

Results

Descriptive Analysis Comparing Alignment Profiles and Non-PreK Attenders

As the team reported in prior studies using this same sample (Authors, 2021; Authors,

2022), public PreK attenders were significantly less likely than their non-attending peers to be Black or eligible for free- or reduced-price lunch. These racial/ethnic and socioeconomic disparities in enrollment in BPS public PreK mirror prior work done with district-wide data (McCormick et al., 2023). In the current study we extended this work to compare the characteristics of children in each of the alignment groups and the children who did not attend the public PreK program. As fully described in Table 1, children who were in the high alignment group were less likely to be eligible for free or reduced price lunch, less likely to be Dual Language Learners, more likely to be White, and had parents with higher levels of education than students in the full sample, and students in the other fidelity profiles.

Descriptive Data on Fidelity to *Focus on Early Learning* Across Time

Fidelity of implementation ranged from moderate to high across different indicators (see Table 2). Depending on grade, instructional coaches were able to see about half to 60% of the total curricular components across their classroom observations. Adherence was consistent across grades – generally about 70% or higher – which is the target that BPS coaches aim to help teachers achieve given the recognition that adaptation of the curriculum will be needed within each unique classroom setting. Overall quality was also consistent and ranged from 3.10 – 3.36 on a scale from 1 – 5. Quality was slightly higher in first grade compared to PreK and kindergarten. First grade classrooms had particularly high scores on exposure to rich vocabulary (3.57) compared to the PreK and first grade classrooms ($p < .001$). And kindergarten and first grade classrooms had higher scores on scaffolding/differentiation than PreK. Content-rich instruction and cognitive demand scores revealed moderate levels of implementation quality.

The high alignment group experienced cross-component fidelity scores above 3.9 on average in each of the three years. The low alignment group experienced lower scores on the

cross-component items across years ranging from 3.03 to 3.47 on a scale of 1 – 5. The difference in PreK fidelity between these high and low groups translates into about .75 SDs. The Mixed: HLH group experienced high fidelity in PreK (mean = 3.75) followed by a marked drop in kindergarten to 3.0 before spiking back up in first grade to 3.7. And the Mixed: GD group experienced high fidelity in PreK (3.75) followed by gradually lower scores by first grade (3.35).

Associations Between Instructional Alignment and Language Outcomes

All results from gains models are presented in Table 3. As illustrated, there were no statistically significant associations between any of the alignment profiles and gains in language skills through the spring of kindergarten. This means that PreK attenders across alignment profiles and non-attenders made similar gains in language across the school year. However, children in the high alignment actually had slower gains through the spring of first grade compared to non-attenders. Coefficients for the other profiles were trend-level or null.

Associations Between Instructional Alignment and Literacy Outcomes

Students in the Mixed: HLH improvement group had slower gains in literacy skills during kindergarten than non-attenders as did students in the low alignment group. These translate into about 7 months and 4 months of learning, respectively and reveal that mis-alignment contributed to catch-up in literacy skills in kindergarten. Although the coefficients for high alignment and gradual decline were also negative one was trend-level and one was not statistically significant. On average, these findings map onto an overall pattern of convergence of literacy skills by spring of kindergarten. The direction of the results differed for first grade. We found that children in the high alignment group had bigger gains in literacy skills through the spring of first grade (predicting WJLWID and adjusting for the DIBELS letter naming score in the fall of kindergarten) than non-attenders as did children in the Mixed: HLH group. The

magnitude of these gains translates into about 4 months of learning in first grade (Lipsey et al., 2013). The other coefficients were null.

Associations Between Instructional Alignment and Math Outcomes

In models predicting the REMA – which considers a broader range of math skills than the WJAP and takes problem solving and strategy use into account when scoring – we found that students in the gradual decline group had slower gains in math through the spring of kindergarten than non-attenders. This translates into about 2 months of learning. All other associations were not statistically significant in kindergarten for both math assessments. In first grade, we found that children who were in the high alignment group experienced larger gains on the REMA than non-attenders. See a visual illustration of this pattern in Figure 2. There were no other statistically significant coefficients for the other alignment groups. This pattern suggests that although high alignment can contribute to sustained associations between BPS PreK and math skills, mis-alignment only really had negative implications through the spring of kindergarten for children who experienced strong PreK followed by a steep drop in fidelity during kindergarten.

Robustness Checks

We used the following checks to assess the sensitivity of substantive findings: 1) compared the findings from the LPA to a different approach for creating profiles using median-splits to create alignment groups; 2) included the three domains of the CLASS (Pianta et al., 2008) as covariates in predictive models to determine whether alignment predicted sustained benefits of PreK over and above the quality of teacher practices; 3) considered different strategies for treating missing data; 4) used school fixed effects in models rather than random effects; and 5) controlled for classroom-level indicators of variation in children's skills by including the SD of the classroom-level skill aggregate in models. Findings were generally

robust with respect to magnitude and statistical significance (see supplemental online Appendix A). We actually found more compelling evidence for the importance of instructional alignment when we used the manual median split approach to create the alignment groups.

Discussion

The field of early care and education continues to tackle the challenge of implementing high-quality PreK programming on a wide-scale and building systems to promote positive developmental outcomes even after children move to elementary school (McCormick et al., 2017). Key to the success of this work is identifying evidence-based strategies for promoting continuity in children's learning across PreK and the transition to the early grades (Stipek et al., 2017). Current findings may hold important lessons for other localities aiming to align instruction across the critical transition from PreK to elementary school by using a curriculum and PD model like *Focus on Early Learning*. For example, one might expect that although this child-centered approach rooted in developmentally appropriate practice for young children can work well in early childhood settings with supports, it could be more challenging to implement in kindergarten or first grade which some research shows have become more teacher-directed and focused on direct academic instruction (particularly first grade) in recent decades (e.g. Bassok et al., 2016). Yet, intervention fidelity – observed using a validated tool measuring adherence and quality (Maier et al., 2022) – was moderate to high across different components, domains, and years. We observed the highest fidelity in first grade, which may reflect those classrooms receiving intensive training and coaching most recently.

Our findings too provide some insights on benchmarks for implementation outside of a tightly controlled trial. That is, the existing literature on educational interventions has traditionally used an 80% adherence rate as a core goal for establishing high levels of

intervention fidelity (e.g., Horner et al., 2004; Pas et al., 2019). These levels, however, are generally used in the context of researcher-directed trials where significant focus and resources are put into maintaining high levels of intervention fidelity. The majority of classrooms in the current study (62% across years) did implement *Focus on Early Learning* with at least 70% adherence for all components and domains, a threshold that is likely meaningful for observing acceptable levels of fidelity under real-world conditions (Hulleman & Cordray, 2009) and is the standard that BPS has set for teachers to achieve when they use this tool for program strengthening purposes. The district does not aim to achieve perfect fidelity because they recognize that teachers need the clear ability to adapt the curriculum as needed for their specific context. Although BPS has been implementing their public PreK model since 2007 and started *Focus on Early Learning* in kindergarten in 2012, the district-wide roll-out of the aligned approach did not begin in the majority of kindergarten and first grade classrooms until 2014 – 2015, just three to four years before the data were collected for the current study. The finding that fidelity was fairly consistent across grades in spite of this short implementation time frame is encouraging for districts interested in making investments related to curricula and alignment, supported by training and on-going coaching.

Yet, given evidence from trials of similar studies finding that effects may be larger when fidelity is stronger (Hulleman & Cordray, 2009; Mendive et al., 2016), there is a continued need to invest in systems that monitor and support high levels of fidelity across grades and time. This is further supported by our finding that only about one quarter of the PreK attenders in the sample made up the "aligned" group and experienced high levels of fidelity – about a 3.5 on a score of 1 – 5 – in each grade. BPS developed the observational measure of intervention fidelity data not just to conduct research but also to provide instructional coaches with a reliable and

valid tool to collect consistent information on model implementation within and across grades. Coaches can use the tool to collect information when observing teachers and then provide that feedback directly in line with the observational system. Teachers would then have access to the observational tool to help reflect on their own practice and improve their ability to implement with higher levels of adherence and quality to the model.

Investing in structures – like curriculum and professional development – that target alignment rather than trying to strengthen the general quality of teacher practices may be key. In a meta-analysis, Bailey et al. (2020) found no systematic evidence that the global quality of subsequent learning environments influenced whether the initial benefits of PreK participation were sustained in elementary school. And additional research has found evidence that pre-existing student characteristics are much more powerful predictors of PreK fadeout/convergence than the quality of subsequent of learning environments (Ansari et al., 2020; Bailey et al., 2016). These studies suggest the need for more *tailored curricular interventions* – rather than general quality supports – to sustain the PreK boost (e.g., Clements et al., 2013; Mattera et al., 2021).

Importantly, however, these findings were not consistent across grades or learning domains. Indeed, stronger instructional alignment was predictive of gains in children's skills through first grade but not through the spring of kindergarten. We only observed complete convergence by the spring of kindergarten for one outcome – the DIBELS literacy measure. The reductions in associations between BPS PreK enrollment and vocabulary and math outcomes were much smaller for outcomes assessed in the spring of kindergarten. At the same time, literacy was the only domain where greater mis-alignment was associated with smaller gains during kindergarten compared to non-attenders. It may be that more constrained skills (see more in Paris, 2005; Snow & Matthews, 2016) – like literacy measured with the DIBELS – are more

sensitive to weaker alignment particularly during kindergarten, when research shows the bulk of convergence traditionally occurs (Weiland et al., 2021). This is the grade that will always include a certain proportion of non-PreK attenders because PreK attendance is not (and is unlikely to be) mandatory. As such, it may be more challenging to align instruction when there will always be a subset of students who need to catch up to their PreK attending peers and learn foundational skills in key domains like literacy in order to meet established goals and benchmarks. In contrast, once students transition into first grade and are all working on a more advanced set of literacy skills, greater instructional alignment – where students are called to build on their earlier PreK and kindergarten experiences –may be better structured to sustain the benefits of PreK over time.

We observed benefits of strong alignment on math skills for the REMA outcome but not the WJAP and through first grade but not kindergarten. It is possible that curricula like *Focus on Early Learning* stand to support a broader range of skills – like problem solving, measurement, knowledge of data – measured with the REMA but not with the WJAP. Indeed, in a study examining the impacts of intentionally aligned math curricula across PreK and kindergarten, Mattera et al. (2021) found impacts on math skills in the spring of kindergarten when measured with the REMA but not with the WJAP. The PreK curriculum in that study was the same as the one implemented in BPS in the current study. Similarly, mis-alignment was related to slower growth on the REMA but not the WJAP. It is possible that the REMA is simply a more sensitive measure due to the broader set of domains assessed. It might take longer for the benefits of instructional alignment to become salient when the initial effects of PreK are sizeable. Findings demonstrate the importance of considering alignment beyond the transition to kindergarten.

And in contrast to findings demonstrating benefits of instructional alignment, we also found that stronger alignment was associated with *smaller* gains in language skills through the

spring of first grade compared to non-attenders. This was surprising in part because vocabulary is an unconstrained skill (Snow & Matthews, 2016) meaning that there is no limit on the extent to which students can develop vocabulary skills and students should be able to continue building on skills within this domain across years when supported by high-quality content and instruction (Pianta et al., 2021). However, in our context, PreK attenders did experience a substantial boost in vocabulary skills, on average, that was largely sustained through the end of kindergarten. It could be that there was inevitably some slowing down in the development of vocabulary in first grade due to the size of the initial boost and the fact that it was largely sustained through the end of kindergarten. In addition, implementation of vocabulary in particular was very strong in first grade and was substantially higher than the other domains of curricular implementation. It is possible that non-BPS PreK attenders were able to increase vocabulary skills faster than would have been expected given more normative levels of implementation (like in kindergarten). This pattern – where exposure to more advanced content is actually linked to faster fadeout/convergence – has been documented in earlier work (e.g., Bierman et al., 2014; Magnuson et al., 2007). High fidelity on vocabulary may not have had the same benefits for students who experienced strong PreK followed by strong subsequent implementation.

Limitations and Directions for Future Research

Although this study is novel and adds significantly to research focused on sustaining the initial benefits of PreK, there are a number of limitations that are important to consider. First, this study uses a descriptive design and we cannot make any causal connections between curricular alignment and child outcomes. Although we used residualized gains models and controlled for a robust range of covariates, it is certainly possible that some unobserved confounding variables could be driving our results, particularly given the documented selection

into alignment profiles. A complementary study leveraging natural lotteries to estimate the causal impact of *Focus on Early Learning* is underway and aims to address this study limitation (see more in Authors, 2022). Second, there were some differences in measures across years that could be affecting the pattern of results. For example, we were only able to collect the WJ Letter Word ID in the spring of first grade to assess literacy skills and had access to the DIBELS letter naming subtest in the fall and spring of kindergarten. We used the DIBELs as a covariate in the model predicting the WJ Letter Word ID. Differences in the measures could potentially explain why there were linkages between alignment and outcomes in first grade but not kindergarten. Third, our measures of fidelity relied on two observations aggregated within each year. Like all observational measures, we may be missing key information that could help to operationalize experiences of instructional alignment. And the measure of alignment relies assess intervention fidelity to an existing curriculum. Further work integrating teacher survey data and a broader number of observations may be helpful for addressing these concerns. Once we broke up our groups into alignment profiles, they were likely too small to use to examine variation by race/ethnicity, family income, and home language and we did not explore variation by subgroup. Conducting further research with a larger sample will be important for unmasking policy-relevant heterogeneity in associations. Finally, results from this study may be specific to the Boston context and may not generalize to a broader group of settings. Work to replicate this research on instructional alignment is needed.

Implications for Policy and Practice

The current study provides some initial evidence that, given sufficient investments in coaching and professional development and structures that provide some oversight over both PreK and elementary school settings, it is feasible to implement aligned curricula from PreK to

first grade with moderate to high levels of intervention fidelity. And collection of systematic, observational data can be used to measure instructional alignment – a critical tool for strengthening program quality at-scale. Through implementation and measurement, it is possible to support children to experience stronger instructional alignment across grades, potentially reducing redundancy, helping students to more effectively build on their skills in each grade, and creating more seamless transitions across years. Districts interested in instructional alignment may look to *Focus on Early Learning* – or other aligned approaches (e.g., Clements et al., 2013; Mattera et al., 2021) – if they are interested in making these investments. Although descriptive, results suggest that efforts to align instruction and promote more advanced content seamlessly across grades in this way may be one lever for sustaining the initial benefits of PreK (for some domains) as children move into the later elementary grades and can potentially address negative implications of mis-alignment, particularly for key foundational skills literacy. Districts and states who choose to take this approach should focus on integrating research and evaluation into their work in order to expand the generalizability of these findings and understand whether findings can be replicated across different settings and contexts. Importantly, BPS is currently working with several state departments of education – including Rhode Island, Maine, and Mississippi – to implement *Focus on Early Learning* across public PreK and elementary school settings so there are current opportunities to determine whether results are generalizable to a broader set of contexts. Further research should also consider how effects of instructional alignment may vary for students from different groups. For example, students with special needs and Dual Language Learners may not benefit similarly from instructional alignment reforms if efforts are not also put in place to differentiate across years and ensure targeted supports specific to the unique needs of those students. Further replication and subgroup work is needed to explore

these questions. And continued research leveraging experimental designs will be important for continuing to strengthen the evidence base for these models and lessons for implementation.

References

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, *36*(3), 320–338. https://dx.doi/10.1177/1098214014557009

Ansari, A., Pianta, R. C., Whittaker, J. V., Vitiello, V. E., & Ruzek, E. A. (2020). Persistence and convergence: The end of kindergarten outcomes of pre-k graduates and their nonattending peers. *Developmental Psychology*, *56*(11), 2027–2039. https://doi/10.1037/dev0001115

Bailey, D. H., Duncan, G., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. https://doi.org/10.1080%2F19345747.2016.1232459

Bailey, D. H., Jenkins, J. M., & Alvarez-Vargas, D. (2020). Complementarities between early educational intervention and later educational quality? A systematic review of the sustaining environments hypothesis. *Developmental Review*, *56*, 100910. https://doi.org/10.1016/j.dr.2020.100910

Bailey, D. H., Nguyen, T., Jenkins, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences?. *Developmental psychology*, *52*(9), 1457. https://doi.org/10.1037/dev0000188

Barrueco, S., Lopez, M., Ong, C., & Lozano, P. (2012). *Assessing Spanish-English bilingual preschoolers: A guide to best approaches and measures*. Paul H Brookes Publishing.

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, *2*(1). https://doi.org/10.1177/2332858415616358

Biancarosa, G., Bryk, A. S., Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal*, *111*(1), 7–34. https://doi.org/10.1086/653468

Bierman, K. L., Nix, R. L., Heinrichs, B. S., Domitrovich, C. E., Gest, S. D., Welsh, J. A., & Gill, S. (2014). Effects of Head Start REDI on children's outcomes one year later in different kindergarten contexts. *Child Development, 85*(1), 140–159. https://doi.org/10.1111/cdev.12117

Bornfreund, L., & Loewenberg, A. (2018). A Focus on Teaching and Learning in Pre-K through 2nd Grade: Lessons from Boston. *New America* Foundation Report. files.eric.ed.gov/fulltext.pdf

Bronfrenbrenner, U. & Morris, P.A. (1998). The ecology of developmental processes. In R. Lerner (Vol. Ed.) Handbook of child psychology: Theoretical models of human development. (5th Ed., Vol. 1, pp 993-1028). New York, N.Y.: John Wiley.

Choi, J. Y., Elicker, J., Christ, S. L., & Dobbs-Oates, J. (2016). Predicting growth trajectories in early academic learning: Evidence from growth curve modeling with Head Start children. *Early*

Childhood Research Quarterly, *36*, 244–258. https://doi/10.1016/j.ecresq.2015.12.017

Choi, J. Y., Jeon, S., & Lippard, C. (2018). Dual language learning, inhibitory control, and math achievement in Head Start and kindergarten. *Early Childhood Research Quarterly*, *42*, 66-78. https://doi.org/10.1016/j.ecresq.2017.09.001

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Maths Assessment. *Educational Psychology*, *28*(4), 457–482. DOI: 10.1080/01443410701777272

Clements, D. H., Sarama. J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness*, *50*(4), 812–850. https://doi.org/10.3102/0002831212469270

Cohen-Vogel, L., Little, M., Jang, W., Burchinal, M., & Bratsch-Hines, M. (2021). A missed opportunity? Instructional content redundancy in Pre-K and kindergarten. *AERA Open*, *7*, 23328584211006163.

Cohen-Vogel, L., Sadler, J. R., Little, M., & Merrill, B. (2020). (Mis)Alignment of instructional policy supports in pre-k and kindergarten: Evidence from rural districts in North Carolina. *Early Childhood Research Quarterly*, *52*(Part B), 30–43. https://doi/10.1016/j.ecresq.2019.11.001

Cummings, K. D., Kaminski, R. A., Good, R. H. III, & O'Neill, M. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of first sound fluency. *Assessment for Effective Intervention*, *36*(2), 94–106. https://doi/10.1177/1534508410392209

Duncan, S. E., & De Avila, E. A. (1998). PreLAS 2000. Monterey, CA: CTB/McGraw-Hill.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). NCS Pearson, Inc. https://psycnet.apa.org/record/9999-15144-000?doi=1

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D. M. C., Crawford, G. M., ... & Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *25*(2), 177-193.

Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behavior Research and Therapy*, *98*, 4–18. https://doi.org/10.1016/j.brat.2016.11.008

Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, *35*(2), 157–178. http://doi.org/10.3102/0162373712461850

Guerrero-Rosada, P., Weiland, C., McCormick, M., Hsueh, J., Sachs, J., Snow, C., & Maier, M. (2021).

Null relations between CLASS scores and gains in children's language, math, and executive function skills: A replication and extension study. *Early Childhood Research Quarterly, 54*(1), 1–12. https://doi.org/10.1016/j.ecresq.2020.07.009

Hamre, B. K., Justice, L. M., Pianta, R. C., Kilday, C., Sweeney, B., Downer, J. T., & Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly, 25*(3), 329–347. https://doi.org/10.1016/j.ecresq.2009.07.002

Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The School-Wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions, 6*(1), 3–12. https://doi.org/10.1177/10983007040060010201

Hulleman, C., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88–110. http://dx.doi.org/10.1080/19345740802539325

Justice, L. M., Jiang, H., Purtell, K. M., Lin, T., & Ansari, A. (2021). Academics of the early primary grades: Investigating the alignment of instructional practices from pre-k to third grade. *Early Education and Development, 33*(7), 1237–1235. https://doi.org/10.1080/10409289.2021.1946762

Kauerz, Kristie. 2010. *PreK-3rd: Putting Full-Day Kindergarten in the Middle. PreK-3rd Policy Action Brief, No. Four*. New York: Foundation for Child Development.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *The Economics of Early Childhood Education, 26*(1), 33–51. https://doi.org/10.1016/j.econedurev.2005.09.008

Maier, M. F., McCormick, M. P., Xia, S., Hsueh, J., Weiland, C., Morales, A., ... & Snow, C. (2022). Content-rich instruction and cognitive demand in Prek: Using systematic observations to predict child gains. *Early Childhood Research Quarterly*, *60*, 96-109.

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods: Statistical analysis* (pp. 551–661). Oxford University Press.

Mattera, S. K., Jacob, R. T., MacDowell, C., & Morris, P. (2021). *Long-term effects of enhanced early childhood math instruction: The impacts of making pre-k count and high 5s on third-grade outcomes*. MDRC. https://www.mdrc.org/sites/default/files/MPC_High5-3yr_Impact_Report.pdf

McCormick, M., Pralica, M., Hsueh, J., Weiland, C., Weissman, A. K., Shapiro, A., ... & Sachs, J. (2023). Going the Distance: Disparities in Pre-K Enrollment in Higher-Quality Schools by Geographic Proximity, Race/Ethnicity, Family Income, and Home Language. *AERA Open*, *9*, 23328584231168867.

McCormick, M. P., Pralica, M., Weiland, C., Hsueh, J., Moffett, L., Guerrero-Rosada, P., ... & Sachs, J.

(2022). Does kindergarten instruction matter for sustaining the prekindergarten (PreK) boost? Evidence from individual-and classroom-level survey and observational data. *Developmental Psychology*, *58*(7), 1298.

McCormick, M. P., Weiland, C., Hsueh, J., Maier, M., Hagos, R., Snow, C., ... & Schick, L. (2020). Promoting content-enriched alignment across the early grades: A study of policies & practices in the Boston Public Schools. *Early Childhood Research Quarterly*, *52*, 57-73.

McCoy, D. C., Koepp, A. E., Jones, S. M., Bodrova, E., Leong, D. J., & Hemenway Deaver, A. H. (2022). An observational approach for exploring variability in young children's regulation-related skills within classroom contexts. *Developmental Science*, 13250. https://doi.org/10.1111/desc.13250

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology, 108*(1), 130-145. https://doi/10.1037/edu0000047

Nelson, M. C., Cordray, D. S., Hulleman, C., Darrow, C., & Sommer, E. C. (2010). A procedure for assessing fidelity of implementation in experiments testing educational interventions. *Society for Research on Educational Effectiveness*.

Nguyen, T., Jenkins, J. M., & Whitaker, A. A. (2018). Are content-specific curricula differentially effective in Head Start or state prekindergarten classrooms? *AERA Open*, *4*(1). https://doi.org/10.1177/2332858418784283

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*(2), 184-202.

Pas, E. T., Ryoo, J. H., Musci, R. J., & Bradshaw, C. P. (2019). A state-wide quasi-experimental effectiveness study of the scale-up of school-wide Positive Behavioral Interventions and Supports. *Journal of School Psychology, 73*, 41–55. https://doi.org/10.1016/j.jsp.2019.03.001

Pianta, R., La Paro, K., & Hamre, B. (2008). Classroom Assessment Scoring System (CLASS) Manual, Pre-K. Baltimore: Paul H. Brookes Pub. Co.

Pianta, R. C., Lipscomb, D., & Ruzek, E. (2021). Coaching teachers to improve students' school readiness skills: Indirect effects of teacher-student interaction. *Child Development, 92*(6), 2509–2528. https://doi.org/10.1111/cdev.13600

Piasta, S. B., Logan, J. A. R., Pelatti, C. Y., Capps, J. L., & Petrill, S. A. (2015). Professional development for early childhood educators: Efforts to improve math and science learning opportunities in early childhood classrooms. *Journal of Educational Psychology, 107*(2), 407–422. https://doi.org/10.1037/a0037621

Sarama, J., Lange, A. A., Clements, D. H., & Wolfe, C. B. (2012). The impacts of an early mathematics curriculum on oral language and literacy. *Early Childhood Research Quarterly*, *27*(3), 489-502.

Schickedanz, J., & Dickinson, D. K. (2004). Opening the world of learning: A comprehensive early literacy program. Pearson Early Learning.

Schrank, F. A., McGrew, K. S., Ruef, M. L., Alvarado, C. G., Munoz-Sandoval, A. F., & Woodcock, R. W. (2005). *Overview and technical supplement*. Rolling Meadows, IL: Riverside Publishing.

Snow, C. E., & Matthews, T. J. (2016). Reading and language in early grades reading and language in the early grades. Future of Children, 26(1), 57–74.

Stipek, D., Franke, M., Clements, D., Farran, D., & Coburn, C. (2017). PK-3: What does it mean for instruction? *Social Policy Report*, *30*(2), 1–23. https://doi.org/10.1002/j.2379-3988.2017.tb00087.x

Vitiello, V. E., Nguyen, T., Ruzek, E., Pianta, R. C., & Whittaker, J. V. (2022). Differences between pre-k and kindergarten contexts and achievement across the kindergarten transition. *Journal of Applied Developmental Psychology*, *80*(2), 1-11. https://doi.org/10.1016/j.appdev.2022.101396

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wechsler, D. (1989). Wechsler Preschool and Primary Scale of Intelligence—Revised. San Antonio, TX: The Psychological Corporation.

Weiland, C., & Guerrero-Rosada, P. (2022). *Widely used measures of pre-k classroom quality: What we know, gaps in the field, and promising new directions*. MDRC. https://www.mdrc.org/sites/default/files/Widely_Used_Measures.pdf

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 trails. *AERA Open*, *4*(1). https://doi.org/10.1177/2332858418757735

Weiland, C., Unterman, R., & Shapiro, A. (2021). The kindergarten hotspot: Literacy skill convergence between boston prekindergarten enrollees and nonenrollees. *Child Development*, *92*(2), 600-608.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. https://doi.org/10.1111/cdev.12099

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock Johnson III Tests of Achievement. Rolling Meadows, IL: Riverside Publishing.

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

Table 1

Student Sample Characteristics Overall and by Alignment Profile

			Fidelity Alignment Profile										
	Full sample		Low alignment		Mixed: high-low-high		Mixed: gradual decline		High alignment		BPS PreK non-attenders		
Characteristic	Mean/%	SD	Mean/%	SD	Mean/%	SD	Mean/%	SD	Mean/%	SD	Mean/%	SD	
Student demographic characteristics													
Child age	5.47	0.30	5.50	0.29	5.48	0.31	5.51	0.26	5.47	0.29	5.44	0.31	*
Eligible free/red. lunch (%)	68.07	46.67	69.52	46.25	55.00	50.38	52.38	50.55	43.48	49.94	79.34	40.57	**
Female (%)	50.40	50.05	49.52	50.24	47.50	50.57	57.14	50.09	42.03	49.72	52.48	50.04	
Dual Language Learner (%)	52.52	49.99	56.19	49.85	57.50	50.06	64.29	48.50	43.48	49.94	50.62	50.10	
Student Race/Ethnicity (%)													
Asian	18.31	38.71	20.95	40.89	25.00	43.85	23.81	43.11	8.70	28.38	17.84	38.37	†
Black	28.17	45.03	21.90	41.56	10.00	30.38	16.67	37.72	20.29	40.51	38.17	48.68	***
Hispanic	30.38	46.04	31.43	46.65	32.50	47.43	33.33	47.71	23.19	42.51	31.12	46.39	†
Other race	3.62	18.70	3.81	19.23	2.50	15.81	2.38	15.43	14.49	35.46	0.83	9.09	
White	19.52	39.67	21.90	41.56	30.00	46.41	23.81	43.11	33.33	47.49	12.03	32.60	*
Attended public BPS preK program (%)	51.41	50.03	-	-	-	-	-	-	-	-	-	-	
Parent demographic characteristics													
Age of mother when child was first born	26.39	6.82	26.57	7.02	28.33	6.93	27.74	7.59	28.81	6.63	24.70	6.17	*
Total household size	4.28	1.42	4.51	1.67	4.05	0.88	4.34	1.04	4.16	0.99	4.23	1.56	
Parent works full-time (%)	85.91	34.83	87.76	32.95	85.00	36.16	90.48	29.71	89.86	30.41	82.72	37.90	
Parent married (%)	53.64	49.92	57.58	49.67	65.00	48.30	64.29	48.50	69.57	46.35	41.05	49.32	
Parent age	36.18	7.27	35.76	8.10	37.31	6.06	36.80	7.68	37.50	6.55	35.54	7.17	
Parent education[a] (%)													
High school diploma/GED or less	32.88	47.03	38.38	48.88	25.00	43.85	21.43	41.53	27.54	45.00	36.13	48.16	
Some college/technical or vocational cert.	29.93	45.85	26.26	44.23	30.00	46.41	28.57	45.72	14.49	35.46	37.70	48.59	**
4-year degree	15.65	36.37	16.16	37.00	20.00	40.51	23.81	43.11	14.49	35.46	13.09	33.82	
> 4-year degree	21.54	41.16	19.19	39.58	25.00	43.85	26.19	44.50	43.48	49.94	13.09	33.82	*
Sample Size	498		105		40		42		69		242		

NOTES: ANOVA tests were conducted to determine whether measures signifificantly differ across grades. *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Table 2

Teacher and Classroom Characteristics and Focus on Early Learning Fidelity of Implementation

	Full sample		PreK		Kindergarten		First grade		
Characteristic	Mean/%	SD	Mean	SD	Mean	SD	Mean	SD	
Teacher race									
Black (%)	16.56	-	26.09	-	13.16	-	12.20	-	
Hispanic (%)	19.02	-	13.04	-	23.68	-	17.07	-	
White (%)	53.37	-	43.48	-	59.21	-	53.66	-	
Asian (%)	8.59	-	8.70	-	3.95	-	17.07	-	†
Other race (%)	2.45	-	8.70	-	0.00	-	0.00	-	**
Teacher experience									
Years teaching	13.57	8.76	14.83	8.86	12.73	8.54	13.73	9.09	
Years teaching at current school	8.01	6.73	8.41	7.26	7.01	5.98	9.43	7.29	*
Classroom size	16.60	3.33	15.21	3.56	17.64	2.90	17.10	2.96	***
Classroom demographic composition									
% Black	30.96	0.24	33.13	0.27	31.23	0.23	28.13	0.23	
% Hispanic	34.15	0.23	32.04	0.21	35.16	0.24	34.73	0.23	
% White	17.87	0.19	17.37	0.19	17.75	0.18	18.60	0.20	
% Asian	13.66	0.23	13.28	0.20	12.92	0.22	15.32	0.27	
% Other race	3.36	0.05	4.18	0.05	2.93	0.04	3.21	0.05	
% eligible for free or reduced price lunch (K and first only)	68.96	0.23	69.44	0.26	67.85	0.20	70.34	0.25	
% Dual Language Learner	49.12	0.26	47.60	0.25	50.04	0.26	49.20	0.28	
% attended BPS public school PreK program	58.76	0.23	-	-	55.42	0.23	64.46	0.22	*
Focus on Early Learning Implementation Fidelity									
Dosage (total minutes of curriculum implemented)	117.90	45.65	83.72	27.53	97.51	32.39	125.78	50.54	***
Percent of *Focus on Early Learning* components observed (%)	54.56	0.15	55.07	0.12	50.49	0.16	61.54	0.14	***
Overall adherence to Focus on Early Learning (%)	72.04	0.14	71.62	0.13	72.54	0.15	71.65	0.14	
Overall quality of Focus on Early Learning implementation (1 - 5)[a]	3.17	0.59	3.12	0.53	3.10	0.62	3.36	0.58	†
Quality of extending/building on knowledge (1 - 5)	2.93	0.88	2.77	0.84	3.06	0.92	2.93	0.86	
Quality of making connections/summarizing information (1 - 5)	3.01	0.78	2.93	0.68	3.13	0.86	2.92	0.79	
Quality of scaffolding and differentiation (1 - 5)	2.79	0.95	2.35	0.97	3.00	0.99	2.86	0.72	**
Quality of exposure to rich vocabulary (1 - 5)	3.14	0.83	3.17	0.82	2.89	0.75	3.57	0.84	***
Content-rich instruction (1 - 5)	2.96	0.84	3.10	0.79	2.93	0.82	2.85	0.93	
Cognitive demand (1 - 5)	2.94	0.78	2.89	0.75	3.01	0.84	2.90	0.71	

NOTES: ANOVA tests were conducted to determine whether measures significantly differed across grades. *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

[a] This shows the overall item average as opposed to the component item average. Correlations between the two versions of the measure for each respective grade range from .91 to .97.

Table 3

Associations Between Instructional Alignment Profiles and Gains in Academic Skills Through the Spring of Kindergarten and First Grade

	Language Skills		Literacy skills		Math skills: WJAP		Math skills: REMA	
PreK attender fidelity profile	γ	SE	γ	SE	γ	SE	γ	SE
Gains through spring of kindergarten								
High alignment	1.82	2.38	-4.57	2.56 †	0.42	0.48	0.02	0.49
Mixed: High-low-high	1.87	2.67	-6.29	2.59 *	0.51	0.55	0.62	0.54
Mixed: Gradual decline	2.04	2.66	-3.23	2.94	0.38	0.54	-1.08	0.54 *
Low alignment	0.53	1.95	-4.91	2.01 *	0.31	0.40	0.06	0.39
Gains through spring of first grade								
High alignment	-5.06	2.12 *	3.91	1.76 *	0.21	0.51	1.04	0.52 *
Mixed: High-low-high	-3.94	2.33 †	3.49	1.71 *	0.74	0.56	0.21	0.56
Mixed: Gradual decline	-2.00	2.36	0.31	1.92	-0.15	0.57	-0.73	0.58
Low alignment	-2.84	1.84	1.10	1.41	0.33	0.44	0.57	0.45

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

Notes: Models control for a robust set of student- and family-level covariates, including the level of the outcome measured in the fall of kindergarten, child age in the fall of kindergarten, child eligiblity for free or reduced price lunch, child female, child dual language learner, child race/ethnicity, mother's age at time of first birth, household size, parental employment, parental marital status, parent age, and parent education.

Literacy skills were assessed using the DIBELS Letter Naming subtest in the fall and spring of kindergarten and the Woodcock Johnson Letter Word ID in the spring of first grade. Language skills were assessed using the PPVT. Math skills were assessed using both the Woodcock Johnson Applied Problems and the Research-based Early Math Assessment.

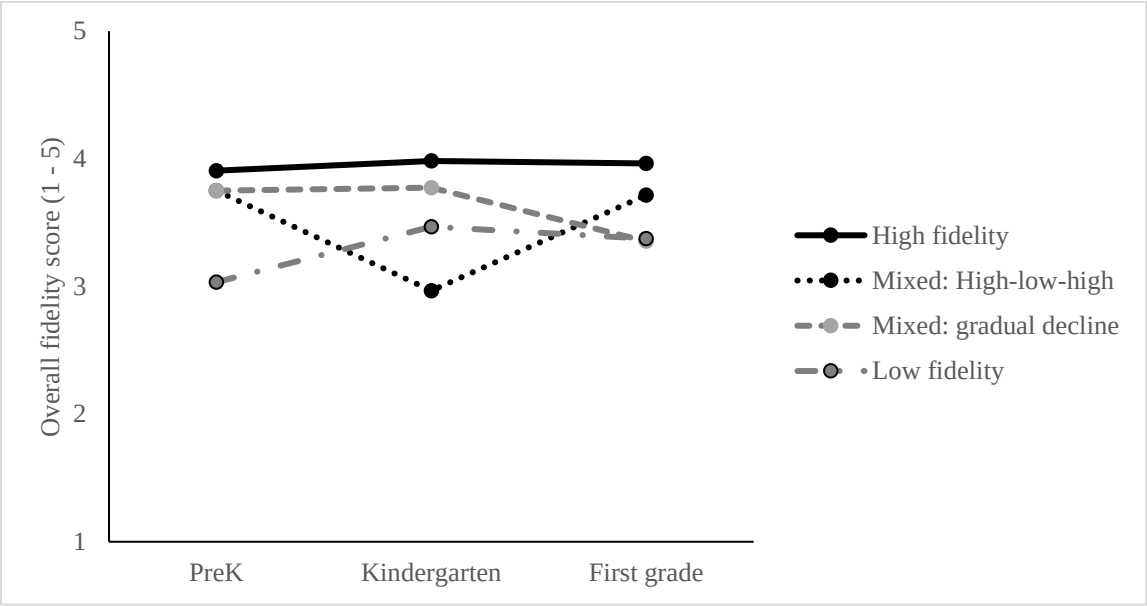


Figure 1

Results from Latent Profile Analysis of Fidelity Scores from PreK to First Grade to Measure Instructional Alignment

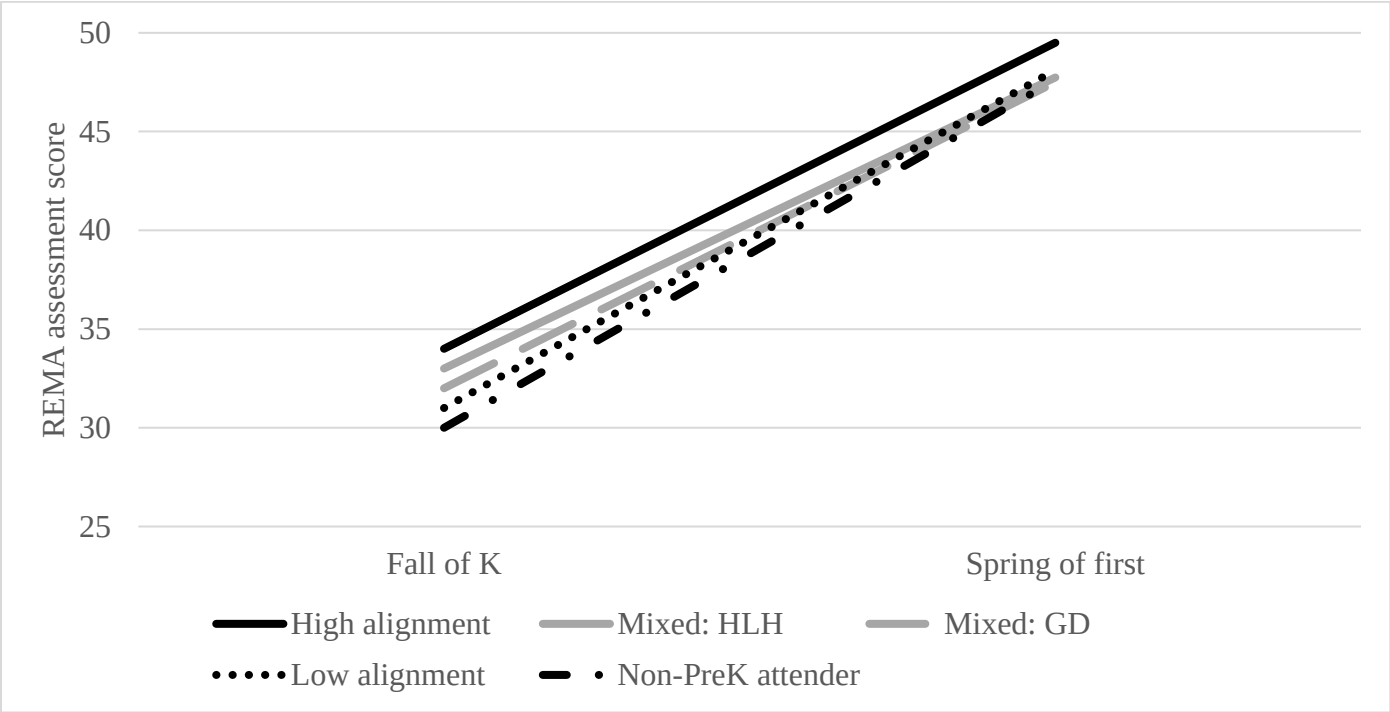


Figure 2

Variation in Gains in REMA Math Scores Between Fall of Kindergarten and Spring of First Grade by Instructional Alignment Profile

Appendix A
Summary of Robustness Checks

Although the current study uses robust observational data to answer research questions and aims to include a number of covariates in predictive analyses, there are threats to both the construct validity of the alignment profiles and the descriptive analysis linking alignment to child outcomes. Below we summarize these issues and report on additional analytic work we did to examine the sensitivity of our substantive results to each of these threats. Across the various checks, we found evidence that findings reported in the main text of the paper were consistent. The substantive results reported in the paper remained consistent across these checks.

Creation of alignment profiles. We used Latent Profile Analysis to empirically derive profiles representing students' exposure to the *Focus on Early Learning* model from PreK to first grade. Although we found that this approach was generally successful and met basic rules of thumb for model fit, probability of profile membership was not perfect. In addition, use of profile analysis may not be as intuitive for future work as a more straightforward approach based on key data "cut-points" representing higher and lower intervention fidelity. Given both of these issues, we recreated the profiles using a manual approach. We discussed this approach in the main text of the paper. For each year we defined higher intervention fidelity as having a scoring above the sample median and lower fidelity as falling below the median. We then defined students as falling in the high alignment group if they experienced fidelity that was above the median in all three years. Students were mixed: high-mixed if they fell above the median for PreK and then were below the median for at least one group in kindergarten or first grade. They were in mixed: low-mixed if they were below the median in PreK and then above the median in either kindergarten or first grade. Finally, students were in the low alignment group if they were in a classroom that scored below the median in each grade.

Although these profiles were correlated with the LPA results, profile membership was not redundant. We were interested to examine whether substantive results for our models – showing benefits of high instructional alignment for sustained literacy and REMA gains through first grade – were consistent when using this approach. Findings are summarized in Appendix A Table A.1. As illustrated, results were generally consistent with the LPA approach with respect to statistical significant. Magnitudes were actually a bit larger using this approach. For example, we found that the association between the "high alignment" group and gains in literacy skills through first grade was 4.87 (SE = 1.95, $p < .05$) and gains in math skills through first grade was 1.32 (SE = .59, $p < .05$). Although neither instructional alignment profile approach is without measurement error, the consistency in the pattern of results improves our confidence on there being some benefits of instructional alignment for some domains of learning in this sample and setting.

CLASS scores as covariates in predictive analyses. Due to the descriptive nature of this study we cannot establish the causal impact of alignment on outcomes. Alignment could be confounded with some other characteristic – like overall quality – and that could really be explaining our findings. For example, it could just be that children experience higher overall quality across time and that explains sustained associations between PreK and child outcomes rather than exposure to the aligned curriculum. To address this concern, we refit all of our predictive models with controls for the three domains of the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) – emotional support, classroom organization, and instructional support. In the kindergarten models we included the three domains measured in kindergarten. And in the first grade models we included the average CLASS domain scores for each child across kindergarten and first grade. As illustrated in Table A.2, we found that the

substantive study results were consistent with respect to both magnitude and statistical significant even after including these controls for overall quality. Although there are other features of general quality – like instructional content – that are not captured by the CLASS, this finding does strengthen our interpretation of instructional alignment predicting gains in child skills over and above the quality of general teacher practices.

Treatment of missing data. In line with prior work for this broader study (e.g., Authors, 2022, Authors, 2021, Authors, 2020), analyses of missing data revealed evidence that data – namely covariate data – in this study were Missing at Random (MAR). As summarized in the main text of the paper we decided to use complete case analysis to deal with missing data. We then examined how sensitive our substantive results were to the treatment of missing data and re-fit all of those models using multiple imputation (N = 100 datasets). As summarized in Table A.3 we found that the two main associations related to high alignment did drop in magnitude but remained statistically significant. In contrast, the bulk of the results showing negative associations between profiles and language skills became much smaller and were not statistically significant.

School fixed effects. To map onto prior work in this study, we fit predictive models with random intercepts to answer our main research questions. However, this approach does not allow us to make comparisons *within* schools and to adjust for unobserved school-level confounders, such as the demographic makeup of the school or the overall program quality. As such, we explored whether we could fit our models using school fixed effects rather than random intercepts for schools. However, when we explored our data we found that because of classroom groupings, we did not have sufficient variation in the profiles within the schools to fit the models

with school fixed effects. As such, we rely on random effects models to address the non-independence of the observations and model the multi-level structure of the data.

Controlling for variation in children's skills at the classroom-level. Lastly, there is a concern about how the skills of the children in the classroom as a whole may affect the extent to which teachers were able to implement *Focus on Early Learning* with fidelity and whether that then translated into better outcomes for students. To address that concern, we fit all of our predictive models and included controls for the standard deviation of children's skills at the classroom-level. Results are included in Table A.4 (LPA) and A.5 (median split approach). As noted there, substantive results were largely unchanged demonstrating that findings were robust to inclusion of these controls.

Table A.1

Associations Between Instructional Alignment Profiles and Gains in Academic Skills Through the Spring of Kindergarten and First Grade Median-Split Approach for Creating Profiles

	Language skills		Literacy skills		Math skills: WJAP		Math skills: REMA	
PreK attender fidelity profile	γ	SE	γ	SE	γ	SE	γ	SE
Gains through spring of kindergarten (compared to non-BPS PreK attenders)								
High alignment	3.58	2.44	-4.37	2.80	0.47	0.51	0.24	0.49
Mixed alignment: high-mixed	3.36	2.25	-9.12	2.36 ***	-0.01	0.46	-0.91	0.46 †
Mixed alignment: low-mixed	0.58	2.11	-2.44	2.18	0.49	0.43	0.02	0.43
Low alignment	-2.00	2.48	-4.75	2.39 *	0.61	0.51	0.55	0.50
Gains through spring of first grade (compared to non-BPS PreK attenders)								
High alignment	-4.68	2.35 *	4.87	1.95 *	0.72	0.56	1.32	0.59 *
Mixed alignment: high-mixed	-3.36	1.97 †	-0.06	1.59	0.07	0.46	0.14	0.48
Mixed alignment: low-mixed	-2.19	2.00	2.76	1.51 †	0.11	0.47	-0.12	0.49
Low alignment	-4.20	2.39 †	0.63	1.68	0.56	0.56	0.79	0.57

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

Notes: Models control for a robust set of student- and family-level covariates, including the level of the outcome measured in the fall of kindergarten, child age in the fall of kindergarten, child eligiblity for free or reduced price lunch, child female, child dual language learner, child race/ethnicity, mother's age at time of first birth, household size, parental employment, parental marital status, parent age, and parent education.

Literacy skills were assessed using the DIBELS Letter Naming subtest in the fall and spring of kindergarten and the Woodcock Johnson Letter Word ID in the spring of first grade. Language skills were assessed using the PPVT. Math skills were assessed using both the Woodcock Johnson Applied Problems and the Research-based Early Math Assessment.

Table A.2

Summary of Associations Between Instructional Alignment Profiles and Gains in Academic Skills Between Fall of Kindergarten and Spring of First Grade Using Latent Profile Analysis Approach to Define Instructional Alignment with Complete Case Analysis Controlling for Classroom Covariates

	Language skills		Literacy skills		Math skills: WJAP		Math skills: REMA	
PreK attender fidelity profile	γ	SE	γ	SE	γ	SE	γ	SE
Gains through spring of kindergarten								
High alignment	0.44	2.53	-4.68	2.78 †	0.26	0.48	-0.08	0.48
Mixed: High-low-high	1.51	2.86	-5.79	2.85 *	0.50	0.54	0.62	0.54
Mixed: Gradual decline	-0.22	2.95	-3.41	3.33	0.23	0.57	-1.43	0.57 *
Low alignment	-0.85	2.11	-4.59	2.22 *	0.11	0.41	-0.09	0.40
Gains through spring of first grade								
High alignment	-4.95	2.14 *	3.45	1.63 *	0.26	0.52	1.05	0.52 *
Mixed: High-low-high	-3.82	2.37	3.09	1.76 †	0.82	0.57	0.17	0.57
Mixed: Gradual decline	-1.59	2.42	-0.16	2.00	-0.05	0.58	-0.85	0.59
Low alignment	-2.51	1.88	0.76	1.44	0.44	0.45	0.46	0.46

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

Notes: Models control for a robust set of student- and family-level covariates, including the level of the outcome measured in the fall of kindergarten, child age in the fall of kindergarten, child eligiblity for free or reduced price lunch, child female, child dual language learner, child race/ethnicity, mother's age at time of first birth, household size, parental employment, parental marital status, parent age, parent education, and Classroom Assessment Scoring System covariates.

Literacy skills were assessed using the DIBELS Letter Naming subtest in the fall and spring of kindergarten and the Woodcock Johnson Letter Word ID in the spring of first grade. Language skills were assessed using the PPVT. Math skills were assessed using both the Woodcock Johnson Applied Problems and the Research-based Early Math Assessment.

Table A.3

Summary of Associations Between Instructional Alignment Profiles and Gains in Academic Skills Between Fall of Kindergarten and Spring of First Grade Using Latent Profile Analysis Approach to Define Instructional Alignment with Multiple Imputation of Covariates and Outcomes

	Language skills		Literacy skills		Math skills: WJAP		Math skills: REMA	
PreK attender fidelity profile	γ	SE	γ	SE	γ	SE	γ	SE
Gains through spring of kindergarten								
High alignment	2.48	3.67	-4.79	3.05	0.45	0.68	0.32	0.74
Mixed: High-low-high	2.61	5.45	-5.17	3.19	0.66	1.02	1.02	1.03
Mixed: Gradual decline	2.25	4.45	-5.62	3.25 †	0.17	0.83	-0.76	0.94
Low alignment	0.33	3.18	-3.26	2.47	0.18	0.60	0.46	0.64
Gains through spring of first grade								
High alignment	-2.19	3.36	3.46	1.65 *	0.38	0.65	1.43	0.68 *
Mixed: High-low-high	0.88	5.12	2.24	1.87	0.84	0.79	0.46	0.89
Mixed: Gradual decline	-0.83	3.41	-0.75	1.90	-0.54	0.74	-1.14	0.75
Low alignment	-1.32	2.41	0.15	1.36	0.14	0.50	0.24	0.60

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

Notes: Models control for a robust set of student- and family-level covariates, including the level of the outcome measured in the fall of kindergarten, child age in the fall of kindergarten, child eligiblity for free or reduced price lunch, child female, child dual language learner, child race/ethnicity, mother's age at time of first birth, household size, parental employment, parental marital status, parent age, and parent education.

Literacy skills were assessed using the DIBELS Letter Naming subtest in the fall and spring of kindergarten and the Woodcock Johnson Letter Word ID in the spring of first grade. Language skills were assessed using the PPVT. Math skills were assessed using both the Woodcock Johnson Applied Problems and the Research-based Early Math Assessment.

Table A.4

Associations Between Instructional Alignment Profiles and Gains in Academic Skills Through the Spring of Kindergarten and First Grade with Complete Case Analysis Controlling for Classroom Skill Heterogeneity

	Language Skills		Literacy skills		Math skills: WJAP		Math skills: REMA	
PreK attender fidelity profile	γ	SE	γ	SE	γ	SE	γ	SE
Gains through spring of kindergarten								
High alignment	1.27	2.43	-4.18	2.69	0.44	0.49	0.13	0.48
Mixed: High-low-high	3.25	2.76	-5.77	2.75 *	0.62	0.56	0.66	0.53
Mixed: Gradual decline	1.10	2.73	-3.40	3.08	0.27	0.56	-0.87	0.54
Low alignment	0.31	2.00	-4.79	2.13 *	0.35	0.41	0.18	0.39
Gains through spring of first grade								
High alignment	-4.96	2.13 *	3.45	1.69 *	0.24	0.51	1.06	0.52 *
Mixed: High-low-high	-3.39	2.36	3.38	1.72 †	0.75	0.56	0.17	0.57
Mixed: Gradual decline	-2.09	2.40	-0.96	1.99	-0.15	0.58	-0.91	0.59
Low alignment	-2.44	1.86	0.69	1.43	0.33	0.45	0.54	0.45

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

Notes: Models control for a robust set of student- and family-level covariates, including the level of the outcome measured in the fall of kindergarten, child age in the fall of kindergarten, child eligiblity for free or reduced price lunch, child female, child dual language learner, child race/ethnicity, mother's age at time of first birth, household size, parental employment, parental marital status, parent age, parent education, and classroom skill heterogeneity covariates.

Literacy skills were assessed using the DIBELS Letter Naming subtest in the fall and spring of kindergarten and the Woodcock Johnson Letter Word ID in the spring of first grade. Language skills were assessed using the PPVT. Math skills were assessed using both the Woodcock Johnson Applied Problems and the Research-based Early Math Assessment.

Table A.5

Associations Between Instructional Alignment Profiles and Gains in Academic Skills Through the Spring of Kindergarten and First Grade with Complete Case Analysis Controlling for Classroom Skill Heterogeneity

	Language skills		Literacy skills		Math skills: WJAP		Math skills: REMA	
PreK attender fidelity profile	γ	SE	γ	SE	γ	SE	γ	SE
Gains through spring of kindergarten (compared to non-BPS PreK attenders)								
High alignment	2.66	2.50	-4.49	2.93	0.56	0.52	0.43	0.49
Mixed alignment: high-mixed	4.15	2.30 †	-8.77	2.49 ***	-0.05	0.47	-0.81	0.46 †
Mixed alignment: low-mixed	0.24	2.16	-2.35	2.31	0.54	0.44	0.15	0.43
Low alignment	-1.91	2.53	-4.40	2.48 †	0.56	0.52	0.55	0.50
Gains through spring of first grade (compared to non-BPS PreK attenders)								
High alignment	-4.64	2.41 †	4.69	2.01 *	0.74	0.58	1.38	0.61 *
Mixed alignment: high-mixed	-2.62	2.01	-0.76	1.60	0.10	0.46	0.11	0.48
Mixed alignment: low-mixed	-2.50	2.03	2.30	1.53	0.12	0.47	-0.15	0.50
Low alignment	-3.57	2.42	0.31	1.68	0.57	0.57	0.74	0.58

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

Notes: Models control for a robust set of student- and family-level covariates, including the level of the outcome measured in the fall of kindergarten, child age in the fall of kindergarten, child eligiblity for free or reduced price lunch, child female, child dual language learner, child race/ethnicity, mother's age at time of first birth, household size, parental employment, parental marital status, parent age, parent education, and classroom skill heterogeneity covariates.

Literacy skills were assessed using the DIBELS Letter Naming subtest in the fall and spring of kindergarten and the Woodcock Johnson Letter Word ID in the spring of first grade. Language skills were assessed using the PPVT. Math skills were assessed using both the Woodcock Johnson Applied Problems and the Research-based Early Math Assessment.

Supplemental Online Appendix B

Detailed Procedure for Collection of Observational Data

Live classroom observations. In the winter of PreK (2017), kindergarten (2018), and kindergarten (2019), trained instructional coaches from the BPS DEC observed each participating classroom for two 2-hour blocks of academic instruction (Mean observation time = 117.90 minutes, *SD* = 61.36). Importantly, coaches only completed fidelity observations in classrooms in which they were not also providing coaching. Observations took place between March and May of the academic year. The focus of the observation was on the full classroom including the lead teacher, the children, and any other assistant teachers and adults in the classroom. Prior to conducting observations, we reviewed all teachers' weekly schedules and identified one 2 – 3 hour block of instructional time that included substantial focus on language/literacy and another 2 – 3 hour block with some focus on mathematics instruction. We worked with teachers to schedule observations during these two blocks of time when we would ideally observe a broad spectrum of curricular components. For example, if we were able to observe literacy activities (e.g., Read Aloud and/or literacy small group sessions) during the first observation, we aimed to capture mathematics activities (e.g., mathematics small groups and whole group) for at least a portion of the second observation. In scheduling these observations, we tried to minimize the amount of time that children spent in meals, naps, and specials. On average, there were about 14 days between classrooms' live observation occasions. If we observed the same curricular component twice across the two measurement occasions, we averaged across the observations to create one score. If we only observed the component once, however, we used the scores from that observation on its own.

BPS instructional coaches trained as observers participated in a three-day training in

January of each academic year to learn how to rate classrooms and teachers on different indicators of intervention fidelity for the each of the components of the curriculum. The training was conducted by a PhD-level member of the research team who also was the lead author on the observational tool. This trainer served as the master-coder for future reliability coding checks. As part of the training, observers also learned how to rate classrooms on a new measure created to capture global indicators of instructional quality, regardless of the curricular components observed. This measure is intended to capture content-rich instruction, cognitive demand, and the extent to which the vocabulary used in the classroom was rich and complex. Observers needed to establish reliability with master-coded videos developed by the observation trainer prior to collecting data in the field. Using a similar approach as the Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2008), the team defined reliability on the observed global indicators of instructional quality (which were all rated on a Likert scale, discussed in more detail below) to be 80% agreement "within 1" across observed items. Throughout each data collection period (March – May of each year), the observers also double-coded twenty percent of observations so that the research team could assess interrater reliability. These observations suggested high levels of agreement, with 89% of fidelity quality double codes – including global items assessing the quality of the full observation -- being reliable "within 1."

Videotaped classroom observations. On different days than the live classroom observations, we also collected two videotaped observations of classroom instruction (Mean = 3.16 hours of total time across the observations, *SD* = .83, min = 2.21 hours, max = 4.62 hours) during the winter of 2017 (discussed in more detail in Authors, under review). The team collected videotapes between late January and late March of each academic year with 89% of tapes collected by the end of March. There were an average of 13.78 days (*SD* = 13.65) between

video observations. We coded these observations using both the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) and an adapted version of the Individualizing Student Instruction (ISI; Connor et al., 2009) tools. We used the CLASS scores as covariates in robustness checks in the current study. Other papers (see Authors, in press; Authors et al., 2022) leveraged data from the ISI measure.

Similar to the live observations, we worked with teachers to identify blocks of observation time when we could observe both language/literacy and mathematics instruction. We used two video cameras during each observation session. One camera focused primarily on the lead teacher (and the teacher's microphone), and the other camera focused on the children in the classroom enrolled in the study to capture as much of the classroom activity as possible. Both videotapes allowed for viewing of children's behavior, faces, and responses to teachers. Research assistants coded teacher-focused videotapes using the CLASS, which is comprised of Emotional Support, Classroom Organization, and Instructional Support domains. All coders participated in a two-day training led by a certified trainer employed by the developer to learn the CLASS measure and then established reliability on a set of master codes created by the developers. Coding of each videotape started once the instructional time began. As recommended by the measure's protocol (Pianta et al., 2008), coders used cycles of 20 minutes for observing and 10 minutes for scoring, which they repeated up to 4 times. We averaged scores across the four segments to calculate observation-specific scores and then we averaged scores across observations to generate one overall score for each classroom. We double-coded 20% of the observations to assess interrater reliability. Results revealed acceptable levels of agreement with 81% of codes across CLASS dimensions from double-coded tapes being reliable "within 1." We also conducted drift checks wherein coders had to code a master tape from the

observation developer every three weeks to ensure they were still reliable.

Supplemental Online Appendix C
Summary of Focus on Early Learning Curricular Components in
PreK, Kindergarten, & First Grade

Table A.6
Overview of Focus on Early Learning Curricular Components in PreK, Kindergarten, and First Grade

Focus on Early Learning curricular component	PreK	Kindergarten	First grade
Introduction to centers	✓	✓	
Centers (PreK & K)/studios (1st grade)	✓	✓	✓
Language & literacy small groups	✓	✓	✓
Literacy whole group & phonics	✓	✓	✓
Vocabulary language			✓
Thinking & feedback (present learning to class)	✓	✓	✓
Let's find out about it	✓		
Read aloud	✓	✓	✓
Songs, word play, & letters	✓	✓	
Math whole group	✓	✓	✓
Math small groups	✓	✓	✓
Math centers	✓		
Math Summary		✓	✓
Number Talks		✓	✓
Student story-telling	✓	✓	
Student story-acting	✓	✓	
Writing		✓	✓

In-depth descriptions of each curricular component are available at:
https://www.bpsearlylearning.org/our-curricula

Supplemental Online Appendix D

Items on Focus on Early Learning Intervention Fidelity Measure by Domain

Summary of Items in Cross-Component Constructs for Focus on Early Learning Intervention Fidelity Tool

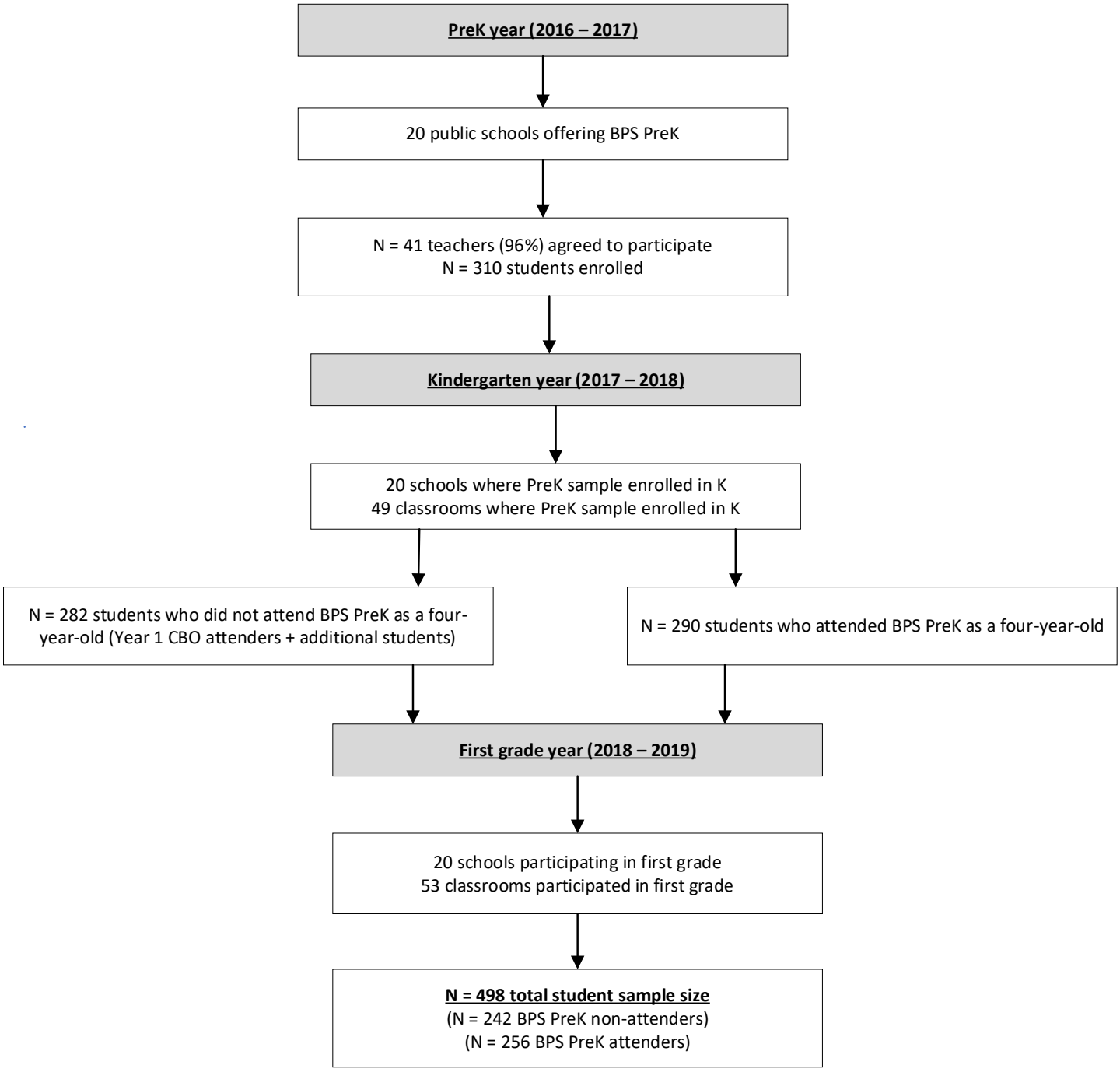
Construct	Coded for all observed components	Global fidelity item (coded once for full observation)
Extending/building on knowledge		
Teacher asks questions that guide children to be more deeply and actively engaged (focused on) in the activity (e.g., Why and How questions).	✓	
The teacher asked children to share, clarify, or justify their ideas.	✓	
Teacher talks to children in ways that encourage them to expand on or think more deeply about ideas.	✓	
Teacher exploits opportunities to describe or comment about children's actions.	✓	
Teacher exploits opportunities to provide process information (e.g., "why", "how" info) to children.	✓	
Teacher builds on and extends children's thinking and understanding in flexible ways.	✓	
Teacher repeats, builds on and extends children's thinking and understanding.	✓	
The teacher supported the "describer's" thinking.	✓	
The teacher extends children's conceptual understanding.	✓	
To what degree does this classroom capitalize on learning opportunities for children?		✓
To what degree are the learning opportunities in this classroom cognitively demanding (i.e., requiring strategic and extended thinking?		✓
The teacher encouraged children to listen to and evaluate others' thinking and discuss ideas.		✓
Teachers and children have sustained, substantive discussions around unit content.		✓
Summarizing/making connections across thematic units		
Teacher incorporates or references prior classwork to inform or reinforce the current day's planned activities.	✓	
Teacher connects or links the activities to the unit or book at least once.	✓	
Teacher connects or links activities to the curriculum unit in explicit and intentional ways.	✓	
To what degree did the teacher make connections between activities to deepen children's understanding of the theme/focal question?		✓
Discussion is focused on target component activities.	✓	

Presentations and feedback are aligned with and reflect a coherent discussion.	✓	
Discussion of the book (at the end of the read) is linked to the story and relatively extended.		✓
Teacher verbally summarizes/reflects on the lesson before transitioning to the next activity.		✓
Teacher summarizes feedback and suggestions at the end of the Thinking & Feedback component		✓
The teacher encouraged mathematical reflection in each math component	✓	
<u>Scaffolding and differentiation</u>		
Teacher's behaviors demonstrate scaffolding of children to help them extend their learning, providing "just enough" support to complete task(s) (e.g., appropriate level of detail, not too little or too much help or information).	✓	
Through instructional strategies, the teacher scaffolds children to help them extend their learning, providing just enough support for children to complete the task(s).	✓	
Teacher adapted the task or discussion according to children's abilities and development by purposefully presenting the content in different ways, varying materials, or providing children with flexibility in how they complete the activity(ies). For example, some children are asked to write a story while others are asked to draw a picture of their story.	✓	
To what degree does this teacher use differentiated learning strategies to make the curriculum accessible to and address the range of children in the classroom, including ELL students and those with special needs?		✓
<u>Rich vocabulary</u>		
Teacher uses center-specific vocabulary for the highlighted centers.	✓	
Teacher uses vocabulary words as related to the unit book(s) and/or small group activity.	✓	
Teacher embeds vocabulary in language	✓	
Teacher defines vocabulary words	✓	
Teacher uses a variety of vocabulary words that are sophisticated or advanced (e.g., Tier 2 or 3).	✓	
Teacher is intentional in which vocabulary words are used and how they are defined.	✓	
How often are relevant vocabulary (related to the Unit or children's work on activities) and rich academic language used and clearly defined throughout the observation (e.g., repetition in conversations, explore meaning through multiple modes)?		✓

Note: Full measure is available upon request.

Supplemental Online Appendix E

Overview of Sample Recruitment from PreK to First Grade



Supplemental Online Appendix F
Equation for Predictive Models

We then regressed each of the outcomes – measured separately at the end of kindergarten and the end of first grade – on the alignment profile dummies. Non-PreK attenders were the reference group. The base equation for these models is as follows:

$$Y_{ij} = \beta_0 + \beta_1 HighAligned_{ij} + \beta_2 Mixed{:}\,HLH_{ij} + \beta_3 Mixed{:}\,GD_{ij} + \beta_4 LowAlign_{ij} + \beta_5 Baseline_{ij} + \gamma_{ij} + \zeta_k + \varepsilon_{ij}, \quad (1)$$

where *i* denotes students and *j* represents elementary schools. *Baseline* is the level of the outcome measured at the start of kindergarten, γ is a vector of child- and family-level covariates (described above), ζ is a school-level random intercept, and ε is the child-level error term. The coefficients on the four alignment indicators ($\beta_1 - \beta_4$) and their associated *p*-values are the parameters of interest. These coefficients represent the difference in gains made in the outcome between that group and the students who did not attend the public PreK program between the start of kindergarten and the outcome period (spring of kindergarten or first grade). We calculated standardized associations for statistically significant associations by dividing the parameter estimates by the standard deviation of the outcome.