Fast-Response Research to Answer Practice and Policy Questions

Christina Weiland, Jason Sachs, Meghan McCormick, JoAnn Hsueh, and Catherine Snow

Summary

Research-practice partnerships often face a fundamental tension: well-designed, high-quality research takes time, but practitioners and policy makers need answers to pressing questions as soon as possible.

In this article, Jason Sachs, Meghan McCormick, JoAnn Hsueh, and Catherine Snow discuss this mismatch between the tight timelines of educational decision makers and the typically longer timelines of researchers who are pursuing rigorous analyses. They tell us how, in a partnership between researchers and the Boston Public Schools Department of Early Childhood, they've worked to make fast-turnaround research as rigorous as they can, while also conducting longer-term causal studies.

Because policy makers and practitioners typically aren't highly trained in study design and causal inference, a key responsibility for researchers is communicating the strengths and limitations of fast-turnaround work in ways that can be easily understood by their partners— and making it clear that fast-response analyses should be viewed as only one piece of evidence for guiding a decision.

www.futureofchildren.org

Christina Weiland, Jason Sachs, Meghan McCormick, JoAnn Hsueh, and Catherine Snow

Well-executed research-practice partnerships can have many benefits. For the practitioner, partnering with researchers can provide data and analysis to guide important decisions within often-short windows of policy attention; create a narrative of challenges and successes that can bring new district leaders up to speed quickly; and bring in scarce resources and expertise for research. For the researcher, partnering with practitioners can lead to more relevant research questions; access to and buy-in from key groups, such as principals, teachers, and parents; opportunities to assess classroom practice and child development in a way that balances theory and implications for real-world practice; and the ability to conduct studies with generalizable samples that would not be available otherwise. However, as this issue of *Future of Children* illustrates, such partnerships share common tensions, such as navigating the different needs and cultural worlds of researchers and practitioners or balancing the slow pace of rigorous (that is, well-designed and high-quality) research against practitioners' desire for timely information.[1]

In 14 years and counting, our partnership has enjoyed many benefits while managing most of the inevitable tensions. The partnership comprises a continuous relationship between Jason Sachs and the Boston Public Schools Department of Early Childhood, on the one hand, and Christina Weiland (formerly at the Harvard Graduate School of Education and now at the University of Michigan), on the other; project-specific relationships with the Wellesley Centers for Women and research and consulting firm Abt Associates; and, in its latest form, an expansion to include nonprofit research organization MDRC and the Harvard Graduate School of Education. In this article, we describe the history of our partnership and then focus on our most difficult challenge—balancing rigor against timeliness in a large public school district.

Balancing rigor and timeliness has been our central challenge in large part because of the complexity of large educational systems. The Department of Early Childhood manages programs for thousands of prekindergarteners through second-graders in Boston and oversees professional development and other support for hundreds of teachers each year. Its oversight responsibilities include deciding where new prekindergarten classrooms will open, what curricula to use, what professional development model to use, and how teachers will assess student learning. Each program element has many possible directions and could be the subject of a careful, years-long academic study.

But the district lacks the luxury of time. Most of these decisions have to be made within a few months and then quickly acted on and communicated. Thus the district often has to make decisions based on imperfect but quickly accessible evidence or no evidence at all. At the same time, rigor still matters. As education research has shown time and again, correlation is not causation.[2] Moreover, districts need measurement approaches that generate reliable and valid data, and research samples that can be generalized to a study's population of interest. With less rigorous studies, we risk getting the answer wrong because of the limitations posed by research design, measurement approaches, and sampling constraints. Further, Boston—and early education researchers and practitioners more broadly—have benefitted from lengthy,

careful studies that have guided policy and practice. We illustrate here how we balance competing demands for rigor and timeliness by trying to make our fast-turnaround work as rigorous as we can; by communicating the strengths and limitations of that work clearly and loudly to policy makers and practitioners; and by mixing fast-turnaround work with longer-term causal studies so that we get the right answers to our most pressing questions.

Background

The BPS Prekindergarten Program

The Boston Public Schools (BPS) created the Department of Early Childhood (DEC) in 2005 to oversee the City of Boston's public prekindergarten programs for three- and four-year-olds. Then-mayor Thomas Menino and then-BPS superintendent Tom Payzant pushed for a city-funded prekindergarten program because they believed it could better prepare children for school and could help attract families to the Boston Public Schools who might otherwise choose other options. Jason Sachs (a co-author of this article) has led the DEC since its inception. Sachs came from a research background. He held a doctorate from the Harvard Graduate School of Education and, as the director of research at the Massachusetts Department of Early Care and Education, he had used data and research extensively.

From a national perspective, the program was unusual. It was based entirely in the public schools, paid teachers on the same scale as K–12 teachers, subjected teachers to the same educational requirements as K–12 teachers (for example, achieving a master's degree within five years of being hired), and was open to any child in the city, regardless of income. The program grew quickly in its early years, from 750 students in 2005 to

1,206 in 2006, 1,467 in 2007, and 1,900 in 2008.[3] In recent years, the DEC's work has grown to cover curriculum and instruction through second grade. The city has also continued to expand the program, which doesn't yet have enough seats for all children in the district who apply.

Early Use of Research

From its first years under Sachs's leadership, data and research were a key part of the DEC's strategic planning and decision making. Most notably, in the program's first year, the DEC hired the Wellesley Centers for Women, a research institute at Wellesley College, to measure quality in a random sample of its classrooms. The study was meant to be a needs assessment—that is, it was meant to provide baseline data to guide the DEC's efforts to increase quality. The *Boston Globe* displayed the study's findings prominently on its front page under the headline "Boston preschools falling far short of goals, study says: Teacher quality, site safety faulted," followed by "Boston's public preschool and kindergarten programs are hobbled by mediocre instruction, unsanitary classrooms, and dangerous schoolyards, according to a first-ever study of the programs."[4]

The very public nature of those findings spurred the district to sharpen its focus on improving quality. DEC leaders asked experts to review the evidence on preschool curricula broadly, determine which curricula were being used around the district, and find out how teachers and principals viewed these curricula.[5] Based on their findings, the DEC decided to use Opening the World of Learning in its prekindergarten classrooms. This curriculum targets children's early language and literacy skills; each unit also

embeds a social-skills component, in which teachers discuss social-emotional issues and introduce emotion-related vocabulary words.[6] The DEC also chose Building Blocks, an early mathematics curriculum that covers both numeracy and geometry, with a heavy focus on verbal mathematical reasoning.[7] In other studies, both curricula have shown positive effects on children's outcomes, though the evidence for Building Blocks is stronger than that for Opening the World of Learning.[8] Teachers received training in both curricula, as well as bi-weekly to monthly coaching focused on helping them troubleshoot problems with classroom management, differentiating instruction for children with special needs and dual language learners, and implementing the curricula. Importantly, this professional development model matches the science of adult learning and the existing evidence on helping teachers improve.[9] Coaches prepared detailed guides that showed teachers how to implement the two curricula in tandem and also made sure classrooms were well stocked with the many materials and supplies necessary to carry out the curricula as intended by the developers.[10] To improve quality in both prekindergarten and kindergarten and to ensure classrooms had adequate start-up funding, the district also sought accreditation from the National Association for the Education of Young Children (NAEYC).

More than a decade since the program's founding, Boston's structural and programmatic choices remain exceptional among the nation's public programs, which tend not to require that teachers have master's degrees, not to pay prekindergarten teachers on the same scale as K–12 teachers, not to use a proven, consistent curriculum, and not to provide coaching.[11]

Partnership History

Our research-practice partnership (RPP) began in the summer of 2007, following the decisions to implement the new curricula and coaching model and pursue NAEYC accreditation. Our partnership was sparked by an internship. Christina Weiland, the first author of this article, had just completed one year of doctoral studies at the Harvard Graduate School of Education and was interested in quantitative methods and early childhood education, particularly in the universal prekindergarten movement. When she approached Sachs about spending the summer working for the DEC, he decided the price was right (Weiland had outside funding that made her work free). Sachs saw value in having someone map out available data and create a research plan. Weiland spent that first summer on two tasks: developing relationships with district staff and coaches, and learning which BPS departments held administrative data relevant to the prekindergarten program, the quirks of the available data, the program's components, and the DEC's goals and questions. Weiland was supported by her adviser, Professor Hirokazu Yoshikawa, who provided guidance and expertise as she learned the ropes (and who continues to contribute to research in Boston).

At the end of that first summer, Weiland and Yoshikawa prepared a memo listing all the data collected by the district that were relevant to the DEC.[12] They also highlighted study designs that could be appropriate for answering different kinds of questions that the DEC wanted to ask. This memo helped create a blueprint and timeline for the questions that our research partnership would address. Some questions were centered on monitoring progress

and improving quality and others on determining how the program affected instruction and children's learning. Some could be answered by reviewing previous research, but others required new data and analysis.

One key early issue was whether the program was ready for an impact study and, if so, how to fund it. The district wanted to identify the program's strengths and find out where it could be improved. The research team wanted to learn about how the impacts of a program that was not typical in the national landscape matched or differed from those of other programs. From other studies, we determined that two years of full program implementation was generally considered enough time to determine whether the new model was working. We identified a potential funding opportunity from the Institute of Education Sciences. Because the district wouldn't allow random assignment of students to the program, we looked for an approximation of experimental conditions in the real world (known to researchers as a *natural experiment*) and found one in the program's age cutoff. As in many districts, a child had to turn four by a given date (on or before September 1 in Boston) to attend the prekindergarten program that year. This created a natural experiment, one previously used in several other contexts, for estimating the causal effect of attending the program.[13] Children just one day apart on either side of the age cutoff for enrollment in prekindergarten are equivalent in their background characteristics. Yet the cutoff assigned children born on September 1 to be eligible for prekindergarten one year earlier than children born on September 2. This effectively randomized children who

were right at the age cutoff. Provided that all statistical requirements of the design were met, the difference in outcomes between children just at or just below the cutoff would represent the program's causal impact on children's school readiness.

Writing the grant was Weiland's course project in Richard Murnane and John Willett's spring 2008 causal inference class at Harvard. These two respected methodologists agreed to be part of the research team for the grant submission, as did Nonie Lesaux, an expert on dual language learners who is also a Harvard professor; Yoshikawa was the lead researcher. We applied in summer 2008 and received funding to start in spring 2009. The study we conducted included 2,018 children enrolled in 238 classrooms at 67 schools. We found that the BPS model had meaningful impacts on language, literacy, math, and socio-emotional skills—precisely the kindergarten readiness outcomes that were directly targeted by the program. It also had positive impacts on children's executive function skills—that is, their working memory, flexible thinking, and response inhibition skills. The Boston program didn't directly target executive function, but it is developmentally linked to growth in other domains, particularly math.[14] These impacts are shown in figure 1.

RPPs necessarily evolve over time. Today, our partnership has broadened to include the DEC, the University of Michigan, MDRC, and the Harvard Graduate School of Education. The work has extended to include the full prekindergarten-to-third-grade (P–3) span in the Boston Public Schools. As our research aims expanded, we needed to be able to collect and analyze more detailed data.

Figure 1. Boston Prekindergarten Impacts on Children's School Readiness
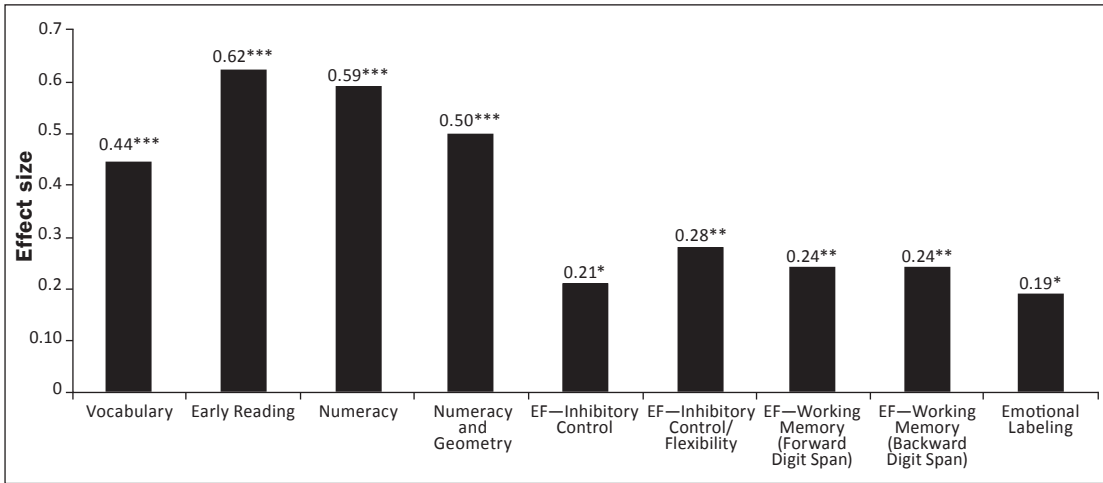
Notes: EF = executive function; *p<.05; **p<.01; ***p<.001; estimates come from a study of 2,018 children who attended Boston's prekindergarten program in 2008–09.

Source: Christina Weiland and Hirokazu Yoshikawa, "Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills," *Child Development* 84 (2013): 2112–30, https://doi.org/10.1111/cdev.12099.

Partnership Principles

Our partnership today has largely the same goals and philosophy as it did in its early days, even as the DEC's reach within BPS has extended through second grade and as the research team has added institutions. Our goal is simple: we use data and research to improve educational experiences and outcomes for young children. Our principles reflect how we do so:

1. *District questions are the priority.* The research team and the DEC work together to identify priority areas for research. Over the years, the DEC has resisted investigating questions it considered "too academic," that is, likely to answer questions that interest scholars but at a cost of too high a burden on the district without direct benefit. We've learned to select research topics at the intersection of the DEC's practical challenges and unanswered questions among researchers.

2. *The DEC is in the driver's seat in designing its programs.* In our partnership, the research partner's role is not to design interventions or tell the DEC what to do programmatically. We view the practitioners as the experts on implementation and the BPS context. The researcher's role is to support the DEC's agenda by offering insights from the broader scholarship when appropriate, while designing studies to guide quality improvement and key decisions in the BPS context.

3. *Coaches' and teachers' voices matter.* We share research findings first with coaches and then with teachers. We ask for their input on whether the findings seem right to them and on the story behind the findings. For example, why is a given aspect of quality lower than other aspects? How could supports for teachers be changed to improve quality in a given area? From our experience, sharing data and findings also improves

buy-in for research and reinforces a quality-improvement culture.

4. *What you don't do is as important as what you do.* The DEC philosophy is that teachers should focus on teaching. The DEC doesn't ask teachers to rate children's skills the way most US prekindergarten programs do. As part of our partnership work, we reviewed research on such rating systems and found scant rigorous evidence that they provide reliable, valid data or that they change teachers' practices.[15] But they do require a lot of time from teachers. Instead, for research, we largely rely on outside assessors who collect data on a sample of children and classrooms. To guide instruction, teachers do collect short direct assessments of children's language and literacy on well-validated, reliable measures. Because formal assessments don't measure skills like critical thinking and problem solving, the DEC also encourages and supports teachers to regularly collect data on student work and learning processes so they have authentic artifacts of learning to shape both their reflections on their practice (alone and with others) and their teaching strategies in real time.

5. *Trust is fundamental.* Our relationship's longevity has allowed us to develop trust. The DEC trusts that the research team understands the BPS context and the history of the program, won't misrepresent the findings, and will prioritize accurate and balanced dissemination of results. The DEC also trusts that when it would like research guidance for new issues or to guide decision-making, the research team will respond quickly. The research

team trusts that the DEC will grapple with the findings and take seriously any implications about changing course.

Overview of Our Work Together

Following these principles, we've collected and used data in a variety of ways.[16] Table 1 offers an overview of the types of data we've used, how frequently and why we collect them, and how we've used them to drive change. The table broadly summarizes these factors so as not to overwhelm readers with every data type and wave of data collection.

Funding and funders' agendas have also influenced our data collection; grants to support programs often require evaluation as well. We've also won several large-scale federal evaluation grants that have funded additional data collection, including work to create and use measures in district classrooms that assess the degree to which curricula are implemented as intended (known as *fidelity*) and to expand child outcome data to important domains beyond language and literacy, such as mathematics and executive function. Our only constant has been administrative data collected routinely by the district, such as child demographics, who enrolls where, teacher characteristics, and children's basic literacy skills.

Our collective knowledge of the BPS data, schools, context, and DEC staff, gained over more than a decade of work, allows us to consolidate different data sources and different findings and to identify what's new. Without an RPP, the district would have had to do more work to launch each new research effort, and less research would have been done. For example, the partnership means that our researchers already know about the district's structure,

Table 1. Types of Data Collected, Frequency of Collection, and Use (Summary)

Data source	When collected	Purpose	Use
Classroom quality and curriculum fidelity observational scores	About every two years	Changes as program evolves; in 2012, for example, data collection focused on K–2 due to concerns about quality of education after prekindergarten	Determine program gaps, needs, and strengths; guide professional development and programmatic decisions
Administrative data	Continuously	Tracking important programmatic data like child attendance, enrollment, demographics, teacher education, certification, and experience	Answer questions about program use and take-up; describe population and how it changes over time; use as control variables in analyses to limit participant burden
Teacher survey	About every two years	Gathering richer data on teacher background, experience of professional development, and opinions/desires related to current offerings	Understand teacher population in more depth; guide professional development and program decisions
Early reading skills and prekindergarten vocabulary	Three times per year by teachers	Monitor children's early literacy and language skill development; identify supports as needed	Describe BPS population; useful as outcomes in evaluation studies
Broader set of child outcomes	When external funding is available or when a research study requires them	Examine children's levels and growth on a broader set of important outcomes, like math, executive function, and socio-emotional skills	Describe BPS population; used as outcomes in evaluation studies

Note: A version of this table appears in Betty Bardige, Megina Baker, and Ben Mardell, *Children at the Center: Transforming Early Childhood Education in the Boston Public School*s (Cambridge, MA: Harvard Education Press, 2018).

what data it collects, and its core model. Sachs has also had five superintendents and eight immediate supervisors in 13 years. Strong research on the program quickly brought these bosses up to speed, showcased the DEC's strengths, and helped persuade BPS leaders to continue the DEC's direction rather than instituting new reforms and risk "reform weariness" among the district's early childhood teachers. For the researchers, the buy-in and support from the district has been invaluable. Launching large-scale, complex studies would

have been harder without a close working relationship with district staff. Also, our research questions have been more relevant to practice and policy because we frequently hear from administrators and practitioners who touch the program daily and we set our research agenda together with the DEC.

Navigating Tension between Rigor and Timeliness

Across projects funded by three large federal grants from the Institute of Education

Science, by small local grants, and by donated time from researchers, we've always had to navigate the natural tension between rigor and timeliness. Rigorous work tends to be painstakingly slow; policy decisions can be head-spinningly fast. The three examples that follow—covering decisions about summer learning, NAEYC accreditation, and professional development for teachers—illustrate how we navigated this tension.

Summer Learning

Research shows that among children from low-income families, growth in academic skills stagnates in the summer, leading to increases in income-based achievement gaps during the summer months.[17] High-quality summer enrichment programs can help stem this problem.[18] Well aware of this research, the BPS offered its own summer reading program for kindergarten and first-grade students. In fall 2010, facing a tight budget, the district considered whether to continue offering this program and, if so, whether to extend it to incoming prekindergarten students. Designing an experimental study for summer 2011 was not an option; the decision had to be made quickly, based on data the district already had. Weiland and Sachs identified data from the 2009 district summer program that could influence the decision (specifically, data on which families chose to send their children to the program, attendance, and student outcomes) and framed the key research questions. Weiland began analyses in mid-November 2010 and worked with Sachs to summarize results by mid-January 2011—a much tighter turnaround than the prekindergarten impact study displayed in figure 1, which took five years from grant writing to publication.

Without an experimental design, we had to solve a central "rigor" challenge caused by student selection into the program. In trying to estimate the effects of the summer program, the risk was that any results, whether positive or negative, could have had to do with characteristics of the students themselves and not those of the program (that is, *selection bias*). For example, imagine that only students who were already strong in reading attended the program—students who might have spent much of their summer choosing skill-building reading experiences on their own, even without the program. Now imagine that we compared their summer gains to those of their peers with weaker reading skills who didn't attend the program. We would find that the stronger students made larger gains than their peers. We would therefore risk concluding that the program was effective, when in fact the stronger students might have shown the same growth even without it.

To help overcome this potential problem, we compared the students who attended the summer program to two other groups of students: 1) students who applied to the program but did not attend; and 2) all other students in the same regular-school-year schools as those who attended the summer program. The students in the first control group were more similar to those who attended summer school, as they were drawn from a pool of those interested in the summer program. The second control group allowed us to compare children who attended the same schools and therefore might have shared background characteristics and classroom reading experiences with students who attended the summer program. Both of these comparison groups were imperfect; because students were not randomly assigned to attend the program or not, we couldn't

definitively conclude that the program itself, rather than other factors, *caused* any student gains. But having two control groups allowed us to examine whether our answer about the program's effectiveness depended on our choice of the control group. If so, our results would have been less trustworthy. If not, we would have greater confidence in our answers to the district.

We found that program attendance was strong. Eighty percent of students had attendance rates of 73 percent or higher, indicating that if it were offered, many district families would send their young children to the program. We also found that the program reached children who needed it more; before the program began, participants had lower literacy skills than their peers, and they were statistically significantly more likely to have repeated a grade. Controlling for their end-of-year literacy scores and background characteristics, students who attended the program showed stronger summer gains in literacy skills than did their peers in either of the two control groups (on one of four measures when compared to appliers and on three out of four measures when compared to all other students in the same schools).

Sachs presented this evidence in district meetings. We were careful to communicate its limitations and also to emphasize other data relevant to the decision, such as feedback from teachers in the program, to ensure that the student-level analysis would be seen as just one piece of evidence and not as decisive by itself. Though it was imperfect, in our view the student-level analysis helped the district make a more informed decision than it could have otherwise. Ultimately, the district decided to continue offering the program and to begin serving entering

kindergarteners as well (particularly those who had not attended preschool). As the summer program has continued and matured, it has become an important place for the DEC to pilot new curricula and has been expanded through second grade. The program has served over 3,000 students and is now part of the district's core summer programming.

And our evaluation work on summer has continued. In summer 2019, members of our RPP evaluated the program's effects on kindergarteners and first-graders using a more rigorous randomized trial. We were able to do so because more families applied for slots in the program than there were seats available. Within our relatively small samples (157 kindergarteners and 114 first-graders), we found benefits in reading skills for those who attended the program, particularly in phonics for kindergarteners.[19] We expect that as the district and the nation chart a way forward after COVID-19, our results point to a concrete and proven approach to address summer learning loss.

NAEYC Accreditation

Another critical decision the district faced was whether to pursue accreditation from the National Association for the Education of Young Children (NAEYC) for all district elementary schools. In the early childhood field, NAEYC accreditation is widely considered to be a marker of quality. The accreditation process is intended to improve program quality by ensuring that participating programs meet 10 standards covering four domains: children, teachers and staff, management and administration, and family and community relations.

To become accredited, a program must complete four steps. First, staff members

evaluate whether their program meets NAEYC's 10 standards and then make changes where they deem necessary. In the second and third steps, the program gathers and submits evidence to prove that it meets the standards. In the final step, an NAEYC assessor observes the program and recommends whether it should be accredited. To retain accredited status, a program must submit yearly reports documenting that it is maintaining quality. Program accreditation must be renewed every five years.[20]

Meeting the 10 standards and going through the four-step process is hypothesized to increase classroom quality and thereby improve the developmental outcomes of enrolled children.[21] Yet when we reviewed the research, we found limited empirical evidence that NAEYC accreditation affects classroom quality and child outcomes. Several studies found that NAEYC-accredited programs were of higher overall quality as measured by a commonly used classroom quality measure, the Early Childhood Environment Rating Scale–Revised (ECERS–R), than were programs not seeking NAEYC accreditation. Compared to unaccredited centers, NAEYC accredited centers have a greater proportion of college-educated teachers and staff with degrees or certifications in early childhood education and lower staff turnover rates. Compared to staff in unaccredited programs, staff in NAEYC-accredited programs have also been found to have interactions with children that are more positive.[22] But these studies weren't experimental, meaning that the results may be subject to selection bias—that is, they may be due to some other factor than the NAEYC accreditation process. Further, in some studies, being accredited didn't necessarily ensure high quality; some

accredited programs had mediocre scores on an observational measure of quality.[23] Only one prior study had examined the relationship between NAEYC accreditation and child outcomes, and that study was not experimental.[24]

The DEC had two primary reasons to pursue accreditation for its public elementary schools: to improve programs' structural elements so that they would be ready for deeper curriculum and instruction work and to have a tool to begin improving kindergarten. Supporting schools to help them achieve and sustain NAEYC accreditation costs BPS roughly $5,000 per classroom per year in coaching, materials, and facility work and takes three years to complete, on average. This is a considerable investment to make without clear-cut evidence that those resources would best be spent on NAEYC accreditation rather than another need in the district.

Using district data, we examined whether undertaking NAEYC accreditation was associated with higher classroom quality and with larger gains in children's vocabulary skills, comparing early adopters of the approach in the district to other district classrooms. Importantly, schools chose whether to seek accreditation, posing a major threat to rigor via selection bias. It could have been the case that the strongest schools or the schools most motivated to change had chosen to be early adopters. Students in these schools might have shown more growth across time due to factors other than NAEYC accreditation. But an experimental study was out of reach, because of time constraints, cost, and other practicalities. Even so, we aimed to maintain rigor by using reliable and valid classroom measures and including a sample of classrooms large enough to make

some generalizations about the district more broadly. Ultimately, we used data collected by the Wellesley Centers for Women in 2008 as part of the district's biannual progress monitoring program.

From a sample of 119 prekindergarten and kindergarten classrooms, the results overall were positive. For example, on the language and reasoning subscale of the ECERS–R (a commonly used observational measure of quality), classrooms in accredited schools scored 0.55 points higher than non-accredited schools (about half of a standard deviation, which is a relatively large difference in quality). However, we found no association between NAEYC accreditation and another measure of classroom quality, the Classroom Assessment Scoring System (CLASS).

We also found that preschool and kindergarten children in BPS schools that were involved in the NAEYC accreditation process had statistically significantly higher vocabulary score gains from fall to spring than did their peers in other BPS schools. This link remained after controlling for children's vocabulary scores in the fall and characteristics such as race/ethnicity, eligibility for the free/reduced-price lunch program, and language spoken at home. When we controlled for the global quality of the classroom as measured by the ECERS–R, preschool and kindergarten children in BPS classrooms undertaking the NAEYC accreditation process significantly outscored children in other classrooms.

Along with feedback from coaches, principals, and teachers, the district used our analysis as one piece of evidence in making the decision to expand NAEYC accreditation to more district schools. We were careful to

explain that our research couldn't identify causal effects and that factors other than NAEYC accreditation might have led to the gains we saw in classrooms that undertook the accreditation process. We did examine whether schools with NAEYC-accredited preschool and kindergarten classroom programs differed from those with unaccredited programs when it came to school-level factors such as third-grade test scores, school size, availability of wraparound services (such as before- and after-school care), principals' participation in an early childhood fellowship program, and the proportion of early childhood teachers with master's degrees. We found no statistically significant differences, though we always underscore that selection bias is a possible explanation for the results of a study like this one. Ultimately, despite the limitations of the research, the district had to make a decision about NAEYC accreditation. Our imperfect evidence was better than none at all.

We've since repeated our NAEYC analyses using data collected in 2010 and 2015 as part of the district's biannual progress monitoring. The analyses using 2010 data largely replicated our 2008 findings. In 2015, however, we found little association between NAEYC accreditation and three separate measures of classroom quality (though we didn't use the ECERS–R in 2015). We discussed these findings with our coaching team, which had expanded considerably since 2008. We found that they were spending much of their time meeting NAEYC requirements and not enough time on instructional quality specifically. These findings led to a shift in the DEC's approach to the NAEYC process; coaches have since placed a high priority on cognitively demanding instruction and tasks for students. Specifically, coaches were asked

to spend at least 50 percent of their NAEYC coaching time on instruction, particularly on supporting teachers' use of cognitively demanding tasks.

The district now has 47 NAEYC-accredited schools (of 80 elementary schools) and has invested $8 million in accreditation. We are once again examining associations between NAEYC accreditation and student gains using data collected in 2017 and 2018. This analysis will be used to make decisions about further NAEYC work in the district.

Professional Development

BPS is part of the IES Early Learning Network, an effort across five states to identify malleable factors at home, in the classroom, and in schools that can increase children's success from prekindergarten through third grade.[25] Like others in the IES network, we've focused on three research questions:

1) What are the district's goals and outcomes for P–3, and which state, local, and district policies either facilitate or hinder this vision?

2) What malleable classroom-level features, processes, and practices can predict within-year gains in students' outcomes?

3) How do students' cumulative experiences in their classrooms, homes, after school, and during the summer influence their P–3 developmental skill trajectories?

Beginning in fall 2016, we recruited a cohort of prekindergarten children to follow through the end of third grade. In the same cohort, we also added a group of children who didn't attend prekindergarten, to be followed

beginning in their kindergarten year. We're assessing the children's language, literacy, mathematics, social-emotional, and executive function skills across time. We're also measuring students' classroom experiences each year using CLASS, the Individualizing Student Instruction (ISI) measure, and curriculum fidelity measures; conducting surveys of parents and teachers; attending teacher training; interviewing key district leaders; conducting teacher focus groups; and reviewing relevant district and state documents. Finally, we're using administrative data on students, families, teachers, and schools, both from the district and, for students who leave BPS, from other districts in Massachusetts. Because of the COVID-19 pandemic, we were unable to assess children in second and third grade and now plan to follow them into fourth grade and possibly beyond.

As we conduct rigorous work to answer our primary research questions, we're also aiming to respond quickly to the DEC's need to identify the strengths and shortcomings of its curriculum reforms that aim to align instruction within and across prekindergarten to second grade, and to help teachers improve their practice. These added goals answer two needs. First, while measures like CLASS have helped the DEC improve overall quality, the feedback they offer teachers is somewhat broad and have not been predictive of children's gains. For example, a teacher might be told that language modeling isn't one of his or her strengths. But this guidance doesn't identify a specific weakness (for example: Too many closed-ended questions? Too little elaborative talk? And if so, when? In small groups? In the whole group?). We wanted to pinpoint more specific feedback that teachers could act on. Second, the

research we reviewed offers few specific recommendations for instruction, including how to allocate instructional time.[27] BPS coaches and teachers report that after subtracting transitions, bathroom breaks, lunch, recess, and special activities like music and art, about 3.5 hours are available for instruction out of a 6.5-hour school day. Administrators and teachers alike feel the need to make sure they use this precious instructional time judiciously. How much small group time should teachers have? How long should students be in centers? How long should we focus on a content-specific area to teach skills, and how much time should we allot for integration and synthesis? Furthermore, should our calculations and practices change for children who have experienced poverty and/or students whose first language isn't English? More detailed data on classroom practices could help with these hard choices.

Finally, for about five years the DEC has been working on creating and implementing its own aligned, play-based, interdisciplinary P–2 curriculum supported by district-developed training and coaching. The DEC undertook this work for three reasons: no P–3 curriculum models have been proven by experimental evaluation to improve students' third-grade outcomes; internal data showed that Boston's kindergarten through third-grade classrooms were of lower quality than its prekindergarten classes; and growing evidence suggests that exposure to high-quality learning environments after prekindergarten can help prevent fadeout of the boost children get from attending preschool.[28] The BPS curriculum, professional development, and coaching models are being revised based on lessons learned during their implementation (that is, via an iterative process), and data from the

IES study could help the DEC with these revisions.

But progress has been slow in meeting our fast-response goals of identifying the strengths and shortcomings of the P–2 curriculum reforms in real time and giving teachers guidance to help improve their practice. It took time to put our data infrastructure and analysis systems in place. It also took time to build cohesion and understanding among the DEC staff about expectations for curriculum adherence and fidelity and for them to agree on an acceptable level of implementation. Our fidelity measure combined measures created by the curriculum developers, fidelity measures from past research in BPS, and new or revised items for the current project.[29] To maximize accuracy, we aimed to collect at least two two-hour observations per classroom.

In August 2017—about a year after data collection began, about three months after fidelity data were collected, and just before our cohort was to start kindergarten—we sent our first results memos to the district. These were essentially a set of means and frequencies of the prekindergarten curriculum fidelity data, and they came with the caveat that the work was very preliminary. To make the data comparable across classrooms and usable to teachers and the district, we undertook an iterative process to identify key constructs for assessing fidelity and to examine the measurement properties of those constructs. We also discussed extensively how to give data back to teachers. Ultimately, for each curriculum component, our measure captured *dosage* (whether a given component was observed and how long teachers engaged in a component),

adherence (the degree to which teachers implemented the curriculum as intended), and *quality* (how well the curriculum was implemented); these were also averaged across components.[30] We also looked for measures that cut across components, grouping items into four constructs that were easy to talk about with coaches and teachers, and for which we found some support in our data analysis: use of rich vocabulary, making connections, scaffolding and differentiation, and building and extending children's thinking.

After an initial analysis of the fidelity data, we first shared key results with coaches. Next, we worked with coaches in large and small groups to figure out the right venue and format for sharing data with teachers. Coaches were worried that teachers with less-than-stellar scores might become more difficult to enlist in improvement efforts. Together, we planned to share the data in September 2018, when our cohort was entering first grade. Sachs would present key study findings to an audience of about 500 prekindergarten and kindergarten teachers at a district-wide training session to start

Box 1. Vocabulary Items from Fidelity Measure Used to Guide Professional Development

- Teacher uses vocabulary words as related to the unit book(s) and/or small group activity.
- Teacher embeds vocabulary in language (from Centers, Intro to Centers, Read Aloud components)
- Teacher defines vocabulary words (from Centers, Intro to Centers, Read Aloud components)
- Teacher uses a variety of vocabulary words that are sophisticated or advanced (from Centers, Intro to Centers, Read Aloud components)
- Teacher is intentional in which vocabulary words are used and how they are defined (from Intro to Centers, Centers, Small Group Read Aloud components)

Box 2. Professional Development Session Descriptions Shared with Teachers

Explicit & Embedded Vocabulary Instruction in Kindergarten Classrooms
In this session, attendees will briefly discuss early childhood vocabulary acquisition before delving into specific opportunities for vocabulary instruction within the K1 and K2 curricula. Through watching classroom videos of successful vocabulary instruction, attendees will become more familiar with best practices for explicit and embedded vocabulary instruction. Finally, attendees will analyze texts and select vocabulary most appropriate for instruction, and conclude the session by drafting preliminary plans for vocabulary instruction during the coming school year.

Making Connections
"Integrated curricula," "holistic approach," and "interdisciplinary learning" describe the instructional practice of making connections. This practice, implemented with quality, consistency, and as part of a classroom's discourse, is correlated with positive outcomes for children's gains in executive function and flexible thinking.

Scaffolding & Differentiating for High Quality Instruction
In this session, teachers will learn and share useful differentiation strategies that address the range of learning needs in the classroom. Teachers will learn appropriate scaffolds for a variety of learners. The session will hone in on scaffolding and differentiating during three key components of the Focus Curricula: Centers, Literacy Small Group and Math Small Group.

Building & Extending Children's Thinking Through Conversations, Questions, and Interactions
This session looks at the teacher's role in helping children think more deeply about their ideas and understanding of the world through the use of conversations and questioning.

the school year. To keep teachers who had weaker reports engaged in this work, we framed the presentation as highlighting strengths and weaknesses for all of BPS, district administrators included. We also worked together to develop individual data reports for teachers, to be shared only with the teachers themselves, consistent with the consent forms teachers had signed.

Coaches also worried that teachers might learn their weaknesses without clear advice about how to improve. Accordingly, they developed professional development sessions guided by our study findings and details of the four constructs (see box 1 for examples of vocabulary items and see box 2 for descriptions of the sessions shared with teachers to guide their selections).

In his presentation to teachers, Sachs started with the good news, particularly that students generally improved from fall to spring and that their gains on many tests averaged above the national norm. Despite these gains, we still saw large differences between white and nonwhite students and between children from low-income and better-off families. He then shared the CLASS findings, which showed teachers' scores for instructional support were substantially lower than for emotional support and classroom organization. This finding is typical nationally, but the data also showed that kindergarten teachers scored lower than prekindergarten teachers on conceptual development and language modeling.[31] Then, Sachs introduced the four cross-cutting fidelity constructs and went over key findings from the fidelity study. He explained that teachers could pick up their individual data reports and discuss them with study team members one on one. He also explained that teachers

should use their reports as just one piece of information for guiding their practice in the classroom. And most immediately, they could use it to choose professional development sessions on the following day, where activities would focus on the four fidelity constructs. He invited them to speak with district administrators about the patterns in the data and what steps could be taken to change them.

Our work on the fidelity measure continues. We have analyzed the kindergarten data collected in 2018 and used them to create measures that assess the extent to which children's instruction was aligned across prekindergarten and kindergarten. We also adapted the fidelity tool for use in first and second grade, incorporating adjustments to the curricular components made for older students. In doing so, we aimed to collect a consistent set of measures across grades while also working collaboratively with DEC staff to make the tools relevant to the district's ongoing need for classroom observation and coaching, in addition to using the data for research purposes. We collected the first-grade data in 2018–19 and a subset of second-grade classrooms in 2019–20 before the start of the pandemic. We aim to be as rigorous as possible—careful about sampling and the quality of the measurement—while also being able to share descriptive findings of interest to the district as soon as they become available. Rather than waiting until the end of the project to use the data to guide instruction, we're applying lessons from the prekindergarten and kindergarten years to improve Boston's Focus on Early Learning curriculum now. (The real-world tension in this work—localities need decisions quickly, but building valid,

reliable measures takes a great deal of time—is also reflected in the article in this issue by Amanda Willford and Jason Downer.)

Takeaways

The three vignettes above offer several broad lessons. First, when it comes to tradeoffs between timeliness and rigor, the potential cost for researchers is underappreciated. Meeting a practice partner's needs will almost certainly result in some work that, on its own, isn't publishable in peer-reviewed journals. Yet such publications are essential for academic researchers' careers. As an example, none of the work we discussed in the vignettes has appeared in a peer-reviewed journal. The summer-school study, for example, used an identification strategy that was too weak for economics-oriented journals, and we had too little valid, descriptive observational data on the components and quality of delivery of the program to satisfy educational research journals. The 2008 results of the NAEYC work were published in NAEYC's non-peer-reviewed publication for practitioners.[32] We tried to include the 2010 work in a paper on quality measures in BPS prekindergarten, but peer reviewers told us to cut the NAEYC work from the paper due to concerns that it wasn't sufficiently rigorous.

When it comes to tradeoffs between timeliness and rigor, the potential cost for researchers is underappreciated.

To be sure, the RPP model also can facilitate scholarship. For example, between 2012 and 2020, members of our RPP team published 19 peer-reviewed articles that used Boston data, and four others are currently under review. Our unpublished work for the district helped us by building trust, giving us important insights into how the district makes decisions, and developing good will that supported some of our slower, more rigorous research. But early career researchers in particular need to carefully balance fast-turnaround RPP work with more publishable, slower work. Early career faculty members may not yet have PhD students who can help with data and analysis (and who in turn can benefit from being trained to apply research methods in real-world contexts). Such students require supervision and training, but they can greatly reduce the amount of time faculty spend analyzing RPP data.

Ultimately, fast-turnaround work—which can have some of the largest impacts on partners and programs—carries a time cost no matter how it's managed. Academic/researcher incentive systems generally don't recognize this cost, discouraging researchers from investing in fast-turnaround work. To encourage more scholars to get involved in RPPs, many academic institutions would need to change their evaluation systems to give such work additional weight, particularly for early career researchers for whom the stakes of producing peer-reviewed publications quickly are particularly high. This is even more true for researchers who work with smaller school districts, where there may not be enough classrooms and children for the kind of large-scale randomized trials that are generally easier to find funding for and publish about. We need research on programs and contexts beyond large cities like Boston and New York to better

guide policy and practice, yet too little such work has been done, particularly on rural prekindergarten programs.

Second, funding structures can exacerbate the tension between rigor and timeliness. Most funders want to support either research or programming, but not both. We've cobbled together funding as best we can, sometimes combining several smaller grants into a common evaluation effort. When a program is funded but research isn't, it's much more difficult to design a carefully planned research study that produces rigorous, usable knowledge. To ensure stronger studies and greater learning, programmatic and research funding need to be better intertwined. (See article in this issue by Jacqueline Jones for a perspective on how funders are thinking through the role of philanthropy in supporting RPPs.)

Third, fast-response approaches are better suited for some research questions than for others. Once we decided that the Boston prekindergarten program was ready for an impact evaluation, for example, we rejected a fast-response approach. The impact evaluation and associated extension studies took five years.[33] In that case, we were evaluating the program as a whole, not just a single component. Less intensive fast-response approaches, in our view, are best for guiding decisions about specific model components or policies. Not every element of a program or program policy can be investigated separately via a rigorous randomized trial; programs and children's and family's needs are simply too complex. Moreover, the whole can be greater than the sum of the parts. Investigating each component separately may not lead to the best answers about building the strongest program.

Fourth, in building RPPs with capacity for both fast-turnaround and longer-term work, we advise starting simple. Faster work on questions like "Is NAEYC Accreditation associated with stronger classroom quality and strong child language gains?" helped build trust in our RPP, helped establish how the members of the partnership would get work done, and trained the researcher in using the district's data systems. Many of the questions asked and answered by RPPs may yield unflattering or null findings, and these can lead to adaptations to improve the program as well as revised research questions. It takes time going back and forth to get to the more nuanced questions that are directly relevant to the challenges faced by the school district. We're now asking questions such as: Under what conditions does the prekindergarten boost last?[34] How does quality vary across children in P–2 classrooms? How do we define and measure alignment from prekindergarten to second grade? Simple analysis of basic data helps build the foundation for a strong RPP. The trust developed through the RPP then allows researchers to disseminate their findings regardless of the results, which is imperative for building actionable knowledge and generating broader learning about early childhood education.

Finally, a core goal of fast-response work is to balance speed with the most rigorous approach to generating results. Accordingly, it's important to keep the limitations of fast-response studies front and center and to combine such quantitative analyses with other data, including interviews with teachers and program administrators. Policy makers typically aren't highly trained in study design and causal inference. The researcher bears the responsibility of making clear that quantitative fast-response analyses should

be viewed as only one piece of evidence for guiding a decision.

The Path Forward

Most of the eventual achievement gap between wealthier and poorer children is in place by the first day of kindergarten.[35] Closing the gap requires our collective best thinking on how to create high-quality early educational programs and improve existing ones.[36] For Boston, our path forward has included a strong RPP that has helped shape teacher practice and district policies while contributing to the broader scholarship. After 13 years of our relationship and counting, we believe deeply that despite their inevitable challenges, RPPs in early childhood are essential to chart the way forward.

Endnotes

1. Cynthia E. Coburn, William R. Penuel, and Kimberly E. Geil, *Research-Practice Partnerships: A Strategy for Leveraging Research for Educational Improvement in School Districts* (New York: William T. Grant Foundation, 2013).

2. Richard J. Murnane and John B. Willett, *Methods Matter: Improving Causal Inference in Educational and Social Science Research* (New York: Oxford University Press, 2010).

3. Betty Bardige, Megina Baker, and Ben Mardell, *Children at the Center: Transforming Early Childhood Education in the Boston Public Schools* (Cambridge, MA: Harvard Education Press, 2018).

4. Tracy Jan, "Boston Preschools Falling Far Short of Goals, Study Says," *Boston Globe*, April 7, 2007, A1; Christina Weiland and Jason Sachs, "The Role of Data, Research, and Evaluation in the Process of Change [appendix]," in Bardige, Baker, and Mardell, *Children at the Center*, 195–210.

5. Bardige et al., *Children at the Center*.

6. Judy Schickedanz and David Dickinson, *Opening the World of Learning* (Iowa City, IA: Pearson, 2005).

7. Douglas H. Clements and Julie Sarama, *SRA Real Math, PreK-Building Blocks* (Columbus, OH: SRA/McGraw-Hill, 2007).

8. Mary K. Ashe et al., "Opening the World of Learning: Features, Effectiveness, and Implementation Strategies," *Early Childhood Services* 3 (2009): 179–91; Douglas H. Clements and Julie Sarama, "Effects of a Preschool Mathematics Curriculum: Summative Research on the Building Blocks Project," *Journal for Research in Mathematics Education* 38 (2007): 136–63; Douglas H. Clements et al., "Mathematics Learned by Young Children in an Intervention Based on Learning Trajectories: A Large-Scale Cluster Randomized Trial," *Journal for Research in Mathematics Education* 4 (2011): 127–66, https://doi.org/10.5951/jresematheduc.42.2.0127.

9. Matthew A. Kraft, David Blazar, and Dylan Hogan, "The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence," *Review of Educational Research* 88 (2016): 547–88, https://doi.org/10.3102/0034654318759268; Christina Weiland et al., "Preschool Curricula and Professional Development Features for Getting to High-Quality Implementation at Scale: A Comparative Review across Five Trials," *AERA Open* 4, no. 1 (2018): https://doi.org/10.1177/2332858418757735.

10. Greg J. Duncan and Richard J. Murnane, *Restoring Opportunity: The Crisis of Inequality and the Challenge for American Education* (New York: Russell Sage and Cambridge, MA: Harvard Education Press, 2014).

11. Allison H. Friedman-Krauss et al., *The State of Preschool 2017* (Rutgers, NJ: National Institute of Early Education Research).

12. Weiland and Sachs, "Role of Data."

13. William T. Gormley Jr. et al., "The Effects of Universal Pre-K on Cognitive Development," *Developmental Psychology* 41 (2005): 872–84, https://doi.org/10.1037/0012-1649.41.6.872; Vivian C. Wong et al., "An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs," *Journal of Policy Analysis and Management* 27 (2008): 122–54, https://doi.org/10.1002/pam.20310.

14. Christina Weiland and Hirokazu Yoshikawa, "Impacts of a Prekindergarten Program on children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills," *Child Development* 84 (2013): 2112–30, https://doi.org/10.1111/cdev.12099.

15. Jaclyn M. Russo et al., "Examining the Validity of a Widely-Used School Readiness Assessment: Implications for Teachers and Early Childhood Programs," *Early Childhood Research Quarterly* 48 (2019): 14–25, https://doi.org/10.1016/j.ecresq.2019.02.003.

16. For more details, see Weiland and Sachs, "Role of Data."

17. Paul T. von Hippel, , Joseph Workman, and Douglas B. Downey, "Inequality in Reading and Math Skills Forms Mainly before Kindergarten: A Replication, and Partial Correction, Of 'Are Schools the Great Equalizer?,'" *Sociology of Education* 91 (2018): 323–57; Meghan P. McCormick et al., "Can Center-Based Care Reduce Summer Slowdown Prior to Kindergarten? Exploring Variation by Family Income, Race/Ethnicity, and Dual Language Learner Status," *American Educational Research Journal* 58 (2020): 420–55, https://doi.org/10.3102/0002831220944908.

18. Geoffrey D. Borman and N. Maritza Dowling, "Longitudinal Achievement Effects of Multiyear Summer School: Evidence from the Teach Baltimore Randomized Field Trial," *Educational Evaluation and Policy Analysis* 28 (2006): 25–48, https://doi.org/10.3102/01623737028001025; Doris R. Entwisle and Karl L. Alexander, "Summer Setback: Race, Poverty, School Composition, and Mathematics Achievement in the First Two Years of School," *American Sociological Review* 57 (1992): 72–84, http://dx.doi.org/10.2307/2096145; Brian A. Jacob and Lars Lefgren, "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis," *Review of Economics and Statistics* 86 (2004): 226–44, http://dx.doi.org/10.1162/003465304323023778.

19. Gloria Maldonado,and Christina Weiland, "The Impacts of Boston's Summer Focus Program on the Literacy Skills of Kindergarten and First-Grade Students," unpublished manuscript, 2020.

20. National Association for the Education of Young Children (NAEYC), "Accreditation Process for Early Learning Programs," accessed June 20, 2019, https://www.naeyc.org/accreditation/early-learning/process; Marcy Whitebook, Laura Sakai, and Carollee Howes, *NAEYC Accreditation as a Strategy for Improving Child Care Quality: An Assessment by the National Center for the Early Childhood Work Force* (Washington, DC: National Center for the Early Childhood Workforce, 1997).

21. Marcy Whitebook.,"NAEYC Accreditation as an Indicator of Program Quality: What Research Tells Us," in *NAEYC Accreditation: A Decade of Learning and the Years Ahead*, ed. Sue Bredekamp and Barbara A. Willer (Washington, DC: NAEYC, 1996), 31–46.

22. Paula J. Bloom, "The Quality of Work Life in Early Childhood Programs: Does Accreditation Make a Difference?," Bredekamp and Willer, *NAEYC Accreditation*, 13–24; Carollee Howes and Ellen Galinsky, "Accreditation of Johnson & Johnson's Child Development Center," in Bredekamp and Willer, *NAEYC Accreditation*, 47–60; Whitebook "NAEYC Accreditation as an Indicator"; Whitebook, Sakai, and Howes, *NAEYC Accreditation as a Strategy*; Marcy Whitebook, Carollee Howes, and Deborah Phillips, *Who Cares? Child Care Teachers and the Quality of Care in America. Final Report, National Child Care Staffing Study* (Oakland, CA: Child Care Employee Project, 1989).

23. Whitebook "NAEYC Accreditation as an Indicator"; Whitebook, Sakai, and Howes, *NAEYC Accreditation as a Strategy.*

24. Minnesota Department of Human Services, *School Readiness in Child Care Settings: A Developmental Assessment of Children in 22 Accredited Child Care Settings* (St. Paul, MN: Minnesota Department of Human Services, Children and Family Services, 2005).

25. Meghan McCormick et al., *The Challenge of Sustaining Preschool Impacts* (New York: MDRC, 2017).

26. Paola Guerrero-Rosada et al., "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 (2020): 1–12, https://doi.org/10.1016/j.ecresq.2020.07.009; Christina Weiland et al., "Associations between Classroom Quality and Children's Vocabulary and Executive Function Skills in an Urban Public Prekindergarten Program," *Early Childhood Research Quarterly* 28 (2013): 199–209, https://doi.org/10.1016/j.ecresq.2012.12.002.

27. Nell K. Duke and Heidi Anne E. Meiser, *Phonics Faux Pas: Avoiding Instructional Missteps in Teaching Letter-Sounds Relationships* (Washington, DC: American Federation of Teachers, 2018); Judith A. Schickedanz and Catherine Marchant, *Inside PreK Classrooms: A School Leader's Guide to Effective Instruction* (Cambridge, MA: Harvard Education Press, 2018).

28. Deborah Phillips et al., *Puzzling It Out: The Current State of Scientific Knowledge on Pre-Kindergarten Effects—A Consensus Statement* (Washington DC: Brookings Institution, 2017).

29. Weiland and Yoshikawa, "Impacts"; Monica Yudron, Christina Weiland, and Jason Sachs, "The Importance of Adjusting for Context: Learning from a Pilot Expansion of the Boston Public School Pre-K Model to Community-Based Settings," unpublished manuscript, 2018.

30. Christopher S. Hulleman, Sara Rimm-Kaufman, and Tashia Abry, "Innovative Methodologies to Explore Implementation: Whole-Part-Whole—Construct Validity, Measurement, and Analytical Issues for Intervention Fidelity Assessment in Education Research" in *Applying Implementation Science in Early Childhood Programs and Systems*, ed. Tamara Halle, Allison Metz, and Ivelisse Martinez-Beck (Baltimore, MD: Brookes Publishing, 2013), 65–93.

31. Ajay Chaudry et al., *Cradle to Kindergarten: A New Plan to Combat Inequality* (New York: Russell Sage, 2017).

32. Jason Sachs and Christina Weiland, "Boston's Rapid Expansion of Public School-Based Preschool: Promoting Quality, Lessons Learned," *Young Children* 65, no. 5 (2010): 74–7,

33. Christina Weiland, M. Clara Barata, and Hirokazu Yoshikawa, "The Co-Occurring Development of Executive Function Skills and Receptive Vocabulary in Preschool-Aged Children: A Look at the Direction of the Developmental Pathways," *Infant and Child Development* 23 (2014): 4–21, https://doi.org/10.1002/icd.1829; Weiland et al., "Associations"; Christina Weiland et al., "Early Mathematics Assessment: Validation of the Short Form of a Prekindergarten and Kindergarten Mathematics Measure," *Educational Psychology* 32 (2012): 311–33, https://doi.org/10.1080/01443410.2011.654190; Christina Weiland and Hirokazu Yoshikawa, "Does Peer Socio-Economic Status Predict Children's Gains in Receptive Vocabulary and Executive Function in Prekindergarten?," *Journal of Applied Developmental Psychology* 35 (2014): 422–32; Weiland and Yoshikawa "Impacts."

34. Rebecca Unterman and Christina Weiland, "Quantifying and Predicting Variation in the Medium-Term Effects of Oversubscribed Prekindergarten Programs," working paper, MDRC, 2019; Christina Weiland et al., "The Effects of Enrolling in Oversubscribed Prekindergarten Programs through Third Grade," *Child Development* 91 (2020): 1401–22, https://doi.org/10.1111/cdev.13308.

35. Chaudry et al., *Cradle*; Sean F. Reardon and Ximena A. Portilla, "Recent Trend in Income, Racial, and Ethnic School Readiness Gaps at Kindergarten Entry," *AERA Open* 2 (2016): https://doi.org/10.1177/2332858416657343.

36. Daphna Bassok and Mimi Engel, "Early Childhood Education at Scale: Lessons from Research for Policy and Practice," *AERA Open* 5 (2019): https://doi.org/10.1177/2332858419828690; Christina Weiland, "Commentary: Pivoting to the 'How': Moving Preschool Policy, Practice, and Research Forward," *Early Childhood Research Quarterly* 45 (2018): 188–92, https://doi.org/10.1016/j.ecresq.2018.02.017.