



When Do the Literacy Skills of Preschool Attendees and Non-Attendees Converge? Evidence from Boston

Christina Weiland, University of Michigan
Rebecca Unterman, MDRC
Anna Shapiro, University of Michigan

Education Policy Initiative
Gerald R. Ford School of Public Policy
735 S. State Street
Ann Arbor, Michigan 48109

This study is funded by the Institute of Education Sciences RFA R305A140059 and R305B150012. Correspondence concerning this article should be addressed to Christina Weiland, School of Education, 610 E. University Ave, Ann Arbor, MI 48104; email: weilandc@umich.edu. Thanks to the Boston Public Schools, Jason Sachs, Brian Gold, the BPS Department of Early Childhood coaches and staff, the BPS Office of Data and Accountability (particularly Nicole Wagner, Erin Cooley, Barry Kaufman, and Peter Sloan), Kamal Chavda, and the Wellesley Centers for Women. Special thanks also to Hirokazu Yoshikawa, Howard Bloom, Richard Murnane, David Deming, Catherine Snow, Caroline Ebanks, Gina Biancarosa, Sarah Kabay, Eleanor Martin, Shana Rochester, and Sara Staszak. EPI Working Papers are circulated for discussion and comment purposes and have not been peer-reviewed or been subject to the review by EPI co-Directors. Any opinions, findings, conclusions, or recommendations expressed are those of the author(s) and do not necessarily reflect the view of the Education Policy Initiative or any sponsoring agency.

When do the literacy skills of preschool attenders and non-attenders converge?

Evidence from Boston

Christina Weiland
University of Michigan

Rebecca Unterman
MDRC

Anna Shapiro
University of Michigan

Authors' note:

This study is funded by the Institute of Education Sciences RFA R305A140059 and R305B150012. Correspondence concerning this article should be addressed to Christina Weiland, School of Education, 610 E. University Ave, Ann Arbor, MI 48104; email: weilandc@umich.edu

Thanks to the Boston Public Schools, Jason Sachs, Brian Gold, the BPS Department of Early Childhood coaches and staff, the BPS Office of Data and Accountability (particularly Nicole Wagner, Erin Cooley, Barry Kaufman, and Peter Sloan), Kamal Chavda, and the Wellesley Centers for Women. Special thanks also to Hirokazu Yoshikawa, Howard Bloom, Richard Murnane, David Deming, Catherine Snow, Caroline Ebanks, Gina Biancarosa, Sarah Kabay, Eleanor Martin, Shana Rochester, and Sara Staszak.

Abstract

Attending preschool improves children’s kindergarten readiness, but the cognitive outcomes of preschool attenders and non-attenders tend to converge partially or fully in elementary school. In older programs, most of the non-attender “catch up” occurs in kindergarten (Li et al., 2016), but evidence from today’s programs is relatively sparse. Using data on approximately 5,000 Boston Public School prekindergarten appliers and a quasi-experimental approach, we examine convergence patterns in the K-3 literacy outcomes of prekindergarten attenders and non-attenders. Consistent with the previous literature, we find that most of the convergence in K-3 literacy outcomes occurs in kindergarten. Our findings suggest that detailed investigations into the kindergarten teaching and learning context may be particularly important for solving the widely noted preschool convergence pattern.

When do the literacy skills of preschool attenders and non-attenders converge?

Evidence from Boston Public Schools

Partial or full convergence of cognitive outcomes between preschool attenders and non-attenders in elementary school is a near-universal finding in the literature, in both older and recent preschool studies, and in contexts in which the benefits of preschool attendance are detected later in the life course (Yoshikawa, Weiland, & Brooks-Gunn, 2016). Existing research shows that much of this convergence occurs very early in elementary school. For example, a recent meta-analysis found that about half of the eventual convergence on cognitive outcomes occurs during kindergarten and then by about half again by the end of second grade (Li et al., 2016). However, most of this evidence is drawn from relatively small programs from decades ago when students' counterfactual options were different than today's and when parents of all income levels invested less time and money in their children's learning (Bassok, Finch, Lee, & Waldfogel, 2016). We do not yet know if the same convergence pattern holds for today's large-scale public preschool programs.

Evidence so far from more recent large-scale programs is sparse. In two of the largest and most rigorous recent studies, convergence was particularly rapid. In the experimental Head Start Impact Study, there were positive and statistically significant impacts after one year on four-year-old treatment group children's language and literacy skills (0.09-0.25 SDs; Puma et al., 2012). By the end of kindergarten, these benefits had declined to -0.02 to 0.06 SDs on the same skills tested at the end of preschool and treatment/control differences were no longer statistically significant – a faster rate of decay than found in the overall literature. Likewise, the Tennessee Voluntary Prekindergarten quasi-experimental study found initial benefits for prekindergarten attenders (0.09 to 0.41 SDs across language, literacy, and math tests) that did not persist through

end of kindergarten (-0.10 to 0.09 SDs; Lipsey, Farran, & Hoffer, 2015). In contrast, in a non-experimental, nationally representative study using ECLS-K data, effects decayed at a slower rate and mirrored the pattern in the older literature for both the 1998 and 2010 cohorts. For example, in the 2010 cohort, preschool attenders outperformed non-attenders by 0.15 SDs in the fall of K and by 0.07 SDs in the spring of K – a decline of approximately half for both reading and math outcomes (Bassok, Gibbs, & Latham, 2015). Data on K-2 cognitive outcomes is not available in most studies of the longitudinal effects of preschool, including in programs with particularly strong results from third grade and beyond (e.g., Bania, Kay, Aos, & Pennucci, 2014; Barnett, Jung, Youn, & Frede, 2013; Cascio & Schanzenbach, 2013; Gormley, Phillips, & Anderson, 2016; Ladd, Muschkin, & Dodge, 2014).

Determining when and how rapidly cognitive convergence occurs in elementary school in today's programs is important for determining the drivers of this phenomenon and for finding approaches to stem it. Elementary-school convergence presently is a puzzle (Bailey, Duncan, Odgers, & Yu, 2017), one that a leading group of early childhood researchers recently emphasized deserves considerable research, practice, and policy attention (Phillips et al., 2017). Further, while in the older literature, kindergarten appears to be the convergence "hot spot" (Li et al., 2016), kindergarten has changed markedly in the last 15 years. As documented in Bassok, Latham, and Rorem (2016), today's kindergarten teachers hold higher academic expectations for children, devote more time to advanced literacy and math content, use a more teacher-directed approach, and devote less time to other learning domains (e.g., science, music, art). These changes are ambiguous in terms of their consequences for convergence patterns – today's kindergarten classrooms may now be better poised to build on preschool attenders' enhanced skills and knowledge (sustaining the prekindergarten boost) or to compensate for any gaps in the

skills and knowledge of non-attenders (contributing to a convergence of student outcomes). Additional research on convergence patterns is needed to understand whether patterns from older programs holds for today's programs and to identify for which grades detailed research into teaching and learning processes may be particularly insightful in solving the convergence puzzle.

We add to this nascent literature by examining the pattern of convergence between Boston Public Schools (BPS) prekindergarten attenders and children who applied to the program but ultimately did not attend. Importantly, in a rigorous study, the program had meaningful impacts on children's language, mathematics, literacy, executive function and socioemotional skills at kindergarten entry (the 2008-2009 cohort; Weiland & Yoshikawa, 2013). Previous research found that, for a more advantaged subset of children who participated in lotteries to attend the program, there were no lasting benefits in third grade on retention, special education placement, and third grade standardized reading and math test scores (the 2007-2011 cohorts; Weiland et al., 2019). However, in the full sample of children who attended the program in those years, quasi-experimental research found a pattern of small persistent benefits on those same outcomes. Notably, for the full sample, the third-grade reading and math benefits (about 0.09 SD) were considerably smaller than those found at the beginning of kindergarten in for the 2008-2009 cohort only (0.60 SD) – mirroring the cognitive convergence pattern found more broadly. In the present study, we use K-3 literacy test scores to add to this set of findings. We examine *when* and *how rapidly* convergence of literacy outcomes occurred between kindergarten and the end of third grade in the Boston context.

Method

Setting and program details. The Boston Public Schools Prekindergarten began in 2005. It is based entirely in the public schools, pays teachers on the same scale as K-12 teachers, subjects teachers to the same educational requirements of K-12 teachers (masters degree within five years), and is open to all children in the city. Also, since 2007, it has utilized a consistent curricula and coaching system. Teachers in our study years implemented both *Opening the World of Learning*, which targets children's early language and literacy skills and includes a social-skills component embedded in each unit (Schickedanz & Dickinson, 2005), and *Building Blocks*, an early mathematics curriculum which covers numeracy and geometry, and has a heavy focus on verbal mathematical reasoning (Clements & Sarama, 2007a). Both curricula have shown positive effects on children's outcomes in other studies (Ashe, Reed, Dickinson, Morse, & Wilson, 2009; Clements & Sarama, 2007b; Clements et al., 2011), though the evidence base for Building Blocks is stronger than that for the OWL (Weiland & Yoshikawa, 2013). In 2007-2009, curricula implementation was supported via trainings and regular coaching, meaning weekly to bi-weekly on-site support from an experienced early childhood coach trained in both curricula (see Weiland & Yoshikawa, 2013 for additional details). In our study years (2009-2011), due to budget cuts, coaching was targeted to new teachers and to prekindergarten and kindergarten teachers in schools undergoing National Association for the Education of Young Children Accreditation, a quality assurance process used in early childhood settings nationally. Overall, Boston's structural and programmatic choices make it fairly unique among public programs nationally which tend not to require masters degrees, usually do not pay prekindergarten teachers on the same scale as K-12 teachers, target slots to children from low-income families or with other risk factors, do not require a proven, consistent curriculum, and do not employ coaching (Barnett et al., 2017). The program has similar structural quality and

emotional support quality as other large-scale program nationally and has the highest average instructional quality of a large-scale program to date (Chaudry, Morrissey, Weiland, & Yoshikawa, 2017).

In our study's focal years, children's K-2 experiences in the district were not as high quality as their prekindergarten experiences (see Weiland et al., 2019). The district implemented the literacy curriculum *Reading Street* and the mathematics curriculum *TERC Investigations* which do not have as strong an evidence base as the district's prekindergarten program curricula, and were not supported by professional development as systematic or as frequent as the pre-k program's supports. Data show that prekindergarten classroom instructional quality was markedly higher on average than K-3 instructional quality (see Weiland et al., 2019). Notably, the district responded to this evidence and other related evidence by subsequently (but not in our study years) developing its own K-2 curriculum and associated professional development program (Boston Public Schools, 2017).

Sample

Our sample comes from the population of students who applied to the Boston prekindergarten program for four year olds in 2009-2010 or 2010-2011 who had at least one K-2 literacy test score ($N=4,971$, or approximately 81% of the full sample of appliers). We focus on these two cohorts because our key literacy outcome measures were collected most consistently in these years. As shown in Table 1, the sample was majority free-reduced-lunch eligible and mostly Black or Hispanic and about 71% ultimately attended at least one day of Boston prekindergarten. Compared to appliers who did not enroll in Boston prekindergarten, enrollee children were on average less likely to qualify for free and reduced price lunch in Kindergarten (-7 percentage points, $p<.001$), less likely to be Hispanic (-13 percentage points, $p<.001$) and more

likely to be Asian (4 percentage points, $p < .001$). Attrition rates were very similar among enrollees and non-enrollees across K-2 (3%-21% range across K-2, with small differences between 2.4 and 3.6 percentage points in attrition by enrollee status; results available upon request).

Outcomes

K-2 literacy skills. We use teacher-collected data from the Dynamic Indicators of Basic Literacy Skills (DIBELS; Good & Kaminski, 2002; Good et al., 2011). Administered subtests measured children's letter knowledge (Letter Naming Fluency; LNF), oral reading fluency (Oral Reading Fluency; ORF), phonological awareness (Initial Sound Fluency and Phoneme Segmentation Fluency; ISF and PSF), and alphabetic principle (e.g., letter-sound correspondence and the ability to blend letters into words in which letters represent their most common sounds; Nonsense Word Fluency; NWF). These subtests have good reliability and good concurrent, predictive, and discriminant validity properties, are widely used, and are sensitive to intervention effects (e.g., Biancarosa, Bryk, & Dexter, 2010; Burke, Hagan-Burke, Kwok, & Parker, 2009; Good et al., 2004; Good et al., 2011). Additionally, content covered by the DIBELS subtests aligns reasonably well with the primary early literacy test used in previous RD studies of public preschool programs – the Woodcock-Johnson Letter-Word Identification subtest (Woodcock, McGrew & Mather, 2001) – and the two tests have good concurrent and predictive validity (Speece, Hills, Ritchey, & Hillman, 2003).

Following developer guidelines, BPS teachers administered different subtests at different time points from K-2. Specifically, the ISF and LNF were administered in kindergarten fall; the PSF and NWF in kindergarten spring; the LNF and PSF in first-grade fall; and the ORF in first-grade spring, second-grade fall, and second-grade spring. The district switched using the

DIBELS Sixth Edition to the DIBELS Next in 2012-2013, when cohort four children were in first grade. The LNF and PSF subtests that first-grade children took in first-grade fall are equatable across versions (Good et al., 2011). However, the ORF subtest used new reading passages (e.g., all test items changed) in the updated version and thus, scores are not equatable across test forms. For this reason, we do not combine the first-grade end-of-year ORF scores for the two cohorts. We z-scored all subtests so that scores could be interpreted in reference to the average BPS DIBELS subtest-taker score at each time point.

Third-grade standardized reading scores. Our 2009-2010 cohort took the Massachusetts Comprehensive Assessment System (MCAS) in third grade, the test used for state accountability purposes in Massachusetts (see Weiland et al., 2019 for psychometric details). In 2015, all but two schools in BPS chose to administer a new test based on Common Core standards, the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment (Massachusetts Department of Elementary and Secondary Education, 2016). We use students' reading scores from these tests, following the state's recommendations that researchers standardize students' estimated theta (i.e., IRT) scores when conducting analyses that require pooling across the MCAS and PARCC exams (Massachusetts Department of Elementary and Secondary Education, 2016). We standardized each student's theta score on the mean and standard deviation of all third graders within Boston Public Schools taking the given exam in that year. Test score data in this paper accordingly can be interpreted as a given group's performance in reading compared to the average BPS third grader.

Covariates. Using administrative records, we constructed a set of student-level covariates. We captured students' race/ethnicity using a set of dichotomous variables that identified whether a student was Asian, Black, Hispanic, White, or Mixed/other. Similarly, we

used a set of dichotomous variables to identify whether the students' home language was English only, Spanish, or another language. Using student birthdates, we calculated students' age as of September 1 of their prekindergarten application year. We also created dichotomous variables that identified student eligibility for free-reduced priced lunch in pre-k or K; whether the student was male; and whether the student's country of origin was the U.S. We also identified which school they lived closest to in their first BPS year as a proxy for home neighborhood using administrative records.

Analytic approach. We used a propensity-score approach to estimate the relationship between BPS prekindergarten enrollment and our key K-3 literacy outcomes. Specifically, we predicted the probability that a student would receive BPS prekindergarten conditional on their background characteristics, their cohort year, and home neighborhood in their first BPS year. We then inverted these propensities to obtain an inverse probability weight (IPW) that we used in our subsequent regression analysis to counteract selection into the program (Imbens & Wooldridge, 2009; Murnane & Willett, 2010). The covariate differences between prekindergarten attenders and non-attenders shown in Table 1 were greatly reduced using IPW (for example, the standardized raw difference for Hispanic was -0.32 and the weighted difference was -0.06; results available upon request).

In addition to the inverse probability weight, our regression equation included, by reading outcome, a Boston prekindergarten enrollment indicator, children's covariates (gender, race/ethnicity, home language, country of origin, cohort), and fixed effects for school they lived closest to in their first BPS year (as a proxy for home neighborhood). For analytic simplicity, we did not impute covariates in our IPW work (98 children – about 2% - were missing at least one covariate value).

Counterfactual. Information on the before-kindergarten settings of non-enrollee children (i.e., the counterfactual to which BPS prekindergarten is being compared) was not available. However, we do have parent-reported data on the before-kindergarten settings for non-enrollee children who enrolled in the two prior cohorts – e.g., whether they attended Head Start, private preschool, family daycare, or parental/relative care. While imperfect, this information helps to clarify to *what* Boston prekindergarten is likely being compared in the present study. In the two prior cohorts, 74% of prekindergarten non-enrollees attended another center-based program. Of those in center-based care, 55% were in private centers, 38% in Head Start, 5% in other non-BPS public settings, and 1% in charter schools. Six percent were in family daycare and twenty percent of non-enrollees were at home. Accordingly, Boston prekindergarten in our study is likely being compared to a mixed counterfactual, in which the majority of children likely attended other preschool programs. For details on counterfactual data in the two prior cohorts, see Weiland and colleagues (2019).

Results

As shown in Figure 1, in kindergarten fall, Boston prekindergarten attenders outscored non-attenders on the LNF subtest by 0.44 SD and for the ISF subtest, by 0.26 SD. By the end of kindergarten spring, these differences declined to 0.17 SD (PSF) and 0.17 SD (NWF). Grade 1 results were comparable for two out of four subtests (0.17 SD for fall LNF and spring ORF). The two other Grade 1 subtests and the two Grade 2 subtests were in the 0.06-0.09 range (one fall and one spring in each grade). In third grade, the difference between Boston prekindergarten attenders and non-attenders was 0.10 SD on the state standardized reading assessment ($p < .0001$).

Robustness checks. We probed the robustness of our results in several ways. We re-estimated results using a threshold of 150 days for defining prekindergarten enrollment given the

distribution in our data (e.g., 89% K1 enrollees were enrolled for at least 150 days). We also fit OLS models controlling for neighborhood fixed effects and student covariates and with imputed missing covariate data. We also refit our primary models with fixed effects for school most attended in kindergarten. The pattern of results was stable across models (available upon request).

Discussion

Consistent with prior literature (Li et al., 2016), most of the test-score convergence in our full sample appears to have occurred during the kindergarten year. At kindergarten entry, Boston prekindergarten enrollees scored 0.44 SD higher than non-enrollees. By spring of kindergarten, differences were about half as large (0.17 SD for PSF and NWF). Interestingly, the second part of the broader pattern in preschool literature – that after kindergarten, the preschool advantage declines by about half again by the end of second grade (Li et al., 2016) – held for two subtests in Grade 1 and for the two Grade 2 subtests. But for two other subtests, Grade 1 magnitudes were identical to kindergarten spring magnitudes. Interestingly, Bassok and colleagues (2015) also found deviation from the broader pattern – e.g., a decline of approximately half in reading and math differences between preschool attenders and non-attenders in the 1998 and 2010 ECLS-K cohorts and very little decline these differences from end of kindergarten to end of first grade. Overall, the convergence rates in our study and in Bassok and colleagues (2015) are markedly slower than those in the recent Head Start Impact Study and Tennessee Voluntary Prekindergarten programs (Lipsey et al., 2015; Puma et al., 2012) in which the effects present at the end of the preschool year had declined to zero by the end of kindergarten.

Our results have several important limitations. Prekindergarten attendance was not randomly assigned. Helpfully, administrative records allowed us to identify our sample as

appliers to Boston prekindergarten in the relevant years; in contrast, longitudinal non-experimental studies of prekindergarten programs tend to lack information on applier behavior and instead all compare children enrolled in a given system K+, regardless of pre-k applier behavior (e.g., Barnett et al., 2013; Jung et al., 2013; Phillips et al., 2017). Limiting our sample to appliers likely helped to reduce selection bias concerns to some degree. We also took steps (e.g., IPW and school fixed effects) to mitigate selection bias, though our covariates were relatively coarse particularly compared to some other recent prekindergarten studies that have used the propensity scores approach (e.g., Gormley et al. 2016; Lipsey et al., 2015; Phillips, Anderson, & Gormley, 2018). Selection bias thus could have played a role in our results. It is worth noting that the beginning-of-K difference between attenders and non-attenders on the LNF test in the present (0.47 SD) is remarkably close to the impact on literacy skills in the previous regression discontinuity study of the Boston prekindergarten's 2008-2009 (Weiland & Yoshikawa, 2013). Specifically, the RD study found an impact of 0.62 SD for one bandwidth around the cohort and 0.47 SD for an alternate bandwidth. Accordingly, we view it as unlikely that our results are due entirely to selection bias. But given our study's design, we interpret our results as *associations* and not causal estimates. In addition, 18% of appliers to Boston prekindergarten had no DIBELS scores available, primarily because of non-enrollment in BPS in the K-2 years. Appliers with DIBELS scores were on average more likely to be eligible for free and reduced price lunch than appliers without (73% compared with 43%), more likely to be Hispanic (46% compared with 39%) and more likely to speak Spanish at home (32% compared with 24%). This selection into BPS limits the external validity of our results to children whose families applied for BPS prekindergarten and who subsequently enrolled their children in at least some K-2 grades in BPS schools. Finally, the DIBELS taps decoding-focused skills and not the

full range of reading skills children must master to become strong readers (Snow & Matthews, 2016). Further, because K-2 classrooms tend to focus attention on decoding skills, the preschool boost on these skills may be particularly susceptible to convergence (Bailey et al., 2017). More work is needed on a broader range of skills, in literacy and in other domains, to inform our understanding of when and how rapidly convergence between preschool attenders and non-attenders occurs.

Despite these limitations, at this early stage in solving the convergence puzzle (Bailey et al., 2017; Phillips et al., 2017), our findings have two major implications. First, in our study and more broadly, kindergarten appears to be the key period in which the prekindergarten boost on cognitive skills fades. This finding is consistent with the older pattern in the literature (Li et al., 2016) but is an important addition to the evidence base given that kindergarten has changed markedly over the last 15 years (Bassok, Lathem, & Rorem, 2016), becoming more academic overall and thus poised to either to better build on gains from preschool or alternatively, to catch up preschool non-attenders. Our results add to prior studies (Li et al., 2016) that point to the kindergarten teaching and learning context as a particularly important one for understanding and stemming convergence. Some such studies already exist and have yielded highly useful information. For example, Engels, Claessans, and Finch (2013) found that kindergarten teachers tend to teach material that the majority of students have already mastered and generally do not differentiate instruction. More such studies are needed. And finally, though there is significant policy and practitioner attention on the prekindergarten through third grade continuum, there are no proven P-3 models (McCormick, Hsueh, Weiland, & Banger, 2017; Stipek, Clements, Coburn, Franke, & Farran, 2017). Our results provide further fodder for the urgency of developing and testing P-3 models.

References

- Ashe, M.K., Reed, S., Dickinson, D.K., Morse, A.B., & Wilson, S.J. (2009). Opening the World of Learning: Features, effectiveness, and implementation strategies. *Early Childhood Services, 3*, 179-191.
- Bailey, D., Duncan, G. J., 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.
- Bania, N., Kay, N., Aos, S., & Pennucci, A. (2014). Outcome evaluation of Washington State's Early Childhood Education and Assistance Program, (Document No. 14-12-2201). Olympia: Washington State Institute for Public Policy.
- Barnett, W.S., Jung, K., Youn, M., & Frede, E.C. (2013). *Abbott Preschool Program longitudinal effects study: Fifth grade follow-up*. New Brunswick, NJ: National Institute for Early Education Research. Retrieved from <http://nieer.org/wp-content/uploads/2013/11/APPLES205th20Grade.pdf>.
- Barnett, W. S., Friedman-Krauss, A. H., Weisenfeld, G. G., Horowitz, M., Kasmin, R., & Squires, J. H. (2017). *The State of Preschool 2016: State Preschool Yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Bassok, D., Finch, J. E., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic Gaps in Early Childhood Experiences: 1998 to 2010. *AERA Open, 2*(3), 2332858416653924.
- Bassok, D., Gibbs, C. R., & Latham, S. (2015). *Do the effects of early childhood interventions systematically fade? Exploring variation in the persistence of preschool effects*. Working paper.
- Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade?. *AERA Open, 2*(1), 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*, 7-34.
- Boston Public Schools. (2017). *Focus on K2*. Retrieved May 30, 2017 from <https://sites.google.com/bostonpublicschools.org/earlychildhood/focus-on-k2?authuser=0>.
- Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy indicators from the middle of kindergarten to second grade. *Journal of Special Education*, *42*, 209-226.
- Cascio, E. U., & Schanzenbach, D. W. (2013). *The impacts of expanding access to high-quality preschool education* (No. w19735). National Bureau of Economic Research.
- Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2017). *Cradle to Kindergarten: A new plan to combat inequality*. New York, NY: Russell Sage.
- Clements, D. H., & Sarama, J. (2007a). *SRA Real Math, PreK-Building Blocks*. Columbus, OH: SRA/McGraw-Hill.
- Clements, D. H., & Sarama, J. (2007b). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education*, *38*, 136-163.
- Clements, D. H., Sarama, J. H., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). 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*, 127-166.

- 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.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for Development of Educational Achievement.
- Good, R.H., Kaminski, R.A., Shinn, M., Bratten, J., Shinn, M., Laimon, D., Smith, S., & Flindt, N. (2004). *Technical Adequacy of DIBELS: Results of the Early Childhood Research Institute on measuring growth and development* (Technical Report, No. 7). Eugene, OR: University of Oregon.
- Good, R.H., Kaminski, R.A., Dewey, E.N., Wallin, J., Powell-Smith, K.A., & Latimer, R.J. (2011). *DIBELS Next Technical Manual*. Retrieved July 15, 2013 from http://oregonrti.org/files/u11/DIBELSNext_TechnicalManual.pdf.
- Gormley, W. T., Phillips, D., & Anderson, S. (2016). The effects of Tulsa's Pre-K Program on middle school student performance. *Journal of Policy Analysis and Management, 37*, 63-87.
- Hagan-Burke, S., Burke, M. D., & Crowder, C. (2006). The convergent validity of the Dynamic Indicators of Basic Early Literacy Skills and the test of word reading efficiency for the beginning of first grade. *Assessment for Effective Intervention, 31*, 1-15.
- Imbens, G.W., & Wooldridge, J.M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*, 5–86. doi: 10.1257/jel.47.1.5
- Jung, K., Barnett, W. S., Hustedt, J. T., & Francis, J. (2013). *Longitudinal effects of the Arkansas Better Chance Program: Findings from first grade through fourth grade*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University. Retrieved from

[http://nieer.org/wp-](http://nieer.org/wp-content/uploads/2016/08/Arkansas20Longitudinal20Report20May2013n.pdf)

[content/uploads/2016/08/Arkansas20Longitudinal20Report20May2013n.pdf](http://nieer.org/wp-content/uploads/2016/08/Arkansas20Longitudinal20Report20May2013n.pdf)

Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. *Journal of Policy Analysis and Management*, 33(1), 162-187.

Li, W., Leak, J., Duncan, G. J., Magnuson, K., Schindler, H., & Yoshikawa, H. (2016). *Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program*. Working Paper.

Lipsey, M. W., Farran, D.C., & Hofer, K. G., (2015). *A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade (Research Report)*. Nashville, TN: Vanderbilt University, Peabody Research Institute.

Massachusetts Department of Elementary and Secondary Education. (2016). *Working with 2015 Massachusetts assessment data: Advisory from the Office of Planning and Research and the Office of Student Assessment Services*. Malden, MA: Massachusetts Department of Elementary and Secondary Education.

McCormick, M., Hsueh, J., Weiland, C., & Banger, M. (2017). *The Challenge of Sustaining Preschool Impacts: Introducing ExCEL P-3, a Study from the Expanding Children's Early Learning Network*. New York: NY: MDRC Policy Brief.

Murnane, R., & Willett, J. (2010). *Method matters: Improving causal inference in educational research*. New York, NY: Oxford University Press.

PRESCHOOL LITERACY CONVERGENCE TIMING

- Phillips, D. A., Anderson, S., & Gormley Jr, W. T. (2018). The dog that didn't bark: Preschool education and middle-school attitudes in Tulsa. *Applied Developmental Science, 22*(3), 200-216.
- Phillips, D., Lipsey, M., Dodge, K.A., Haskins, R., Bassok, D., Burchinal, M.R., Duncan, G.J., Dynarski, M., Magnuson, K.A., & Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects*. Washington, DC: Brookings Institution. https://www.brookings.edu/wp-content/uploads/2017/04/consensus-statement_final.pdf
- Puma, M., S. H. Bell, R. Cook, C. Heid, Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third grade follow-up to the Head Start Impact Study Final Report*, OPRE Report # 2012-45, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Schickedanz, J., & D. Dickinson. (2005). *Opening the World of Learning*. Iowa City, IA: Pearson Publishing.
- Snow, C. E., & Matthews, T. J. (2016). Reading and language in the early grades. *The Future of Children, 26*(2), 57-74.
- Speece, D.L, Mills, C., Ritchey, K.D., & Hillman, E. (2003). Initial evidence that letter fluency tasks are valid indicators of reading skill. *Journal of Special Education, 36*, 223-233.
- Stipek, D., Clements, D., Coburn, C., Franke, M., & Farran, D. (2017).PK-3: What does it mean for instruction? *Social Policy Report, 30*(2).
- Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2019). *The effects of enrolling in oversubscribed prekindergarten programs through third grade*. Under review.

PRESCHOOL LITERACY CONVERGENCE TIMING

Weiland, C. & Yoshikawa, H. (2013). The impacts of an urban public prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills:

Evidence from Boston. *Child Development, 84*(6), 2112-2130.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Itasca, IL: Riverside Publishing.

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter?. *The Future of Children, 26*(2), 21-35.

Tables and Figures

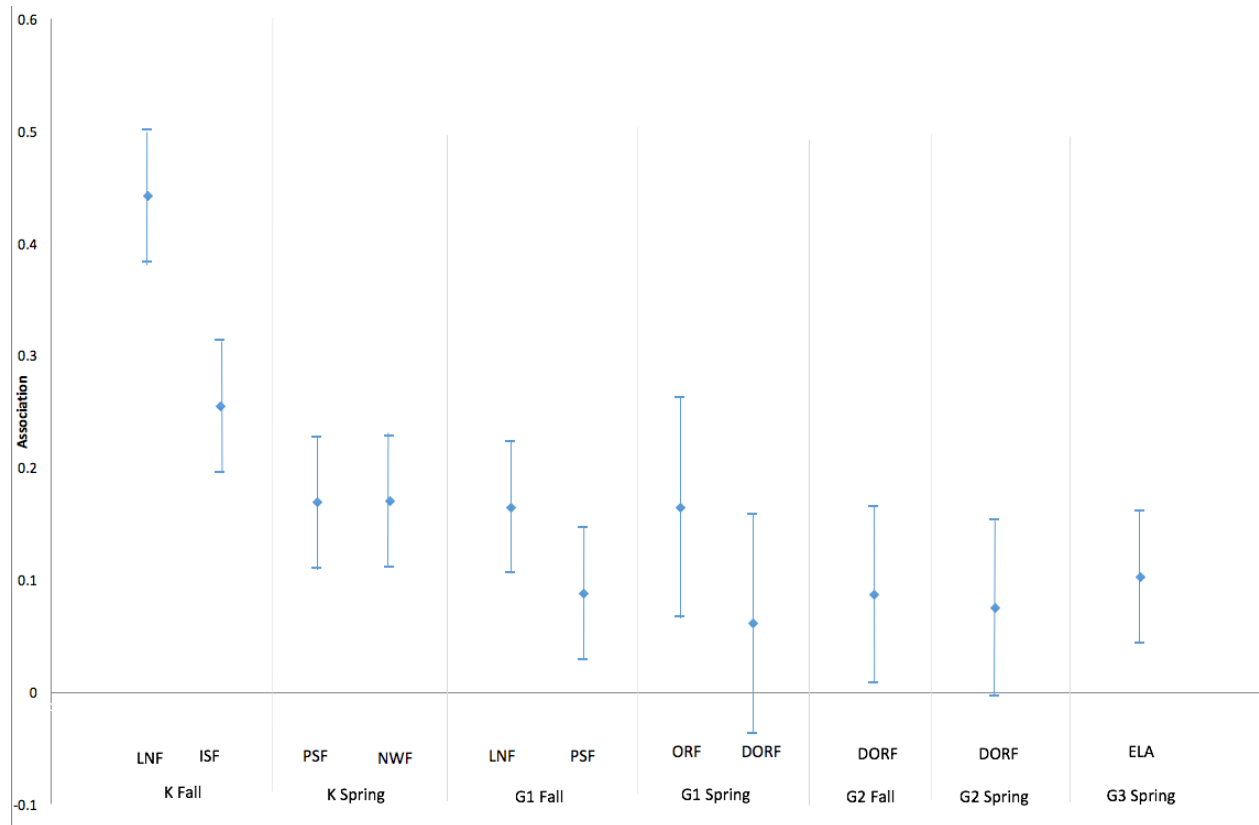
Table 1: Background characteristics of Boston prekindergarten enrollees versus non-enrollees

	Did Not Enroll	Enrolled	Difference
Age (PK year)	4.52	4.53	0.01 (0.09)
Eligible for FRPL	77.03	69.90	-7.13*** (1.31)
% Male	51.48	50.07	-1.41 (1.59)
% Black	26.87	28.34	-1.47 (1.30)
% Hispanic	55.60	42.23	-13.37*** (1.48)
% Asian	4.61	8.86	4.25*** (0.78)
% Mixed/Other	2.35	2.72	0.37 (0.51)
Home Language- % Spanish	32.09	30.85	-1.24 (1.35)
Home Language- % Other	19.36	21.65	2.29 (1.25)
Country of Origin- % USA	92.80	95.13	2.33** (0.72)
K1 eligible year 2010	53.47	49.45	-3.55* (1.59)

Note: Means were estimated using OLS models with school fixed effects. Statistical significance levels are indicated as: ***=.1 percent ** = 1 percent; * = 5 percent. Standard errors in parentheses. N for full sample was ~4971 (N enrollees=3551, N non-enrollees=1420).

PRESCHOOL LITERACY CONVERGENCE TIMING

Figure 1: Associations between BPS prekindergarten enrollment and children’s K-2 literacy scores and grade 3 standardized test scores



Note: LNF=Letter Naming Fluency; ISF=Initial Sound Fluency; PSF=Phoneme Segmentation Fluency; ORF=Oral Reading Fluency (DIBELS 6th Edition); DORF=Oral Reading Fluency (DIBELS NEXT); ELA=MCAS/PARCC English Language Arts; Math= MCAS/PARCC Math. Grade 1 Spring models were fit separately for cohorts 3 and 4 because the ORF and DORF subtests are not equatable (cohort 3 took the ORF and cohort 4, the DORF in grade 1). We standardized each student’s theta score on the mean and standard deviation of all third graders within BPS taking the given exam in that year. ELA models were restricted to children who had at least one non-missing DIBELS subtest score. Statistical significance levels are indicated as: ***=.1 percent ** = 1 percent; * = 5 percent.