Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: http://www.tandfonline.com/loi/uree20

Does Early Mathematics Intervention Change the Processes Underlying Children's Learning?

Tyler W. Watts, Douglas H. Clements, Julie Sarama, Christopher B. Wolfe, Mary Elaine Spitler & Drew H. Bailey

Does Early Mathematics Intervention Change the Processes Underlying Children's Learning?

Tyler W. Watts[a], Douglas H. Clements[b], Julie Sarama[b], Christopher B. Wolfe[c], Mary Elaine Spitler[d], and Drew H. Bailey[a]

[a]University of California, Irvine, Irvine, California, USA;

[b]University of Denver, Denver, Colorado, USA;

[c]Saint Leo University, St. Leo, Florida, USA;

[d]University of Buffalo, SUNY, Buffalo, New York, USA

CONTACT Drew H. Bailey dhbailey@uci.edu School of Education, 3200 Education, University of California, Irvine, Irvine, CA 92697-5500, USA.

Abstract

Early educational intervention effects typically fade in the years following treatment, and few studies have investigated why achievement impacts diminish over time. The current study tested the effects of a preschool mathematics intervention on two aspects of children's mathematical development. We tested for separate effects of the intervention on "state" (occasion-specific) and "trait" (relatively stable) variability in mathematics achievement. Results indicated that, although the treatment had a large impact on state mathematics, the treatment had no effect on trait mathematics, or the aspect of mathematics achievement that influences stable individual differences in mathematics achievement over time. Results did suggest, however, that the intervention could affect the underlying processes in children's mathematical development by inducing more transfer of knowledge immediately following the intervention for students in the treated group.

 ACCEPTED MANUSCRIPT

Keywords

early childhood; interventions; mathematics

 ACCEPTED MANUSCRIPT

Well-controlled correlational studies show a strong, and persistent, relation between children's early mathematics skills and their later achievement (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Bailey, Siegler, & Geary, 2014a; Byrnes & Wasik, 2009; Claessens & Engel, 2013; Duncan et al., 2007; Geary, Hoard, Nugent, & Bailey, 2013; Jordan et al., 2009; Watts, Duncan, Siegler, & Davis-Kean, 2014; Watts et al., 2015). Theoretically, the link between early and later mathematics achievement is thought to be straightforward, as earlier knowledge in mathematics is necessary for building later knowledge. For example, understanding single-digit arithmetic and place value is essential for gaining competence in multi-digit arithmetic (Rittle-Johnson & Siegler, 1998). Moreover, the idea that "skill begets skill" is a hallmark of current theories of development (e.g., Cunha & Heckman, 2008).

Viewed through this theoretical perspective, previous correlational findings imply that if interventions can boost the early mathematics achievement of at-risk children, the effects of such efforts may last for many years. Indeed, multiple reports and position papers from various educational advocacy groups have supported this notion, as they have called for investments in early math instruction with the hope of setting children on a higher-achieving trajectory throughout school (National Council of Teachers of Mathematics, 2000; National Mathematics Advisory Panel, 2008; National Research Council, 2006). These conclusions are primarily based on rigorously conducted correlational studies on the longitudinal links between early and later mathematics achievement, many of which used a broad and robust set of control variables to approximate causal effects (e.g., Claessens et al., 2009; Duncan et al., 2007; Geary et al., 2013; Watts et al., 2014). For example, using a nationally representative dataset, Duncan and

 ACCEPTED MANUSCRIPT

colleagues found that even when controlling for approximately 80 variables that included measures of general and domain-specific cognitive skills, family background characteristics, and socio-emotional skills, early mathematics achievement was the strongest predictor of later mathematics achievement. Further, this result replicated across 5 other large-scale datasets.

However, recent evidence suggests that the correlational estimates of the effects of early mathematical skills on later achievement may overstate the long-run benefits of early intervention. Bailey, Watts, Littlefield, and Geary (2014b) observed that the long-term predictive strength of much later mathematics achievement by early mathematical skills does not strongly diminish as the distance in time between measures of early- and later-mathematics achievement increases. For example, in a large, diverse U.S. sample, the correlation between children's mathematics achievement scores in grades 1 and 3 was .72. This correlation hardly decreased to .66 between grade 1 and age 15, suggesting substantial stability in the correlation between early and later measures of mathematics achievement. This contrasts with findings from studies of the effects of early childhood interventions on children's later skills, which typically show clearly diminishing treatment effects over time (e.g. Puma et al., 2012; for review see Bailey, Duncan, Odgers, & Yu, 2015). Indeed, a recent meta-analysis of early childhood interventions found an average initial treatment effect across 117 studies of approximately .27 standard deviations, but this average effect faded completely by 2-3 years following the end of treatment (Leak et al., 2010). Similarly, a meta-analysis of early phonological awareness training programs, which have been thought to teach critical skills for the development of early reading, found that large initial effects typically faded by over 60% when follow-up assessments were collected (e.g., Bus & van IJzendoorn, 1999).

Bailey and colleagues (2014b) hypothesized that this apparent discrepancy between correlational and experimental findings stems from the implausibility of fully controlling for the stable factors affecting children's mathematics learning in correlational studies. These factors are likely numerous, include both environmental and personal characteristics (e.g., low-resource communities, ability, motivation, parental support), and are difficult to perfectly measure. If these unmeasured, stable, factors consistently contribute to individual differences in mathematics achievement, then even correlational studies that include a large set of control variables are still likely to yield upwardly biased estimates of the effect of early mathematics achievement on later mathematics achievement.

Bailey and colleagues (2014b) investigated whether unaccounted factors explain the variance in long-run mathematics achievement by partitioning the variance in repeated math measures into two components: "state" and "trait" variability (Steyer, 1987). In this model, "trait variation" captures the aspects of long-run mathematics achievement that are stable over time. Conceptually, trait-mathematics can be thought of as a collection of factors, both personal and environmental, that consistently influence a given student's mathematics achievement throughout their development. Such factors might include domains of personality (e.g., conscientiousness), cognition (e.g., working memory capacity), and environments (e.g., poverty) that show some degree of inter-individual stability during development. Statistically, Bailey and colleagues modeled trait-mathematics by estimating a single factor that accounted for all of the stable variation in 4 consecutive measures of mathematics achievement taken over time.

In contrast, "state variation" is comprised of within-individual variation in individual differences in children's mathematics achievement, and effects of earlier states on later states imply a unique influence of previous mathematics achievement on later mathematics achievement. More formally, state effects can be thought of as the impact of changes in an early mathematics test score on a later test score, which approximates the causal interpretation of the early- to later-mathematics achievement effects reported by correlational studies (e.g., Bailey et al., 2104a; Claessens et al., 2009; Watts et al., 2014). Statistically, Bailey and colleagues (2014b) modeled state effects by simply regressing a later measure of mathematics achievement on the immediately preceding measure, controlling for stable, between-individual differences (i.e., those comprising trait mathematics).

The "state-trait" model of mathematics achievement helps elucidate the specific processes that lie behind a correlation between an early and later measure of mathematics achievement. If the knowledge learned during an earlier period has a substantial causal impact on later mathematics achievement, then the state-trait model would show large state-mathematics effects, and smaller trait-mathematics effects. Conversely, if the strong correlation between two mathematics measures is merely a product of stable individual differences in mathematics achievement over time (which are likely caused by unmeasured factors operating at both the personal and environmental level) then trait-mathematics effects would be much larger than state-mathematics effects.

In the results reported by Bailey and colleagues (2014b), the trait factor explained much more of the variation in children's mathematics achievement than state effects. Further, Bailey and colleagues also examined whether substantial variation in trait-mathematics could be accounted for by commonly used statistical controls. They found that measures of cognitive abilities, SES, and reading achievement were all correlated with trait-mathematics, and these controls explained approximately 60% of the variation in trait-math. Even after including these controls, trait-math continued to explain far more of the variation in children's mathematics achievement than did previous mathematics achievement. These results were consistent with the hypothesis that skill begets skill, but also suggested that regression analyses of non-experimental studies overestimate the size of these effects because of their inability to completely control for stable underlying factors. In other words, interpreting a correlation between two measures of mathematics achievement (even one conditional on commonly used statistical controls) as the causal effect of time-1 achievement on time-2 fails to take into account that any two observed measures of math ability are likely to be correlated because of other factors maintaining the stability of children's mathematics achievement over time. After adjusting for latent factors influencing this stability, the model predicts that a change in children's mathematics achievement at a single timepoint early in schooling is unlikely to substantially affect their long-run math achievement.

Although their study may help explain discrepant findings between correlational and experimental studies of long-run mathematics achievement, Bailey and colleagues' (2014b) models were only tested using non-experimental data. Consequently, they could not observe whether mathematics interventions that fall outside of the range of environments children usually experience in school could change the structure of the model. Further, these non-experimental data did not allow for an investigation of whether high-quality instruction in children's early mathematical knowledge could boost children's scores on the latent trait factor. For example, if some children learn a deep understanding of the subtraction procedure early, the children who never fully learn this concept could be at a disadvantage across their mathematical development.

If this were true, then children's early mathematics knowledge per se constitutes some of the variance in children's trait mathematics achievement, and high-quality early mathematics interventions could affect children's long-run ability to learn mathematics via changes in these stable factors.

Bailey and colleagues (2014b) concluded that this was unlikely due to the limited amount of overlap in the items accounting for variance in children's scores at different times during development. On the other hand, the correlations between some basic mathematical or numerical skills and children's mathematics achievement remain substantial long after children are expected to have learned these skills. For example, children's ability to accurately place numbers on a number line correlated .54 with their mathematics achievement even in third grade (Jordan et al., 2014). Further, though different meta-analyses have reached varying conclusions on whether the association between mathematics achievement and non-symbolic number acuity differs by age (Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014), it is clear that individual differences in these measures remain correlated even in adults. Therefore, perhaps these skills underlie children's general mathematics achievement throughout their development. In summary, it is not clear whether trait mathematics is influenced by children's early mathematics knowledge per se. However, this is a testable hypothesis: if an intervention targeted at children's early mathematics skills boosts trait mathematics, estimated from children's subsequent mathematics achievement scores, this suggests that early mathematics knowledge is a component of trait mathematics. If the effect of an early mathematics intervention on children's mathematics achievement operates only via a state-dependent process, this suggests that early

 ACCEPTED MANUSCRIPT

mathematics knowledge is not a component of trait mathematics. Testing these hypotheses is a primary goal of the current study.

Even if specific early mathematics skills do not comprise a significant component of children's trait mathematics, early mathematics interventions could induce other changes likely to result in long-term benefits. Specifically, high-quality, conceptually rich early mathematics interventions could affect the extent to which children utilize their previous knowledge to learn new information. Empirical work suggests that conceptual and procedural knowledge in mathematics develop through an iterative process, but conceptual knowledge is thought to lead to a deeper understanding of mathematics and more transfer of learning (Clements & Sarama, 2014; Hiebert & Gouws, 2007; Perkins & Salomon, 1992; Rittle-Johnson, Siegler, & Alibali, 2001; Rittle-Johnson, 2006; Rittle-Johnson & Schneider, in press)**.** Thus, an intervention that emphasizes conceptual understanding of mathematics may increase children's ability to connect previous knowledge to new learning, consequently leading to higher long-run achievement outcomes. In this case, the intervention would not impact trait-mathematics, but would instead alter the structure of the model reported by Bailey and colleagues (2014b). More specifically, this would result in larger state-mathematics effects, and smaller trait-mathematics effects, because children's later achievement would be more directly influenced by the knowledge gained in an earlier period.

Likewise, this raises the possibility that high-quality instruction can make low-achieving children learn mathematics more like higher-achieving children. Children with low cognitive performance tend to show less differentiation across cognitive skills, especially crystallized intelligence (Kane, Oakland, & Brand, 2006; Reynolds & Keith, 2007; Tucker-Drob, 2009). One possible partial explanation for this lack of differentiation is the inability to connect previous knowledge to new learning. In the context of children's mathematical development, perhaps higher-achieving children are better able to use the mathematical knowledge gained at a lower level to learn new mathematics at later grades. This process, referred to as transfer (see Barnett & Ceci, 2002; Rittle-Johnson, 2006), could be encouraged by an intervention that provides clear connections among foundational concepts, and between concepts and procedures.

Intervention-induced changes in transfer of learning may have implications for practice. Although the control group from a successful early mathematics intervention would still be predicted to catch up to the treatment group eventually if the intervention targeted skills likely to be learned by the control students, this would not necessarily imply that early interventions could not boost children's long-term learning outcomes. An intervention that heavily promoted conceptual understanding and successfully induced transfer of learning in the period immediately following the early intervention could help treated children maintain a higher-level achievement trajectory. However, successful transfer of learning would still depend on high-quality instruction in the years following the intervention, as transfer would be unlikely to occur without adequate exposure to new mathematical concepts and procedures. Thus, this would raise the possibility that with extended high-quality mathematics intervention, children at risk for persistently low mathematics achievement might have substantially improved long-term mathematics outcomes.

Current Study

To test whether improvements in early mathematical skills can impact the stable characteristics that influence mathematics achievement over time, and whether a conceptually-rich early mathematics intervention can affect children's transfer of learning, we analyzed the results of an experimental dataset. We investigated how the *Building Blocks* scale-up intervention (described in detail in method section), impacted state- and trait-mathematics achievement in a sample of children from predominantly low-income families. The *Building Blocks* curriculum was designed to promote both conceptual and procedural knowledge of mathematics through activities that helped children "find the mathematics in, and develop mathematics from, [their] experiences and interests" (Clements & Sarama, 2007). Following Bailey and colleagues (2014b), we predicted that the intervention would have little or no effect on trait-mathematics achievement, but would have a sizable effect on occasion-specific mathematics achievement. Although the intervention was found to have large effects on children's mathematics achievement at the end of pre-school (Hedges g = .71), subsequent analyses found substantial fadeout of the average treatment effect after two years, especially in the absence of post-treatment instructional support (Clements et al., 2011; 2013).

We also investigated whether the amount of transfer of learning, estimated by the strength of the state paths, was larger for higher-achieving children and children assigned to the treatment condition than for lower-achieving children and children in the control condition. Although we expect the variance in mathematics scores to be primarily explained by trait-mathematics across all groups, we hypothesized that the state and trait effects could differ between groups in key ways. First, we hypothesize that the treatment may have produced more transfer of knowledge, and larger state paths, for students in the treatment group when compared with control students. Second, we expect that the state-trait model for the treatment group more closely resembles the state-trait model for higher-achieving students, as the model for higher-achieving students should also include larger state effects and smaller trait paths.

Method

Additional information regarding the study data can be found in the supporting information document. Here, we provide a description of key study components.

Data

Data were drawn from the TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development) evaluation study (Clements et al., 2011; 2013; Sarama et al., 2012), which evaluated the scale-up of the *Building Blocks* preschool curriculum. Study operators contacted 10 urban school districts in low-income areas of two northeastern U.S. cities; two met the criteria of (a) serving ethnically diverse populations who live mainly in poverty, (b) having a large number of prekindergarten (pre-K) classrooms within elementary schools, with self-contained feeder patterns (a history of preschoolers continuing their education in that school), (c) willingness to have schools randomly assigned to treatments. Participating schools (n = 42) were then grouped into 8 blocks based on their state-administered achievement test scores, and schools were randomly assigned within block to one of three conditions: 1) preschool curriculum intervention; 2) preschool curriculum intervention with follow-through; 3) control condition (business as usual).

Schools assigned to either of the preschool curriculum intervention conditions (i.e., conditions 1 and 2) implemented the *Building Blocks* curriculum in their preschool classrooms. Implementation was primarily carried out through 13 professional development (PD) sessions offered to preschool teachers working in treated schools. These PD sessions emphasized both mathematical content and pedagogical strategies for teaching conceptually focused mathematics to 4-year old children. Schools assigned to the treatment with follow-through condition (condition 2) received additional PD sessions for kindergarten and first grade teachers, and these sessions were designed to help early-grade teachers build on the mathematics that students learned during the *Building Blocks* program in preschool. Unfortunately, analytic limitations (described in the supplementary information) prevented us from using this treatment condition in our study. Thus, the current analysis only considered students attending schools assigned to either the preschool curriculum intervention group (condition 1) or control group (condition 3).

The *Building Blocks* preschool curriculum focused on building numeric/quantitative and geometric/spatial knowledge in preschool, and the curriculum was designed to take approximately 15 to 30 minutes of each school day. Curriculum materials focused on a wide range of mathematical topics thought to be foundational to later mathematics achievement, including counting, operations, geometry, measurement and patterning, and processes emphasized including communicating, reasoning, and problem solving within each mathematical domain. The curriculum was organized into distinct learning trajectories, and students progressed through each trajectory by participating in structured whole and small group activities, and through using the complementary *Building Blocks* software. The learning trajectories would typically begin with non-symbolic, heavily conceptual, content, designed to help students build foundational knowledge relating to a certain mathematical topic (e.g. composition of number). Each learning trajectory would then provide practice in skills while also building to incorporate more advanced strategies and procedures for problem solving.

Implementation of the curriculum was assessed with the *Building Blocks Fidelity instrument*. This includes 52 items, responses to most of which are on 5-point Likert scales (-2 to +2), organized according to the curriculum's components, whole-group activities, small-group activities, learning centers, and computer centers. Visits were approximately one hour in duration to allow observing all components. The resulting average of 0.77 in fall and 0.86 in spring (that is, near "Agree") was considered adequate fidelity (Clements et al., 2011). For a full description of the curriculum and study procedures see Clements & Sarama (2007) and Clements et al. (2011).

A random sample of 1375 preschool students from the 42 study schools were recruited for student-level data collection. From within the 30 schools included in our analysis (i.e., schools assigned to either condition 1 or 3), 880 students were recruited and consented for data collection. These students were tracked from the beginning of preschool through the end elementary school. Student mathematics achievement was first measured at the beginning of the preschool year (pretest), and students were again tested at the end of preschool (posttest), and in the springs of kindergarten and first grade, and fall of fourth grade.

We restricted our analytic sample to only include students that had at least one non-missing mathematics achievement measure (n = 834). Table 1 presents sample characteristics for the current study. Descriptive statistics for baseline characteristics are displayed for both the treatment (n = 465) and the control (n = 378) groups. Baseline equivalence on all observable attributes between the treatment and control groups was found. The sample was approximately 50% African American, 22% Hispanic, and 85% of students qualified for free or reduced price lunch. Further, approximately 75% of the sample continued on at the same school after completing preschool, and this did not differ between the treatment and control groups ($p = .83$).

Measures

Mathematics achievement.

The TRIAD study used the Research-based Early Math Assessment (REMA: Clements, Sarama, & Liu, 2008; Clements, Sarama, & Wolfe, 2011) to measure the mathematics achievement of study participants at the beginning and end of preschool, as well as during follow-up assessments in the spring of kindergarten and first grade. The REMA, designed to measure the mathematics achievement of children from ages 3-8, assessed a wide variety of mathematical concepts and procedures including counting, arithmetic, geometry and measurement. Study administrators delivered the test through a one-on-one interviews with participants, and the interviews were taped and subsequently coded for strategy use and correctness. Many items involved the use of manipulatives, and study administrators structured the test to become more difficult with each question. A stop-rule dictated that the test would end after a student incorrectly answered four consecutive questions.

REMA designers extensively validated the measure through repeated testing in three different samples, which produced an overall reliability of .94 (Clements et al., 2008). Further, Clements and colleagues (2008) reported that the REMA had a .89 correlation with the *Applied Problems* subtest of the Woodcock Johnson III (WJ-R). Study administrators scored the REMA using a Rasch-IRT model that accounted for correctness, strategy use, and item difficulty. The Rasch-IRT scores were age-normed to a mean of 0 and standard deviation of 1, with a mean of 0 indicating the average score for a typically-achieving first grader. As Table 1 shows, at preschool entry, students in the current sample scored approximately 3.2 standard deviations below 0.

In the fall of fourth grade, students took the TEAM 3-5, a measure similar to the REMA designed to assess the mathematics achievement of students in third, fourth, and fifth grade. The TEAM 3-5 was a multiple-choice test, and it did not include a stop rule, as students were asked to answer every item of the exam. Study administrators also scored the TEAM 3-5 using a Rasch-IRT model, and the measure assessed students on their knowledge of fractions, division and multiplication, and geometry (among other topics). For the current sample, the TEAM 3-5 had good internal reliability (Cronbach's α = .91), and it was found to have strong correlations with state achievement tests in New York ($r(351) = .82$, $p < .001$) and Massachusetts ($r(110) = .76$, $p < .001$).

Fidelity of administration.

Explicit protocols and procedures were used for administration, videotaping, coding, and scoring and for staff training on all dependent measures in pre-K to first grade. After practicing administering the instrument with classrooms and children not involved in the study, assessors had to be certified on their respective instruments by submitting video recordings of two consecutive sessions of errorless administration; thereafter, about every tenth administration (randomly selected---all assessments were video recorded) was checked for accurate administration for the child outcome measures. Each child outcome item was coded by two trained coders for accuracy and, when relevant, for solution strategy. Any discrepancies were resolved via consultation with the senior researchers. Continuous coder calibration mitigated against drift.

Covariates.

The TRIAD study used two letter-recognition exams to measure pre-reading skills at the end of the preschool: the PALS-PreK (Invernizzi, Sullivan, Swank & Meier, 2004) or the MCLASS:CIRCLE (Landry, 2007). Study schools administered one of the two tests, and both tests involved naming letters or words. We standardized scores such that a given student was only compared with other students who took the same exam, and both tests have strong reliability (MCLASS:CIRCLE α = .90 - .93, Landry, 2007; PALS-PreK, α = .75- .93, Invernizzi et al., 2004).

The TRIAD study measured language skills at the beginning of the kindergarten year with the Renfrew Bus Story- North American Edition (RBS: Glasgow & Cowley, 1994). The exam asked students to learn a story then retell the story using their own words, and study administrators coded the child's recounting of the story for indicators of language skills. In the current sample, the measure was found to have a reliability of .79 (Sarama et al., 2012). District offices reported information regarding whether the child qualified for free or reduced price lunch, and parents reported the mother's education through a parent survey administered during the preschool year.

Analysis

The current study examines the extent to which a preschool mathematics curricula impacts two separate types of mathematics achievement: trait- and state-mathematics [see Kenny & Zautra (1995) and Steyer (1987) for thorough descriptions of the state-trait model]. We define trait-mathematics as a group of characteristics that exert a stable influence on a child's mathematics ability throughout development. State-mathematics describes the time-varying aspects of long-run mathematics achievement, which could be affected by changes in instruction or specific mathematical knowledge. Following the analytic design set forth by Bailey et al. (2014b), we modeled state- and trait-mathematics by using a latent-factor design in which end-of-school-year mathematics achievement was measured four times: 1) preschool (end of treatment), 2) kindergarten, 3) first grade, and 4) fourth grade. To account for measurement error, these four measures were used to create four factors of mathematics achievement (subsequently called state factors), and each factor was regressed on the corresponding achievement measure with the path set to the square root of the reliability for that measure.

Trait-mathematics was operationalized by modeling a single-factor that accounted for the variance in mathematics achievement that was stable at each of the four mathematics measurement points. The standardized factor loadings of the latent-trait on each of the four state factors can be interpreted as the trait-mathematics effects. To model state-mathematics effects, each of the four state factors was regressed on the immediately preceding factor. The resulting standardized coefficients represent the various state-mathematics effects. The treatment impacts on trait- and state-mathematics were estimated by regressing the latent-trait factor and first state factor on the treatment assignment dummy, and model paths were weighted by the baseline (preschool entry) mathematics measure.

We also accounted for the trait variance explained by control variables commonly used in analyses of long-run mathematics achievement: kindergarten measures of language and literacy skills, mother's education, and whether the student qualified for free or reduced price lunch. The inclusion of control variables tests whether stability in individual differences in mathematics achievement can be explained by commonly-used controls. Bailey and colleagues (2014b) reported that simple control variables, including domain-general cognitive ability, did not account for all of the variation in trait-mathematics. By including these controls, we also test whether these variables, or constructs correlated with these variables, could be plausible contributors to trait-mathematics. Such a relationship would be observed through statistically significant paths from these variables to trait-mathematics, and good fit indices for the models that assume that these constructs contribute to between-individual differences in mathematics. Further, including these controls allowed us to test the extent to which these variables account (or fail to account) for variation in trait-mathematics. The results of this test can be useful for future research, as these models allow us to judge how much unobserved variation in trait-mathematics might be accounted for by commonly-used control variables. Notably, the purpose of including these controls was not to create a complete causal model of children's trait mathematics, which would ideally require experimental longitudinal datasets in which plausible contributors were manipulated.

Because the *Building Blocks* curriculum has been shown to also affect language skills (Sarama et al., 2012), we also regressed kindergarten language skills on treatment status. Finally, we estimated the state-trait model in four key subgroups of interest: above-median achievers, below-median achievers, treatment group, and control group.

Results

Table 2 contains correlations among key analysis variables. In the supplementary file, we present the same correlations for our key subgroups: treatment and control, and above-median and below-median achieving students. As expected, these correlations suggest a combination of both state and trait effects; mathematics achievement exams were highly correlated across all measures (average $r = .65$), but the highest correlations occurred between measures assessed at the closest intervals.

Missing Data and Modeling Specifications.

All models presented were estimated in Mplus 7.0, with maximum likelihood robust estimator accounting for missing data. Data were missing due to non-response on measures of language (20%), literacy (19%), free or reduced price lunch (16%), and mother's education (30%). Due to the longitudinal nature of the data, study attrition also accounted for missing test scores at kindergarten (7%), first grade (14%) and fourth grade (44%). The high rate of attrition by fourth grade was due to students moving schools and losing contact with TRIAD study administrators over time, as no follow-up interview occurred between the spring of first grade and the fall of fourth grade. To test whether attrition related to treatment assignment, we regressed a dummy variable indicating whether the student was still present in the study by the fall of fourth grade on treatment status. We found that the attrition rate differed between the treatment and control groups by a statistically insignificant 1% ($p = .79$).

To test whether missing on any of these variables was related to treatment assignment, which could potentially bias results, we ran a series of regressions in which we regressed a dummy variable indicating whether a student was missing on each of the aforementioned variables, respectively, on the treatment assignment indicator. These regressions produced no statistically significant results, suggesting that the missing data patterns in the current study were not related to treatment assignment.

Further, because the treatment was administered at the school level, and our analysis depends on a study design in which students were clustered within schools, we tested models in

ACCEPTED MANUSCRIPT

which standard-errors were adjusted to account for school-level clustering. This adjustment was calculated using the “type = complex” and “cluster” commands in Mplus 7.0. The resulting models included only marginally larger standard errors than the models presented here, and the process did not alter the statistical significance of any coefficient or factor loading.

Treatment Impact Estimates

Our primary results are displayed in Figure 1. In this model ($\chi^2(18) = 67.25$, $p < .001$; RMSEA = .06; CFI = .98), we tested the effects of the preschool mathematics intervention on both state and trait mathematics achievement. Similar to Bailey et al. (2014b), we found much larger trait effects than state effects at each time-point assessed. Standardized trait loadings were statistically significant ($p < .001$), and ranged from .76 to .94. Conversely, state effects were only statistically significantly different from zero at time-points 2 and 4, and ranged from .04 to .25. As with Bailey et al. (2014b), we also tested the extent to which common control variables account for the variance in trait-math. Kindergarten measures of language and literacy skills, as well as mother’s education level and whether free or reduced price lunch accounted for approximately 44% of the variance in trait mathematics. Among the controls tested, language ($\beta = .35$, $p < .001$) and literacy ($\beta = .46$, $p < .001$) skills were the strongest correlates of trait-mathematics, whereas free or reduced price lunch status and mother’s education did not explain unique variation in trait-math.

We found that at each time-point, relatively stable characteristics influenced children’s performance on the next mathematics achievement measure more than time-varying characteristics from the previous time-point. Here, we compare the sizes of state- and trait-effects on correlations among measures of children’s mathematics achievement at the second

 ACCEPTED MANUSCRIPT

wave of measurement, as in Bailey et al. (2014): the proportion of variance in Time 2 mathematics achievement that was explained by direct trait effects was .58. (trait loading × trait variance × trait loading, or .76 × 1 × .76). The proportion of variance in Time 2 mathematics achievement that was explained by Time 1 state mathematics achievement (including the indirect effect of the latent trait via Time 1 state mathematics achievement) was .06 (state effect^2, or $.25^2$). These values were similar to those reported by Bailey et al. (2014) in two different datasets: in those data, the proportions of variance accounted for at the second time-point were .55 and .36 by trait factors and .07 and .16 by state factors.

However, we did find statistically significant and positive state effects between the end of preschool and kindergarten (β = .25, $p < .001$), and first grade and fourth grade (β = .13, $p < .01$). Interestingly, we did not find a significant state effect between first-grade and kindergarten mathematics achievement (β = .04, p = .24). Yet, this effect was not particularly precisely estimated, as the 95% confidence interval would include effects as large as .08, which would lie within the confidence interval of the statistically significant effect of fourth grade achievement on first grade achievement. Given that the current study and Bailey and colleagues (2014b) found ubiquitous significant state paths across other time periods and samples, our null effect should be interpreted with caution, pending replication.

We found that the *Building Blocks* scale-up intervention had a strong effect on state mathematics achievement, assessed at the end of preschool (β = .53, p < .001), but had no effect on trait mathematics (β = -.04, p = .59). This suggests that an intervention narrowly focused on building preschool mathematical competencies can alter aspects of achievement that change over time (e.g. acquisition of specific math skills), but not the underlying characteristics that exert a relatively stable influence on individual differences in children's achievement throughout development. As reported by Sarama and colleagues (2012), we also found that the intervention had a statistically significant effect on language skills measured during kindergarten (β = .20, p < .01). Although this effect was modest, the impact of *Building Blocks* on language skills indicates that a math curriculum intervention could affect trait math through the promotion of other skills. However, the current study only included one measure of language skills, thus we could not test the state-trait model on language ability. Further, the treatment effect on time-2 mathematics through language in the current model would be only .05 (treatment effect X language effect on trait X time-2 trait loading).

Subgroup Models

To test whether the state-trait model differed for certain groups of students, we estimated the model for four groups: above-median and below-median achieving students, and treatment and control students. We hypothesized that over time, above-median achieving students would make more connections with the mathematical knowledge they had previously learned, and this would lead to lower trait loadings, and larger state effects. Further, we expected that the state-trait model for above-median achieving students may resemble the model for the treatment group, as the treatment emphasized mathematical content thought to promote transfer of learning.

Consistent with our hypotheses, a Satorra-Bentler scaled chi-squared difference test revealed that a model in which the paths for above-median and below-median achieving groups (defined by a median-split on the preschool entry test) were freely-estimated had substantially better model-fit than a model that constrained the paths for both groups to be equal ($\chi^2(23)$ = 110.06, p < .001). Models 1 and 2 of Table 3 present the trait loadings and state paths for above-

 ACCEPTED MANUSCRIPT

median and below-median achieving students, respectively. Across both groups, trait loadings were much larger than state effects. On average, trait effects were over 5 times larger than state effects for below-median achieving students and 2.5 times larger than state effects for above-median achieving students. In the above-median achievement group, state effects were consistently statistically significant and ranged from .25 - .29 (p < .01). For below-median students, state paths were more varied and smaller, as they ranged from a statistically non-significant -.01 to .27 (p < .01). Although trait effects far outweighed state effects for both groups, our models do suggest that the developmental processes may differ for below-median and above-median students, as higher-achieving students showed more evidence of transfer.

To compare the treatment and control group estimates with models estimated for above-median and below-median achieving groups, we first tested whether the state-trait model was significantly different for the treatment and control students. A Satorra-Bentler test again revealed that the model had substantially better fit when treatment and control group paths were freely-estimated ($\chi^2(21) = 187.21$, $p < .001$). Models 3 and 4 of Table 3 display the results of models estimated separately for treatment and control group students. As expected, we observed stronger state paths for treatment group students (ranging from .11 (p < .05) to .30 (p < .01)) than control group students (state paths were non-significant and ranged from -.04 to .10). However, average trait paths in the treatment group were still over four times larger than state paths for treatment students. The largest difference between state paths was found between the first two time-points, and the average state path in the treatment group (.19) was lower than the average state path for above-median achieving students (.28). Thus, we found some indication that the state-trait model for treatment group students more closely resembled the model for higher-achieving students, but these differences were not as heavily pronounced as the differences between the above-median and below-median students, and were largest immediately following the intervention.

Discussion

The current study tested the impact of a high-quality preschool mathematics curriculum on state- and trait-mathematics achievement. As hypothesized, we found a large and significant effect of the treatment on the time-varying component of mathematics achievement, state-mathematics, and we found no treatment effect on trait-mathematics (the underlying factor that consistently contributes to mathematics achievement). However, we did find that the intervention temporarily changed the structure of the state-trait model by increasing state-mathematics effects in the period immediately following the treatment. This indicates that students relied more heavily on their previous mathematics knowledge when learning new material, and suggests that conceptually rich instruction may help students learn more mathematics following a high-quality intervention. Yet, this change in state-mathematics effects was not sustained past kindergarten, implying that more high-quality instruction is needed in the years following an early intervention to meaningfully encourage the persistence of early treatment gains.

The null-finding regarding the treatment impact on trait-mathematics suggests that individual differences in early mathematics knowledge *per se* do not comprise trait-mathematics, as a curriculum shown to strongly boost early knowledge of counting, geometry, measurement, and operations (among other topics) did not affect the latent trait. How might we reconcile this finding with the persistence of correlations between children's basic mathematical or numerical skills and their mathematics achievement throughout development? We hypothesize that, above a certain threshold, children's number line understanding, nonsymbolic magnitude acuity, and basic arithmetic skill do not influence their mathematics achievement as much as they do in younger children. Individual differences may still be associated with domain general cognitive abilities, such as processing speed (Ackerman, 2007), which remain associated with children's mathematics achievement throughout development. However, prior knowledge may increasingly account for variance in children's general mathematics achievement; indeed, that is what the significant state paths in the state-trait model suggest.

This finding does not suggest that such early knowledge is not crucial to learning later mathematics, rather, it suggests that differences in early attainment of this knowledge do not explain long-run differences in achievement. This is probably because a large number of students in the control group eventually learned these early competencies in later grades; for this reason, some have proposed that the term "catch-up" is more appropriate than the term "fade-out" for describing decreasing treatment post-intervention effects on children's academic skills over time (Clements et al., 2013), especially if those later grades emphasized simple or basic mathematics skills (Engel, Claessens, and Finch, 2013; see also Clements & Sarama, 2014).

Further, we cannot rule out whether other kinds of mathematics knowledge, encountered by students later in school, could manifest as trait-mathematics. For example, studies suggest that many students never fully master procedural and conceptual understanding of fractions (e.g. Schneider & Siegler, 2010). Thus, an intervention that teaches fractions understanding to low achieving children might impact trait-mathematics, as this knowledge might not be typically learned by students in the control condition. If this is true, the structure and malleability of trait-mathematics may vary across children's ages, levels of previous knowledge, cultural contexts,

ACCEPTED MANUSCRIPT

and learning experiences. This possibility deserves further investigation, and predicts greater persistence of mathematics intervention on older children.

These findings also raise additional questions about which factors explain stable long-run longitudinal correlations among measures of mathematics achievement. It is likely that a combination of environmental and personal characteristics contribute to these differences, but most studies of mathematics achievement are unlikely to include all the relevant measures required to completely account for these latent processes. Consistent with Bailey and colleagues (2014b), we found that scores on tests from other academic domains - literacy and language exams - predicted trait-mathematics. This suggests that trait-mathematics may be somewhat general to academic domains. The predictive power of these cognitive tests does not rule out the role of environmental factors in shaping trait-mathematics, whether directly or indirectly. Although we found no unique statistical relation between environmental factors and trait-math, this is likely partially due the lack of heterogeneity in the current sample (i.e. students were recruited from similar schools and neighborhoods by design), and does not rule out indirect effects of environments on mathematics via relatively stable personal characteristics. Further, Bailey and colleagues, in a more diverse sample, found that SES predicted variance in trait mathematics above and beyond measures of domain general cognitive abilities and reading achievement.

Our findings contribute to our theoretical understanding of why early academic intervention effects fade. Although the *Building Blocks* intervention focused on knowledge thought to be crucial to long-run achievement trajectories, this boost in knowledge was unable to affect the stable factors affecting individual differences in mathematics achievement over time.

Thus, our findings suggest that early mathematics interventions alone are likely insufficient to produce substantial effects on children's long-term achievement outcomes. However, we do not mean to suggest that all early academic interventions cannot produce long-term effects, and some evidence suggests that interventions complemented by high-quality instruction during the years after initial treatment can produce longer-lasting impacts (e.g., Clements et al., 2013). Consistent with this possibility, we found suggestive evidence that the preschool treatment facilitated more transfer of learning, as trait loadings were slightly smaller, and state effects were slightly larger, for students in the treatment group. The model of math learning for the treatment group resembled the model for above-average achievers, but these effects were short lived, as strong transfer effects for treatment children could only be detected immediately following the end of treatment.

Nevertheless, this suggests that conceptually rich and intensive intervention might allow low-achieving children to learn more like their higher-achieving peers. The inducement of transfer effects in the treatment group could lead to long-run achievement effects if more environmental supports were present to take advantage of transfer potential in the years following the initial treatment. Indeed, some limited empirical research has found that additional instructional support following an early intervention might help sustain early gains. The TRIAD study included a second treatment group, in which schools implemented the *Building Blocks* curriculum during preschool, but kindergarten and first grade teachers also received some professional development to help them build upon the gains students made during the preschool year. Clements and colleagues (2013) reported that although the follow-through condition was less intensive and extensive than the pre-K intervention (i.e., fewer PD sessions were attended by

 ACCEPTED MANUSCRIPT

kindergarten and first grade teachers, no curricular intervention), students assigned to follow-through schools still showed significant gains over treated students that did not attend follow-through schools by the end of first grade. Thus, although students in the follow-through condition still showed substantial treatment effect fadeout after preschool, their fadeout pattern was diminished compared with treated students who did not receive the follow-through supplement. Unfortunately, it is unclear whether this supplement would have demonstrated a treatment effect for children who had not received the initial treatment, and analytic limitations described in the supplementary file prevented us from testing the state-trait model on students assigned to the follow-through condition. Although we were unable to directly test this hypothesis using the follow-through condition, our results suggest that such efforts may help sustain treatment impacts as they are likely to facilitate more transfer of learning following the intervention.

Certainly, attaining a better understanding of what comprises trait-mathematics should be a goal of researchers hoping to better understand how mathematics achievement develops in the long run. We have hypothesized that a combination of both personal and environmental factors likely contribute to long-run stability in individual differences in mathematics achievement, but more research should seek to identify the specific constructs that strongly influence trait-mathematics. A large number of possible characteristics such as parenting and environmental support, interest and motivation, and IQ were not measured as part of the TRIAD study and could all comprise trait-mathematics. Following Bailey and colleagues (2015), we think that the most promising interventions for influencing trait mathematics will be those that influence skills that are fundamental to future learning, readily malleable, and unlikely to develop quickly under counterfactual conditions. For reasons reviewed in that paper, the list of such skills is difficult to generate, but we hope this list of criteria will help guide research toward skills that might have persistent effects on children's mathematics achievement. Regardless, we believe that much could be learned by using the state-trait approach with other skill-promoting interventions that include multiple follow-up assessments. By testing if various types of academic (e.g., reading achievement) and cognitive (e.g., executive function) interventions operate at the state- or trait-level, we could better understand how these interventions impact children's development in the long-term.

Our findings should be considered in the context of a few important limitations. Although we found some evidence suggesting that the state-trait model differed between high- and low-achieving students, these differences were not clear at every time-point and should be replicated in other samples. Such differences might be more detectable in a more heterogeneous sample. Further, our models were only tested on one particular intervention evaluated with the participation of one particular sample of children. Thus, generalizing our results to larger populations should be done with caution, and this limitation presents the need for further replication. Finally, although we found no evidence that attrition patterns differentially affected the treatment or control groups, our overall rate of attrition by fourth grade was still high (44%), and should be considered when interpreting our results.

In sum, our findings suggest that early mathematics interventions can help lower-achieving students learn more like higher-achieving students, but this impact is likely to be limited to the period following the end of the treatment. Developing short-term early interventions that produce persistent effects on long-run achievement would be ideal, but the conditions under which such interventions are plausible may be limited (see Brooks-Gunn, 2003). If designing interventions that can affect the long-run achievement trajectories of students at highest risk for academic failure is the goal, then a more promising approach may be to intervene throughout development with methods known to promote mathematics learning and transfer of learning (e.g. curriculum interventions like *Building Blocks*). Although our results suggest that a single early intervention may be insufficient to substantially alter children's long-term mathematics achievement, these interventions may be valuable if researchers and practitioners can find ways to help students continually build upon the early gains interventions produce.

Acknowledgments

The authors wish to express appreciation to the school districts, teachers, and students who participated in this research. We would like to acknowledge the helpful contributions of Greg Duncan, Tutrang Nguyen, and Katerina Schenke.

Funding

This research was supported by the Institute of Education Sciences, U.S. Department of Education through Grants R305K05157 and R305A120813 and the NICHD-supported Irvine Network on Interventions in Development (HD065704 P01). The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

 ACCEPTED MANUSCRIPT

References

Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Science*, *16*, 235-239.

Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental Dynamics of Math Performance From Preschool to Grade 2. *Journal of Educational Psychology*, *96*(4), 699.

Bailey, D. H., Duncan, G., Odgers, C., & Yu, W. (2015). Persistence and fadeout in the impacts of child and adolescent interventions (Working Paper No. 2015-27). Retrieved from the Life Course Centre Working Paper Series Website: http://www.lifecoursecentre.org.au/working-papers/persistence-and-fadeout-in-the-impacts-of-child-and-adolescent-interventions

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014a). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*(5), 775-785. doi: 10.1111/desc.12155

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014b). State and trait effects on individual differences in children's mathematical development. *Psychological Science, 25*(11), 2017-2026. doi: 10.1177/0956797614547539

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612.

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report*, *17*(1), 1; 3-14.

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), 403.

Byrnes, J. P., & Wasik, B. A. (2009). Factors predictive of mathematics achievement in kindergarten, first and third grades: An opportunity–propensity analysis. *Contemporary Educational Psychology, 34*(2), 167-183.

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta psychologica*, *148*, 163-172.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*, 060306.

Clements, D. H., & Sarama, J. (2007). 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. (2014). *Learning and teaching early math: The learning trajectories approach.* New York, NY: Routledge.

Clements, D. H., Sarama, J., Layzer, C., Unlu, F., Germeroth, C., & Fesler, L. (under review). *Effects on executive function and mathematics learning of an early mathematics curriculum synthesized with scaffolded play designed to promote self-regulation versus the mathematics curriculum alone.*

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

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics

 ACCEPTED MANUSCRIPT

learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education*, *42*, 127-166.

Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM—Tools for early assessment in mathematics.* Columbus, OH: McGraw-Hill Education.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies Persistence of effects in the third year. *American Educational Research Journal*, *50*(4), 812-850.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources, 43*(4), 738-782.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428-1446.

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*, 157-178.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123,* 53-72.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS One, 8*(1), e54651.

 ACCEPTED MANUSCRIPT

Glasgow, C., & Cowley, J. (1994). *Renfrew Bus Story test (North American Edition).* Centreville, DE: Centreville School.

Hiebert, J. C., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 371-404). New York, NY: Information Age Publishing.

Invernizzi, M., Sullivan, A., Swank, L., & Meier, J. (2004). *PALS pre-K: Phonological awareness literacy screening for preschoolers (2nd ed.).* Charlottesville, VA: University Printing Services.

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology*, *116*(1), 45-58.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850–867. doi:10.1037/a0014939

Kane, H. D., Oakland, T. D., & Brand, C. R. (2006). Differentiation at higher levels of cognitive ability: Evidence from the United States. *The Journal of Genetic Psychology*, *167*(3), 327-341.

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology, 63*(1), 52.

Leak, J., Duncan, G. J., Li, W., Magnuson, K., Schindler, H., & Yoshikawa, H. (2010, November). *Is timing everything? How early childhood education program impacts vary*

 ACCEPTED MANUSCRIPT

by starting age, program duration and time since the end of the program. Paper presented at the Fall Conference of the Association for Public Policy Analysis and Management, Boston, MA.

Landry, S. (2007). *MClass:CIRCLE*. New York, NY: Wireless Generation.

National Council of Teachers of Mathematics (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence.* Washington D.C.: NCTM. Retrieved from: http://beck.cherryhill.k12.nj.us/nctm.pdf

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* Washington D.C.: U.S. Department of Education.

National Research Council (2009). *Mathematics Learning in Early Childhood.* Washington D.C.: The National Academies Press.

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In T. Husen & T. N. Postlethwait (Eds.), *International encyclopedia of education, 2nd ed.* (Vol. 11, pp. 6452-6457). Oxford, England: Pergamon Press.

Puma, M., Bell, S., Cook, R., Heid, C., 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.

Reynolds, M. R., Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical intelligence for children and adolescents. *Intelligence*, *35*, 267-281.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77*(1), 1-15.

Rittle-Johnson, B. & Schneider, M. (in press). Developing conceptual and procedural knowledge of mathematics. To appear in R. Kadosh & A. Dowker (Eds.), *Oxford handbook of numerical cognition.* Oxford Press.

Rittle-Johnson, B., & Siegler, R. S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 75–110). East Sussex, England: Psychology Press.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346.

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

Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1227.

Steyer, R. (1987). Konsistenz und Spezifitaet: Definition zweier zentraler Begriffe der Differentiellen Psychologie und ein einfaches Modell zu ihrer Identifikation [Consistency and specificity: Definition of two central concepts of differen- tial psychology and a simple model for their identification]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *8*, 245–258.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology, 45*(4), 1097.

Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P. E., Duckworth, K., Engel, M., Siegler, R. S., Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal mathematics achievement associations. *Child Development, 86*(6), 1892-1907. doi: 10.1111/cdev.12416

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352-360. doi: 10.3102/0013189X14553660

Wilson, S. J., & Farran, D. C. (2012, March). *Experimental evaluation of the Tools of the Mind preschool curriculum*. Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Table 1 *Sample Characteristics*

	Treatment	Control	β	p-value
PreK Entry Math	-3.25	-3.16	-0.08	0.47
	(0.86)	(0.80)		
Site				
New York	0.70	0.76	-0.05	0.76
Massachusetts	0.30	0.24	0.05	0.76
Female	0.50	0.50	0.00	0.89
African American	0.52	0.49	0.03	0.81
Hispanic	0.20	0.26	-0.06	0.52
White	0.25	0.17	0.07	0.51
Age at PreK Entry Test	4.33	4.39	-0.06	0.38
	(0.35)	(0.35)		
Limited Eng Prof.	0.13	0.23	-0.10	0.28
Special Education	0.17	0.16	0.01	0.68
Free/Reduced Lunch	0.82	0.88	-0.06	0.25
Mother's Education				
High School or Above	0.86	0.85	0.01	0.90
Continue at Same School	0.73	0.75	0.02	0.83
Observations	456	378		

Note. Values in the "β" column represent the effect size of the difference between the treatment and control groups. For continuous variables, values in the "β" can be compared with Cohen's *d*, for dummy variables they are equivalent to proportions. Positive values indicate a larger value on a particular variable for the treatment group. Values in the p-value column represent the results of regressions in which each characteristic was regressed on a dummy-variable indicating treatment status. P-values less than .05 would indicate a significant difference on a given baseline characteristic. The preschool entry math test was scaled to have a mean of 0 for the typically-achieving student in first grade.

 ACCEPTED MANUSCRIPT

Table 2 *Correlations Between Key Analysis Variables*

	PreK Entry Math	PreK Post Math	Spring Kind. Math	Spring 1st G. Math	Fall 4^{th} G. Math	Treatment	Kind. Language Skills	Kind. Literacy Skills	Free/Reduced Lunch
PreK Post Math	0.58***								
Spring Kind. Math	0.56***	0.77***							
Spring 1st G. Math	0.54***	0.71***	0.79***						
Fall 4th G. Math	0.43***	0.62***	0.73***	0.75***					
Treatment	-0.05	0.25***	0.09*	0.03	-0.01				
Kind. Language Skills	0.44***	0.46***	0.44***	0.44***	0.34***	0.12**			
Kind. Literacy Skills	0.36***	0.46***	0.44***	0.39***	0.39***	-0.09*	0.32***		
Free/Reduced Lunch	-0.12**	-0.08*	-0.06	-0.12**	-0.07	-0.08*	-0.11**	-0.01	
Mom Ed- HS or Above	0.16***	0.12**	0.15***	0.17***	0.15**	0.01	0.14**	0.14**	-0.13**
N = 834									

Note.

* p < .05,

**p < .01,

***p < .001

Table 3 *State and Trait Effects for Key Subgroups*

	Treatment Group		Control Group		High Achievement		Low Achievement	
	Model 1		Model 2		Model 3		Model 4	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.
Trait Effects								
S1 on Trait	0.88	0.03***	0.89	0.02***	0.76	0.04***	0.80	0.03***
S2 on Trait	0.70	0.11***	0.91	0.14***	0.74	0.09***	0.72	0.09***
S3 on Trait	0.87	0.05***	1.01	0.05***	0.75	0.06***	0.95	0.06***
S4 on Trait	0.74	0.04***	0.78	0.05***	0.59	0.05***	0.71	0.05***
State Effects								
S2 on S1	0.30	0.11**	0.08	0.14	0.29	0.09**	0.27	0.10**
S3 on S2	0.11	0.05*	-0.04	0.05	0.25	0.06***	-0.01	0.06
S4 on S3	0.15	0.05**	0.10	0.05	0.29	0.06***	0.16	0.06*
Tx Effects								
Tx on Trait					0.03	0.11	-0.05	0.11
Tx on State 1					0.67	0.09***	0.64	0.09***
RMSEA	0.05		0.07		0.05		0.07	
CFI	0.99		0.98		0.98		0.96	
Observations	456		378		416		418	

Note. Each model was estimated separately in Mplus 7.0, as nested model chi-squared tests indicated significantly better model fit if the state-trait model for individual group models.

*p < .05, **p < .01, *** p < .001

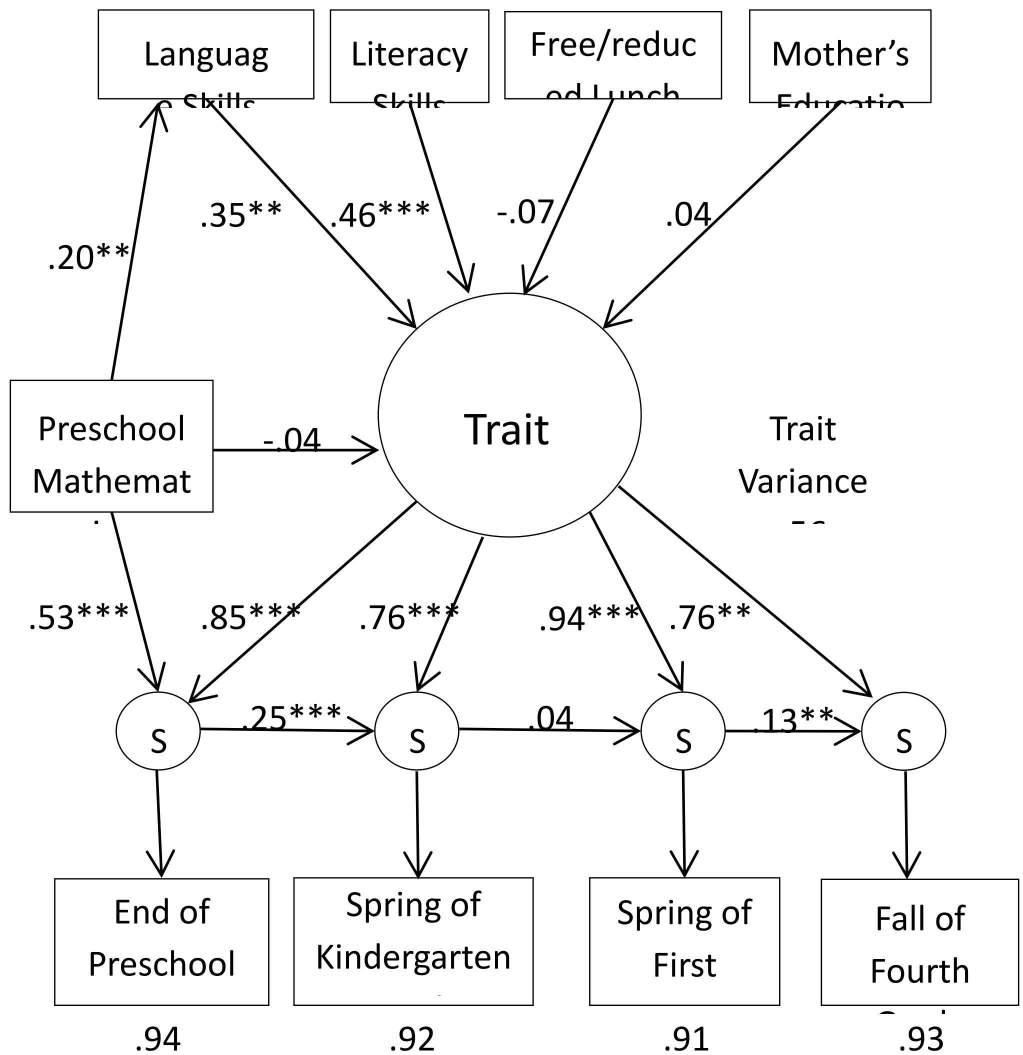

Figure 1 *Treatment Impacts on State and Trait Mathematics Note.* Observations: 834. The model shown had good model-fit ($\chi^2(18) = 67.25$, $p < .001$; RMSEA = .06; CFI = .98). All coefficients presented represent standardized effects, which were derived from the STDYX standardization procedures in Mplus 7.0. Treatment impacts were standardized using the STDY standardization procedures, as is common with dummy variable effects. State factors (S1 -- S4) were derived by setting the paths equal to square root of the reliability for each measure (shown at the bottom of each separate state factor).