Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

*Dev Psychol*. Author manuscript; available in PMC 2017 September 01.

Published in final edited form as:

*Dev Psychol*. 2016 September ; 52(9): 1457–1469. doi:10.1037/dev0000188.

# Fadeout in an Early Mathematics Intervention: Constraining Content or Pre-existing Differences?

**Drew H. Bailey**,
University of California, Irvine

**Tutrang Nguyen**,
University of California, Irvine

**Jade Marcus Jenkins**,
University of California, Irvine

**Thurston Domina**,
University of North Carolina at Chapel Hill

**Douglas H. Clements**, and
University of Denver

**Julie S. Sarama**
University of Denver

## Abstract

A robust finding across research on early childhood educational interventions is that the treatment effect diminishes over time, with children not receiving the intervention eventually catching up to children who did. One popular explanation for fadeout of early mathematics interventions is that elementary school teachers may not teach the kind of advanced content that children are prepared for after receiving the intervention, so lower-achieving children in the control groups of early mathematics interventions catch up to the higher-achieving children in the treatment groups. An alternative explanation is that persistent individual differences in children's long-term mathematical development result more from relatively stable pre-existing differences in their skills and environments than from the direct effects of previous knowledge on later knowledge. We tested these two hypotheses using data from an effective preschool mathematics intervention previously known to show a diminishing treatment effect over time. We compared the intervention group to a matched subset of the control group with a similar mean and variance of scores at the end of treatment. We then tested the relative contributions of factors that similarly constrain learning in children from treatment and control groups with the same level of post-treatment achievement and pre-existing differences between these two groups to the fadeout of the treatment effect over time. We found approximately 72% of the fadeout effect to be attributable to pre-existing differences between children in treatment and control groups with the same level of achievement at post-test. These differences were fully statistically attenuated by children's prior academic achievement.

**Corresponding Author Contact Information:** Drew H. Bailey, School of Education, University of California, Irvine, 3200 Education Building, Irvine, CA 92697-5500, TEL: (949) 824-4329, dhbailey@uci.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

**Keywords**

Mathematics achievement; early childhood; interventions

## Fadeout in an Early Mathematics Intervention: Constraining Content or Pre-existing Differences?

Early childhood educational interventions aim to provide children, often those who are economically disadvantaged, with essential skills for academic and school readiness. Many of these interventions have raised disadvantaged children's academic achievement in the short-term. Unfortunately, a consistent finding across early childhood interventions targeting mathematics and other achievement-related skills that is that initial treatment effects fade over time, with children not receiving the intervention catching up to children who did (Bailey, Duncan, Odgers, & Yu, 2015; Currie and Thomas, 1995; Clements, Sarama, Wolfe, & Spitler, 2013; Leak, Duncan, Li, Magnuson, Schindler, & Yoshikawa, 2010; Puma et al., 2010). Understanding the causes of this fadeout pattern may offer key insight into how to increase the persistence of early childhood educational intervention treatment effects. However, because the factors that likely lead treatment effects to diminish or persist over time – most notably children's post-treatment skills and their post-treatment school environments – are seldom randomly assigned, the mechanisms underlying fadeout of the treatment effect on children's achievement are not well understood.

### The "Constraining Content" Hypothesis

The current study differentiates between two broad explanations of fadeout. First, a prominent explanation is that fadeout is attributable to school environmental factors that fail to successfully build on the skills children gained during the intervention. Early childhood educational interventions primarily target low-income children who, after the intervention, often enter lower resourced elementary schools with lower quality instruction (Crosnoe & Cooper, 2010; McLoyd, 1998; Pianta, Belsky, Houts, & Morrison, 2007; Stipek, 2004). Specifically, this means that many teachers may be unaware that some of their students have already mastered the material that they intend to cover in class (Sarama & Clements, 2015). Therefore, to ensure that all children receive basic skills training, elementary school teachers may not teach the kind of advanced content that children are prepared for after receiving an effective early mathematics intervention (Bodovski & Farkas, 2007; Crosnoe et al., 2010; Engel, Claessens, & Finch, 2013). Engel and colleagues (2013) found that most kindergarten teachers taught mathematics content at a level appropriate only for the lowest achieving students. Teachers reported spending more time on basic skills, such as counting and shape recognition, than any other skills, even though the majority of the children had already mastered these skills at school entry.

Taken together, these ideas suggest that a lack of exposure to advanced content may impose a ceiling on higher achieving children's subsequent achievement trajectories. Thus, children who benefit from early mathematics interventions may experience flatter mathematics achievement trajectories in subsequent years. In contrast, children who did not receive the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

intervention and thus enter kindergarten with low levels of early mathematics skills may benefit more from basic instruction and catch up to their higher achieving peers. We refer to this hypothesis for fadeout of treatment effects in early mathematics interventions as the *constraining content* hypothesis: although children's school readiness skills may improve from high quality preschool interventions, elementary school instruction remains the same, thus constraining higher achieving children's subsequent learning opportunities.

However, evidence for the constraining content explanation of fadeout is ambiguous. Inconsistent with a strong equalizing role of mathematics instruction implied by the constraining content hypothesis, gaps between socioeconomically advantaged children's and socioeconomically disadvantaged children's mathematics achievement scores do not undergo large changes during the school years in kindergarten and first grade (Downey, Von Hippel, & Broh, 2004). In this large correlational study, a standard deviation of SES was associated with 2.40 additional points on a mathematics achievement test on the first day of kindergarten. This gap grew by only .05 points per month during kindergarten and shrank by only .05 points per month during first grade. These associations between SES and growth rates were statistically significant in both grades, but were small compared to the initial SES gap and cancelled each other out.

One correlational analysis indicated that children who subsequently received higher quality instruction or lower class sizes showed *less* persistent positive effects of attending preschool than their peers with lower quality subsequent school environments (Magnuson, Ruhm, & Waldfogel, 2007). Most relevantly, a recent analysis using stronger experimental data found no evidence that treatment effects for two early childhood educational interventions were more persistent for children in higher quality kindergarten and first grade classrooms (Jenkins et al., 2015). Still, it remains possible that a large majority of early elementary mathematics instruction is aimed at a low enough level that children in control conditions of early intervention studies will catch up to their peers who received the treatment. Of course, the current study cannot address whether substantial improvements in the quality of schools would change the long-term effectiveness of early academic interventions. Thus, the focus of the current study is on the narrower question of whether fadeout of the treatment effects of an early mathematics intervention can be attributed to the low-level (and therefore constraining) content children encounter in their subsequent schooling.

### The Pre-Existing Differences Hypothesis

As noted above, the constraining content hypothesis posits that fadeout is attributable to environmental factors that reduce the skill variation among treatment and control students after the intervention ends. Alternatively, fadeout might result from relatively stable differences between children that cause them to revert back to their previous individual achievement trajectories after an effective early childhood intervention. These factors likely include a combination of domain-general cognitive abilities, relatively stable academic skills, motivation, home conditions and other environmental conditions that vary between children, and other factors that affect children's mathematics achievement across time.

One recent study provides empirical support for the pre-existing differences hypothesis. The authors found that relatively stable factors across development explained the correlations

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

among children's mathematics achievement scores at different time points, as opposed to the unique effects of children's previous achievement on their later achievement (Bailey, Watts, Littlefield, & Geary, 2014). However, both appeared to influence children's mathematical development. If this is true, then children's higher mathematics achievement at the end of an effective early mathematics intervention might subsequently drift to levels that would be predicted based on their pre-existing skills, motivation, and environmental conditions. Put differently, this work suggests that even though children's skills are susceptible to early school-based intervention, their longer term mathematics outcomes are primarily influenced by factors that are outside the scope of early mathematics interventions, and thus will not be substantially altered by them. We refer to this hypothesis for fadeout of treatment effects in early mathematics interventions as the *pre-existing differences* hypothesis.

Indeed, a number of child characteristics that are not specifically targeted by early mathematics interventions also likely contribute to children's mathematical development. These characteristics include family income (Duncan, Morris, & Rodrigues, 2011), and other aspects of children's home environments (Blevins-Knabe & Musun-Miller, 1996; Levine, Suriyakham, Rowe, Huttenlocher, & Gunderson, 2010), working memory (Geary, Hoard, Nugent, & Bailey, 2012; Szücs, Devine, Soltesz, Nobes, & Gabriel, 2014), attention (Zentall, 2007), beliefs and expectancies (Meece, Wigfield, & Eccles, 1990), motivation (Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2013), and general intelligence (Deary et al., 2007). If these factors generate most of the variance in individual differences in children's later mathematics learning, effects of early childhood interventions that do not permanently affect these factors will decay over time.

The constraining content and pre-existing differences hypotheses generate different predictions about children's subsequent achievement trajectories following the conclusion of an effective early mathematics achievement intervention. The constraining content hypothesis posits that a lack of advanced instruction in early schooling imposes a ceiling on higher achieving children's mathematics learning. This ceiling will be present both for children whose skills were boosted by the intervention as well as control-group children with similar math skills to treatment-group children following the intervention.

The pre-existing differences hypothesis presumes that children in the control group with similar scores on the post-intervention assessment to children in the treatment group are advantaged on pre-existing differences, having reached the same level of achievement without the benefit of the effective early mathematics intervention. If this is true, the relatively high-achieving control students who enter kindergarten with similar mathematics achievement to their peers in the treatment group are predicted to experience greater achievement gains during the kindergarten year. We illustrate the conflicting predictions of the constraining content hypothesis and the pre-existing differences hypothesis in Figure 1.

The existence of fadeout does not imply that both of these hypotheses must be true. In principle, either factors that make children more similar after the intervention ends (specifically, the constraining content) or pre-existing differences between children could account for the entire fadeout effect. Based on prior research, we hypothesize that fadeout is attributable to some combination of these factors, but that the majority of the fadeout effect

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

can be explained by pre-existing differences between children influencing children's long-term mathematics outcomes.

We tested these two hypotheses using data from an effective early mathematics intervention with an initial treatment effect that faded out over time. Specifically, we compared children from the control group with similar post-intervention achievement scores to children in the treatment group, testing whether their subsequent learning trajectories differed.

## Method

### Participants

The current study used data from the Technology-enhanced, Research-based, Instruction, Assessment, and professional Development (TRIAD) intervention evaluation (Clements et al., 2011; Clements et al., 2013) designed to investigate the effects of an early mathematics curriculum tested in 42 low-resource schools across two sites. The intervention in the pre-kindergarten year included the *Building Blocks* mathematics curriculum and professional development for teachers implementing the curriculum. Schools were rank-ordered on their school-wide standardized mathematics scores within each site, creating blocks with three schools with similar scores. They were then randomly assigned within block to one of three treatment groups: (1) pre-kindergarten only treatment; (2) pre-kindergarten to first grade treatment; (3) control condition. In pre-kindergarten, the two treatment groups received identical intervention, but in the pre-kindergarten to first grade treatment conditions, teachers had knowledge of the pre-kindergarten intervention and were provided with strategies to build upon that knowledge. Our primary analyses included children who were in the pre-kindergarten only treatment and the control condition, since the largest fadeout effect occurred between these groups.

Professional development for teachers in the treatment group was designed to inform them of the rationale for the instructional design of the curriculum's activities, each of which were based on empirically supported developmental progressions for children learning mathematics. Children and teachers in the control group received alternative curricula that also included professional development[1]. Children were assessed immediately after the intervention and then one year later at the end of the kindergarten year. Our resulting sample included 396 children in the control group and 484 children in the treatment group who had valid mathematics achievement scores in pre-kindergarten and kindergarten for a total of 779 children. Students came predominantly from low-income households: 85% of students qualified for free or reduced price lunch. Descriptive statistics for this sample are presented in Table 1.

**Missing data—**Only children with mathematics achievement scores at all three time points were included in the analyses. This requirement dropped 101 children who were missing scores at the end of kindergarten due to attrition. We used the Full Information Maximum Likelihood (FIML) procedure in Stata 13.0 to account for missing data (Enders, 2001). We

[1]The Massachusetts and New York research sites used a different curriculum for the control group. For more detail on curricula in the control classrooms, see Clements et al. (2011).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

compared the descriptive statistics of covariates for the included and excluded sample and found that there were significant differences between the samples on end of pre-kindergarten mathematics achievement, age, literacy and language achievement scores, and whether the mother attended some college. This information is presented in Appendix Table A. As a robustness check and to ensure that our missing data strategy did not bias our final results, we alternatively used missing dummy variables in our models and found similar results. These analyses are included in Appendix Table B.

Our analyses referred to three different key time points: the pre-test is just before the start of the pre-kindergarten intervention; post-test is the end of pre-kindergarten in the spring; and follow-up is one year after the post-test in the spring of kindergarten. The curriculum produced substantial gains in the treatment group with an effect size of .72 (Clements et al., 2011); this treatment effect persisted into the third year of the study, but the effect diminished over time (Clements et al., 2013).

## Measures

**Mathematics achievement—**Children's mathematics achievement was assessed using the Research-based Early Mathematics Assessment (REMA). The measure is validated and specifically designed for early mathematics (Clements, Sarama, & Liu, 2008). It contained two sections: Part A assessed children's counting, number recognition, and addition and subtraction skills and Part B assessed children's patterning, measurement, and shape recognition skills. This measure defined mathematics achievement as a latent trait using the Rasch model, which produced a score with a consistent metric that places children on a common ability scale (Bond & Fox, 2001; Clements et al., 2011; Linacre, 2005; Watson, Callingham, & Kelly, 2007). The Rasch model allowed for accurate comparisons of scores across ages and has been particularly useful for when pre-intervention scores differ between groups (Clements et al., 2011; Wright & Stone, 1979). Test items are ordered by difficulty, and testing is concluded after four consecutive incorrect answers.

**Pre-existing differences—**To test which pre-existing differences statistically attenuated matched control children's higher follow-up scores, we included measures of socioeconomic status and academic skills as covariates. We describe these measures briefly here. For a more thorough discussion of these measures, see Sarama et al. (2012).

**Socioeconomic status:** Measures of children's socioeconomic status included whether the child received free or reduced price lunch measured in the kindergarten year (dummy coded as 0 or 1) and mother's highest level of education. For mother's level of education, we created dummy coded variables indicating the highest level of formal schooling the mother attained: no high school, high school, some college, or college (with the mother attaining a high school diploma as the reference group).

**Literacy skills:** Language and literacy skills were included as measures of children's reading achievement, which may directly influence students' mathematics learning, and also serve as proxies for students' more general cognitive skills. Literacy skills were measured in the spring of Pre-K with the Phonological Awareness Literacy Screening – PreK edition

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(PALS – PreK; Invernizzi, Sullivan, Swank, & Meier, 2004) and MCLASS: CIRCLE (Landry, 2007)[2]. Both are commonly used literacy measures for early childhood programs, measuring recognition of upper and lowercase letters. We used standardized scores for these measures in all analyses.

**Language skills:** Language skills were measured in the fall of kindergarten with the Renfrew Bus Story – North American Edition (RBS; Glasgow & Cowley, 1994). This measure assessed oral language using narrative retell. The narrative retell required children to remember the key concepts of the story, know the meaning of the words representing the concepts to correctly use them in the retell, and understand the story structure to use the words and concepts in the appropriate sequence. Scores from this assessment provided indicators of aspects of children's oral language, such as sentence length and complexity of utterances (Sarama et al., 2012). Trained coders transcribed and coded the responses based on three primary subtests: information, complexity, and sentence length. Three additional subtests included two inferential reasoning items and independence. The information subtest is scored by how many of the 32 concepts from the story the children used in their narrative retell. The complexity subtest is scored as the number of complex utterances used in children's narrative retell. Sentence length is scored as the mean of the child's five longest utterances. The inferential reasoning subtest asked children to answer two questions at the end of the assessment. Responses were scored based on the degree to which the child demonstrated causal reasoning, made reference to the story, showed evidence of empathizing with characters in the story, and showed evidence of practical and moral reasoning in the retell. The independence score is based on the amount of guidance or prompting the child needed from the test administrator to retell the story (Sarama et al., 2012). The raw score was then converted to a standardized score and a Rasch score. We used children's Rasch scores in our analyses.

**Schools—**We controlled for unmeasured preschool and elementary school factors (e.g., quality) by including fixed effects for school random assignment block.

## Analytic Strategy

The analysis plan involved comparing the treatment group to a subset of the control group selected by matching children with similar scores on the post-test assessment immediately following the intervention to test both the pre-existing differences and constraining content hypotheses. Each treated child was matched to a corresponding child in the control group who had a nearly identical post-test score, within a random assignment block[3].

The reason we matched children on their post-test score was to induce differences between treatment and control groups on both observed and likely unobserved characteristics. This manufactured differences in mathematics achievement (e.g., pretest scores), along with

[2]Language and literacy were measured after the intervention, and a previous analysis reported an effect of the intervention on children's language skills (Sarama et al., 2012). Though this is not ideal, we note that any treatment effect on these covariates will make us *less* likely to find that children's skills statistically attenuate post-test matched control children's higher follow-up scores, as they will decrease measured skill advantages of matched control children over the matched treatment children.

[3]We use the user created module PSMATCH2 (Leuven & Sianesi, 2003) in Stata 13.0 to match children. Our results were robust to a variety of matching specifications displayed in Appendices C and D.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

socioeconomic factors, such as mother's education and literacy. Because there was a treatment effect, the post-test matched controls were also higher achieving at the pre-test than the matched treatment children (see Figs. 1 and 2). Table 2 displays the means of the matched treatment and control groups.. All children had the same level of achievement at post-test since they were matched on scores at that time point, but children in the matched control group had higher average mathematics achievement at the follow-up.

We then were able to test whether students' pre-existing differences explain fadeout by comparing the achievement of students in these matched treatment and control groups at follow-up. To test for the specific factors accounting for differences in these groups' trajectories between post-test and follow-up, we regressed follow-up achievement on the matched group (matched treatment = 1; matched control = 0) and different sets of predictors, adjusting for child age at pre-test. If the constraining content hypothesis explained fadeout, there will be no differences between the matched treatment and control groups at follow up. This is because the instruction children receive following the intervention will similarly affect children in the control and treatment groups with the same level of achievement at the post-test.

We conceptualized the fadeout effect as a combination of two sets of factors:

1. Pre-existing differences (likely influenced by a combination of genes, early environments, and relatively stable home environmental conditions, although this cannot be directly tested in the current study), which *differ* between children in the treatment and control groups with the same level of achievement at the post-test, and would subsequently cause the matched control children to outperform the matched treatment children, and
2. the low level of mathematics instruction children receive following the intervention, which would similarly affect children in the control and treatment groups with the same level of achievement at the post-test.

The pre-existing differences hypothesis posits that the first set of factors accounts for the fadeout of early academic intervention effects, whereas the second explanation is a restatement of the constraining content hypothesis.

The logic and equations underlying the following calculations are displayed in Figure 2. To estimate the percent of variance attributable to stable factors that differ between children and factors that similarly affect children, we first estimated the fadeout effect. A major advantage of the measure of mathematics achievement used in the current study is that performance is characterized by a Rasch score, which allows us to directly compare scores across years. Thus, the fadeout effect can be calculated as the Rasch score achievement difference between the unmatched original treatment and control groups at the post-test minus the Rasch score achievement difference between the unmatched groups at the follow-up assessment. Then, we calculated the difference between the post-test matched control group and the post-test matched treatment group at the follow-up assessment. Finally, we divided the latter group difference by the total fadeout effect to estimate the proportion of fadeout attributable to pre-existing differences.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To test whether subsequently aligned instruction reduced the proportion of fadeout explained by the constraining content effect, we replicated the analysis and matched group comparisons described above, comparing the control group to the second treatment group, who received the follow-through treatment.

As an additional test of predictions of the constraining content hypothesis, we replicated the analysis and matched group comparisons described above on children who performed above and below the median score at the pretest. If a ceiling imposed by the low level of instruction holds back the higher achieving students in both groups (depicted in Figure 1), two additional predictions follow:

1. Because the constraining content hypothesis predicts disproportionately negative effects for the highest achieving children, the effect of pre-existing differences (depicted in Figure 2) will be smaller among the higher achieving children. Therefore, *higher* achieving, post-test matched treatment and control children will differ less from each other at the follow-up assessment than *lower* achieving, post-test matched treatment and control children. Pre-existing differences will advantage matched control children in both the higher and lower achieving groups; however, the highest achieving children of all – the higher achieving, post-test matched control children – will be hindered most by this low level of instruction to the extent that they learn the most difficult material to which they are exposed and lack opportunities to learn more difficult material. In contrast, the higher achieving, post-test matched treatment children may be hindered some by this ceiling, while lower achieving children from both groups will not.

2. The lower-achieving matched control group will catch up to the higher achieving matched control group between the posttest and the follow-up assessments. Indeed, the constraining content hypothesis, as depicted in Figure 1, predicts no fadeout during this developmental period unless lower achieving control children are catching up to the higher achieving *treatment and control* children[4]. The prediction unique to the constraining content hypothesis, though, is that the lower achieving control children will catch up to the higher achieving *control* children during this time (both hypotheses make the same prediction about lower achieving control children learning more than higher achieving *treatment* children).

To indirectly examine which pre-existing differences (SES, achievement related skills, preschool quality, or some combination) most contributed to fadeout, we estimated models represented by the following equations at the time of the follow-up assessment:

$$\text{Achievement}_{\text{iFollow-up}} = a_i + b_1 \text{Matched}_{\text{iPost-test}} + b_2 \text{Age}_{\text{iPre-test}} + e_{\text{it}} \quad (1)$$

[4]We thank an anonymous reviewer for pointing this out.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

where $Achievement_{iFollow-up}$ is the math achievement of child $i$ at the time of follow-up; $Matched_{iPost-test}$ is the child's group (post-test matched treatment or control, with treatment coded as 1); $Age_i$ serves as a control variable for child age at first test administration at pre-test[5]. $a_i$ is a constant and $e_{it}$ is an error term. $b_1$, the difference between the matched control and treatment groups at post-test, is our key coefficient in all models.

In subsequent models, the covariates that attenuate the matched group difference are most highly correlated with the pre-existing differences that contribute to fadeout. We first tested the extent to which pre-existing differences in socioeconomic status attenuate the matched group difference in the following model:

$$\text{Achievement}_{\text{iFollow-up}} = a_1 + b_1\text{Matched}_{\text{iPost-test}} + b_2\text{Age}_{\text{iPre-test}} + b_3\text{Child}_i + b_4\text{Family}_i + e_{\text{it}} \quad (2)$$

where $Child_i$ and $Family_i$ are child and family characteristics all measured prior to or at the time of pre-test. These characteristics include whether the child qualifies for free or reduced price lunch and mother's level of education. Socioeconomic status is not an exogenous source of variation, so we were not able to estimate the effect of socioeconomic status *per se* on fadeout. However, this model allowed us to test how much the pre-existing differences that contribute to fadeout are associated with socioeconomic status.

Next, we tested how much children's academic skills, including measures of literacy and language skills and children's scores on the mathematics pre-test, attenuated the matched group difference:

$$\text{Achievement}_{\text{iFollow-up}} = a_1 + b_1\text{Matched}_{\text{iPost-test}} + b_2\text{Literacy}_{\text{iPost-test}} + b_3\text{Language}_{\text{iPost-test}} + b_4\text{Math}_{\text{iPre-test}} + e_{\text{it}} \quad (3)$$

where $Literacy_{iPost-test}$ and $Language_{iPost-test}$ are the collection of literacy and language skills that child $i$ has acquired as of the post-test in the fall of kindergarten. $Math_{iPre-test}$ represents children's mathematics skills at the time of the pre-test[6], before pre-kindergarten entry. Again, we do not consider these measures as exogenous; they probably also indicate individual differences in more general cognitive abilities, along with motivation, and may be influenced by the home environment as well. For these models we hypothesized that the coefficient on the matched group would be close to zero and that the coefficients on the pre-existing differences, $b_2$, $b_3$, and $b_4$ would have significant and positive effects for children in the matched treatment group. We estimated an additional model with preschool block fixed

[5]While individual differences in age are perfectly stable and likely to influence individual differences in learning, especially early in development, children of different ages must be placed in the same classrooms for logistical reasons. Therefore, we statistically control for age in subsequent analyses to attempt to isolate the effects of pre-existing differences in more specific skills and environments that might contribute to fadeout.

[6]We also estimate models without controlling for children's mathematics skills at the time of the pre-test. These results can be found in Appendix Table E.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

effects (Model 4; equation not shown). Standard errors have been adjusted to account for lack of independence from the block clustering of schools.

The final model we estimated is the following with the inclusion of all variables described above:

$$\begin{aligned}\text{Achievement}_{\text{iFollow-up}} = {} & a_1 + b_1\text{Matched}_{\text{iPost-test}} \\ & + b_2\text{Age}_i + b_3\text{Child}_i + b_4\text{Family}_i + b_5\text{Literacy}_{\text{iPost-test}} \\ & + b_6\text{Language}_{\text{iPost-test}} \\ & + b_7\text{Math}_{\text{iPre-test}} \\ & + \mu_1\text{Block}_{\text{iPre-test}} + e_{\text{it}} \end{aligned} \quad (5)$$

where $Block_{iPre-test}$ represent fixed effects for preschool assignment block. We hypothesized that the matched group difference would be close to zero in this model.

## Results

### Matched Group Comparison

Figure 3 presents the trajectories of children in the treatment and control groups before matching. In the original, unmatched sample, children in the treatment group outperformed children in the control group at the post-test assessment ($p < .001$). This effect faded out but remained statistically significant at the follow-up assessment ($p = 0.014$).

Figure 4 presents the trajectories of children in the treatment and control groups after matching. In the matched sample, children in both groups performed identically at the post-test assessment, and the matched control group significantly outperformed the treatment group at the follow-up assessment ($p = 0.001$). This pattern of results is consistent with the pre-existing differences model shown in Figure 1. This finding falsifies the constraining content hypothesis, which predicts that the fadeout effect is primarily due to school and classroom characteristics (i.e., instruction) that similarly affect children in the control and treatment groups with the same level of achievement at the post-test.

We calculated the fadeout effect based on the equation in Figure 2. As described above, the fadeout effect can be calculated as the Rasch score achievement difference between the *unmatched* original, complete treatment and control groups at the post-test minus the Rasch score achievement difference between the unmatched groups at the follow-up assessment ((−1.86 – −2.27) – (−1.07 – −1.19), or .29; values displayed in Figure 3). This represents the treatment effect fadeout at follow-up from the original randomized trial (reported in Clements et al., 2013). Then, we calculated the difference between the post-test *matched* control group and the post-test matched treatment group at the follow-up assessment (−.90 – −1.11, or .21; Figure 4). This difference (minus the group difference at the post-test, which is 0, because the groups are matched for post-test mathematics achievement) is the part of the fadeout effect generated by pre-existing differences, as children began kindergarten with the same mean level of math skills but “induced” pre-existing differences in family and child factors. This analysis estimates how much these induced pre-existing differences at post-test

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

explain the divergence of the matched groups' scores at follow-up, and thus estimates the effect of pre-existing differences on fadeout. Finally, we divided this matched group difference by the fadeout effect from the original treatment and control groups (.21/.29, or .72). Using this calculation, we estimate that approximately 72% of the fadeout effect may be attributable to pre-existing differences between children in the treatment and control groups with the same level of achievement at the post-test. [7]

## Robustness Tests

**Comparison with Follow-through Sample—**We replicated the analysis and matched group comparisons described above, comparing the control group to the second treatment group who received the follow-through treatment. Group means before and after matching are displayed in Appendix Tables F and G. A similar pattern of results was observed, with successful random assignment before matching and a significant matched control group advantage at the pretest and follow-up assessments. These differences, along with the fadeout effect and the matched group difference at the follow-up, are displayed in Table 3. The matched group difference was similar to, but slightly smaller than the matched group difference in the treatment-control comparison described above (.16 vs. .21; Table 3), consistent with the hypothesis that pre-existing differences contributed to fadeout in both comparisons. The fadeout effect was also smaller in the treatment with follow-through vs. control comparison than in the treatment vs. control comparison (.22 vs. .29, Table 3). Therefore, the proportion of fadeout attributable to pre-existing differences was similar in the treatment with follow through vs. control comparison to the value observed in the treatment vs. control comparison (.73 vs. .72, Table 3).

**Comparison within Higher and Lower Achieving Groups—**We also replicated the analysis and matched group comparisons described above on children in the treatment and control groups who performed above and below the median score at the pretest. Group means before and after matching are displayed in Appendix Tables H and I. A similar pattern of results was observed in both groups (as with the previous two comparisons), with successful random assignment before matching and a significant matched control group advantage at the pretest and follow-up assessments in both higher and lower achievers. The treatment effects, follow-up differences, and fadeout effects after matching were very similar across achievement levels, shown in Table 3 and illustrated in Figure 5. Contrary to predictions of the constraining content hypothesis, the matched group difference at the follow-up assessment was actually larger among higher- than lower-achieving students (.24 vs. .16, Table 3). Also contrary to predictions of the constraining content hypothesis, we did not observe evidence of the lower achieving matched control children catching up to the higher achieving matched control children in the period between the posttest and follow-up assessments: the matched higher achieving control children gained .98 Rasch score units between the posttest and the follow-up assessment, while the matched lower achieving control children gained .99 Rasch score units (Figure 5; Appendix Tables H and I).

[7]Note that this calculation is a rough estimate of proportion of fadeout due to preexisting differences because it is composed of fadeout estimates from the entire study sample and a matched subsample.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

### Which Pre-Existing Differences Matter?

Table 4 displays the results of the regression models testing for which variables attenuate the difference between the matched treatment and control groups at the follow-up assessment. All models correspond to the equations written in the analysis plan above. Model 1 shows that adjusting for age the matched groups differed, and that age was positively related to children's follow-up mathematics achievement. Older children had higher pre-test mathematics achievement scores, so older children in the control group performed more similarly to children in the treatment group at the post-test mathematics achievement assessment. The magnitude of the matched group difference was reduced by 29% after adding measures of socioeconomic status (Model 2). Qualifying for free or reduced price lunch was negatively associated with children's achievement, while having a mother with a college education was positively associated with achievement. Compared to the baseline model, socioeconomic status explained an additional 6% of the variance in children's follow-up mathematics achievement scores.

When we controlled for language, literacy, and mathematics achievement in Model 3, the group difference was reduced by 85% relative to the baseline model, and was no longer statistically significant. Compared to the baseline model, the achievement measures by themselves explained an additional 27% of the variance in children's follow-up mathematics achievement scores. Thus, the fadeout effect seems to be largely explained by pre-existing differences correlated with children's academic achievement.

Adjusting for preschool block in Model 4 yielded very similar results to Model 3; the matched group effect and parameter estimates for the achievement variables were very similar in both models, suggesting that public school quality was not driving the relations between the matched treatment and control difference and children's later mathematics achievement. This model explained an additional 32% of the variance in children's follow-up mathematics achievement scores relative to the baseline model.

When socioeconomic status, language, literacy, mathematics, and preschool block were all included in the same model (Model 5), the difference between the matched groups was completely eliminated. Literacy, language, and mathematics scores and whether the child's mother had a college education had statistically significant effects on children's follow-up mathematics achievement. Taken together, these results suggest that factors related to children's language, literacy, and mathematics scores accounted for much of the difference between the matched groups' mathematics achievement at the follow-up assessment.

## Discussion

In the present study, we sought to test both the constraining content hypothesis and the pre-existing differences hypothesis to understand the processes underlying the fadeout of the treatment effect of an early mathematics intervention. We found that pre-existing differences contribute substantially to fadeout (Table 3; Figures 2, 4, and 5). Children in the matched control group outperformed children in the matched treatment group a year after the intervention's end, and this difference was approximately 3/4 of the size of the fadeout effect. This suggests that pre-existing differences in children and their home environments

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

explain more of the fadeout effect than school environmental factors like generally low levels of classroom instruction. To the extent that groups were "under-matched" on their post-test mathematics achievement due to measurement error (see Shadish, Cook, & Campbell, 2002, p. 120), the treatment group's true post-test mathematics achievement is higher than the control group's, and the subsequent matched group trajectory difference ($m_2$ in Figure 2) is under-estimated. Therefore, our estimate may be an under-estimate of the proportion of the fadeout effect attributable to pre-existing differences between children. These results are not inconsistent with the possibility that the constraining content hypothesis explains some or all of the remaining fadeout effect.

There was less total fadeout in the comparison of the treatment group with a follow-through intervention, but the proportion of fadeout explained by pre-existing differences was almost identical in this comparison as in the comparison of the other treatment group to the control group. Thus, we did not find evidence that the follow-through intervention decreased the proportion of the fadeout effect attributable to the constraining content. Furthermore, the consistency of the effect of pre-existing differences among higher and lower achievers, along with the lack of catch-up by the lower achieving control children between the post-test and follow-up assessments ran contrary to predictions of the constraining content hypothesis.

The matched group difference at the follow-up assessment was almost fully eliminated after controlling for children's literacy, reading, and mathematics achievement test scores, consistent with the hypothesis that individual differences in relatively stable achievement-related skills are the primary influences on children's mathematics learning across time. The current study does not allow us to identify which specific factors have the largest effects on children's mathematics learning: for example, do individual differences in literacy *per se* influence children's mathematics learning, or are they both influenced by domain general cognitive abilities and stable environmental factors, such as socioeconomic status? We speculate based on previous work that all of these factors matter, but our results speak only to the net importance of these relatively stable factors unaffected by early mathematics interventions for later mathematics learning: Even for groups matched on post-test mathematics achievement, subsequent mathematics achievement was strongly predicted by previous mathematics achievement (Table 4, Models 3–5). It is difficult to conceptualize how earlier mathematics achievement would directly impact later mathematics achievement, after matching on intermediate mathematics achievement. We cannot rule out the possibility of certain pathways through which this might happen; for example, if there are sensitive periods for the development of children's motivation, then children's mathematics motivation could be substantially impacted by their mathematical skills prior to preschool entry, independently of whether their skills are substantially boosted during preschool. However, the pre-existing achievement effects likely indicate the importance of pre-existing differences in skills and environments unaffected by the intervention for mathematics learning across development.

## Implications for Research on Fadeout and Mathematical Development

Our study has important implications for future research on the mechanisms underlying fadeout and persistence of early childhood educational intervention effects. Primarily, the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

current study highlights the importance of going beyond reporting changes in treatment effect sizes over time, and investigating patterns in well-defined groups of children's post-intervention trajectories to test specific hypotheses about why treatment effects decline post-intervention. We are not the first to have made this point; for example, previous researchers have used longitudinal data analysis to demonstrate that fadeout of early mathematics intervention treatment effects is not attributable to decreasing achievement in the treatment group (Clements et al., 2013; Smith et al., 2012). However, the current study's findings point to the usefulness of subgroup analyses for testing theories about learning and development.

Our results point to the importance of factors other than children's immediately preceding mathematics achievement for explaining children's mathematics learning. However, the exact set of skills and environments contributing to individual differences in children's mathematics achievement across development is not fully understood. Correlational studies do not always include the same set of control measures, and include measures of cognitive skills, non-cognitive skills, and socioeconomic status that differ substantially in quality across datasets. We emphasize a need for further clarification of the mechanisms underlying the stability of individual differences in children's mathematics achievement across time, along with an understanding of the malleability of these factors at different points in development.

Understanding the cognitive processes underlying pre-existing differences in children's later mathematics achievement might have especially practical implications for educators. For example, perhaps children from the treatment and control groups, matched on their post-test achievement scores, differ in the depth and fragility of their mathematics knowledge. Children in the treatment group had steeper learning trajectories during the treatment period, on average. This may imply that their knowledge at the post-test is more recent, less consolidated, and therefore perhaps more prone to forgetting and interference (Wixted, 2004) than the knowledge of post-test matched controls. If so, both additional practice and more explicit connections between previously and later learned material may help children to compensate for pre-existing differences putting them at risk for persistently low mathematics achievement.

The possibility of finding subsequent environments that might lead to the persistence of early treatment effects is appealing, and we encourage future work in this area. One condition that might lead to impact persistence is when subsequent environments are intentionally designed to build on knowledge children gain in a previous intervention: this follows from the idea that children's learning trajectories should guide mathematics pedagogy (Clements & Sarama, 2004, 2014). Indeed, previously reported findings that children's mathematics achievement remained somewhat higher in the Building Blocks with follow-through intervention condition than the group that only received Building Blocks in preschool (Clements et al., 2013) suggest the possibility that treatment effects of early interventions will persist longer when interventions are integrated with a curriculum designed to be cumulative across years.

However, because there was no group that received a cumulative curriculum without the initial treatment, we cannot test whether the professional development designed to facilitate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a cumulative curriculum moderated the treatment effect, produced a second treatment effect on its own, or both. Future work that re-randomizes children from the control group and treatment group into different types of subsequent environments would allow for a direct test of whether different types of subsequent environments contribute to treatment effect fadeout or persistence.

## Implications for Practice

A practical implication of the present study is that subsequent higher-quality learning experiences from preschool into elementary may be necessary for children who received an early intervention to sustain their relative achievement advantages. In other words, interventions, which temporarily compensate for children's pre-existing differences in skills and environments that put them at risk for persistently low achievement, may need to be sustained to produce lasting effects. However, there are trade-offs to intensive, sustained targeted interventions aimed at a specific population (e.g., children at risk for persistently low mathematics achievement) as opposed to universal interventions. Some have argued that targeted programs are less efficient than universal programs in that targeted programs may not reach all the children they seek to serve (Barnett, Brown, & Shore, 2004). The ideal policy decision depends on the costs and benefits associated with boosting only at-risk children's achievement, compared with also typically achieving children's achievement. We will not attempt a cost-benefit analysis here, but hope that an understanding of the mechanisms underlying fadeout might help inform practitioners about which programs and policies are likely to lead to the most persistent effects on children's mathematics achievement.

We stress that we are not arguing that schools do not matter. Improving school quality, across schooling, may improve academic achievement for all children, regardless of their previous math achievement. Further, even if early mathematics intervention may be *insufficient* for demonstrating substantial long-term mathematics achievement gains, it may still be *necessary* (Bailey et al., 2014).

## Limitations & Future Directions

The current study has several limitations that should be addressed. First, our results apply to the observed range of environments – changes in school quality mean or variance could change these estimates. For example, perhaps greatly improved schooling would substantially decrease the magnitude of fadeout. Indeed, Jenkins' and colleagues' (2015) failure to find moderation of treatment effect persistence by school quality also relied on existing variation in school quality for disadvantaged children, and is subject to the same criticism. Certainly, to the extent that it would be economically feasible, we support policies that would raise school quality to levels that fall outside of the existing range for disadvantaged children in the U.S.

Second, our sample included a somewhat homogenous group of children, as the TRIAD intervention was aimed at improving the mathematics achievement and learning of children in low-resource communities who were at risk for persistently low levels of achievement. It is possible that in a more representative sample, a different proportion of fadeout (possibly

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

more) would be attributable to pre-existing differences. Additionally, we found significant differences between children who remained in our sample throughout the study time period and those who attrited before the kindergarten follow-up assessment. Because we matched all treatment children to control children based on post-intervention mathematics skills, differential attrition in the treatment group may bias our results. As shown in Appendix A, attrited treatment group children were more disadvantaged than those who remained in our analysis sample. Therefore, our results may underestimate the proportion of fadeout due to pre-existing differences.

Because of the possible non-representativeness of the schools and children in the current study, we encourage others to test the generalizability of our findings by replicating our analysis across different interventions, samples, and age groups. The analysis requires a dataset with at least three time points (pre-test before the intervention, post-test immediately following the intervention, and follow-up assessment sometime after); there must be an initial treatment effect, and this effect must have faded out over time; and the outcome measure must not show a ceiling effect at the follow-up assessment (a measurement artifact that would create the illusion of a constraining content effect). Ideally, the assessment given at the post-test and follow-up should be on the same scale, which allows for the estimation of a fadeout effect and pre-existing differences effect (Figure 2) on the same scale.

Finally, as noted above, our measures of pre-existing differences are heavily focused on prior achievement and skills. We did not have any measures of time-varying differences in children's family experiences (e.g., job loss) that may have contributed to mathematics achievement learning trajectories. A better understanding of exactly which pre-existing differences contribute to the stability and change in individual differences in children's mathematics achievement over time and the malleability of these factors has important implications for theories and practice related to children's mathematics learning.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This research was supported by the Institute of Education Sciences, U.S. Department of Education through awards R305K05157 and R305A120813 to Douglas H. Clements, Julie Sarama, and (the latter) Carolyn Layzer, and award R305B120013 to Greg Duncan. This research was also supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) to Greg Duncan and George Farkas (P01HD065704). The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education or the NICHD. The authors wish to express appreciation to the school districts, teachers, and students who participated in this research. We would also like to acknowledge the helpful contributions of Greg Duncan, Maria Rosales Rueda, Mary Elaine Spitler, Tyler Watts, and Christopher Wolfe to this project.

## References

Bailey, DH.; Duncan, G.; Odgers, C.; Yu, W. Persistence and fadeout in the impacts of child and adolescent interventions (Working Paper No. 2015–27). 2015. 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Bailey DH, Watts TW, Littlefield AK, Geary DC. State and trait effects on individual differences in children's mathematical development. Psychological Science. 2014; 25(11):2017–2026. [PubMed: 25231900]

Barnett, WS.; Brown, K.; Shore, R. The Universal Vs. Targeted Debate: Should the United States Have Preschool for All?. New Brunswick, NJ: NIEER; 2004.

Blevins-Knabe B, Musun-Miller L. Number use at home by children and their parents and its relationship to early mathematical performance. Early Development and Parenting. 1996; 5(1):35–45.

Bodovski K, Farkas G. Do instructional practices contribute to inequality in achievement? The case of mathematics instruction in kindergarten. Journal of Early Childhood Research. 2007; 5(3):301–322.

Bond, TG.; Fox, CM. Applying the Rasch model: Fundamental measurement in human sciences. Mahwah, NJ: Erlbaum; 2001.

Clements DH, Sarama J. Learning trajectories in mathematics education. Mathematical Thinking and Learning. 2004; 6(2):81–89.

Clements, DH.; Sarama, J. Learning and teaching early math: The learning trajectories approach. Routledge; 2014.

Clements DH, Sarama JH, Liu XH. Development of a measure of early mathematics achievement using the Rasch model: the Research- Based Early Maths Assessment. Educational Psychology. 2008; 28(4):457–482.

Clements DH, Sarama J, Spitler ME, Lange AA, Wolfe CB. Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. Journal for Research in Mathematics Education. 2011; 42(2):127–166.

Clements DH, Sarama J, Wolfe CB, Spitler ME. 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. 2013; 50(4):812–850.

Crosnoe R, Cooper CE. Economically disadvantaged children's transitions Into elementary school: Linking family processes, school contexts, and educational policy. American Educational Research Journal. 2010; 47(2):258–291. [PubMed: 20711417]

Crosnoe R, Morrison F, Burchinal M, Pianta R, Keating D, Friedman SL, Clarke-Stewart KA. Instruction, teacher-student relations, and math achievement trajectories in elementary school. Journal of Educational Psychology. 2010; 102(2):407. [PubMed: 20657743]

Currie J, Thomas D. Does Head Start make a difference? American Economic Review. 1995; 3(85): 341–364.

Currie J, Thomas D. School quality and the longer-term effects of Head Start. Journal of Human Resources. 2000; 4(35):755–774.

Deary IJ, Strand S, Smith P, Fernandes C. Intelligence and educational achievement. Intelligence. 2007; 35(1):13–21.

Downey DB, Von Hippel PT, Broh BA. Are schools the great equalizer? Cognitive inequality during the summer months and the school year. American Sociological Review. 2004; 69(5):613–635.

Duncan GJ, Morris PA, Rodrigues C. Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. Developmental Psychology. 2011; 47(5):1263. [PubMed: 21688900]

Duncan GJ, Yeung WJ, Brooks-Gunn J, Smith JR. How much does childhood poverty affect the life chances of children? American Sociological Review. 1998; 63:406–423.

Enders CK. The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. Psychological Methods. 2001; 6(4):352. [PubMed: 11778677]

Engel M, Claessens A, Finch MA. Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. Educational Evaluation and Policy Analysis. 2013; 35(2):157–178.

Geary DC, Hoard MK, Nugent L, Bailey DH. Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five year prospective study. Journal of Educational Psychology. 2012; 104:206–223. [PubMed: 27158154]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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

Imbens GW. Matching Methods in Practice: Three Examples. Journal of Human Resources. 2015; 50(2):373–419.

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

Jenkins, JM.; Watts, T.; Clements, D.; Sarama, J.; Wolfe, CB.; Spitler, M. Preventing preschool fadeout through instructional intervention in kindergarten and first grade. Paper presented at the Spring Conference of the Society for Research on Educational Effectiveness; Washington, DC. 2015 Mar.

Landry, S. MClass: CIRCLE. New York, NY: Wireless Generation; 2007.

Leak, J.; Duncan, GJ.; Li, W.; Magnuson, K.; Schindler, H.; Yoshikawa, H. Is timing everything? How early childhood education program impacts vary 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. 2010 Nov.

Lee VE, Loeb S. Where do Head Start attendees end up? One reason why preschool effects fade out. Educational Evaluation and Policy Analysis. 1995; 17(1):62–82.

Leuven, E.; SianesI, B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. 2003. http://ideas.repec.org/c/boc/bocode/s432001.html

Levine SC, Suriyakham LW, Rowe ML, Huttenlocher J, Gunderson EA. What counts in the development of young children's number knowledge? Developmental Psychology. 2010; 46(5): 1309–1319. [PubMed: 20822240]

Linacre, JM. A user’s guide to Winsteps/Ministep Rasch-model computer program. Chicago: Winsteps.com; 2005.

Magnuson KA, Ruhm C, Waldfogel J. The persistence of preschool effects: Do subsequent classroom experiences matter? Early Childhood Research Quarterly. 2007; 22:18–38.

McLoyd VC. Socioeconomic disadvantage and child development. American Psychologist. 1998; 53(2):185–204. [PubMed: 9491747]

Meece JL, Wigfield A, Eccles JS. Predictors of math anxiety and its consequences for young adolescents' course enrollment intentions and performances in mathematics. Journal of Educational Psychology. 1990; 82:60–70.

Murayama K, Pekrun R, Lichtenfeld S, vom Hofe R. Predicting Long- Term Growth in Students' Mathematics Achievement: The Unique Contributions of Motivation and Cognitive Strategies. Child Development. 2013; 84(4):1475–1490. [PubMed: 23278807]

Pianta R, Belsky J, Houts R, Morrison F. Opportunities to Learn in America’s Elementary Classrooms. Science. 2007; 315(5820):1795–1796. [PubMed: 17395814]

Puma M, Bell S, Cook R, Heid C, Shapiro G, Broene P, Spier E. Head Start Impact Study. Final Report. Administration for Children & Families. 2010

Puma MJ, Olsen RB, Bell SH, Price C. What to Do when Data Are Missing in Group Randomized Controlled Trials. NCEE 2009–0049. National Center for Education Evaluation and Regional Assistance. 2009

Sarama, J.; Clements, DH. Scaling up early mathematics interventions: Transitioning with trajectories and technologies. In: Perry, B.; MacDonald, A.; Gervasoni, A., editors. Mathematics and transition to school. Singapore: Springer; 2015. p. 153-169.

Sarama J, Lange AA, Clements DH, Wolfe CB. The impacts of an early mathematics curriculum on oral language and literacy. Early Childhood Research Quarterly. 2012; 27(3):489–502.

Shadish, WR.; Cook, TD.; Campbell, DT. Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton-Mifflin; 2002.

Smith TM, Cobb P, Farran DC, Cordray DS, Munter C. Evaluating Math Recovery Assessing the Causal Impact of a Diagnostic Tutoring Program on Student Achievement. American Educational Research Journal. 2012; 50(2):397–428.

Stipek D. Teaching practices in kindergarten and first grade: Different strokes for different folks. Early Childhood Research Quarterly. 2004; 19(4):548–568. doi:http://dx.doi.org/10.1016/j.ecresq.2004.10.010.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Szücs D, Devine A, Soltesz F, Nobes A, Gabriel F. Cognitive components of a mathematical processing network in 9-year-old children. Developmental Science. 2014; 17:506–524. [PubMed: 25089322]

Watson JM, Callingham RA, Kelly BA. Students' appreciation of expectation and variation as a foundation for statistical understanding. Mathematical Thinking and Learning. 2007; 9(2):83–130.

Wixted JT. The psychology and neuroscience of forgetting. Annual Review of Psychology. 2004; 55:235–269.

Wright, BD.; Stone, MH. Best test design: Rasch measurement. Chicago: MESA Press; 1979.

Zentall, SS. Math performance of students with ADHD: Cognitive and behavioral contributors and interventions. In: Berch, DB.; Mazzocco, MMM., editors. Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities. Baltimore, MD: Paul H. Brookes; 2007. p. 219-243.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

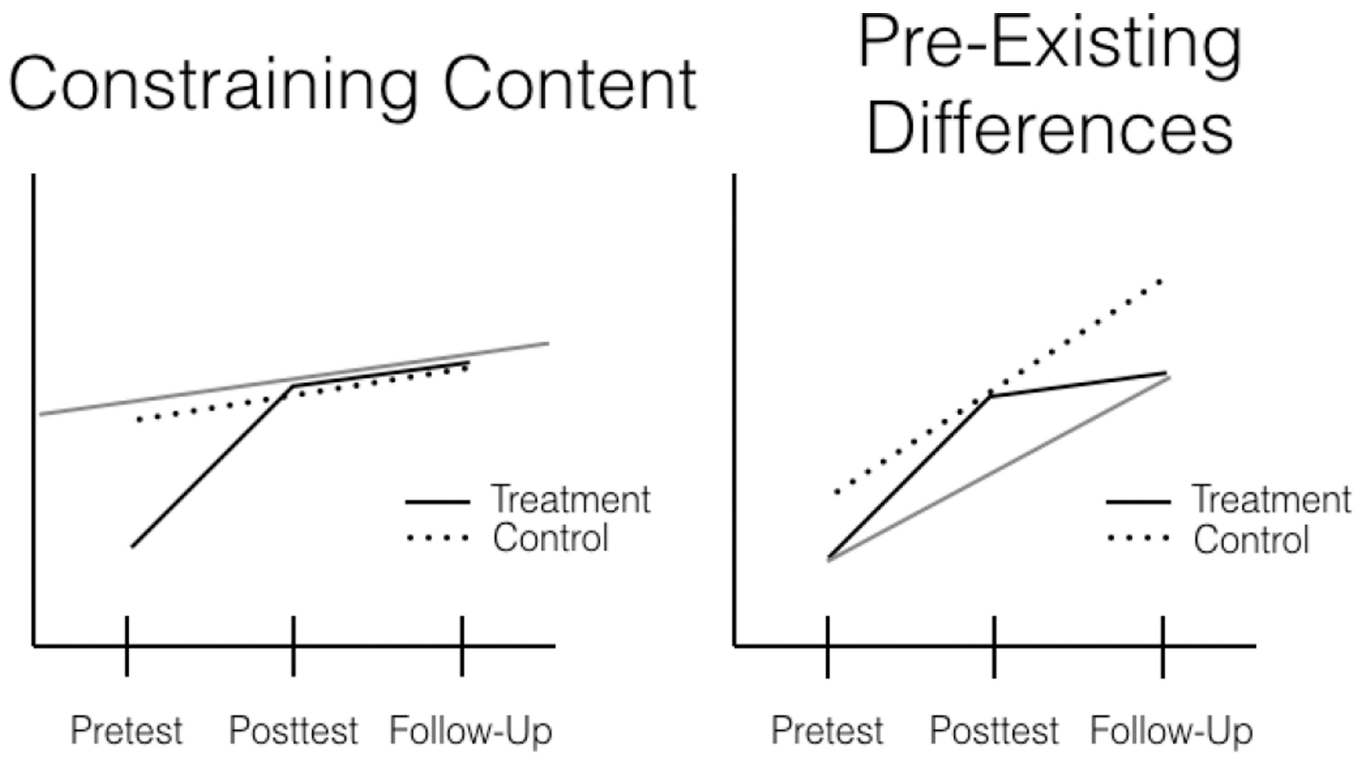


**Figure 1. Predicted Trajectories for Treatment and Higher Achieving Control Children Based on the Constraining Content and Pre-existing Differences Hypotheses**

The gray line in the left panel represents an artificial ceiling imposed by the low-level content of instruction, which constrains the later trajectories of children in the treatment group, according to the constraining content hypothesis. The gray line in the right panel represents the trajectory the treatment group would have been predicted to follow, had they not received treatment, according to the pre-existing differences hypothesis. The key difference between the two hypotheses is that the pre-existing differences hypothesis predicts that the groups will differ in their trajectories following the post-test, whereas the constraining content hypothesis predicts that the groups will have similar trajectories.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

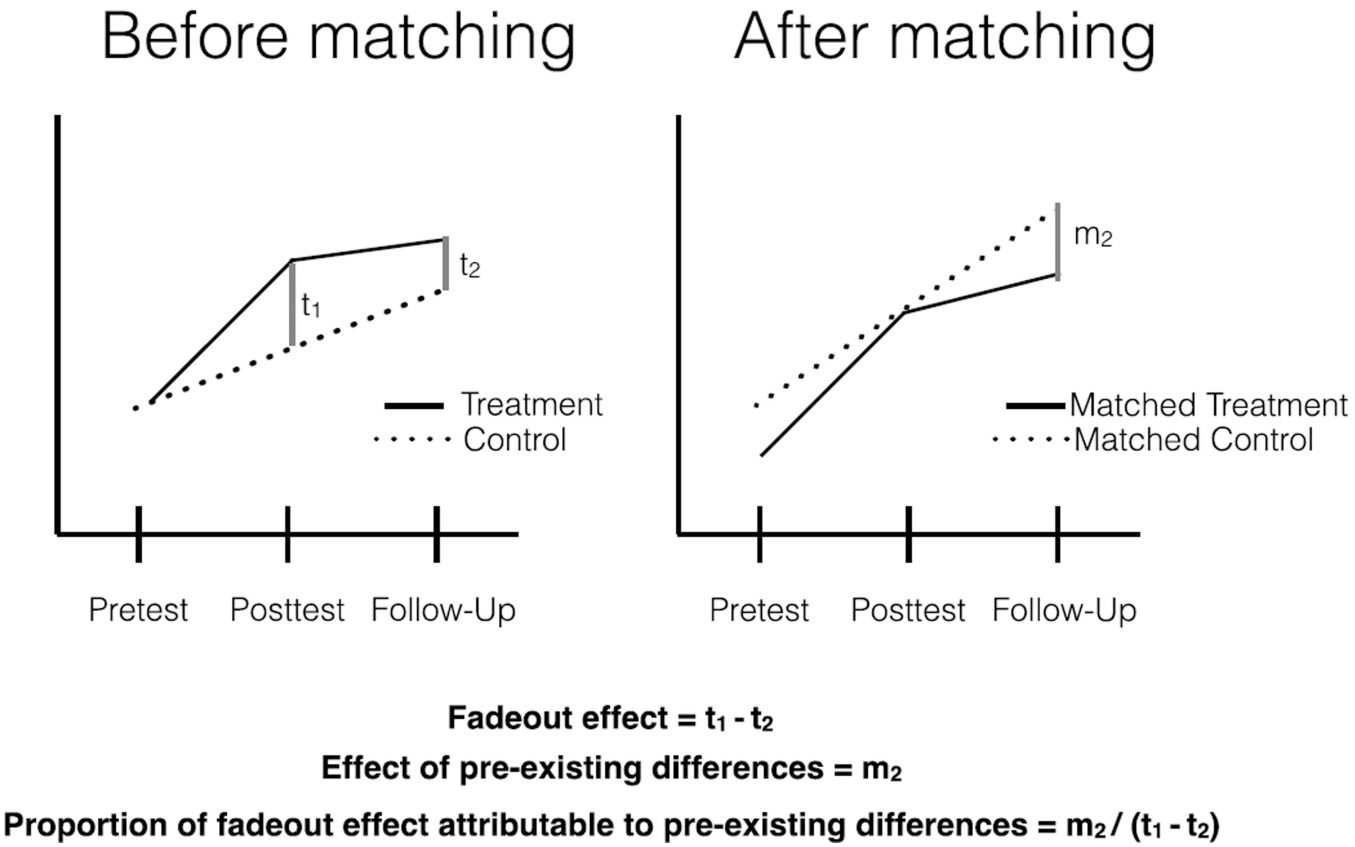


**Figure 2. Graphical Representation of Calculations of the Fadeout Effect and Effect of Pre-Existing Differences**

The first graph in the left panel displays differences between the treatment and control children before they were matched. The difference between the treatment children and control children at post-test (represented by t1) is larger than the difference at follow-up (represented by t2). The fadeout effect is can be found by taking the difference between the two time points. The second graph in the right panel displays the trajectories of children's achievement once they have been matched. Children who were matched to be in the control group outperformed children who were matched to be in the treatment group. The difference between the two groups at follow-up (represented by $m_2$) is the effect of children's pre-existing differences. The proportion of the fadeout effect attributable to pre-existing differences can be calculated by dividing the effect of pre-existing differences by the fadeout effect.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

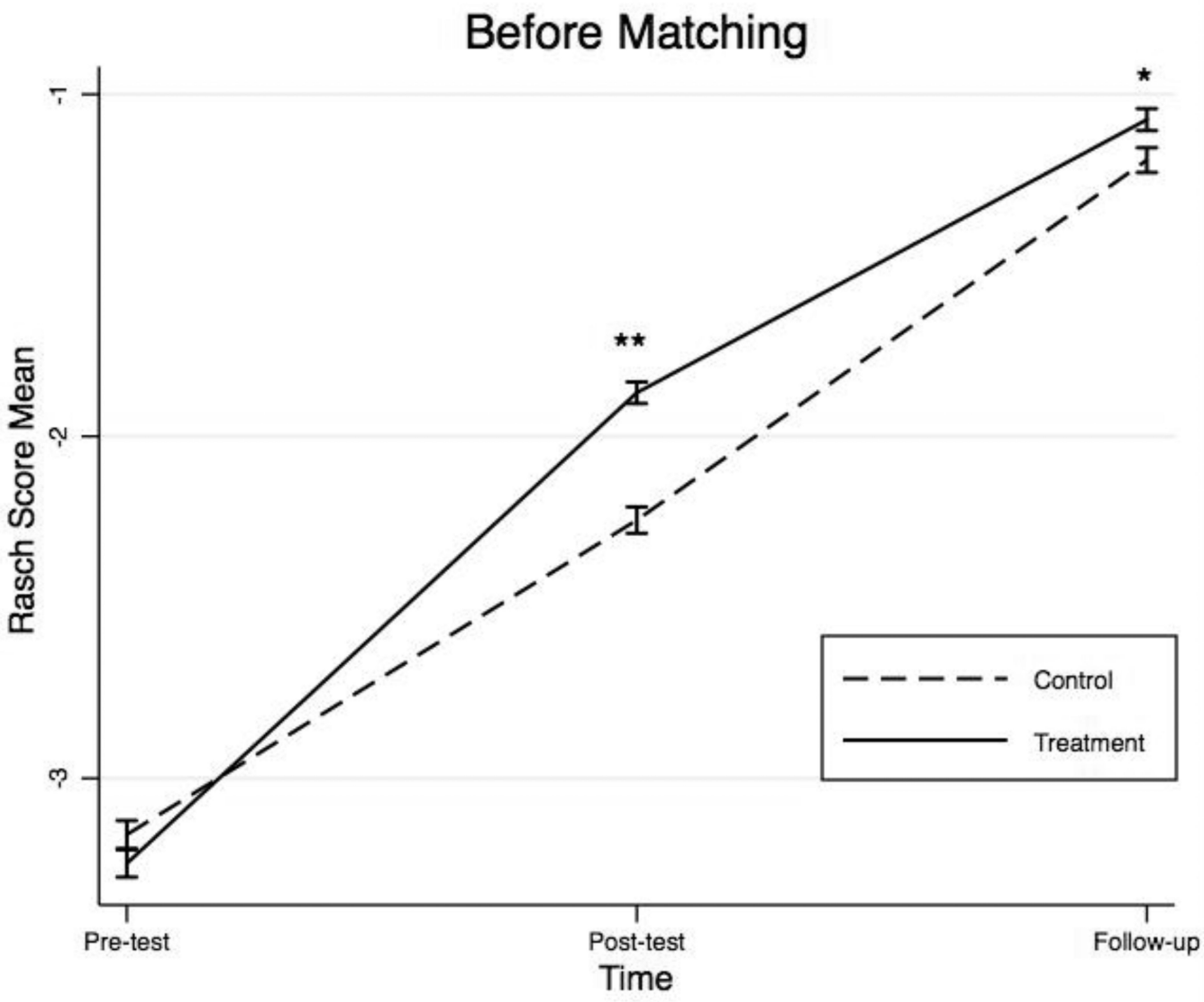


**Figure 3. Observed Mathematics Achievement Trajectories of the Intervention Groups before Matching on Post-Test Mathematics Achievement**

The graph above displays the mathematics achievement trajectories of the children in the treatment and control groups before they were matched. Children in the treatment group outperformed the children in the control group at post-test and follow-up. + p<.10; * p<.05; ** p<.01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

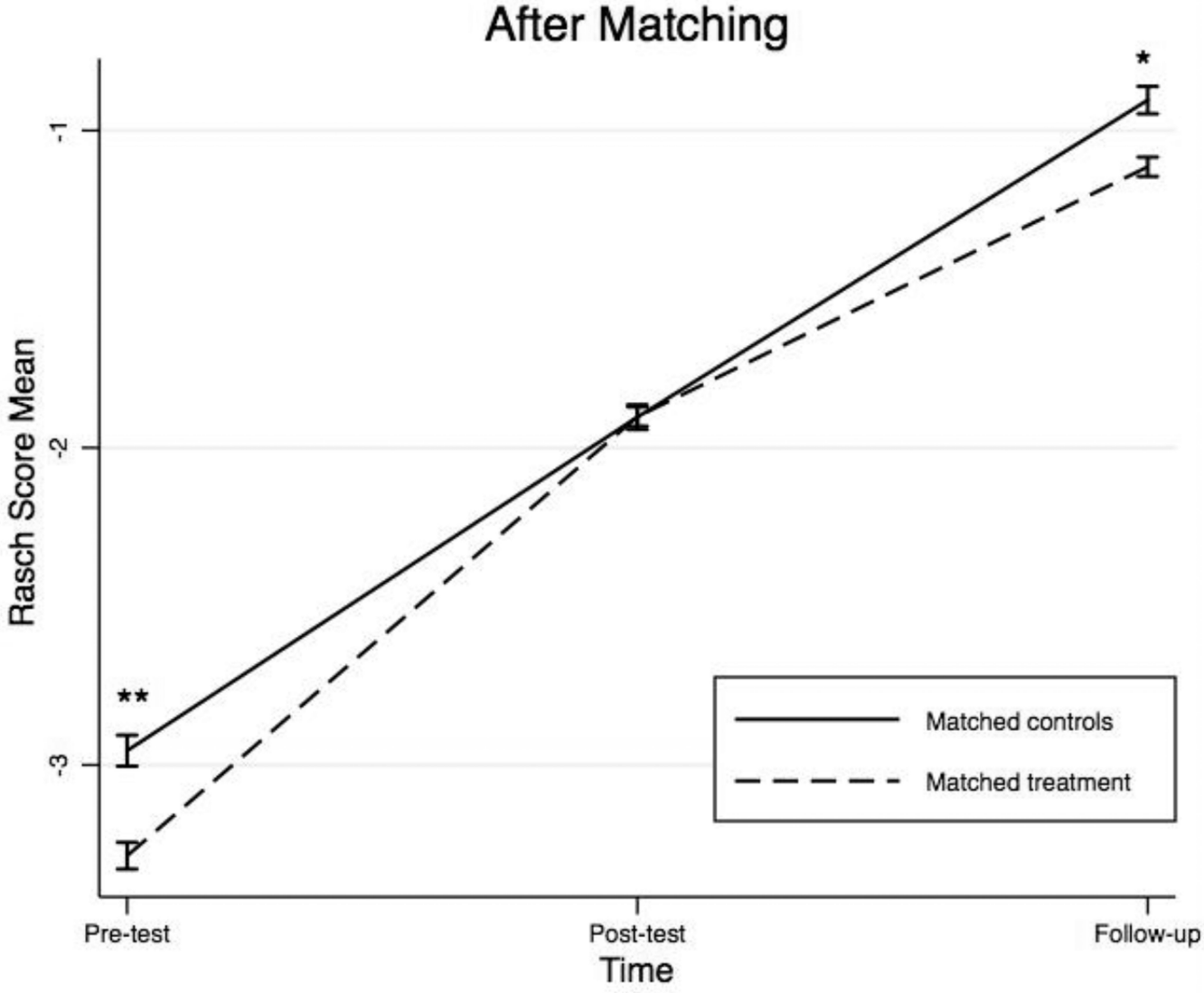


**Figure 4. Observed Mathematics Achievement Trajectories of the Intervention Groups after Matching on Post-Test Mathematics Achievement**

The graph above displays the mathematics achievement trajectories of the children in the treatment and control groups after they were matched. Children in the matched control group outperformed children who were matched in the treatment group. Groups had the same level of achievement at post-test because they were matched on scores at that time point. + p<.10; * p<.05; ** p<.01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

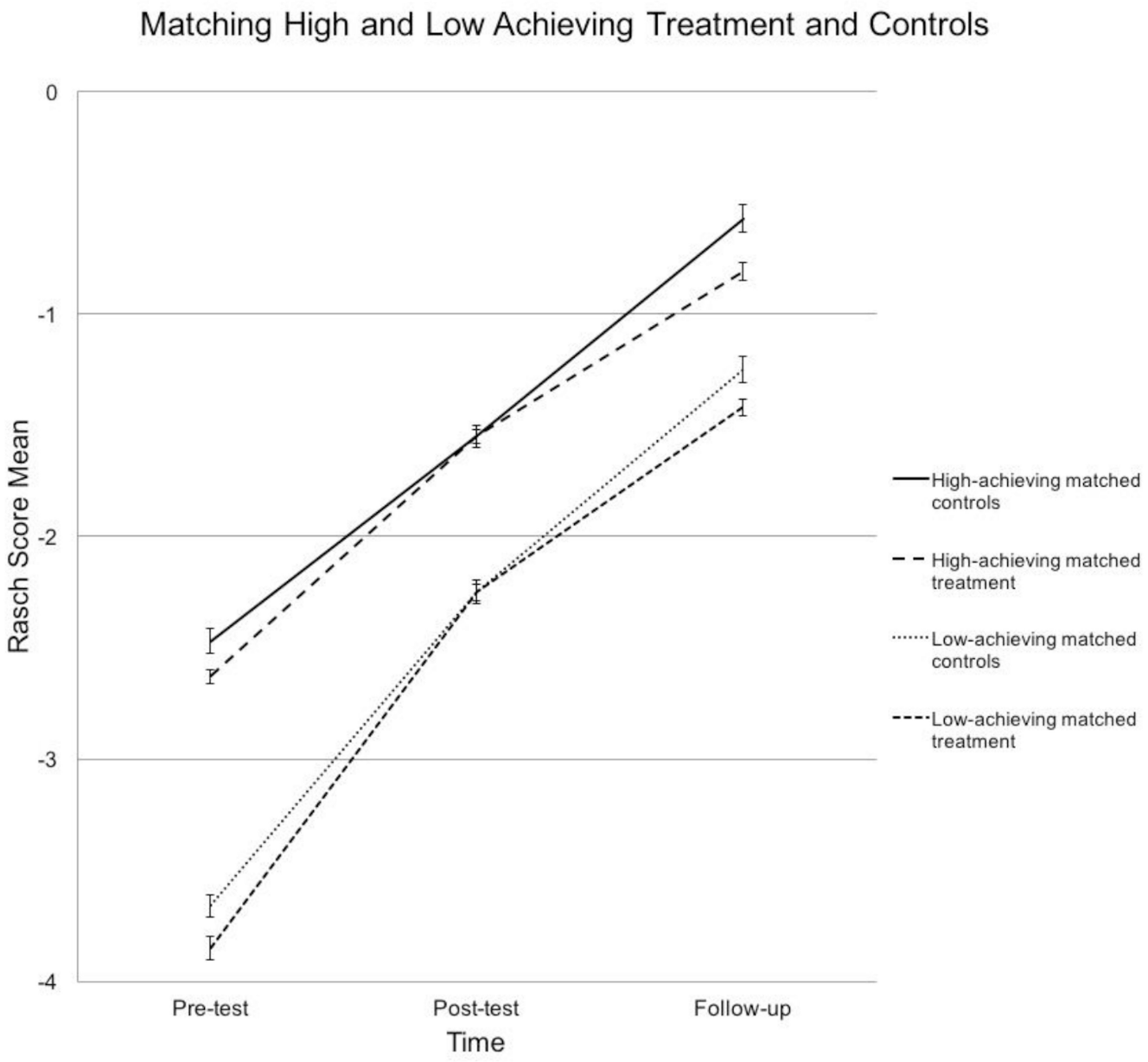


**Figure 5. Observed Mathematics Achievement Trajectories of Pre-Test Split Intervention Groups after Matching on Post-Test Mathematics Achievement**

The graph above displays the mathematics achievement trajectories of the children above and below the pre-test median score in the treatment and control groups after they were matched. For both achievement level subsets, children in the matched control group outperformed children who were matched in the treatment group at the follow-up. Groups within each achievement level had the same level of achievement at post-test because they were matched on scores at that time point.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



*Dev Psychol*. Author manuscript; available in PMC 2017 September 01.

**Table 1**

Sample characteristics before matching on post-test mathematics achievement

| | Before Matching | | | | |
|---|---|---|---|---|---|
| | Treatment | | Control | | |
| | Mean/% of sample | SD | Mean/% of sample | SD | Difference in Means |
| *Mathematics Achievement* | | | | | |
| Pre-test (before pre-K) | −3.25 | 0.86 | −3.17 | 0.80 | |
| Post-test (end of pre-K) | −1.86 | 0.65 | −2.27 | 0.75 | ** |
| Follow-up (end of K) | −1.07 | 0.66 | −1.19 | 0.69 | * |
| *Child Characteristics* | | | | | |
| Male | 0.50 | | 0.49 | | |
| Age | 4.32 | 0.35 | 4.38 | 0.35 | * |
| Free and reduced price lunch | 0.71 | | 0.74 | | |
| Language | 0.08 | 1.01 | −0.13 | 1.07 | * |
| Literacy | −0.01 | 1.02 | 0.06 | 1.01 | * |
| *Maternal Education* | | | | | |
| No high school | 0.11 | | 0.10 | | |
| High school | 0.23 | | 0.21 | | |
| Some college | 0.30 | | 0.23 | | * |
| College | 0.11 | | 0.13 | | |
| *N* | 418 | | 361 | | |

*Note*. Mathematics achievement scores are Rasch scores. Language and literacy scores are *z* scores. Age is in years, and all other variables are proportions.

+ p<0.10;

* p<.05;

** p<.01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



**Table 2**

Sample characteristics after matching on post-test mathematics achievement

| | After Matching | | | | |
|---|---|---|---|---|---|
| | Treatment | | Control | | |
| | Mean/% of sample | SD | Mean/% of sample | SD | Difference in Means |
| *Mathematics Achievement* | | | | | |
| Pre-test (before pre-K) | −3.29 | 0.94 | −2.95 | 0.68 | ** |
| Post-test (end of pre-K) | −1.90 | 0.69 | −1.90 | 0.54 | |
| Follow-up (end of K) | −1.11 | 0.60 | −0.90 | 0.68 | * |
| *Child Characteristics* | | | | | |
| Male | 0.50 | | 0.49 | | |
| Age | 4.32 | 0.39 | 4.45 | 0.33 | ** |
| Free and reduced price lunch | 0.72 | | 0.69 | | |
| Language score | 0.07 | 1.02 | 0.03 | 1.04 | |
| Literacy score | −0.03 | 1.01 | 0.33 | 0.97 | ** |
| *Maternal Education* | | | | | |
| No high school | 0.11 | | 0.09 | | |
| High school | 0.23 | | 0.23 | | |
| Some college | 0.31 | | 0.22 | | * |
| College | 0.10 | | 0.17 | | * |
| *N* | 406 | | 251 | | |

*Note*. Mathematics achievement scores are Rasch scores. Language and literacy scores are *z* scores. Age is in years, and all other variables are proportions .

+ p<0.10;

* p<.05;

** p<.01.

*Dev Psychol*. Author manuscript; available in PMC 2017 September 01.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



**Table 3**

Summary of estimates for key group differences before and after matching

| Comparison | Posttest difference (t1) | Follow-up difference (t2) | Fadeout effect (t1-t2) | Matched follow-up difference (m2) | Proportion of fadeout explained by m2 |
|---|---|---|---|---|---|
| Treatment vs. Control | 0.41 | 0.12 | 0.29 | 0.21 | 0.72 |
| Follow-through vs. Control | 0.39 | 0.17 | 0.22 | 0.16 | 0.73 |
| Higher-achieving treatment vs. Control | 0.40 | 0.14 | 0.26 | 0.24 | 0.92 |
| Lower-achieving treatment vs. Control | 0.42 | 0.16 | 0.26 | 0.16 | 0.62 |

*Note*. Values can be re-calculated by hand from Tables 1 and 2, and Appendix Tables F–I. Variable names (t1, t2, and m2) and formulae used to calculate them correspond to those shown in Figure 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



**Table 4**

Models testing for attenuation of the effect of matched group on follow-up mathematics achievement score

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| Matched treatment group | −0.21 * (0.09) | −0.15 + (0.08) | −0.03 (0.07) | −0.02 (0.08) | 0.00 (0.07) |
| Age at pre-test | 0.28 ** (0.04) | 0.28 ** (0.04) | 0.16 ** (0.04) | 0.13 ** (0.03) | 0.13 ** (0.03) |
| Free or reduced price lunch | | −0.22 * (0.10) | | | −0.06 (0.08) |
| Mother's education: no high school | | −0.17 (0.12) | | | −0.11 (0.10) |
| Mother's education: some college | | 0.00 (0.09) | | | −0.04 (0.07) |
| Mother's education: college | | 0.59 * (0.17) | | | 0.37 ** (0.13) |
| Literacy score | | | 0.17 ** (0.04) | 0.22 ** (0.04) | 0.21 ** (0.04) |
| Language score | | | 0.15 ** (0.04) | 0.14 ** (0.04) | 0.12 ** (0.04) |
| Pre-K entry math score | | | 0.38 ** (0.05) | 0.37 ** (0.05) | 0.34 ** (0.04) |
| Block fixed effects | | | | Inc. | Inc. |
| $R^2$ | 0.08 | 0.14 | 0.35 | 0.40 | 0.41 |
| Observations | 657 | 657 | 657 | 657 | 657 |

*Note*. Standard errors in parentheses. Full Information Maximum Likelihood (FIML) was used in Stata 13.0 to account for missing data independent variables. Alternative methods to handle missing data are presented in the appendix. Mathematics achievement outcome, literacy, and language scores were standardized to have a mean of 0 and standard deviation of 1. Reference group for mother's education was having high school diploma. Children were matched using a nearest-neighbor matching procedure with replacement and two nearest neighbors 0.05 caliper.

+ p<.10;

* p<.05;

** p<.01.