Abstract Title Page

Not included in page count.

Title: Effects of TRIAD on Mathematics Achievement: Long-term Impacts

Authors and Affiliations:
Douglas H. Clements, University of Denver
Julie Sarama, University of Denver
Carolyn Layzer, Abt Associates
Fatih Unlu, Abt Associates
Christopher B. Wolfe, St. Leo University
Mary Elaine Spitler, University of Buffalo, SUNY
Daniel Weiss, Abt Associates

Abstract Body

Background / Context: Although some research-based educational practices have shown promise, many fail to be implemented at a scale that affects more than a small proportion of children. Further, research on interventions for young children includes mixed results, with most documenting "fadeout" of effects after several years, but some showing lasting effects. In this study, we evaluated the long-term impacts of a model for scaling up early interventions, in this case a successful early mathematics curriculum, testing to see whether the originally-sustained impacts persisted up to 7 years beyond baseline.

Purpose / Objective / Research Question / Focus of Study: The original evaluation of TRIAD (**T**echnology-enhanced, **R**esearch-based, **I**nstruction, **A**ssessment, and professional **D**evelopment) employed a cluster randomized trial (CRT) design to test the effectiveness of the TRIAD scale-up approach, using the Building Blocks® early mathematics curriculum (BB) in preschool. In this evaluation, 42 schools were randomly assigned to three groups. One of the groups was a business-as-usual control (CTRL). In preschool, the other two conditions were identical and entailed TRIAD scale-up (intensive training, coaching focused on both BB curriculum and linked assessment and online professional development). After preschool, one of the treatment conditions included follow-through professional development through 1st grade, including knowledge of the intervention and ways to build upon that knowledge using learning trajectories (TRIAD Follow-Through or TRIAD-FT) while the other condition implemented business-as-usual control (TRIAD Non-follow-through or TRIAD-NFT). Students in the two treatment conditions significantly outperformed their peers in the control condition through 1st grade. The primary purpose of the current investigation is to examine the long-term effects of the early mathematics curriculum, but our secondary aim is to explore whether the TRIAD scale-up approach itself, which included aligned professional development for kindergarten and 1st grade teachers, also enhanced the effects of the pre-K curriculum beyond 1st grade. Specifically, we address the following research questions:

RQ1. What are the differences in the average math achievement of the three original study groups (TRIAD-NFT, TRIAD-FT, and CTRL) in 3rd, 4th, and 5th grade?
RQ2. Do the differences in the 3rd, 4th, and 5th grade math achievement of the original study groups vary by the socio-economic and racial/ethnic backgrounds of the students in the groups?
RQ3. What are the differences in the 3rd, 4th, and 5th grade math achievement of the subset of children who remained in the group to which their original school was randomly assigned throughout the intervention period (pre-K through 1st grade)?
RQ4. Do the differences in the in the 3rd, 4th, and 5th grade math achievement of the subset of children who remained in the group to which their original school was randomly assigned throughout the intervention period (pre-K through 1st grade) vary by the students' socio-economic status and racial/ethnic backgrounds?

Note that the first question keeps the three original groups intact (regardless of whether students stayed in those conditions throughout the original study) to preserve the experimental nature of the contrasts between the three groups while the third question focuses on students who stayed in the original groups and received the full dosage of the treatment provided to their condition (also called "stayers") as an attempt to isolate the program effects from the potential effect of student mobility on students' math achievement. The second and fourth questions explore whether the long-term effects pf the TRIAD-NFT and TRIAD-FT conditions estimated for the first and third questions vary by student characteristics.

Setting: Participants in the present study, conducted in schools with diverse racial/ethnic and socioeconomic student bodies, included 1,305 children in two cities (Buffalo, NY and Boston, MA), their 106 pre-K teachers, and their Kindergarten (270) and 1st grade teachers (345). Table 1 shows the demographic characteristics of the participating schools.

Population / Participants / Subjects: Table 2 shows the number of students for whom we have valid outcome measures separately for the three conditions and each grade-level. While Table 2 suggests that there was considerable attrition between pre-K and 5th grade, the three study conditions were fairly comparable at all time-points as demonstrated by Table 3 which displays the standardized differences between TRIAD-NFT and CTRL and TRIAD-FT and CTRL in the pretest measure that captured children's mathematical competence at baseline.

Intervention / Program / Practice: The Follow Through treatment was provided by the TRIAD staff to all kindergarten and (the following year) first grade teachers in the 12 originally randomized follow-through schools. For each cohort, the professional development consisted of approximately 7 half-day sessions (approximately a total of 24 instructional hours). At the beginning of each academic year, teachers were informed that some of the students enrolled in their classrooms had experienced the Building Blocks prekindergarten curriculum, and were now participating in a longitudinal study. These teachers were expected to teach their district-assigned curriculum, differentiating their instruction as needed to facilitate continued growth in mathematical learning.

In pre-K, one group was a business-as-usual control, and the other two conditions were identical and entailed TRIAD scale-up (intensive training, coaching focused on both BB curriculum and linked assessment and online professional development). After preschool, one of the treatment conditions included follow-through professional development through 1st grade, including knowledge of the intervention and ways to build upon that knowledge using learning trajectories (TRIAD Follow-Through, TRIAD-FT) while the other condition implemented business-as-usual control (TRIAD Non-follow-through, TRIAD-NFT). The control groups in both Buffalo and Boston employed their districts' mathematics curriculum, including: DLM Childhood Express (whose math component is actually the first published version of the pre-K *Building Blocks* curriculum) and *Investigations in Number, Data, and Space* (Kindergarten and 1st grade).

Research Design: Participating schools were randomly assigned to one of three conditions described above (TRIAD-NFT, TRIAD-FT, and Control) separately within each district.

Data Collection and Analysis: At pre-K through 1st grade, the Research-based Elementary Mathematics Assessment (REMA; Clements, Sarama, & Liu, 2008) assessed children's mathematical competence. Children were assessed individually by trained interviewers; assessment sessions were videotaped and monitored for quality control. At 4th and 5th grade, the Grade 3-5 Tools for Elementary Assessment in Mathematics (Grade 3-5 TEAM; Clements, Sarama, & Wolfe, 2011), which was designed to be self-administered by upper-elementary children (using paper and pencil), was used to measure the age-appropriate competencies in learning trajectories measured by the REMA.

We estimated the long-term effects of TRIAD-NFT and TRIAD-FT using a 2-level logistic regression model that can be considered as an elaboration of the basic Rasch item response model. The level-1 model is a random intercept logistic regression model that specifies the scored response for each item as a function of the random intercept representing the latent unobservable knowledge level (or math ability) of a student and a vector of dummy variables indicating the test item to which a response belongs. The level-2 model specifies the latent student knowledge as a function two indicators for the NFT and FT conditions and other student and school-level covariates, including: student-level measures for pretest scores at the start of Pre-K, free/reduced price lunch (FRPL) eligibility, gender, and race/ethnicity and school-level percentages of students eligible for FRPL and with English Learner status. This model is estimated separately for each follow-up assessment administered in 3rd, 4th, and 5th grade using the GLLAMM module in Stata (Rabe-Hesketh, Skondral, and Pickles, 2005) and standard errors of the estimated coefficients adjusted for the clustering of students in the original study schools. We applied this methodology with the assessment data collected in pre-K, K, and 1st grade so that we could compare the long-term effects to the effects realized during the original study.

In these models, we interpreted the coefficients on the TRIAD-NFT and FT indicators as the average differences in math achievement between each of the treatment conditions and the CTRL condition or the effects of the FT and NFT conditions, respectively, and converted them to effect sizes using the estimated standard deviation of the random intercept. The only difference between the analyses conducted for RQ1 and RQ3 is that the latter did not include students who did not stay in the original study schools during the study period. Analyses conducted for RQ2 and RQ4 included interactions between the indicators for the subgroups of interest (e.g., high SES dummy) and the indicators for TRIAD-NFT and TRIAD-FT.

Findings / Results: Figure 1 presents results addressing the first research question that pertain to the differences between the three groups at the end of pre-K through 5th grade. We see large and statistically significant effects for TRIAD-FT and TRIAD-NFT at pre-K (0.86 and 0.75 standard deviations (sds) or effect sizes); impacts decay in K and 1st grade by similar amounts for the two groups (about 0.4 sds between pre-K and K and about 0.2 sds between K and 1st grade). Effects in K remain statistically significant while the 1st grade effect is significant at the $p<0.1$ level only for the TRIAD-FT group. The differences between the FT and NFT effects at K and 1st grade may be attributed to the FT component but they could also reflect the larger impact observed for the FT group at pre-K. While effects for the FT and NFT conditions are not statistically distinguishable from zero in 3rd and 4th grade, the effects are statistically significant at the end of 5th grade – about 0.26 sds for FT and about 0.21 sds for NFT.

Figures 2a and 2b inform the second research question and show the effects of the FT and NFT conditions separately for African-American, Hispanic, other minority, and white students. For African-American students in FT, the estimated effects are large and statistically significant at all-time points (between 0.23 and 1.15 sds). For African-Americans in NFT, effect estimates were smaller and only significant at pre-K and K and marginally significant ($p<0.10$) at 5th grade. Again, the observed differences between the FT and NFT effects for African-Americans may reflect the initial difference in the effects at pre-K and/or the FT intervention component. We observe that White students experienced the second largest impacts in pre-K, with a slightly larger initial impact for those in NFT. For whites in FT, decay between pre-K and K was the smallest, but the impact lost its significance thereafter (only significant at $p<0.1$ in 4th and 5th grades). For whites in NFT, the impact was large and significant in pre-K, K, 3rd grade, and 5th grade. Specifically, the impact at 3rd grade and 5th grade was about 0.5 sds and the largest among all subgroups and conditions. For Hispanic students, we observe significant (but smaller impacts than African-American or white students) in pre-K, which decayed and lost its statistical significance at subsequent grades with or without FT. The other racial/ethnic groups (including Asian students) were the smallest in size, therefore, the effects were very imprecise.

Figures 3a and 3b show the FT and NFT effect estimates separately for the low and high SES subgroups formed based on students' eligibility for FRPL. For lower SES students, we observe statistically significant effects in pre-K and K for both FT and NFT, but the impacts for both conditions are not significant in later grades. We see significant effects for higher SES students in NFT at all-time points; effects in 4th and 5th grades are especially large (about 0.6-0.7 sds). For higher SES students in FT, only the pre-K, K, and 5th grade impacts are statistically significant.

Figures 4, 5a, 5b, 6a, and 6b inform the 3rd and 4th research questions and show parallel results for the stayers (students who stayed in the original conditions between pre-K and 1st grade). One consistent observation is that the effects for a given condition and group is generally larger but less precise for stayers than the original study groups. For example, Figure 4 shows that the FT and NFT effects for stayers were generally larger than the effects for the full analytic sample and statistically significant at all-time points except for the FT effect in 4th grade. In addition, effects for higher SES stayer students in both FT and NFT were considerably larger in 3rd, 4th, and 5th grades than the effects estimated for high SES students in the original sample.

Summary and Discussion: Impacts at the end of 5th grade were statistically significant and considerably larger than impacts in 3rd and 4th grade for both conditions and all subgroups. The pattern of re-emergence of the effect size at two critical points in elementary education—the transitions to the increasing demands of the curricula of first and fifth grades suggests that long-term effects of early development of a full range of mathematical proficiencies are more likely to be evinced in periods in which expectations in mathematical complexity and sophistication increase substantially (CCSSO/NGA, 2010; Powell et al., 2013; Westat, 2001).

We see larger impacts for African American students in FT than NFT at almost all time points (which led to a similar pattern for the full sample), while the opposite pattern is observed for White students and students from higher SES families. As mentioned above, these patterns may reflect the larger impacts at Pre-K for African-American students in FT than NFT and for White students in NFT than FT even though the two treatment conditions were identical during that first year. But it may also reflect that the value-added of the FT component for African-American students. For example, centering instruction around learning trajectories focuses teachers' attention on students' thinking and learning of mathematics, and what children can learn to do, may have helped reduce biases and change teachers' views of African-American students' learning from a deficit perspective; and/or the TRIAD intervention may have promoted a conceptual and problem-solving approach infrequently emphasized in schools serving low-income children, explicitly supporting African-American students' participation in increasingly sophisticated forms of mathematical communication and argumentation).

Finally, and perhaps not surprisingly, we observe larger effect estimates at most of the time points for students who stayed in the original study conditions for full three years compared with the original study sample, which implies that effects for "leavers" (students who transferred from the original schools and conditions but were followed and administered the follow-up tests to the extent possible) were smaller. This finding may indicate the importance of receiving the "full dose" of an intervention, but it may very well reflect the potential adverse influence on student achievement of factors that also led to the mobility of these students.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

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

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

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. Journal of Econometrics 128: 301-323.

Appendix B. Tables and Figures

Not included in page count.

Appendix B. Tables and Figures

Table 1. Demographic Characteristics of the Original Study Schools

	Enrollment	SES	LEP	Race/ethnicity					
Research Group	**Average school enrollment**	**% Eligible for free/ reduced lunch**	**% English Language Learning**	**NA/AI**	**AA/AA**	**AP/AS**	**H/HI**	**W/WH**	**Other**
All	475.7	84.53	14.17	1.2%	51.28%	3.87%	26.20%	16.59%	0.77%
TRIAD-FT	464.8	82.73	14.84	2.55%	52.44%	5.11%	24.85%	14.40%	0.625%
TRIAD-NFT	479.1	86.45	12.18	0.63%	52.63%	2.33%	25.33%	18.19%	0.873%
Control	480.9	84.21	15.4	0.86%	49.22%	4.32%	27.99%	16.82%	0.79%

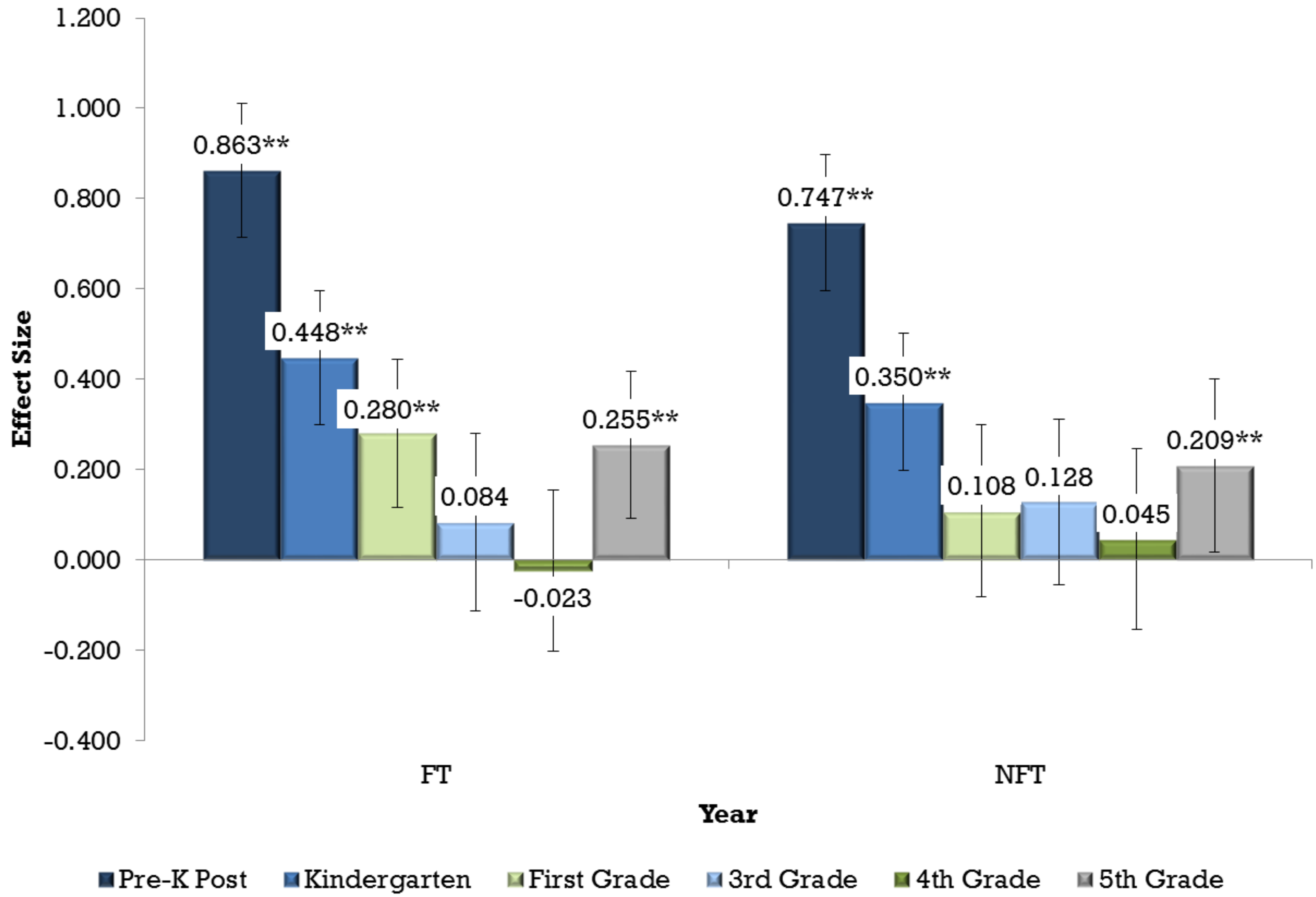
Table 2. Number of Students in the Analytic Samples

Grade-Level	TRIAD-NFT	TRIAD-FT	CTRL
Panel 1: Original Sample			
Pre-K Baseline	484	495	396
Pre-K Posttest	456	471	378
Kindergarten	418	439	361
1st Grade	385	405	336
3rd Grade	266	260	210
4th Grade	298	288	243
5th Grade	279	260	242
Panel 2: Stayers Subsample			
Pre-K Baseline	253	262	235
Pre-K Posttest	253	262	235
Kindergarten	253	262	235
1st Grade	253	262	235
3rd Grade	153	165	142
4th Grade	166	184	160
5th Grade	153	172	158

Table 3. Standardized Baseline Differences Between Study Groups on the Pretest

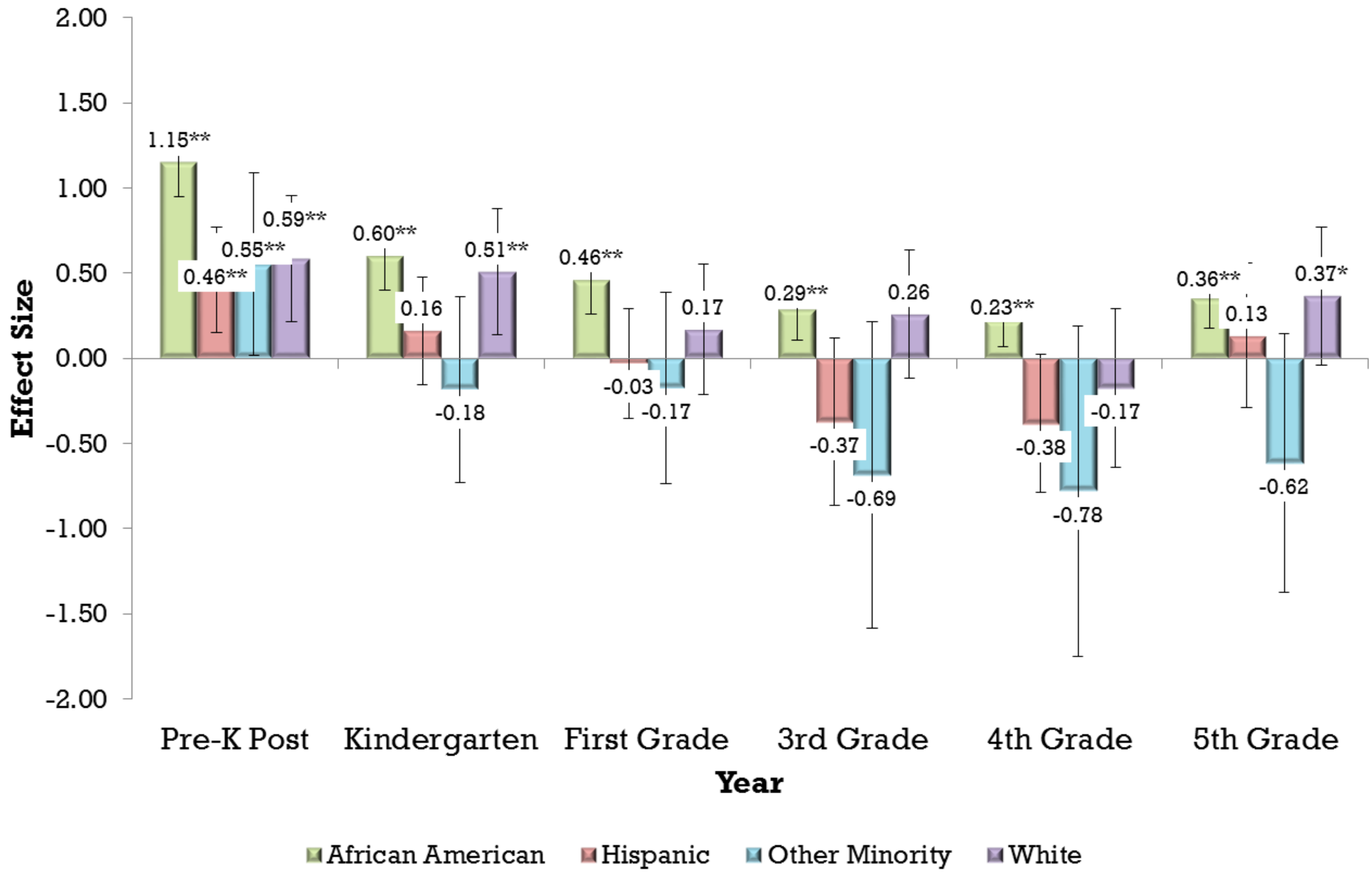
Grade-Level	TRIAD-NFT vs. CTRL	TRIAD-FT vs. CTRL
Panel 1: Original Sample		
Pre-K Posttest	-0.086	-0.108
Kindergarten	-0.076	-0.099
1st Grade	-0.048	-0.095
3rd Grade	-0.095	-0.022
4th Grade	-0.062	0.007
5th Grade	-0.046	-0.006
Panel 2: Stayers Subsample		
Pre-K Posttest	-0.085	-0.039
Kindergarten	-0.085	-0.039
1st Grade	-0.085	-0.039
3rd Grade	-0.171	0.041
4th Grade	-0.136	0.085
5th Grade	-0.178	0.042

Figure 1. Impacts of TRIAD-FT and TRIAD-NFT (Original Study Groups)



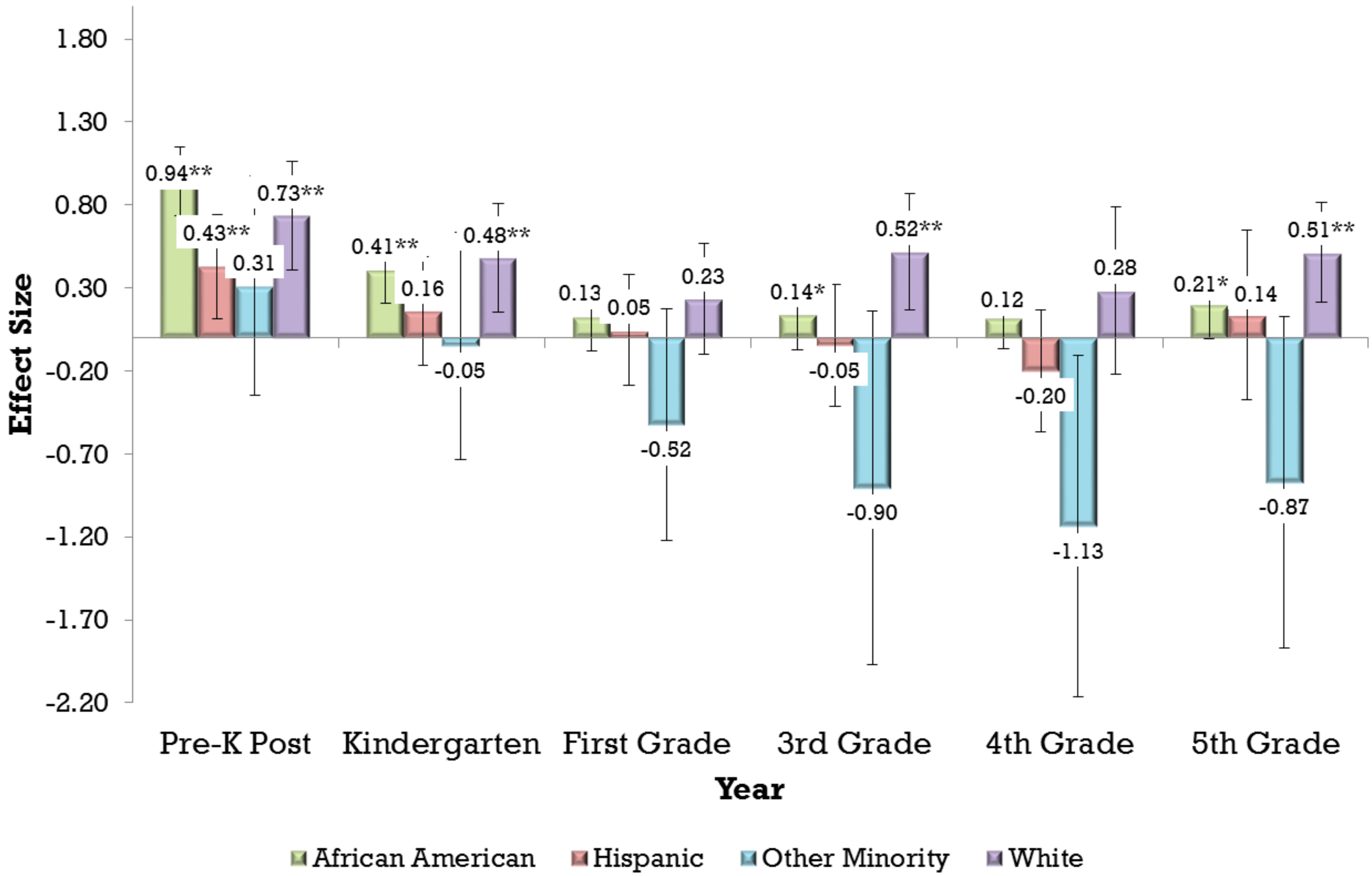
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 2a. Race/Ethnicity Subgroup Impacts for TRIAD-FT (Original Study Groups)



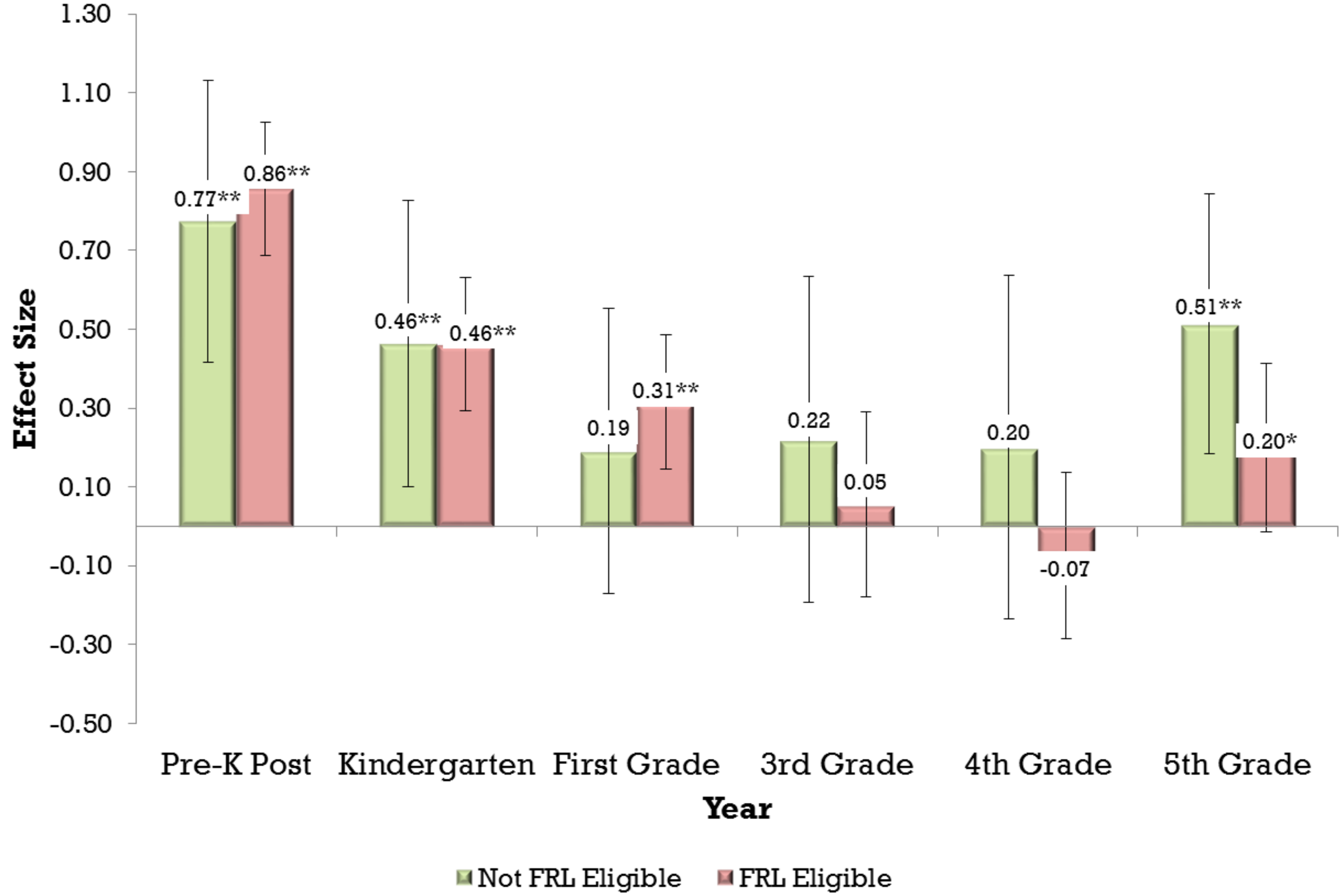
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 2b. Race/Ethnicity Subgroup Impacts for TRIAD-NFT (Original Study Groups)



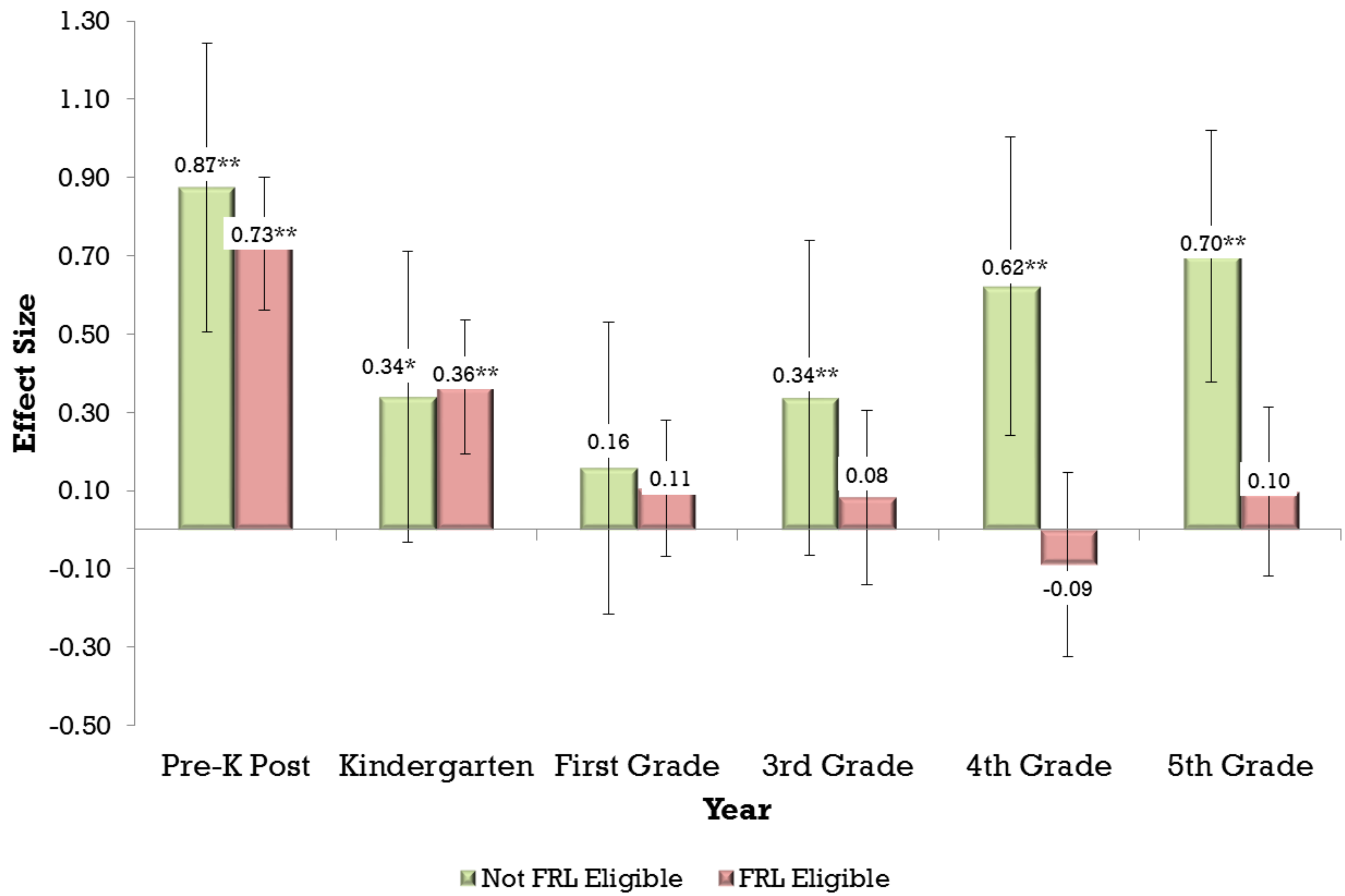
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 3a. SES Subgroup Impacts for TRIAD-FT (Original Study Groups)



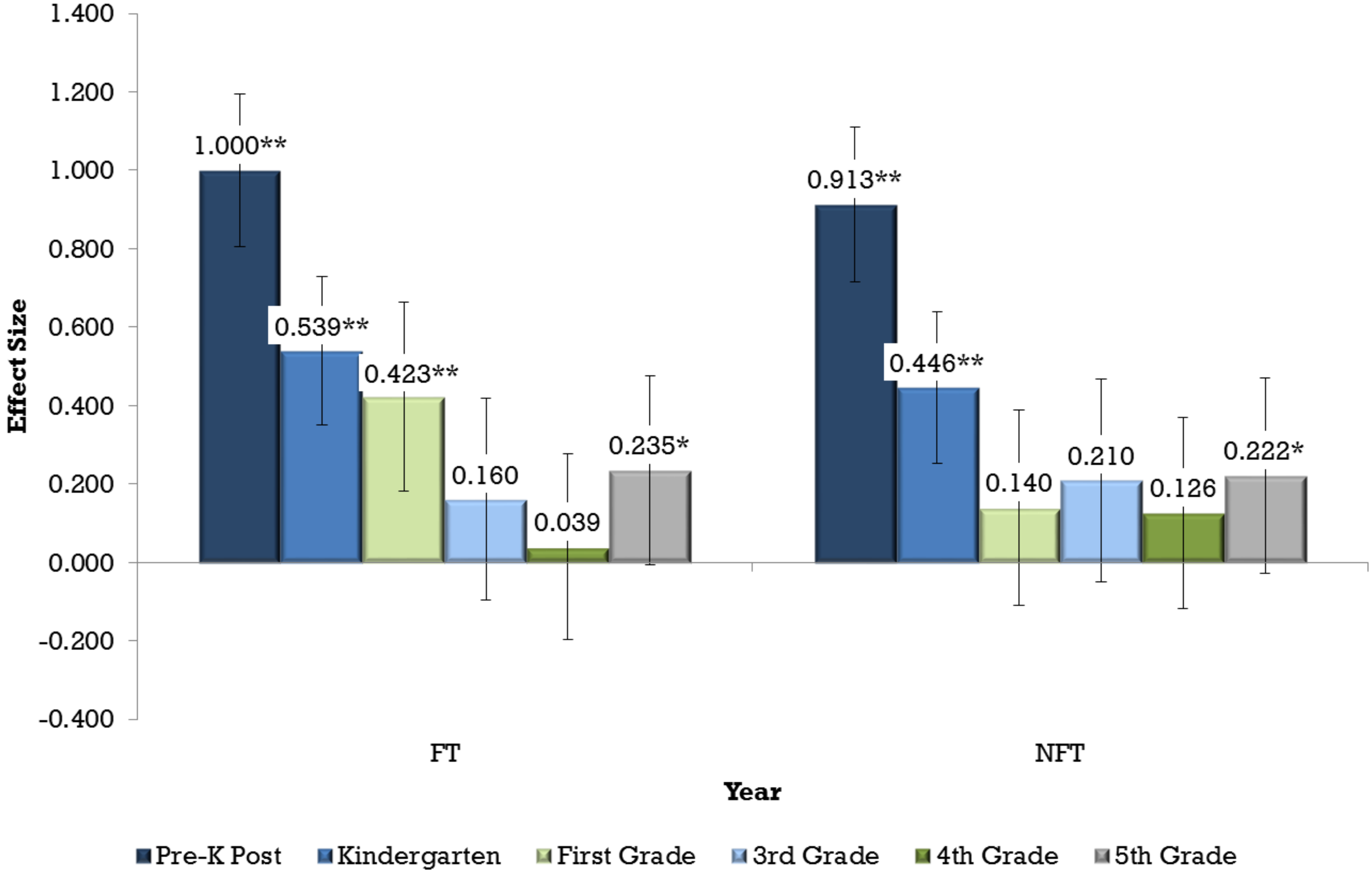
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 3b. SES Subgroup Impacts for TRIAD-NFT (Original Study Groups)



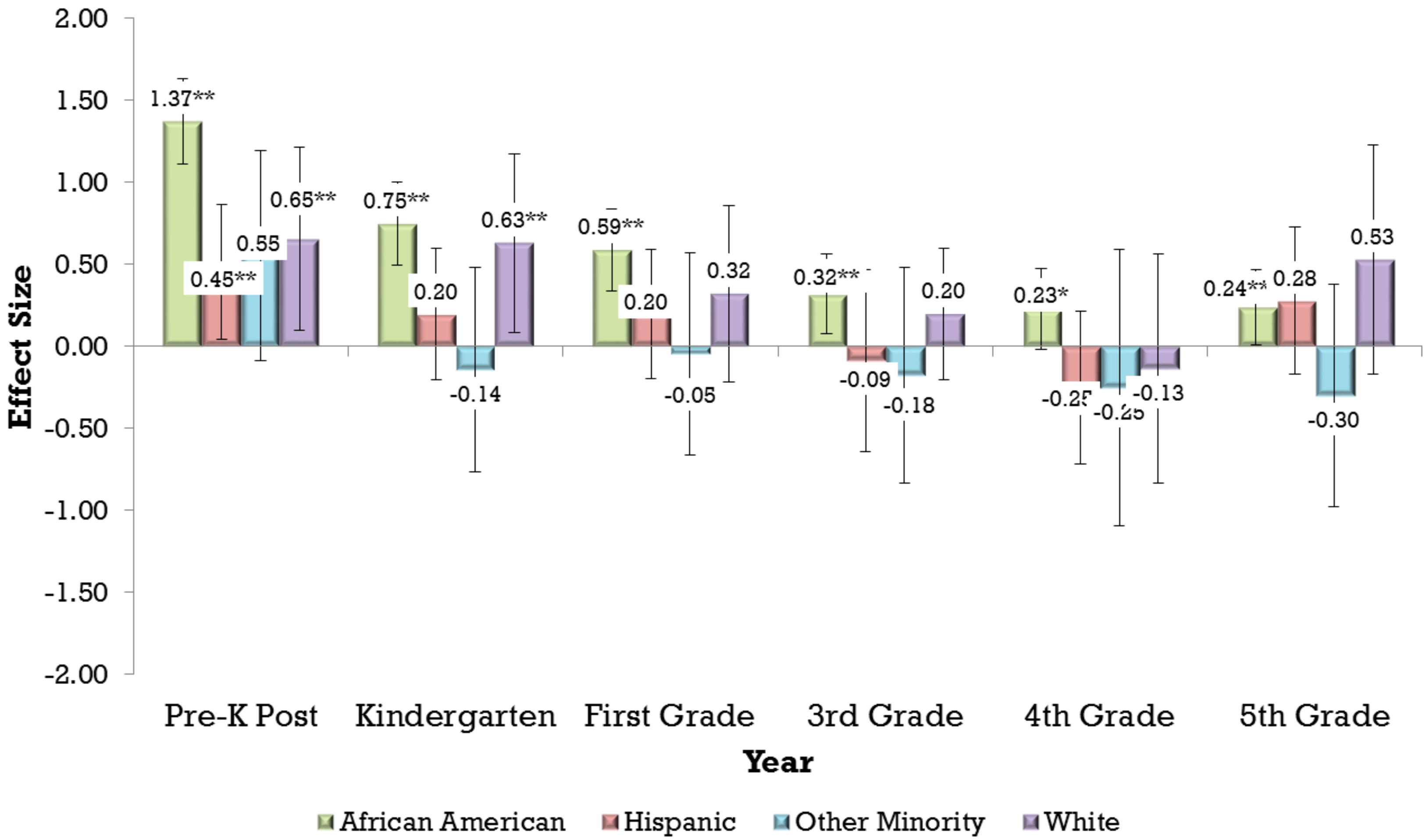
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 4. Impacts of TRIAD-FT and TRIAD-NFT (Stayers)



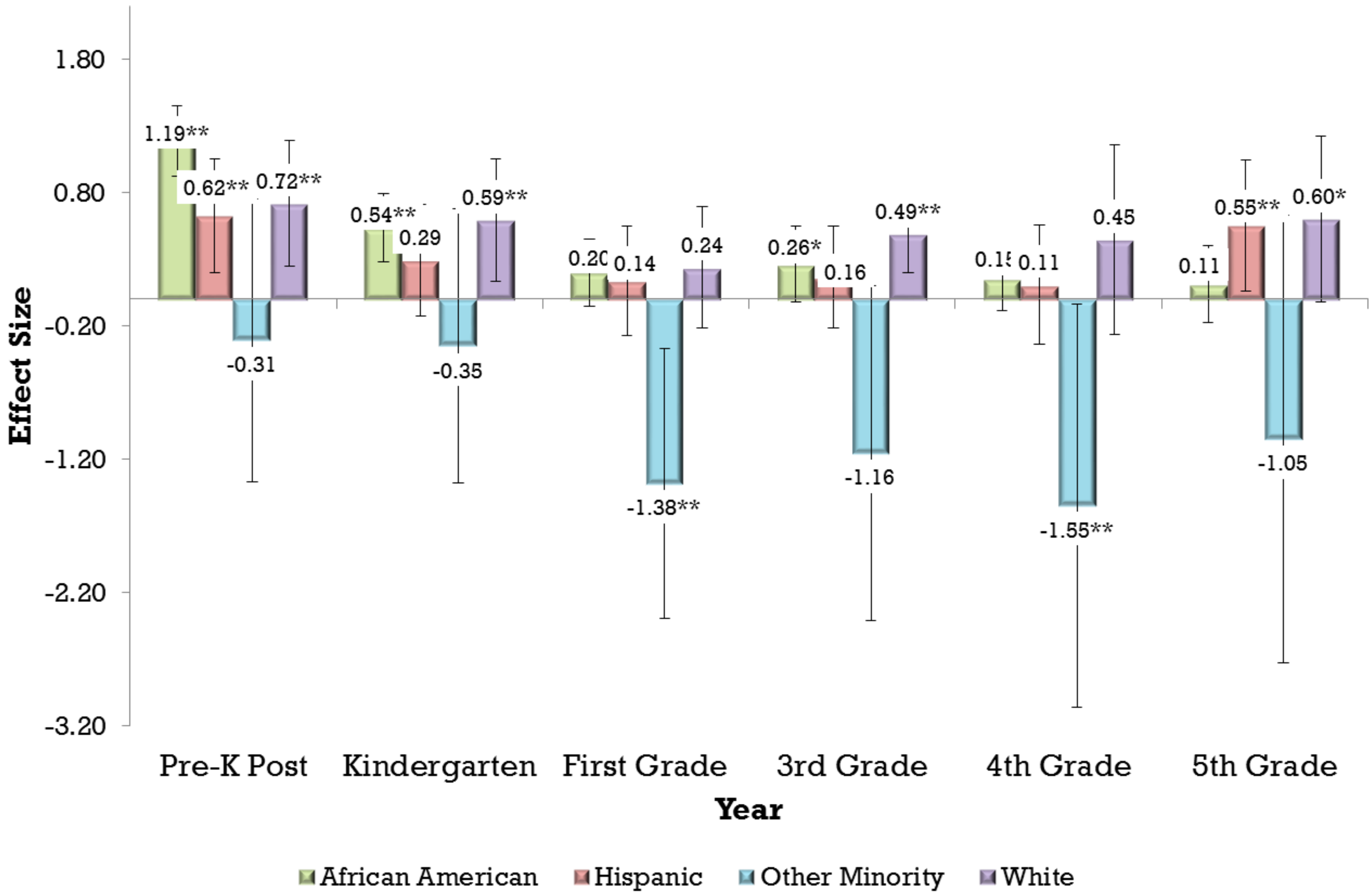
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 5a. Race/Ethnicity Subgroup Impacts for TRIAD-FT (Stayers)



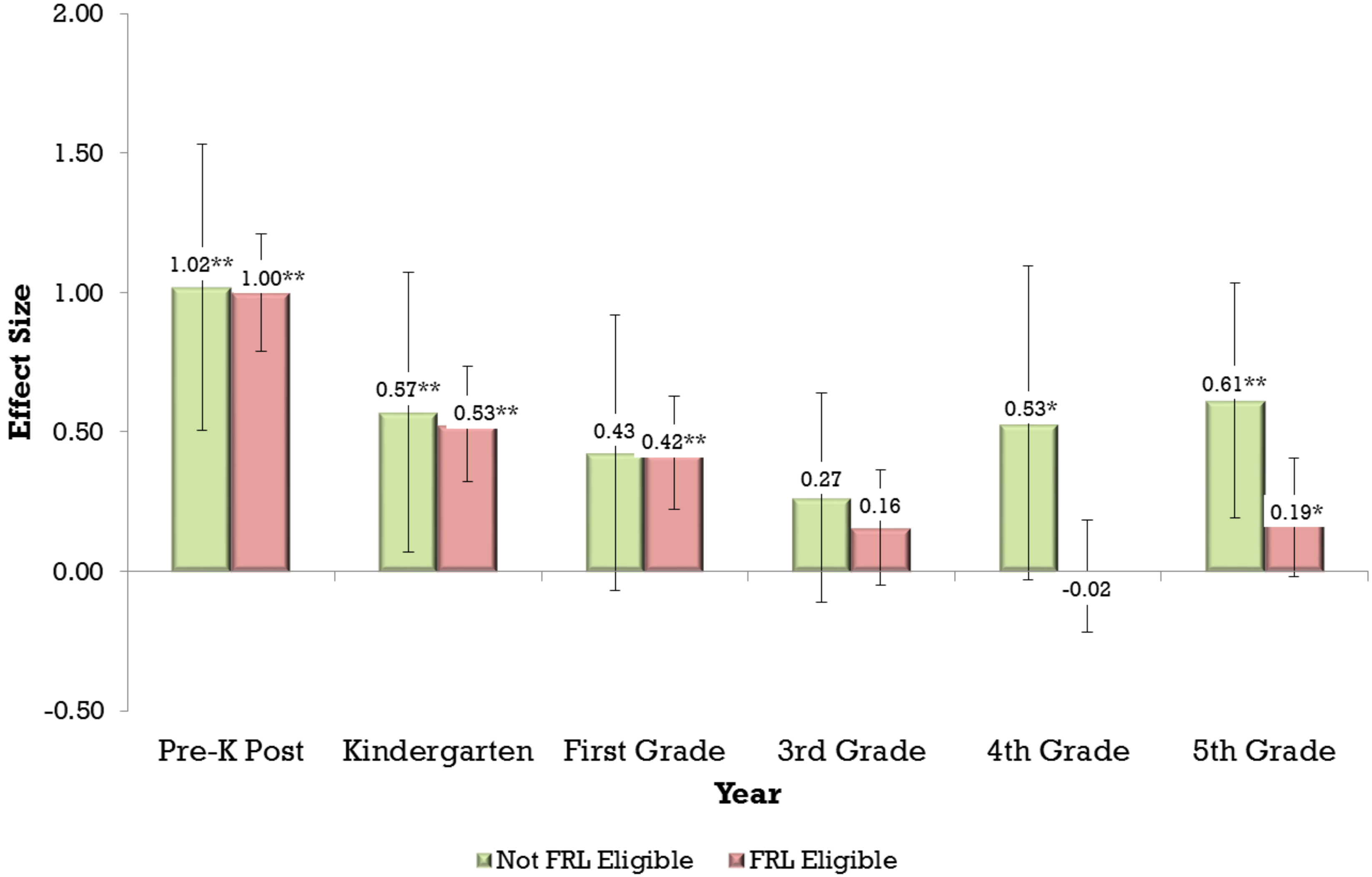
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 5b. Race/Ethnicity Subgroup Impacts for TRIAD-NFT (Stayers)



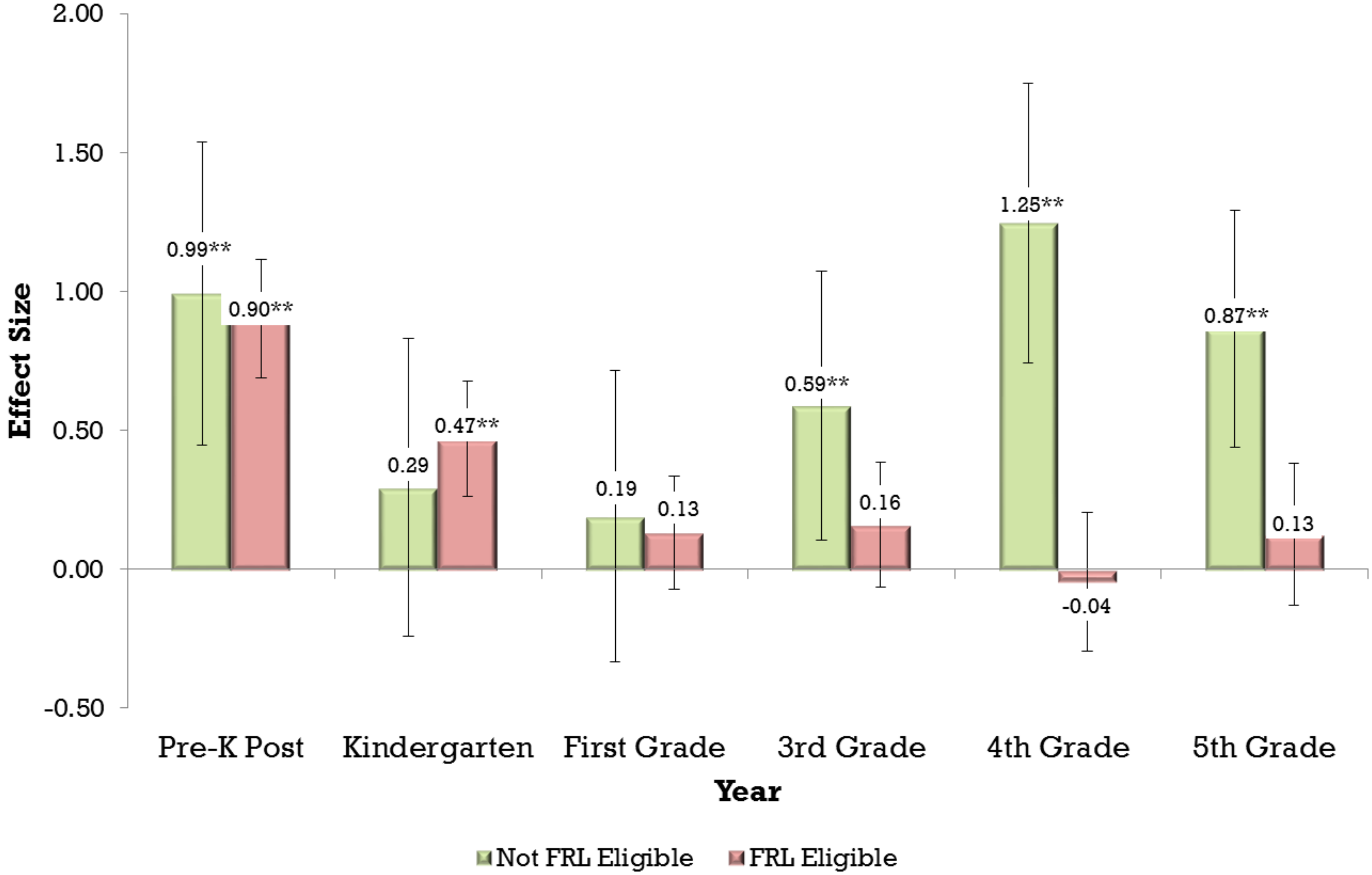
** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 6a. SES Subgroup Impacts for TRIAD-FT (Stayers)



** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.

Figure 6b. SES Subgroup Impacts for TRIAD-NFT (Stayers)



** and * denote statistically significant differences at the p<0.05 and p<0.1 levels respectively.