# NURTURES*
# Longitudinal Summary of Project Impact on Students' Mathematics, Reading, & Science Learning

**Gale A. Mentzer, PhD**
**Peter Paprzycki, PhD**
**May 6, 2020**
**Minor Revisions 1/5/24**
**based in WWC**
**inquiries.**

*NSF Grant # 1102808

## Table of Contents

## Abstract

Because standardized tests in science are not given to PreK-3 students in Ohio, this report examined the longitudinal effects of learning from a teacher who had participated in the NURTURES professional development program. Specifically, it looked at the effects on students' mathematics and reading learning in grades 2-5 and science learning in 5th grade in 2017. Students who were in 5th grade at that time could have had a NURTURES-trained teacher at any time between kindergarten and 3rd grade. Thus, the study followed students up to 5 years after having a NURTURES teacher.

The sample included the population of students enrolled in the 41 elementary schools in the Toledo Public School District. Students who never learned from a teacher who participated in NURTURES served as the control group. The data came from the 2017 administration of the Ohio Measure of Academic Progress (MAP) (NWEA, 2019) for mathematics and reading and the Ohio Achievement Test in Science for science (Ohio Achievement Assessment, 2015).

The total number of students from these schools who took the May 2017 reading MAP was 6759 and the total number who took mathematics was 6703. The number of those students who had at least one NURTURES teacher was 2801 (41.4%) for reading and 2707 (41.6%) for mathematics. Analysis of the reading scores showed 2.14 advantage points for NURTURES students as compared to the average non-intervention student to an annual growth rate of 7.02 units ($p < .001$). The treatment effect size (Hedges' $g$) was 0.12. For mathematics there were 1.55 advantage points to an annual growth rate estimated to be 8.17 units ($p < .001$) as compared to the average non-intervention student. The treatment effect size (Hedges' $g$) was 0.09.

Analysis of the 5th grade Ohio Achievement Science Subtest showed that students associated with at least one NURTURES-trained teacher was modeled to have a 5.86 advantage points as compared to the average non-intervention student. The treatment effect size (Hedges' $g$) was 0.08, which is to be interpreted as a treatment group having, on average, 0.08 higher scores in standard deviation units as compared to the scores of the control cohort.

When compared with our earlier evaluation report (2016; revised in 2018), we see that students who had a NURTURES-trained teacher, on average, continued to show greater gains compared to students who did not. In addition, the achievement gaps between non-minority and minority students in reading and mathematics were reduced when the minority students had a NURTURES-trained teacher and the non-minority students did not. In science, the impact of the intervention roughly compensated for the attainment gap between boys and girls and partially ameliorated the gap between minority and non-minority children's scores associated with these demographic factors.

## l) Impact on student reading and mathematics ability

This evaluation study was conducted in the 41 elementary schools in the Toledo Public school district, which has a high degree of racial diversity and 64.8% of students received free or reduced lunch. Students who completed the Ohio Measure of Academic Progress (MAP) (NWEA, 2019) assessment for 2nd through 5th grade in 2017 were selected. The MAP measures student achievement in mathematics, reading, language usage, and science (assessed only at 5$^{th}$ grade). MAP tests are scaled vertically across grades to allow for the measurement of growth. In this section, we describe the analyses conducted with MAP scores in reading and mathematics.

All students from the school district who completed the 2nd through 5th grade reading and/or mathematics test were selected to examine long-term effects (the total population). The total number of students from these schools who took the May 2017 reading MAP was 6759 and the total number who took mathematics was 6703. The number of those students who had at least one NURTURES-trained teacher anytime between kindergarten and 3$^{rd}$ grade (treatment group) was 2801 (41.4%) for reading and 2707 (41.6%) for mathematics. Students who never had a NURTURES-trained teacher between kindergarten and 3$^{rd}$ grade made up the control group. Each school included both treatment and control group students and teachers.

**Baseline.** Standardized scale scores from the STAR assessment (Renaissance Learning, 2013) administered in the Fall of the first year of the program (when students were at the beginning of Kindergarten) were used as a baseline covariate. This assessment is nationally-normed and deemed reliable and valid by several independent groups. It was selected because it measures three broad early skills (Word Knowledge and Skills, Comprehension Strategies and Constructing Meaning, and Numbers and Operations) using 27 items derived from state standards, the Common Core standards, and current research. This baseline measure was available for 4549 students (67.3%) for the reading sample and 4549 students (67.3%) and 4507 (67.2%) for the mathematics cohort.

Because the student population was transient in this urban district, there were students who joined the school district after the K-grade, consequently 32.7% and 32.8% of the examined reading and mathematics samples, respectively, had missing baseline measures. As a result, multiple imputation (MI) was performed to estimate missing values with the use of pan (Zhao & Schafer, 2018) and mitml (Grund, Robitzsch, & Lüdtke, 2018) R packages (R Core Team, 2018) using the multivariate empty model. Outcome measures were not imputed.

The baseline equivalence for reading and mathematics was tested based on an imputed sample scenario in which the treatment and control means and corresponding standard deviations were averaged from 100 multiple imputed samples. Only baseline data had missing cases. All other variables used in the analyses were complete, including the outcome measures. With each analytic sample used in the reading and mathematics analyses, we calculated the four bounds corresponding to Equations B.7.0, B.7.1, B.7.2, and B.7.3 on page B-3 of the What Works

Clearinghouse Standards Handbook, Version 4.1 (WWC, 2020). The estimated bounds are presented in Table 1, which shows that all four bounds for the baseline differences are below the maximum threshold of absolute value of 0.25 for both Reading and Mathematics achievement outcome measures; a requirement outlined by the WWC (2020).

***Table 1. Baseline differences (in effect sizes) and corresponding bounds for analytic samples used in the reading and mathematics analyses***

| | | Reading | Mathematics |
|---|---|---|---|
| Baseline effect size on imputed data | $g_{xi}$ | 0.065 | 0.051 |
| Bounds | | | |
| B.7.0 | | 0.065 | 0.051 |
| B.7.1 | | 0.115 | 0.095 |
| B.7.2 | | 0.074 | 0.072 |
| B.7.3 | | 0.106 | 0.075 |

**Predictor and outcome variables.** Three analyses models were performed, one for each of the Ohio MAP outcome domains (reading, mathematics, and science measures). Each analysis model included the respective outcome measure, treatment group indicator, as well as the following covariates: students' baseline STAR standardized scale score (reading, mathematics, or science), minority status, gender, grade level. School id was used as a grouping factor in a two-level hierarchical model. To elaborate, the predictor variables in the model were students' STAR baseline measure; minority status (levels: 0 = minority; 1 = non-minority or white), gender (0 = female; 1 = male), grade level (levels: grades 2 through 5), and treatment condition (0 = no NURTURES-trained teachers in grades the following academic years: 2012-13, 2013-14, 2014- 15; (1 = at least one NURTURES-trained teacher in these academic years). The predictor variable of interest was the treatment condition variable (coded as 1 = treatment, 0 = control), which indicated that a 5th-grader had been placed in a classroom with a NURTURES-trained teacher at least once in first-through-third grade at least once. The outcome variable was student performance on the 5th grade MAP reading and mathematics assessments.

**Analyses - First and second level equations**. A two-level random slope model was used for reading achievement data, where the intervention variable was modeled to be random across schools. Mathematics achievement was modeled as a random intercept model due to a convergence problem. The first-level equation included the student's intercept (mean value of student science achievement), student's baseline STAR standardized scale score, minority status, gender, grade level, and the treatment group indicator, random intercept related to schools and finally a random error or a residual variation. The intercept value was modeled to be random as a

function of second-level units (schools). The model was unconditional at the school-level; it did not include school-level variables. Analysis was conducted using lme4 R package (Bates, Maechler, Bolker, & Walker 2015) using a pooled analysis across 100 imputed samples. The equations were

**Level-1 Equation Model**

$$\text{Test Score}_{ij} = \beta_{0j} + \beta_{1j}\cdot(\text{Baseline Test Score}_{ij}) + \beta_{2j}\cdot(\text{Minority Status}_{ij}) + \beta_{3j}\cdot(\text{Gender}_{ij}) + \beta_{4j}\cdot(\text{Intervention}) + r_{ij}$$

**Level-2 Equation Model**

$$\beta_{0j} = \gamma_{00} + u_0$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40}$$

$$\beta_{5j} = \gamma_{50} + u_{5j}$$

**Mixed Model**

$$\text{Test Score}_{ij} = \gamma_{00} + \gamma_{10}\cdot\text{Baseline Test Score}_{ij} + \gamma_{20}\cdot\text{Minority Status}_{ij} + \gamma_{30}\cdot\text{Gender}_{ij} + \gamma_{40}\cdot\text{Grade}_{ij} + \gamma_{50}\cdot\text{Intervention}_{ij} + u_{0j} + u_{5j}\cdot\text{Intervention}_{ij}^{*} + r_{ij}$$

Note. "*i*" denotes a student and "*j*" denotes a school. *Mathematics achievement was modeled as a random-intercept model only.

**Results.** Table 2 provides the raw or unadjusted posttest standard deviations. Table 3 presents a model summary for students' Spring 2017 reading scores using the symbols conventions adopted by Raudenbush, Bryk, Cheong, Congdon, & Du Toit (2011). The intercept value of 191.42 designated as $\gamma_{00}$ represents a minority (non-White), non-intervention (no exposure to a NURTURES-trained teacher) female student with an average STAR baseline measure. Of central interest is the student intervention slope ($\gamma_{50}$), which indicates that a student associated with at least one NURTURES-trained teacher was modeled to have a 2.14 advantage points as compared

to the average non-intervention student described above. An annual growth

***Table 2. Unadjusted posttest (outcome) standard deviations***

| Outcome | Sample Size | Unstandardized Regression Coefficient | Unadjusted Standard Deviation |
|---|---|---|---|
| Reading | 6,759 | 2.14 | 17.82 |
| Mathematics | 6,703 | 1.55 | 17.14 |
| Science | 1,588 | 5.85 | 39.85 |

rate ($\gamma_{40}$) was estimated to be 7.02 units. The treatment effect size (Hedges' *g*) was 0.12, which is to be interpreted as a treatment group having, on average, 0.12 higher scores in standard deviation units as compared to the scores of the control cohort.

***Table 3. Model summary for reading test score analysis (*N = 6759)****

| Fixed Effects | B | *SE* B | *t*-ratio | Approx. *df* | *p* |
|---|---|---|---|---|---|
| INTRCPT1, $B_0$ | | | | | |
| INTRCPT2, $\gamma_{00}$ | 191.42 | 1.01 | 188.85 | 7344864 | <.001 |
| BASELINE slope, $B_1$ | | | | | |
| INTRCPT2, $\gamma_{10}$ | 0.07 | 0.00 | 29.53 | 1487 | <.001 |
| MINORITY STATUS slope, $B_2$ | | | | | |
| INTRCPT2, $\gamma_{20}$ | 2.17 | 0.37 | 5.91 | 19108 | <.001 |
| GENDER slope, $B_3$ | | | | | |
| INTRCPT2, $\gamma_{30}$ | -1.32 | 0.33 | -3.95 | 17996 | <.001 |
| GRADE slope, $B_4$ | | | | | |
| INTRCPT2, $\gamma_{40}$ | 7.02 | 0.15 | 45.42 | 10946 | <.001 |
| INTERVENTION, $B_5$ | | | | | |
| INTRCPT2, $\gamma_{50}$ | 2.14 | 0.45 | 4.72 | 91171 | <.001 |

***** *Final Parameter Estimates and Inferences Obtained from 100 Imputed Data Sets.*

There were other findings of interest. Non-minority students statistically significantly outperformed minority students by 2.17 points ($\gamma_{20}$), and male students statistically significantly underperformed in comparison to female counterparts by 1.32 points ($\gamma_{30}$). *However, for minority students, having a NURTURES-trained teacher reduces their achievement gap with non-minority students without a NURTURES-trained teacher.* The conditional inter-class correlation coefficient was 0.188, indicating that once level-1 predictor variables were specified, approximately 18.8% of modeled variance is due to school effects.

Table 4 summarizes a model summary for students' Spring 2017 mathematics score. The intercept value for this cohort was 192.50 designated as $\gamma_{00}$ represents a minority (non-White), non-intervention (no exposure to a NURTURES-trained teacher) female student who had an average STAR baseline measure. The intervention slope ($\gamma_{50}$) was estimated to have a 1.55

advantage points as compared to the average non-intervention student described above. An annual growth rate ($\gamma_{40}$) was estimated to be 8.17 units. The treatment effect size (Hedges' *g*) was 0.09, which is to be interpreted as a treatment group having, on average, 0.09 higher scores in standard deviation units as compared to the scores of the control cohort.

***Table 4. Model summary for mathematics test score analysis (N = 6703)****

| Fixed Effects | B | *SE* B | *t*-ratio | Approx. *df* | *p* |
|---|---|---|---|---|---|
| INTRCPT1, $B_0$ | | | | | |
| INTRCPT2, $\gamma_{00}$ | 192.50 | 0.93 | 206.57 | 3959802 | <.001 |
| BASELINE slope, $B_1$ | | | | | |
| INTRCPT2, $\gamma_{10}$ | 0.07 | 0.00 | 35.42 | 1619 | <.001 |
| MINORITY STATUS slope, $B_2$ | | | | | |
| INTRCPT2, $\gamma_{20}$ | 2.42 | 0.32 | 7.48 | 8935 | <.001 |
| GENDER slope, $B_3$ | | | | | |
| INTRCPT2, $\gamma_{30}$ | 1.68 | 0.29 | 5.76 | 11921 | <.001 |
| GRADE slope, $B_4$ | | | | | |
| INTRCPT2, $\gamma_{40}$ | 8.17 | 0.13 | 61.70 | 12068 | <.001 |
| INTERVENTION, $B_5$ | | | | | |
| INTRCPT2, $\gamma_{50}$ | 1.55 | 0.30 | 5.24 | 29442 | <.001 |

***** *Final Parameter Estimates and Inferences Obtained from 100 Imputed Data Sets.*

There were other findings of interest. Non-minority students statistically significantly outperformed minority students by 2.42 points ($\gamma_{20}$), and male students statistically significantly outperformed female counterparts by 1.68 points ($\gamma_{30}$). The conditional inter-class correlation coefficient was 0.210, indicating that once level-1 predictor variables were specified, approximately 21.0% of modeled variance can be attributed to school effects.

**Conclusions.** Our analysis provides evidence for the efficacy and longitudinal effects of the NURTURES program on student outcomes in reading and mathematics after student variables, including gender and ethnicity were considered, and after properly accounting for the school context or between-schools variation and adjusting for baseline equivalence. Additionally,

baseline equivalence was established for all students for literacy and mathematics using students' STAR assessment scores.

**Results from our previous Evaluation Report (Mentzer, G.A. & Paprzycki (2016).** ***Longitudinal Effects of Participation in the NURTURES Project at The University of Toledo.*** This report, which can be accessed at: http://nurtures.utoledo.edu/reports/NURTURES_ExternalSummaryEvaluation.pdf indicated that students who had a NURTURES-trained teacher showed net gains of 8.6 points on STAR Early Literacy, 17.0 points on STAR Mathematics, and 41.4 points on STAR Reading on the spring assessments compared to students who never had a NURTURES-trained teacher. An effect size of 0.25 was found with the STAR Reading score gains, *which is considered substantively important by the What Works Clearinghouse evidence standards* (What Works Clearinghouse, 2013).

The data in this longitudinal study was drawn from the same student population. The results of interest, students' reading and mathematics scores on MAP tests, indicate that *two years later, students who had a NURTURES-trained teacher, on average, continued to show greater gains compared to students who did not*. Students associated with at least one NURTURES-trained teacher in their academic career, prior to the 2017-2018 school year, were found to have 1.55 advantage points in mathematics and 2.14 advantage points in reading on the MAP. These advantage points, given the annual (nine-month) growth of 8.17 points and 7.02 points respectively, *correspond to approximately 1.7 months developmental advantage in mathematics and 2.8 months in reading for students who had a NURTURES-trained teacher. This* is an important finding, considering the majority of students in the study only spent nine months with a teacher who had gone through the program.

In terms of effect sizes (Hedges' *g*), the treatment effect was 0.09 for mathematics and 0.12 for reading based on student MAP test scores. These are small effect sizes. However, they indicate that students were impacted in ways that had long-term effects in subsequent elementary years.

With minority and non-minority students, there were additional results of interest related to achievement. *These data indicate that the achievement gaps between non-minority and minority students in reading and mathematics were reduced when the minority students had a NURTURES-trained teacher and the non-minority students did not.* The schools included in this study had high racial diversity and 64.8% of the students received free or reduced lunch. This is promising evidence, indicating that reading and mathematics achievement gaps can be addressed, in part, by providing Framework-aligned science instruction in early elementary classrooms. *Having a NURTURES-trained teacher can potentially have a strong impact on students from racially diverse backgrounds, years after exposure*.

## 2) Impact on science

Our examination of the longitudinal impact on students' science learning used the same 41 schools in the Toledo Public School district. It included any student who had been in the school system in grades 1-3 (ages 6-9) during the academic years of 2012–13, 2013–14, 2014–15, or 2015–16 and were currently in the 5th grade. The total number of students was 1588 with 434 (27.3%) of the 5th-grade students having had a NURTURES-trained teacher at least once in grades 1–3 (treatment group) with the remaining 1154 (72.7%) serving as control students. Males represented 52.6% and females represented 47.4% of the sample, and 60.6% of students self-identified themselves as minorities (non-White). Each school included both treatment and control group teachers and students. As with the examination of MAP scores, STAR (Renaissance Learning, 2013) standardized scale scores were used as the primary baseline covariate.

**Baseline.** The baseline equivalence for science analyses was tested based on an imputed sample scenario in which the treatment and control group means and corresponding standard deviations of the STAR scores were averaged from 100 multiple imputed samples. Multiple imputation was conducted using a combined treatment and comparison sample. As per WWC Procedures and Standards Handbook Version 4.1 used at the time of the 2020 study, the treatment variable (having a NURTURES-trained teacher) was included in the imputation model. Similar to the analyses conducted with reading and mathematics test scores, we calculated the four bounds corresponding to Equations B.7.0, B.7.1, B.7.2, and B.7.3 on page B-3 of the What Works Clearinghouse Standards Handbook, Version 4.1 (WWC, 2020). These bounds are reported in Table 4. We see that all the four bounds for the baseline effect sizes are below the maximum threshold of 0.25 for science required by WWC (2020).

***Table 4. Analytic full imputation reading and mathematics baseline effect sizes and corresponding bounds***

| | | Science |
|---|---|---|
| Baseline effect size on imputed data | $g_{xi}$ | 0.107 |
| Bounds | | |
| B.7.0 | | 0.107 |
| B.7.1 | | 0.073 |
| B.7.2 | | 0.010 |
| B.7.3 | | 0.170 |

**Predictor and outcome variables.** The predictor variables in the model were students' STAR baseline measure; minority status (levels: 0 = minority; 1 = non-minority or white), gender (0 =

female; 1 = male) and treatment condition (0 = no NURTURES teachers in grades the following academic years: 2012-13, 2013-14, 2014-15; 1 = at least one NURTURES-trained teacher in these academic years). The predictor variable of interest was the treatment condition variable which indicated that a 5th-grader had been placed in a classroom with a NURTURES-trained teacher in first-through-third grade at least once. The outcome variable in this study was student performance on the 5th grade Ohio Achievement Science Subtest.

**Science measure.** The science achievement data used in this study were collected as part of the district-wide Ohio Assessment Testing program. The structure of the Grade 5 Science Test consisted of two parts that were administered near the end of the 5th grade school year. Both parts of the test were fixed forms that were administered in an online format. After a student completed both parts of the test, the district combined the scores to yield a comprehensive view of each student's progress. Science content covered in the test included earth science (15-17 points), life science (19-21 points), and physical science (19-21 points) (Ohio Achievement Assessment, 2015).

**First and Second Level Equations.** In the current analysis, a two-level random slope model was used. The first-level equation included the student's intercept (mean value of student science achievement), student's baseline STAR standardized scale score, minority status, gender, a treatment variable, a random intercept and random treatment slope variation related to schools and finally a random error interpreted as a residual variation. The intercept value and the student-level treatment variable were modeled to be random as a function of second-level units (schools). The model was unconditional at the school-level; it did not include school-level variables. Minority status and gender variables were assumed to be invariant between schools. Analysis was conducted using lme4 R package (Bates et al., 2015) using a pooled analysis across 100 samples with the imputed baseline STAR variable (no outcome data were imputed). The equations were:

**Level-1 Equation Model**

$$\text{Test Score}_{ij} = \beta_{0j} + \beta_{1j} \cdot (\text{Baseline Test Score}_{ij}) + \beta_{2j} \cdot (\text{Minority Status}_{ij}) + \beta_{3j} \cdot (\text{Gender}_{ij}) + \beta_{4j} \cdot (\text{Intervention}) + r_{ij}$$

**Level-2 Equation Model**

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40} + u_{4j}$$

**Mixed Model**

$$\text{Test Score}_{ij} = \gamma_{00} + \gamma_{10}\cdot\text{Baseline Test Score}_{ij} + \gamma_{20}\cdot\text{Minority Status}_{ij} + \gamma_{30}\cdot\text{Gender}_{ij} + \gamma_{40}\cdot\text{Intervention}_{ij} + u_{0j} + u_{4j}\cdot\text{Intervention}_{ij} + r_{ij}$$

Note. "*i*" denotes a student and "*j*" denotes a school.

**Results.** Table 4 presents a model summary for 5th-grade Ohio spring 2017 science score data. The intercept value of 680.80 designated as $\gamma_{00}$ represents a minority, non-intervention (no exposure to a NURTURES-trained teacher) student (average kindergarten STAR baseline measure), and who obtained an average 5th-grade Ohio Science Achievement Assessment score in an average school. Of central interest is the student intervention slope ($\gamma_{40}$), which indicates that a student associated with at least one NURTURES-trained teacher was modeled to have a 5.85 advantage points as compared to the average non-intervention student described above. The treatment effect size (Hedges' *g*) was 0.08, which is to be interpreted as the treatment group having, on average, 0.08 higher scores in standard deviation units as compared to the scores of the control group.

**Table 4.** ***Summary of pooled fixed effects for 5th-grade Ohio spring 2017 science achievement data (N = 1588)****

| Fixed Effects | B | *SE* B | *t*-ratio | Approx. *df* | *p* |
|---|---|---|---|---|---|
| INTRCPT1, $B_0$ | | | | | |
| INTRCPT2, $\gamma_{00}$ | 680.80 | 3.31 | 205.62 | 567753 | <.001 |
| BASELINE slope, $B_1$ | | | | | |
| INTRCPT2, $\gamma_{10}$ | 0.16 | 0.01 | 14.31 | 519 | <.001 |
| MINORITY STATUS slope, $B_2$ | | | | | |
| INTRCPT2, $\gamma_{20}$ | 8.15 | 1.81 | 4.50 | 6227 | <.001 |

| | | | | | |
|---|---|---|---|---|---|
| GENDER slope, $B_3$ | | | | | |
| INTRCPT2, $\gamma_{30}$ | 5.70 | 1.65 | 3.45 | 6732 | .001 |
| INTERVENTION, $B_4$ | | | | | |
| INTRCPT2, $\gamma_{40}$ | 5.85 | 2.18 | 2.69 | 34898 | .007 |

***** *Final Parameter Estimates and Inferences Obtained from 100 Imputed Data Sets.*

In addition, non-minority students statistically significantly outperformed minority students by 8.15 points ($\gamma_{20}$), and male students significantly outperformed female counterparts by 5.70 points ($\gamma_{30}$). Comparing the size of the advantage imparted by a NURTURES-trained teacher relative to these effects, *having a NURTURES-trained teacher completely reversed the gender gap and partially reversed the minority student gap*. In other words, if other variables are equivalent, a female student with a NURTURES-trained teacher should perform as well as a male student without a NURTURES-trained teacher. For minority students, having a NURTURES-trained teacher could reduce their achievement gap with non-minority students without a NURTURES-trained teacher by over 72%. The conditional inter-class correlation coefficient was 0.292, indicating that once level-1 predictor variables were specified, approximately 29.2% of modeled variance in the 5th-grade Ohio Spring 2017 Science sample could be attributed to school effects.

**Conclusions.** This study found that students from NURTURES treatment groups scored significantly higher on science tests given at 5th grade (ages 10-11) as compared to their peers. Overall, this study suggests that providing Framework-aligned science instruction, coupled with parent support and community-wide science learning opportunities, during early years improves science skills in later elementary grades. Although our effect size is to be generally interpreted as small, it is noteworthy that the effect of the intervention slope ($\beta_4$ = 5.85) is approximately equivalent to a statistically significant female lower-than-hoped for score ($\beta_3$ = 5.70) and almost equivalent in terms of minority status lower-than-hoped-for-score ($\beta_2$ = 8.15). *This suggests that the impact of the intervention roughly compensated for the attainment gap between boys and girls and partially ameliorated the gap between minority and non-minority children's scores associated with these demographic factors.*

**Summary**. These longitudinal findings provide evidence that access to quality science instruction in the early grades (NURTURES) can help level the playing field for at-risk learners and mitigate the factors that contribute to persistent lower science, mathematics and reading achievement levels of at-risk youth.

## References


Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48.

Grund, S., Robitzsch, A., & Luedtke, O. (2018). mitml: Tools for multiple imputation in multilevel modelling. R package version 0.3-6. URL: https://cran.r-project.org/web/packages/mitml/index.html. Last accessed January 19, 2019.

Mentzer, G.A. & Paprzycki, P. (2016). NURTURES Evaluation Report: 5 Year Summary, 2011-2016. Accessed at: http://nurtures.utoledo.edu/reports/NURTURES_ExternalSummaryEvaluation.pdf

NWEA. (2019). MAP® Growth™ technical report. Portland, OR.

Ohio Achievement Assessment (2015). *Test Specifications: Grade 5 Science*. Retrieved from https://oh.portal.airast.org/core/fileparse.php/3094/urlt/Test_Specifications_G5_Science.pdf. Last accessed January 19, 2019.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL: https://www.R-project.org.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R., & Du Toit, M. (2011). HLM 7: Hierarchical linear and nonlinear modelling. [Computer software]. Lincolnwood, IL: Scientific Software International.

Renaissance Learning. (2013). The research foundation for STAR Assessments: The science of STAR. Wisconsin Rapids, WI: Author. URL: http://doc.renlearn.com/KMNet/R001480701GCFBB9.pdf. Last accessed January 19, 2019.

What Works Clearinghouse. (2020). What Works Clearinghouse: Standards handbook (Version 4.1).Washington, DC:U.S. Department of Education, Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/wwc/handbooks.

Zhao, J. H., & Schafer, J. L. (2018). Package 'pan': Multiple imputation for multivariate panel or clustered data. R package version 1.6. URL: https://cran.r-project.org/web/packages/pan/pan.pdf Last access January 19, 2019.