The Middle School Algebra Readiness Initiative

An Evaluation of Teacher Outcomes and Student Mathematics Achievement and Gains

West Virginia Department of EDUCATION
Office of Research

The Middle School Algebra Readiness Initiative

An Evaluation of Teacher Outcomes and Student Mathematics Achievement and Gains

Nate Hixson

West Virginia Department of EDUCATION
Office of Research

West Virginia Department of Education
Division of Teaching and Learning
Office of Research
Building 6-Room 722 State Capitol Complex
1900 Kanawha Boulevard East
Charleston, WV 25305
http://wvde.state.wv.us/
October 2012

Jorea M. Marple
State Superintendent of Schools
West Virginia Department of Education

Robert Hull
Associate Superintendent
West Virginia Department of Education

Larry J. White
Executive Director
Office of Research

Keywords
middle school, algebra, mathematics, student achievement, teacher outcomes

Suggested Citation

Hixson, N. (2012). The *Middle School Algebra Readiness Initiative: An analysis of teacher outcomes and student mathematics achievement and gains*. Charleston, WV: West Virginia Department of Education, Division of Teaching and Learning, Office of Research.

Content Contact
Nate Hixson
Assistant Director
Office of Research
nhixson@access.k12.wv.us

Abstract

We conducted an evaluation study of the Middle School Algebra Readiness Initiative, a middle school mathematics intervention that was implemented in two West Virginia school districts during the 2011–2012 school year. In participating middle schools, the Carnegie Learning MATHia® software intervention and accompanying classroom curriculum were used as a total replacement for the districts' alternative mathematics curriculum for Grades 6, 7, and 8. A cohort of teachers was trained by Carnegie Learning in mathematical content and pedagogy as well as in the proper implementation of the software and classroom curriculum materials.

Our evaluation tested five hypotheses. Our first was related to the impact of the initiative on teacher-level outcomes, specifically teachers' content and pedagogical knowledge in the areas of patterns, functions, and algebra. This hypothesis was tested by using a pretest/posttest assessment of teacher knowledge. We used the research-validated, *Learning Mathematics for Teaching* (LMT) assessment. Our statistical analysis of teacher pretest/posttest differences revealed that, for the 20 teachers who completed both a pretest and posttest, there was only a marginal gain. This gain was not statistically significant. As such, we rejected our first hypothesis.

The remaining four study hypotheses tested the impact of the initiative on students' mathematical achievement and year-to-year mathematics gains as measured by the Grade 6, 7, or 8 mathematics subtest of the West Virginia Educational Standards Test 2 (WESTEST 2). We used propensity score matching (PSM) to match students in a variety of implementation scenarios to select a comparison group of students. The comparison groups for our investigations of Hypotheses 2–4, included students who used the MATHia software/curriculum at various levels of implementation, matched to their grade-level peers who used some other curriculum during the 2011–2012 school year. For Hypothesis 5, we compared students who used the MATHia program for at least 1 hour per week—meeting the vendor's definition of *adequate use*—to a comparison group of students who used the program for less time. In all cases, we rigorously matched the two groups of students using a variety of covariates including sex, free/reduced-price lunch eligibility, special education eligibility, grade, and prior academic achievement in both mathematics and reading/language arts.

We then conducted two student-level analyses. First, we examined mean differences in students' standardized test scores and mathematics gains, determining if the treatment or comparison group scores differed by a statistically significant margin. Second, we used linear regression to determine, after controlling for the aforementioned covariates, what level of impact the treatment had on student achievement and gains. We found in most cases that students who were in the treatment group underperformed when compared with their grade-level peers who used an alternate curriculum. With a few exceptions, the differences were statistically significant. However, the results of the linear regressions illustrated that, after controlling for important covariates, the negative relationship among treatment and student achievement/gains was relatively small, but still statistically significant.

Several limitations impair our ability to make conclusions based on these results. Most critically, we had very little information about the degree to which the teachers and students implemented the intervention components with fidelity. We know very little about the quality or content of the training provided by Carnegie Learning. We experienced considerable attrition among educators in the period between the pretest and posttest LMT administration. Only 55% of teachers completed both assessments. We also received very different numbers of students from Carnegie Learning and the school districts when we requested this information for our analyses. Finally, we found that very few students met the implementation criteria recommended by Carnegie Learning. This finding, in particular, points to a potential lack of fidelity in implementation, which makes us very reluctant to allow this evaluation to stand as a fair trial of the efficacy of the MATHia software/curriculum. In fact, we recommend strongly against using our report in this manner. It should be seen as an evaluation of an entire initiative rather than any curriculum or software program alone.

In light of these and other limitations described in this report, we make only two recommendations. First, we suggest future program implementations of this type take substantial measures to collect critical qualitative implementation data so that the results of quantitative analyses can be more readily interpreted. This can be accomplished, among other strategies, by devoting greater resources to the program evaluation component of such projects. Second, in districts where similar programs are currently underway or in the planning stages, we recommend continuous monitoring and technical assistance to ensure that the program components are delivered as intended. Doing so may help prevent a potential negative impacts on student outcomes.

Contents

List of Figures

List of Tables

Introduction

This study evaluates the impact of the Middle School Algebra Readiness Initiative (MSARI). Our analysis focuses on the extent to which the initiative increased teachers' mathematics content and pedagogical knowledge and students' achievement and growth on the mathematics subtest of the West Virginia Educational Standards Test 2 (WESTEST 2). The WESTEST 2 is the state's summative assessment, required under the federal No Child Left Behind Act. The mathematics subtest of WESTEST 2 is administered to all students in Grades 3–11 in the spring of each school year. This measure was chosen because of its availability and its focus on concepts aligned to the proposed intervention curriculum.

The MSARI was implemented in two West Virginia school districts during the 2011–2012 school year. These districts committed to using Carnegie Learning's MATHia® software and accompanying classroom curriculum as a total replacement for the standard mathematics curriculum for Grades 6, 7, and 8. Teachers in these schools were trained by Carnegie Learning in the use of the MATHia program via a series of mathematics teacher academies. We hypothesized that teachers who underwent this training would exhibit increased mathematical and pedagogical knowledge and that students who used MATHia during the 2011–2012 school year would achieve greater mathematics achievement and gains when compared with a matched sample of students who either used an alternate mathematics curriculum or who did not use MATHia with fidelity. The following hypotheses were tested:

H1 Teachers who participate in training provided as part of the MSARI will exhibit significantly greater posttest scores on the Learning Mathematics for Teaching (LMT) patterns, functions, and algebra assessment.

H2 Students who use MATHia software during the 2011–2012 school year, regardless of their level of exposure, will score significantly higher on the WESTEST 2 mathematics subtest than students who do not use the software.

H3 Students who are continuously enrolled in a classroom where MATHia is being used for at least 210 days during the 2011–2012 school year will score significantly higher on the WESTEST 2 mathematics subtest than students who do not use the software.

H4 Students who use MATHia software during the 2011–2012 school year for the recommended minimum of at least 1 hour per week will score significantly higher on the WESTEST 2 mathematics subtest than students who do not use the software.

H5 Students who use the MATHia software during the 2011–2012 school year for the recommended minimum of at least 1 hour per week will score significantly higher than students who use MATHia for less than the recommended minimum of at least 1 hour per week.

Methods

Participant Characteristics

Teachers in this study attended the training provided by Carnegie Learning and also took the Learning Mathematics for Teaching (LMT) assessment. Students in the study used Carnegie Learning's MATHia software and curriculum as their mathematics curriculum during the 2011–2012 school year; were in Grades 6, 7, and 8; and were included in the state's 2010–2011 and 2011–2012 WESTEST 2 assessment data files. These students were compared to a matched sample of students from the population of all Grade 6, Grade 7, and Grade 8 students who were known not to have used the MATHia software program as their mathematics curriculum during the 2011–2012 school year.

Sampling Procedures

Teachers

We used all available records for those teachers who completed the LMT assessment at the conclusion of two mathematics academies, one at the outset of the 2011–2012 school year and the other at the end. For our pretest and posttest analyses, we could only include those teachers who had a matched pretest and posttest assessment, limiting our analysis of teacher outcomes to the 20 teachers who met this condition.

Students

We received an initial spreadsheet containing 2,265 unique student records along with software usage statistics from Carnegie Learning. We then requested a list of students from the two county school systems, which we used to cross reference and identify only those students who persisted in classrooms using MATHia for the majority of the academic year. To accomplish this, our county contacts asked teachers in each of the middle schools to verify on their currently active course rosters students who were enrolled in classrooms where MATHia was being utilized. The lists provided by the counties included 1,561 unique student names representing 70 classrooms and 22 teachers[1].

Using the 1,561 students provided by the counties, we then queried the software usage statistics from the data file provided by Carnegie Learning. The query returned 1,605 records because some students were enrolled in multiple courses where MATHia was being utilized. After merging those valid duplicate cases and deleting all remaining duplicates, 1,535 unique student records remained. As a final step prior to matching, we then queried assessment and demographic data for these students from WVEIS. We required 2 years of assessment data as well as a full set of covariate demographic variables in order to conduct

[1] It is unclear why there was such a large discrepancy in the numbers of students reported by Carnegie Learning and by the counties.

the matching and final analyses. Therefore, any student for whom we could not locate this information was removed from the sample. Our final sample included 1,276 students or 82% of those records provided by the counties. These students were then matched using the population of remaining Grade 6, Grade 7, and Grade 8 students (approximately 60,000 students).

We used propensity score matching (PSM) to select a set of matched comparison groups to test each hypothesis. PSM is a methodology that uses a logistic regression model to match samples based on a single score that is based on a variety of observed covariates.

Matching procedures for student outcome analyses

We created a binary indicator for whether a student did or did not participate in the MSARI (hereafter referred to as treatment or comparison students, respectively). We then used propensity scores to match each treatment student to a suitable comparison student. The propensity score is the conditional probability of being assigned to the treatment group given a vector of observed covariates. The goal of PSM is to model equivalent selection bias in both groups, thus exercising some degree of control over the impact of the observed covariates on the outcome variable of interest.

In this study, we sought primarily to control for prior academic achievement in both reading/language arts and mathematics, but specified up to 7 total covariates in the propensity score models including, (1) 2010–2011 WESTEST 2 mathematics achievement, (2) 2010–2011 WESTEST 2 reading/language arts achievement, (3) sex, (4) race, (5) free/reduced price lunch eligibility, (6) special education eligibility, and (7) grade level. Thus, the propensity score we generated was the predicted probability of being assigned to the treatment condition obtained from a binary logistic regression including the listed covariates as predictors (Rosenbaum & Rubin, 1983). Once propensity scores were calculated, we used nearest neighbor matching without replacement in the *R* statistical analysis software program to select comparison group members. Table 1 provides a description of each sampling frame, how the binary indicator of treatment/comparison was defined, and which hypothesis the sampling frame was used to test.

Once matching was complete, we examined balance statistics for the samples to ensure the matching algorithm resulted in samples that were comparable on the measured covariates. Tables A1–A8 in Appendix A (page 31) illustrate, for each of the eight sampling frames, the pre-/post-matching means for each covariate and the percentage of improvement in balance for each covariate after matching. Note that, because we did not discard treatment units, the treatment post mean is identical to the treatment pre mean; this is indicated with an asterisk (*). When matching is successful, the post mean difference should be as close as possible to zero indicating the groups do not differ on the observed covariate. For each sampling frame, we observed a remarkable improvement in the balance post-matching. Furthermore, we verified using chi-square analyses, that for all sampling frames the covariate distributions were not statistically significantly different among the two groups. We found that after matching, with one exception[2], there were no statistically significant differences in these covariates at baseline. As a result, we were very confident going into our analyses of student achievement and gains.

Table 1. Description of Sampling Frames Used to Test Study Hypotheses

Sampling frame	Description	Hypothesis tested
SF16, SF17, and SF18	A set of data frames containing all students who were enrolled in Grade 6 (SF16), Grade 7 (SF17), and Grade 8 (SF18) in West Virginia during the 2011–2012 school year, including a binary indicator for whether or not the student was in the treatment group. Treatment group students were identified by virtue of their having been located in the rosters provided by the participating school district and Carnegie Learning. All treatment group students were included in these sampling frames, regardless of the number of hours/sessions in the software program.	H2
SF26, SF27, and SF28	A set of data frames containing all students who were enrolled in Grade 6 (SF26), Grade 7 (SF27), and Grade 8 (SF28) in West Virginia during the 2011–2012 school year, including a binary indicator for whether or not the student was in the treatment group. Treatment group students were identified by virtue of their having been located in the rosters provided by the participating school district and Carnegie Learning. Treatment group students with less than 210 days of continuous enrollment from first to last program session were removed from these sampling frames.	H3
SF3	A data frame containing all students who were enrolled in Grade 6, Grade 7, and Grade 8 in West Virginia during the 2011–2012 school year, including a binary indicator for whether or not the student was in the treatment group. Treatment group students were identified by virtue of their having been located in the rosters provided by the participating school district and Carnegie Learning. Treatment group students with less than 1 hour per week of program use according to usage statistics provided by Carnegie Learning were removed from the sampling frame.	H4
SF4	A data frame containing ONLY treatment group students who were enrolled in Grade 6, Grade 7, and Grade 8 in West Virginia during the 2011–2012 school year, including a binary indicator for whether or not the student exhibited at least 1 hour of program use according to usage statistics provided by Carnegie Learning. Treatment group students with less than the recommended 1 hour per week of program use according to usage statistics provided by Carnegie Learning were coded as the comparison group for this sampling frame.	H5

[2] Special education eligibility was not equally distributed across groups for SF18. The treatment group included 11.2% while the control group included 7.2%. This difference was statistically significant X^2 (1, N= 750) = 4.49, p = .03.

Sample Size, Power, and Precision

Teachers

As mentioned above, due to attrition between the pretest and posttest administration of the LMT, our sample size for teacher-level outcomes was only 20. Thus, we did not have adequate power to detect small or moderate effects in this study. It is possible that our failure to detect statistically significant differences in this study was due to this issue. We discuss this further in the results and limitations sections of this report.

Students

Sample sizes varied within each of the aforementioned sampling frames. Tables B1–B4 in Appendix B (page 37) provide an overview of the final sample sizes for each sampling frame by hypothesis tested. These tables also include an investigation of whether or not we had enough power to detect moderate effects of the magnitude observed in this study (d = .30). Notably, the final sample sizes for this study were adequate to detect these effects with 95% confidence only for Hypothesis 2. However, this is somewhat of a moot point given that we did observe at least some statistically significant differences for Hypotheses 2–5. This information is provided here only to illuminate the fact that our failure to detect significant differences in some cases (H2, Grade 6, H3, Grade 7, and H5) may have been due to low sample sizes.

Measures and Covariates

Independent variables

Independent variables are those that serve as the predictors of some outcome variable (the outcome is often called a dependent variable). In this study, our independent variables differed by level of analysis.

Teacher-level analysis

The independent variable in our teacher-level analysis was time. Our analysis examined the level of change that occurred in teachers' LMT assessment scores between the administration of the pretest and the posttest. That is, we expected that, over time, the average LMT assessment score would increase by virtue of teachers' accumulation of content and pedagogical knowledge as a result of the training they received from Carnegie Learning.

Student-level analyses

In this study, the independent variables for student-level analyses included various levels of exposure to the MATHia curriculum and software intervention in place of the traditional curriculum. Carnegie Learning provided us with the following information about each student's use of MATHia during the 2011–2012 school year:

1. Date of first session
2. Date of last session
3. Total number of seconds of MATHia use between first and last session

In consultation with Carnegie Learning, we calculated two metrics that would serve as independent variables in this study: (a) number of hours of MATHia use per week and (b) total number of days enrolled in a MATHia classroom. Carnegie Learning recommended hours per week as one potential indicator of implementation fidelity, with the following levels:

1. Low fidelity—less than 1 hour of MATHia use/week
2. Moderate fidelity—at least 1 hour of MATHia use/week
3. High fidelity—at least 1.5 hours of MATHia use/week

We calculated hours/week using the steps detailed in Table 2.

Table 2. Procedure for Calculating Hours/Week Fidelity Measure

Step	Formula
Determine the total number of calendar days each student was an active participant in a classroom implementing MATHia.	Last session date–First session date
Convert the total number of calendar days to the total number of calendar weeks.	Session days ÷ 7
Determine the total minutes of actual MATHia use between the start and end date for each student.	Total seconds of use ÷ 60
Determine the total minutes of actual MATHia use per calendar week available for the student.	Total minutes of use ÷ Total weeks of exposure
Determine the total hours of actual MATHia use per calendar week available for the student.	Total minutes actual MATHia use ÷ 60

We then used the criteria illustrated in Table 3 to create four fidelity categories based on the recommendations of Carnegie Learning—three mutually exclusive groups and a fourth category representing both *adequate* and *high* implementers.

Table 3. Fidelity Categories Developed for This Study

Fidelity category	Cut points
Low	0.00 - .9999 hours/week
Adequate	1.00–1.499 hours/week
High	1.50 and up hours/week
Adequate or high	>1.00

Table 4 presents the number of students that met each of the aforementioned fidelity conditions by grade level. It should be noted that these data were summarized prior to the implementation of the PSM algorithm. Some minor attrition did occur during matching.

Table 4. Fidelity of Implementation for the Sample Used in this Study

Grade	Average hours/week	Range	# of students low fidelity range (0.00 - .9999)	# of students adequate fidelity range (1.00–1.499)	# of students high fidelity range (1.50 and up)	# of students adequate or high fidelity (>.999)
6	.64	.04–2.82	388 (88.6%)	35 (8.0%)	15 (3.4%)	50 (11%)
7	.57	.04–2.29	448 (93.3%)	27 (5.6%)	5 (1.0%)	32 (6.7%)
8	.65	.05–2.09	321 (83.8%)	43 (11.2%)	19 (5.0%)	62 (16.2%)

Upon examining these data, it immediately became clear that using Carnegie Learning's recommended criteria, very few students actually reached the levels of implementation that would be considered *adequate* or *high*.

Covariates

There were no covariates included in teacher-level analyses. For student-level analyses, we matched treatment and comparison cases on 7 covariates using PSM. We also used these covariates as predictors in a series of linear regression models. Each of the covariates is described below.

Prior reading/language arts achievement

For Hypotheses 2 and 3, which employed individual grade-level matching and analyses, we used students' standardized 2010–2011 WESTEST 2 reading/language arts (RLA) scores as a measure of their RLA ability prior to the 2011–2012 school year. For hypotheses 4 and 5, which aggregated students across grade levels, we used students' WESTEST 2 RLA performance levels for the 2010–2011 school year. This covariate was included in the matching model to ensure that the treatment and comparison groups comprised students' with similar RLA skills prior to the intervention.

Prior mathematics achievement

For Hypotheses 2 and 3, which employed individual grade-level matching and main analyses, we used students' standardized 2010–2011 WESTEST 2 mathematics scores as a measure of their mathematical ability prior to the 2011–2012 school year. For Hypotheses 4 and 5, which aggregated students across grade levels, we used students' WESTEST 2 mathematics performance levels for the 2010–2011 school year. This covariate was arguably the most important variable included in the matching model because it ensured that the treatment and comparison groups comprised students' with similar mathematics skills prior to the intervention. Had we not accounted for this variable it may have imparted significant bias in our analysis of 2011–2012 mathematics achievement and gains. The correlation between students' prior and current mathematics achievement is known to be statistically significant and of great magnitude.

Sex

Student biological sex is known to be associated with academic achievement such that male students are often significantly lower performing than their female peers in both mathematics and reading/language arts. Thus, it was included in all matching models.

Race

Student race was operationalized as a binary indicator denoting whether or not students were White. Caucasian students represent approximately 92% of all students in West Virginia.

Free and reduced-price lunch eligibility

Students' socioeconomic status was operationalized using a proxy measure, free and reduced-price lunch eligibility. This indicator was binary, indicating whether or not the student was eligible. This variable is known to possess a negative and statistically significant relationship with student achievement.

Special education eligibility

Special education eligibility was operationalized as a binary indicator, which indicated whether or not a student had an individualized education program (IEP). Special education eligibility is known to possess a negative and statistically significant relationship with student achievement.

Grade

Grade level was controlled for in Hypotheses 2 and 3 by conducting the PSM matching within each grade-level band. That is, there was no variability in grade level for these analyses; students were only matched to other students in the same grade level. With respect to Hypotheses 4 and 5, grade level was operationalized as three binary indicators. Each variable indicated whether or not the student was in Grade 6, Grade 7, or Grade 8 during the implementation year.

Dependent variables

Teacher content and pedagogical knowledge

Teachers' gains in content and pedagogical knowledge were measured in this study via pretest and posttest administration of the *Learning Mathematics for Teaching* (LMT) assessment (Hall, Schilling, & Ball, 2004). The LMT is a teacher assessment that includes a battery of diverse assessments appropriate to measure mastery of multiple mathematical concepts at various programmatic levels. The measures have been extensively validated via multiple research studies and were developed with ongoing support from the National Science Foundation.[3]

For this study, we selected the 2007 revision of the Middle School Patterns, Functions, and Algebra subtest (PFA). This subtest consists of two equated forms, each including 33 items. The items assess the extent to which teachers have the ability to solve mathematics problems of the types typically assigned to their students and how well they are able to evaluate students' knowledge of mathematics. The results of this assessment are normed based on a large and geographically representative sample of middle school mathematics teachers.

---

[3] For more information about the LMT project, readers are referred to http://sitemaker.umich.edu/lmt/home.

Raw scores are converted to standardized scores and gains across multiple forms can be analyzed to determine statistical significance.

We administered the LMT to participating teachers prior to the first mathematics content academy (pretest) and at the conclusion of the final content academy of the year (posttest). During the administration of the pretest, we assigned both of the equated forms of the assessment in a staggered fashion such that the first teacher was assigned Form A, the second Form B, the third Form A, and so forth. Each teacher was then assigned the alternate form during the posttest. For example, if a teacher was assigned Form A at pretest, he or she was assigned Form B at posttest. Because both forms were equated, this allowed us to analyze gains on the assessment from pretest to posttest without worrying about test-retest bias.

Student mathematics achievement and gains

We assessed the effect of the intervention on both mathematics achievement and mathematics gains. Math achievement was operationalized as students' standardized math assessment scores from the 2011–2012 administration of the WESTEST 2—the assessment administered at the end of the school year during which the intervention took place. Scores were standardized within each grade level so that the state mean score for each grade was zero and the standard deviation was 1. This allowed for easy interpretation of scores (e.g., a score of .25 is the equivalent of one quarter standard deviation above the state mean) and also for valid aggregation of assessment results across grade levels to increase effective sample sizes for some tests. Readers should keep in mind that standardized test scores indicate a student's relative position within the distribution of her/his grade-level peers. Conversely, students' scale scores are relatively nebulous quantities that have little interpretive value except as they relate to a cut score that expresses a policy expectation (e.g., proficiency).

Math gains were operationalized as the difference in students' 2011–2012 and 2010–2011 standardized math assessment scores. That is, for each student, we subtracted his or her 2010–2011 standardized score from his or her 2011–2012 standardized score. For example, if a student exhibited a 2011–2012 score of 1.0 and a 2010–2011 score of .70, her math gain score would be 1.0 minus .70 or .30. Positive scores represent increases in relative standing from one year to the next while negative gain scores represent regression in standing from one year to the next.

Importantly, regression in standardized scores may not necessarily correspond to lower scale scores. That is, a student's actual test score may increase from one year to the next while their standardized score decreases.

Research Design

Teacher-level analyses

We used a dependent samples paired *t* test to determine if the average difference in pretest and posttest scores for teachers was statistically different from zero. If this were found to be true, we would accept our hypothesis that participation in the MSARI led to increased content and pedagogical knowledge. In the results section of this report we present the results of this analysis and also compare the average pretest and posttest scores using descriptive statistics.

Student-level analyses

We tested Hypotheses 2 to 5 first by conducting a series of independent samples *t* tests. These simple tests were used to identify the presence or absence of statistically significant differences in math achievement and gains among students in the treatment and comparison groups. The *t* tests illustrate when the two groups differed and descriptive statistics illustrated the amount and direction of those differences. We posited that, if the *t* tests returned significant results and those results were in the predicted direction, we could accept our study hypotheses that students who used the MATHia software and curriculum exhibited higher math achievement and gains than students who used the alternative curriculum.

The *t* tests are useful to illustrate where statistically significant differences exist, but they are not sufficient to accurately estimate the impact of the treatment when accounting for other important covariates that have an impact on the outcome such as prior academic achievement. To address this, we also employed a series of linear regression models. These models allowed us to estimate the proportion of variance accounted for in math achievement/gains by each covariate including a binary indicator of each students' status as either a treatment or comparison group member. We conducted each linear regression in two sequential blocks. During the first block, we simultaneously entered all of the covariates used in the PSM models as predictors of the dependent variable under examination. This first model allowed us to determine our ability to predict the dependent variable without the treatment variable having been accounted for. Our second model included the same covariates, but added the treatment variable. Comparing the output of both models allowed us to calculate the unique contribution of the treatment to students' math achievement/gains after accounting for the impact of the measured covariates.

Table 5 and Table 6 provide an overview of the general structure of the models we used. The reader will notice that the models we used to test H4 and H5 were slightly different. This is because, in these models, we had to account for grade level due to aggregation (see above). These models also differed in that we accounted for prior academic achievement using students' prior performance levels rather than their standardized assessment scores.

Table 5. Overview of Linear Regression Model Structures for H2 and H3

Model	Structure
1	2011–2012 math achievement/gain = sex + free and reduced price lunch eligibility + race + special education eligibility + 2010–2011 math achievement + 2010–2011 RLA achievement
2	2011–2012 math achievement/gain = sex + free and reduced price lunch eligibility + race + special education eligibility + 2010–2011 math achievement + 2010–2011 RLA achievement + *treatment*

Table 6. Overview of Linear Regression Model Structure for H4 and H5

Model	Structure
1	2011–2012 math achievement/gain = sex + free and reduced price lunch eligibility + race + special education eligibility + 2010–2011 math performance level + 2010–2011 RLA performance level + Grade 6 + Grade 7 + Grade 8
2	2011–2012 math achievement/gain = sex + free and reduced price lunch eligibility + race + special education eligibility + 2010–2011 math performance level + 2010–2011 RLA performance level + Grade 6 + Grade 7 + Grade 8 + ***treatment***

Results

Hypothesis 1

Hypothesis 1 stated, "Teachers who participate in training provided as part of the Middle School Algebra Readiness Initiative (MSARI) will exhibit significantly greater post-test scores on the *Learning Mathematics for Teaching* (LMT) patterns, functions, and algebra assessment." Thirty-six teachers completed the LMT pretest assessment. Their standardized LMT scores ranged from -1.21 to 1.66, with a mean score of .19 (sd = .75) for the group. This corresponds to, on average, answering approximately 18 of the 33 questions correctly at pretest.

Because we required both a pretest and posttest record to complete the gains analysis, those teachers who completed a pretest, but not a posttest, were necessarily excluded from the sample (n = 16). With this adjustment, the final sample size for our gains analysis was 20. The average pretest score for the 20 teachers who completed both a pretest and posttest was .43 (sd = .60), which indicates, on average, the 20 teachers who completed both a pretest and posttest answered approximately 20 of the 33 questions correctly at pretest, slightly higher than for the full sample.

The posttest scores for these 20 teachers ranged from -1.06 to 1.36 with a mean of .51 (sd = .65). This illustrates an average pretest to posttest gain of only .08. Put another way, on average, teachers answered approximately 20 of the 33 questions correctly at posttest. As we indicated earlier, this was the same number of pretest questions answered correctly. Therefore, there was no discernible difference in the average number of correct responses between pretest and posttest.

Table 7 presents the results of the paired samples *t* test analysis used to determine if pretest and posttest scores differed significantly. This difference was not statistically significant $t(19) = .513$, $p = .61$. This result indicates that, for the 20 teachers who completed both a pretest and posttest, their content and pedagogical knowledge of patterns, functions, and algebra, though increasing slightly, was not significantly greater at the conclusion of the project. Figure 1 provides a graphical representation of the average pretest and posttest scores.

Figure 1. LMT Patterns, Functions, and Algebra Pretest and Posttest Scores for MSARI

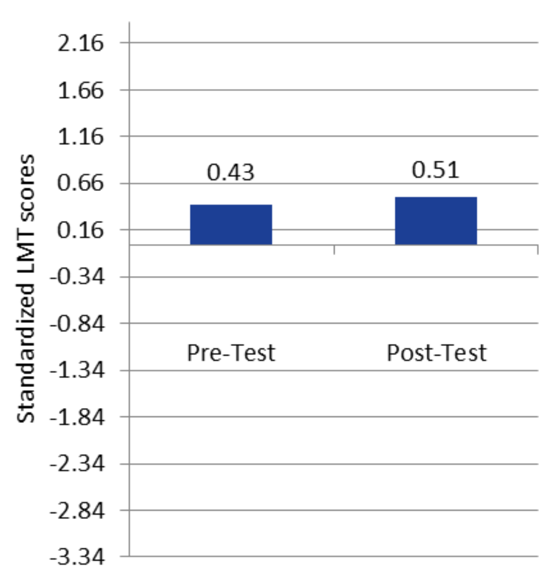


Table 7. Significance Test for Difference in LMT Pretest and Posttest Scores

Pair	Mean	*sd*	Standard error of the mean	*t*	*df*	*p*
Post–Pre	.08	.71	.16	.513	19	.61

Hypothesis 2

Hypothesis 2 stated, "Students who use the MATHia software during the 2011–2012 school year, regardless of their level of exposure, will score significantly higher on the WESTEST 2 math subtest than students who do not use the software." Table 8 presents the results of the *t* test analyses by grade level.

Table 8. *T* Test Results for Hypothesis 2

Grade	T mean (*sd*)	C mean (*sd*)	*t*	*df*	*p*	Mean difference	Significant?
2011–2012 math achievement							
6	.20 (.91)	.30 (.93)	-1.632	868	.103	-.10	NO
7	.13 (.95)	.26 (.97)	-2.087	930	**.037**	-.13	YES
8	-.27 (.96)	-.00 (.92)	-4.017	748	**.000**	-.27	YES *
2010–2011 to 2011–2012 math gains							
6	-.14 (.63)	-.03 (.64)	-2.550	868	**.011**	-.11	YES
7	-.06 (.70)	.03 (.66)	-2.250	930	**.025**	-.10	YES
8	-.14 (.77)	.09 (.77)	-4.33	748	**.000**	-.24	YES *

*Recall that, for these analyses, the treatment group included a significantly greater proportion of students who were special education eligible than the comparison group. As such, we recommend caution interpreting these results.

As is evidenced in Table 8 and Figure 2 through Figure 7, despite starting at comparable points in 2011 math and RLA achievement and having remarkably similar demographic characteristics, students in the treatment group in Grades 7 and 8 scored significantly lower than students in the comparison group on the WESTEST 2 math assessment in 2012. Additionally, treatment group students in all three grades exhibited significantly lower math gains from 2010–2011 to 2011–2012 than students in the comparison group. These findings were counter to our hypothesis.

Figure 2. **Grade 6 Mathematics Achievement Treatment vs. Control—All MATHia Users**

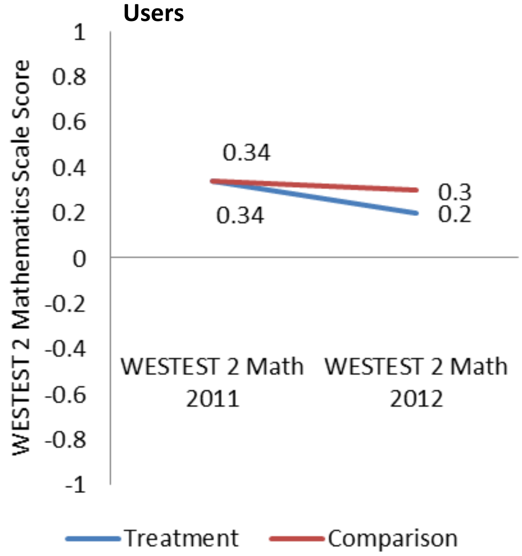


Figure 3. **Grade 5-to-6 Mathematics Gain Treatment vs. Control—All MATHia Users**

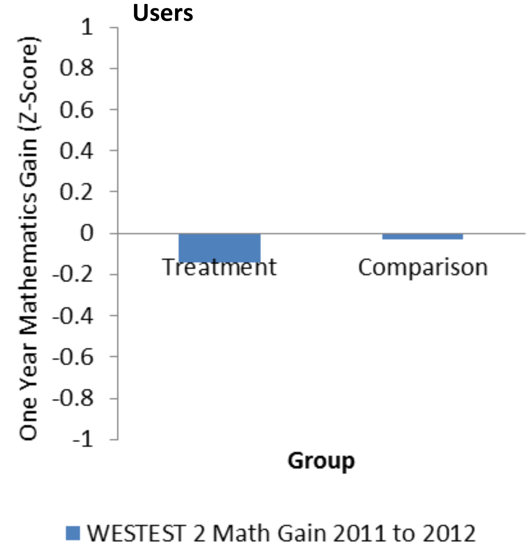


Figure 4. **Grade 7 Mathematics Achievement Treatment vs. Control—All MATHia Users**

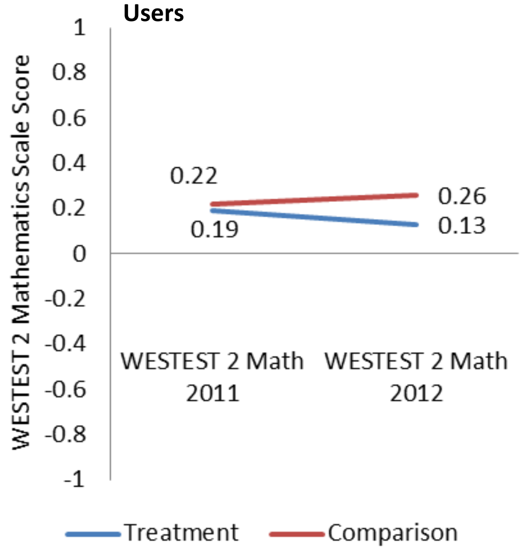


Figure 5. **Grade 6-to-7 Mathematics Gain Treatment vs. Control—All MATHia Users**

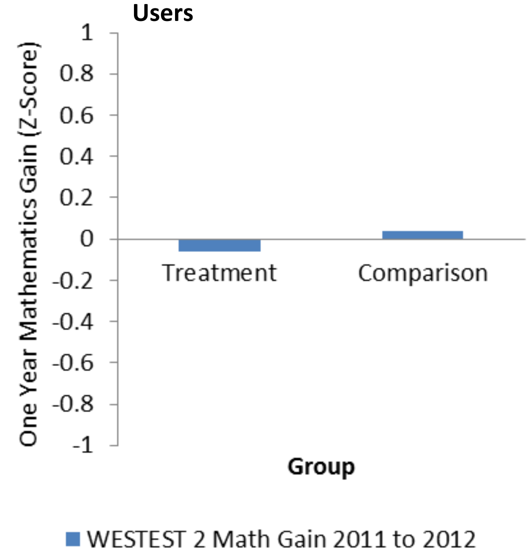


Figure 6. Grade 8 Mathematics Achievement Treatment vs. Control—All MATHia Users

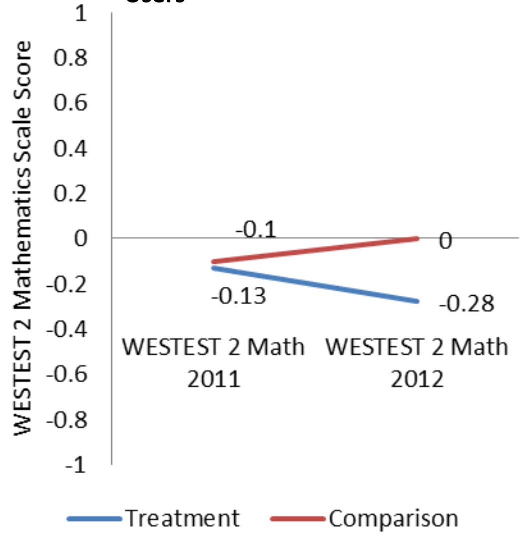
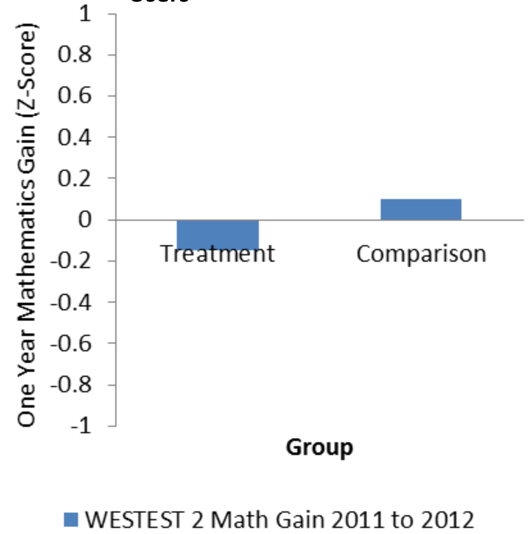


Figure 7. Grade 7-to-8 Mathematics Gain Treatment vs. Control—All MATHia Users



Tables C1–C18 in Appendix C (page 39) present detailed results of the six general linear models we used to test the explanatory power of the treatment on students' math achievement and gains after accounting for all measured covariates. Table 9 below provides a summary of those models. As displayed below, the treatment coefficient was a statistically significant and negative predictor in all six models. However, it should be noted that treatment only contributed minor explanatory power after accounting for covariates (i.e., between .3% and 2% of the variance).

Table 9. Abbreviated Linear Model Summaries for Hypothesis 2

Model	Model 1 adj. R^2	Model 2 R^2 change	*p* value for model 2	*p* value for treatment coefficient	Standardized β for treatment coefficient	Interpretation
Grade 6 math achievement	.623	.003	.000	**.015**	-.051	The treatment was a statistically significant and negative predictor in all models, but only contributed between .3% and 2% explanatory power after accounting for covariates.
Grade 6 math gains	.222	.005	.000	**.015**	-.073	
Grade 7 math achievement	.595	.003	.000	**.008**	-.055	
Grade 7 math gains	.203	.006	.000	**.008**	-.078	
Grade 8 math achievement	.435	.017	.000	**.000**	-.129	
Grade 8 math gains	.166	.024	.000	**.000**	-.157	

Hypothesis 3

Hypothesis 3 stated, "Students who have been continuously enrolled in a classroom where MATHia is being utilized for at least 210 days during the 2011–2012 school year will score significantly higher on the WESTEST 2 math subtest than students who do not use the software." Table 10 presents the results of the *t* test analyses by grade level.

Table 10. ***T*** **Test Results for Hypothesis 3**

Grade	T Mean (*sd*)	C Mean (*sd*)	*t*	*df*	*p*	Mean difference	Significant?
2011–2012 math achievement							
6	-.03 (.79)	.22 (.84)	-2.80	316	.005	-.26	YES
7	.02 (.94)	.04 (.94)	-.268	478	.789	-.02	NO
8	-.31 (.93)	.06 (.87)	-3.92	334	.000	-.38	YES
2010–2011 to 2011–2012 math gains							
6	-.30 (.60)	-.08 (.60)	-3.33	316	.001	-.22	YES
7	-.03 (.67)	-.03 (.77)	.048	478	.962	.00	NO
8	-.19 (.81)	.14 (.81)	-3.92	334	.000	-3.4	YES

As is evidenced in Table 10 and Figure 8 through Figure 13, despite starting at comparable points in 2011 math and RLA achievement and having remarkably similar demographic characteristics, students in the treatment group in Grades 6 and 8 scored significantly lower than students in the comparison group on the WESTEST 2 math assessment in 2012. Additionally, treatment group students in Grades 6 and 8 exhibited significantly lower math gains from 2010–2011 to 2011–2012 than students in the comparison group. There were no differences in grade 7. These findings were counter to our hypothesis.

Figure 8. **Grade 6 Mathematics Achievement Treatment vs. Control—210 Days in MATHia**

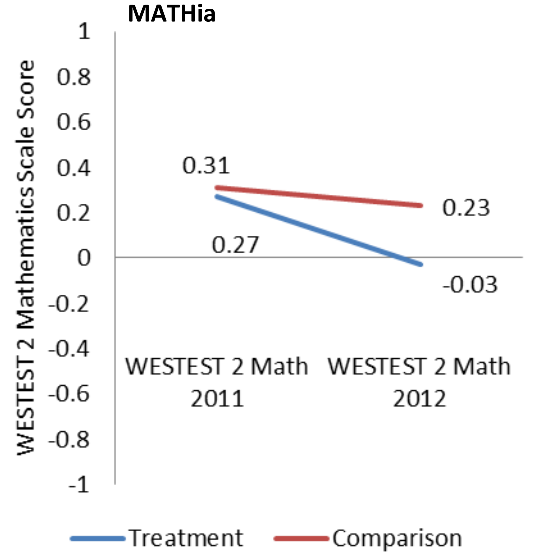


Figure 9. **Grade 5-to-6 Mathematics Gain Treatment vs. Control—210 Days in MATHia**

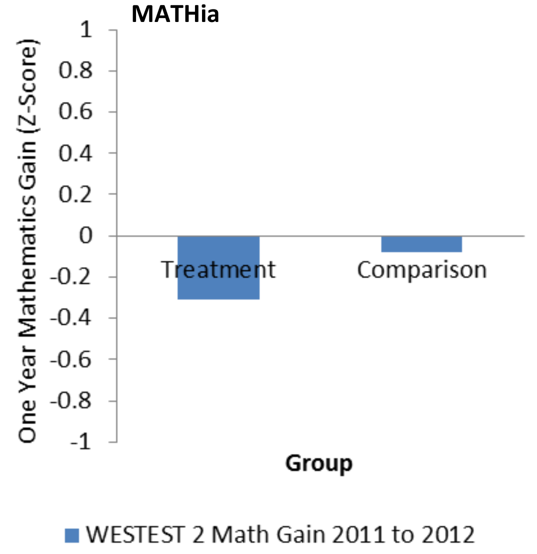


Figure 10. Grade 7 Mathematics Achievement Treatment vs. Control—210 Days in MATHia

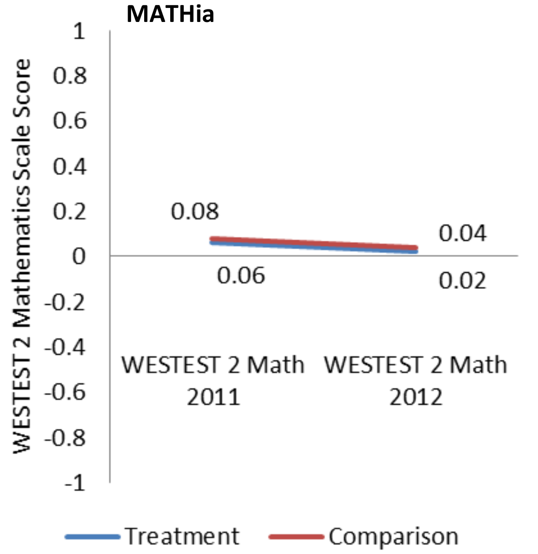


Figure 11. Grade 6-to-7 Mathematics Gain Treatment vs. Control—210 Days in MATHia

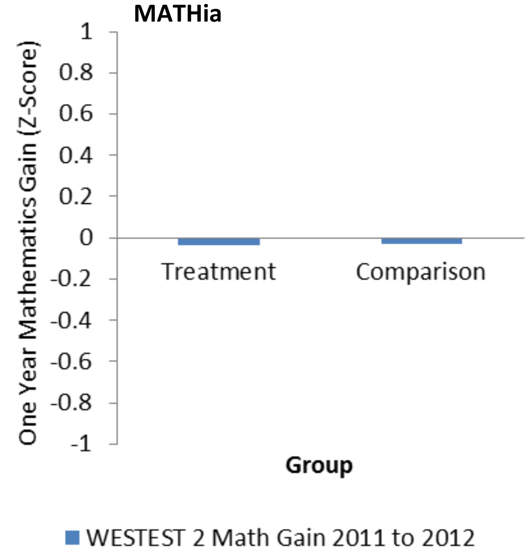


Figure 12. Grade 8 Mathematics Achievement Treatment vs. Control—210 Days in MATHia

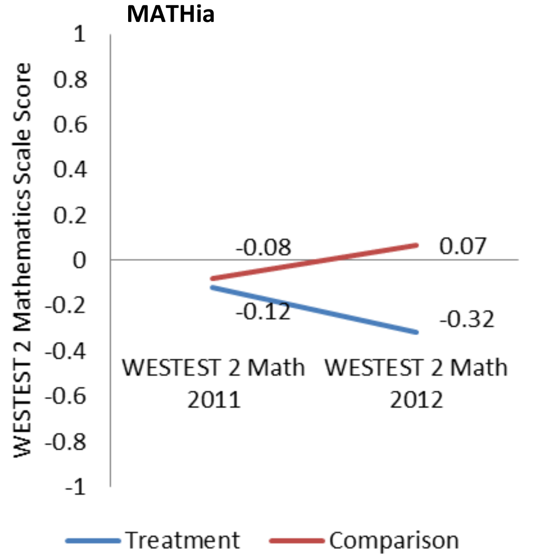
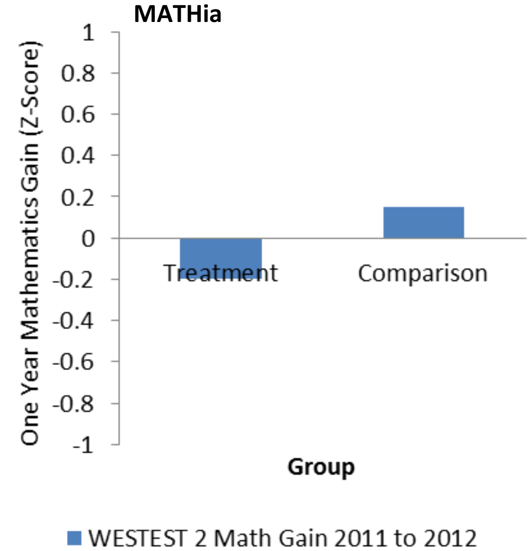


Figure 13. Grade 7-to-8 Mathematics Gain Treatment vs. Control—210 Days in MATHia



Tables C19–C36 in Appendix C present detailed results of the six general linear models we used to test the explanatory power of the treatment on students' math achievement and gains after accounting for all measured covariates. Table 11 below provides a summary of those models. As displayed below, the treatment coefficient was a statistically significant and negative predictor in four of the six models—it was not a significant predictor of Grade 7 math achievement or gains. However, it should be noted that, when significant, the treatment coefficient only contributed minor explanatory power to the model after accounting for covariates (i.e., between 1.5% and 4.4% of the variance).

Table 11. Abbreviated Linear Model Summaries for Hypothesis 3

Model	Model 1 Adj. R^2	Model 2 R^2 change	*p* value for model 2	*p* value for treatment coefficient	Standardized β for treatment coefficient	Interpretation
Grade 6 Math Achievement	.618	.015	.000	**.000**	-.121	The treatment was a statistically significant and negative predictor in all but two models. When significant, the treatment only contributed between 1.5% and 4.4% explanatory power after accounting for covariates.
Grade 6 Math Gains	.266	.028	.000	**.000**	-.168	
Grade 7 Math Achievement	.525	.000	.000	.734	-.011	
Grade 7 Math Gains	.191	.000	.000	.734	-.014	
Grade 8 Math Achievement	.349	.036	.000	**.000**	-.191	
Grade 8 Math Gains	.203	.044	.000	**.000**	-.211	

Hypothesis 4

Hypothesis 4 stated, "Students who have used the MATHia software during the 2011–2012 school year for the recommended minimum of at least 1 hour per week will score significantly higher on the WESTEST 2 math subtest than students who have not used the software." Table 12 presents the results of the *t* tests, aggregating all three grade levels.

Table 12. *T* Test Results for Hypothesis 4

Grade	*T* Mean (*sd*)	*C* Mean (*sd*)	*t*	*df*	*p*	Mean difference	Significant?
2011–2012 math achievement							
6, 7, and 8	-.25 (.85)	.05 (.76)	-3.327	294	**.001**	-.31	YES
2010–2011 to 2011–2012 math gains							
6, 7, and 8	-.16 (.76)	.08 (.60)	-3.103	279.092	**.002**	-.25	YES

As is evidenced in Table 12 and in Figure 14 and Figure 15, despite starting at comparable points in 2011 mathematics and reading/language arts achievement and having remarkably similar demographic characteristics, students in the treatment group scored significantly lower than students in the comparison group on the WESTEST 2 mathematics assessment in 2012. Additionally, treatment group students exhibited significantly lower mathematics gains from 2010–2011 to 2011–2012 than students in the comparison group. These findings were counter to our hypothesis.

Figure 14. All Grades Mathematics Achievement, Treatment vs. Control—At Least 1 Hour/Week

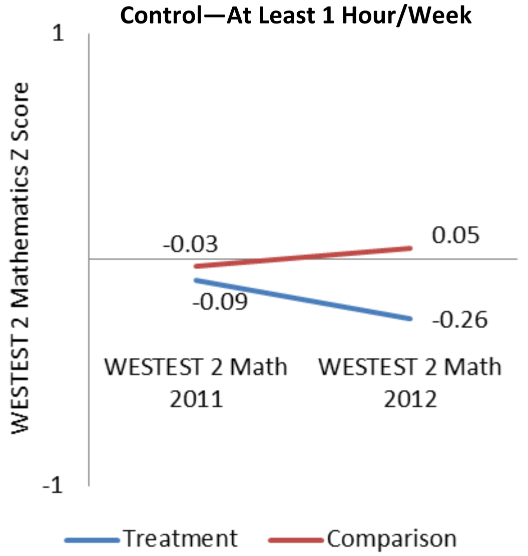
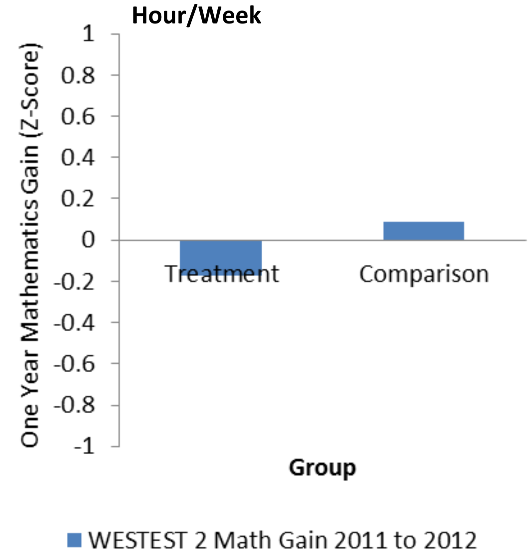


Figure 15. 2011-to-2012 Mathematics Gain, Treatment vs. Control—At Least 1 Hour/Week



Tables C37–C42 in Appendix C present detailed results of the two general linear models that we used to test the explanatory power of the treatment on students' math achievement and gains after accounting for all measured covariates. Table 13 below provides a summary of those models. As displayed below, the treatment coefficient was a statistically significant and negative predictor in both models. Though significant, the treatment coefficient only contributed minor explanatory power to the model after accounting for covariates (i.e., between 2.6% and 3.4% of the variance).

Table 13. Abbreviated Linear Model Summaries for Hypothesis 4

Model	Model 1 Adj. R^2	Model 2 R^2 change	*p* value for model 2	*p* value for treatment coefficient	Standardized β for treatment coefficient	Interpretation
All grades math achievement	.476	.026	.000	**.000**	-.161	The treatment was a statistically significant and negative predictor in both models, but only contributed between 2.6% and 3.4% explanatory power after accounting for covariates.
All grades math gains	.120	.034	.000	**.001**	-.186	

Hypothesis 5

Hypothesis 5 stated, "Students who have used the MATHia software during the 2011–2012 school year for the recommended minimum of at least 1 hour per week will score significantly higher than students who have used MATHia for less than the recommended minimum of at least 1 hour per week." Table 14 presents the results of the *t* test analyses, aggregating all three grade levels.

Table 14. *T* Test Results for Hypothesis 5

Grade	T mean (*sd*)	C mean (*sd*)	*t*	*df*	*p*	Mean difference	Significant?
2011–2012 math achievement							
6, 7, and 8	-.24 (.83)	-.14 (.93)	-.951	286	.340	-.09	NO
2010–2011 to 2011–2012 math gains							
6, 7, and 8	-.15 (.75)	-.10 (.73)	-.618	286	.537	-.05	NO

As is evidenced in Table 14 and in Figure 16 and Figure 17, there were no statistically significant differences between high and low use students with respect to 2012 math achievement or math gains from 2010–2011 to 2011-12.

Tables C43–C48 in Appendix C present detailed results of the two general linear models we used to test the explanatory power of the treatment on students' math achievement and gains after accounting for all measured covariates. Table 15 below provides a summary of those models. As displayed below, the treatment coefficient was not statistically significant in either model.

Figure 16. All Grades Mathematics Achievement High Treatment vs. Low Treatment

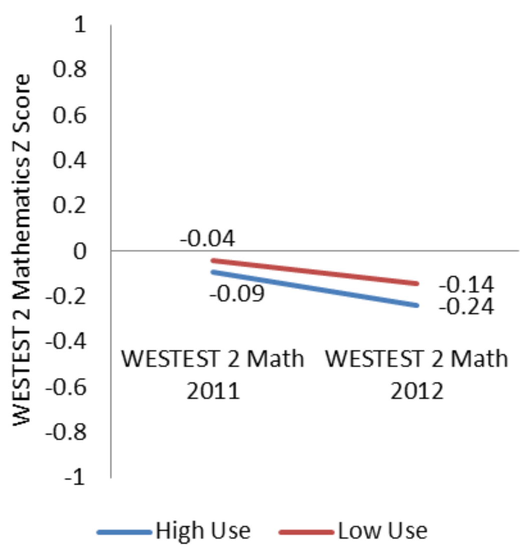


Figure 17. 2011-to-2012 Mathematics Gain High Treatment vs. Low Treatment

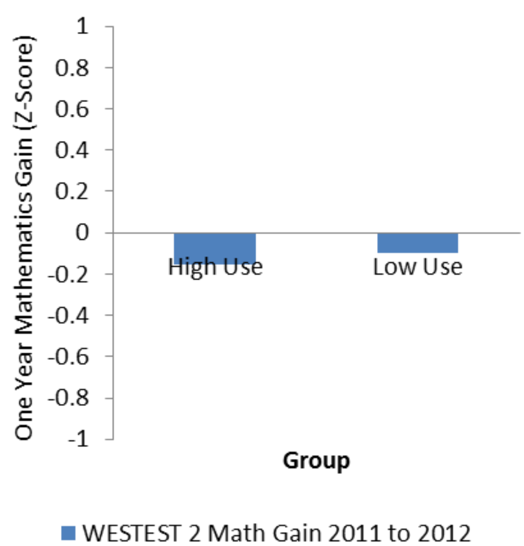


Table 15. Abbreviated Linear Model Summaries for Hypothesis 5

Model	Model 1 Adj. R^2	Model 2 R^2 change	*p* value for model 2	*p* value for treatment coefficient	Standardized β for treatment coefficient	Interpretation
All grades math achievement	.462	.001	.000	.415	-.036	The treatment was not a statistically significant predictor in either model.
All grades math gains	.081	.003	.000	.355	-.053	

Discussion

Hypothesis 1 posited that, "Teachers who participate in training provided as part of the Middle School Algebra Readiness Initiative (MSARI) will exhibit significantly greater posttest scores on the *Learning Mathematics for Teaching* (LMT) patterns, functions and algebra assessment." Based on the results of our analyses, we rejected this hypothesis. We found that teachers' performance on a rigorously developed and research validated teacher assessment of content and pedagogical knowledge remained virtually static from pretest to posttest. There was a negligible gain for the participating teachers who completed both the pretest and posttest, but statistical tests revealed that this gain was statistically insignificant.

There are multiple potential explanations for this finding including a possible lack of quality in the training provided, poor retention of the material on the part of participating teachers, or a low degree of alignment between the training and the content that appears on the LMT assessment. Without additional contextual knowledge, we can only speculate. We must acknowledge several limitations in our ability to thoroughly test Hypothesis 1. First, our final analysis was limited by the fact that we received completed LMT posttests for only 20 of the original 36 teachers who completed the pretest at the outset of the study (55%). Because we required a completed pretest and posttest record for our analyses, this limited us to 20 cases for analysis. Approximately 44 cases would be required to have confidence in our ability to detect statistically significant, but small effect sizes. It is unclear from this study why the remaining 16 teachers did not complete the posttest. It is possible that they simply opted out of taking the assessment given its voluntary nature, but it is equally possible that they ceased participating in the initiative altogether. It was clear from a post-hoc examination of the pretest results that the average score for the 20 teachers who ended up persisting throughout the entire initiative and who ultimately completed a posttest was higher than the pretest score for all 36 teachers—.43 versus .19, a difference of .24. In terms of a raw score, the average for all 36 pretest completers was two points lower than the score for those pretest completers who persisted long enough to complete a posttest. Given these findings, it appears that those teachers who persisted in the initiative were potentially more knowledgeable in the concepts measured by the LMT than those who did not persist. This uncertainty raises questions about the degree to which the outcomes we observed in our study would be different if all pretested teachers were included.

Hypotheses 2 through 5 were concerned with ascertaining the impact of participating in the MSARI on students' math achievement and gains. In all cases, the results indicated to us that we should reject our conjecture that treatment group students would outperform comparison group students. In fact, in most cases, students in the treatment condition exhibited lower math achievement and gains than students in the comparison condition. In almost all instances, these differences were statistically significant. This unanticipated and negative relationship persisted when examining only those students who met the limited fidelity criteria recommended by Carnegie Learning and after accounting for the impact of multiple covariates. The relationship was strongest in our analyses of Grade 8 outcomes. However, we must acknowledge that once covariates were controlled for, the negative relationship between treatment and students' math achievement and gains was very small, accounting for less than 5% of the total variance in all tested models.

The persistence of this negative relationship is quite troubling. Consider the results of Hypothesis 4, which illustrate that prior to the intervention, both groups of students exhibited math achievement that placed them just below the state average when compared with their grade-level peers. The difference between these groups at baseline was not statistically significant. However, after the intervention year, students in the treatment condition, all of whom used the MATHia software for one or more hours a week, declined in achievement to land approximately a quarter of a standard deviation below the state average—a considerable deficit. Meanwhile, their peers in the comparison group who did not implement MATHia managed to achieve a small gain, which placed them at the state average when compared with their grade-level peers[4]. This same trend was evidenced with respect to Grade 8 in Hypothesis 2 and Hypothesis 3. Also consider the results of Hypothesis 2, where both groups of sixth grade students started approximately one third of a standard deviation above the state average prior to the intervention, a considerable advantage. Again, this difference at baseline was not statistically significant. However, at the conclusion of the intervention year, students in the comparison condition had regressed to land nearly a quarter standard deviation above the state average while students in the treatment condition, all of whom had at least 210 days between their first and last MATHia session, regressed to the state average. To be absolutely clear, we are not stating that the intervention caused these negative effects, but nevertheless, this is what we have observed in a defensibly designed quasi-experiment.

We must caution readers of this report against interpreting these results as definitive evidence of the general efficacy of the MATHia software and accompanying classroom curriculum for multiple reasons. First, our evaluation was never intended nor was it adequately designed to make judgments about the quality of the MATHia software program or curriculum itself. Rather, our goal was to ascertain the impact of two districts' individual implementations of that curriculum on teacher knowledge and student achievement on the state summative assessment, WESTEST 2. It would require a complex experimental design study with random assignment to fairly evaluate the program itself. Second, for our evaluation to stand as a fair trial of the program's efficacy, we would require detailed information about the degree to which the program was implemented with fidelity in participating classrooms. Unfortunately, a major limitation of this study is that very little is known to us about the quality of implementation in these two districts. The fidelity metric available to us only addressed the *quantity* of time students spent using the software program. We did not have access to any data that would stand as a suitable proxy measure of the *quality* of time spent in either the computer lab or the classroom. What we do know is that, based on data provided by Carnegie Learning, very few students met the recommended 1.5 hours per week spent using the software program. In fact, even once we relaxed our fidelity criteria to 1 or more hours per week spent with the software program, we found that there were very few students that met this criterion. This is a strong indication that there was a significant gap in imple-

[4] We must stress that, as noted earlier, despite confirming matching across these two groups, this analysis does not take into account the influence of covariates on student achievement.

mentation fidelity. Also, the fact that we observed no significant differences in the mathematics achievement/gains of two matched groups of students, both of which used the MATHia software/curriculum, but which either did or did not meet the fidelity criteria rec-recommended by the vendor, illustrates that the quality of implementation in the high use group may have been less than ideal.

Given these limitations we must restate that we do not support using the results of this evaluation as a broader evaluation of the MATHia software program or classroom curriculum. Instead, we suggest interpreting these results as evidence regarding the importance of consistent and careful monitoring when implementing an intervention of this nature. We can only speculate that failure to implement with fidelity is what contributed to the lackluster student outcomes we observed.

As we alluded to above, our interpretation of these results is also exacerbated by a lack of knowledge about how well each classroom implemented the accompanying curriculum. Based on a cursory review of the standards assessed by the WESTEST 2 and the curriculum materials available from Carnegie Learning, we believe the curriculum selected has a generally quite reasonable alignment with the content assessed on the Grade 6, Grade 7, and Grade 8 math assessments. Therefore, it was a reasonable assumption that this curriculum, implemented with fidelity, would contribute to increase math knowledge on the part of students. However, if teachers did not progress fully through the curriculum prior to the administration of WESTEST 2 or if the quality of their implementation of the curriculum was suspect, that could certainly explain some of the results we observed. Without answers to these critical questions, we are left wondering and have very little conclusive knowledge of the reasons behind our results.

Other outstanding questions from this study include why there was such a discrepancy in the number of students provided to us by the school districts vs. Carnegie Learning. It is possible that districts systematically excluded some classrooms from the lists they provided or dropped classes from the initiative early on. Without additional contextual information, we cannot be sure. We also know very little about the quality of the training provided to teachers and if this training was focused on content knowledge and pedagogy alone or also on appropriate implementation of the software program and accompanying curriculum.

Recommendations

We make only two recommendations based on these results. First, future evaluations of this nature should include a great deal more data collection related to fidelity of implementation. Without this information, we are left to speculate the context surrounding the results we observed. It is possible that Carnegie Learning collects nuanced information in this regard, but it was not available for this report. Even if such information were provided, we did not have the manpower sufficient to analyze these data or to collect additional qualitative data regarding this aspect of the project. Future projects should devote at least some portion of the study budget to program evaluation so that it is not completely undertaken as an in-kind effort. Second, it is clear that close monitoring and technical assistance are critical to ensuring that this type of program is implemented appropriately. Deviations from appropriate implementation have unintended effects. It is apparent that some monitoring did take place throughout the project, but continued close observation and technical assistance are necessary to ensure the program is successful. If other school districts are implementing this program, we recommend some level of ongoing monitoring take place. This monitoring should include measuring the amount and quality of time spent in the computer laboratory and the level of individual classroom-level progress through the curriculum, as well as the extent to which teachers deliver the curriculum as intended.

References

Hall, H.C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105* (1), 11- 30.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Appendix A. Covariate Balance Summaries for Student-level Analyses

Table A1. Covariate Balance Summary for SF16

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	.0262	.0224	.0037	*	0.0261	0.001	99.47
2010–2011 RLA	466.11	452.60	13.50	*	466.96	-0.852	93.68
2010–2011 MATH	618.41	601.31	17.09	*	617.96	0.441	97.41
FRPL	.446	.547	-.102	*	0.448	-0.002	97.74
SPED	.09	.13	-.046	*	0.057	0.029	35.47
RACE	.046	.08	-.034	*	0.048	-0.002	93.26

*Treatment post mean is identical to treatment pre mean because no cases were discarded.

Table A2. Covariate Balance Summary for SF17

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	0.0285	0.0245	0.003	*	0.0285	0.000	99.99
2010–2011 RLA	482.37	467.46	14.90	*	482.40	-0.038	99.74
2010–2011 MATH	627.03	619.01	8.02	*	628.41	-1.38	82.77
FRPL	0.437	0.524	-0.086	*	0.448	-0.010	87.55
SPED	0.049	0.119	-0.070	*	0.036	0.012	81.69
RACE	0.021	0.078	-0.056	*	0.023	-0.002	96.21

*Treatment post mean is identical to treatment pre mean because no cases were discarded.

Table A3. Covariate Balance Summary for SF18

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	0.0207	0.0196	0.001	*	0.0207	0.000	99.97
2010–2011 RLA	478.17	477.24	0.934	*	478.30	-0.128	86.30
2010–2011 MATH	625.39	633.37	-7.977	*	626.95	-1.562	80.41
FRPL	0.517	0.512	0.005	*	0.504	0.013	-156.50**
SPED	0.117	0.122	-0.004	*	0.072	0.045	-830.31**
RACE	0.061	0.076	-0.015	*	0.053	0.008	47.77

*Treatment post mean is identical to treatment pre mean because no cases were discarded.
**The model decreased the balance across groups for this variable. However, post mean differences were close to zero indicating this may be of little concern.

Table A4. Covariate Balance Summary for SF26

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	0.0093	0.0083	0.001	*	0.0093	0.000	99.94
2010–2011 RLA	457.77	452.60	5.17	*	431.11	-3.339	35.47
2010–2011 MATH	614.74	601.31	13.43	*	616.69	-1.94	85.48
FRPL	0.528	0.547	-0.019	*	0.534	-0.006	67.76
SPED	0.113	0.133	-0.020	*	0.075	0.037	-84.42**
RACE	0.025	0.080	-0.054	*	0.031	-0.006	88.55

*Treatment post mean is identical to treatment pre mean because no cases were discarded.
**The model decreased the balance across groups for this variable. However, post mean differences were close to zero indicating this may be of little concern.

Table A5. Covariate Balance Summary for SF27

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	0.0153	0.0128	0.0025	*	0.0153	0.000	99.99
2010–2011 RLA	480.71	467.46	13.24	*	480.07	0.633	95.21
2010–2011 MATH	620.81	619.01	1.80	*	621.71	-0.900	50.14
FRPL	0.537	0.524	0.013	*	0.554	-0.016	-23.10**
SPED	0.062	0.119	-0.057	*	0.058	0.004	92.71
RACE	0.016	0.078	-0.061	*	0.012	0.004	93.22

*Treatment post mean is identical to treatment pre mean because no cases were discarded.
**The model decreased the balance across groups for this variable. However, post mean differences were close to zero indicating this may be of little concern.

Table A6. Covariate Balance Summary for SF28

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	0.0104	0.0089	0.0015	*	0.0104	0.000	99.99
2010–2011 RLA	483.04	477.24	5.79	*	483.76	-0.726	87.47
2010–2011 MATH	626.01	633.37	-7.35	*	627.88	-1.86	74.66
FRPL	0.541	0.512	0.029	*	0.506	0.035	-20.93
SPED	0.065	0.122	-0.056	*	0.053	0.011	79.01
RACE	0.053	0.076	-0.023	*	0.041	0.011	48.42

*Treatment post mean is identical to treatment pre mean because no cases were discarded.
**The model decreased the balance across groups for this variable. However, post mean differences were close to zero indicating this may be of little concern.

Table A7. Covariate Balance Summary for SF3

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	0.0031	0.0026	0.0005	*	0.0031	0.000	99.99
2010–2011 RLA	466.54	465.73	0.807	*	466.20	0.337	58.15
2010–2011 MATH	614.09	617.85	-3.76	*	617.26	-3.16	15.72
FRPL	0.587	0.528	0.059	*	0.567	0.020	66.10
SPED	0.087	0.125	-0.037	*	0.067	0.020	45.80
RACE	0.020	0.078	-0.058	*	0.020	0.000	100.00
GRD6	0.337	0.336	0.001	*	0.331	0.006	-466.34**
GRD7	0.229	0.329	-0.099	*	0.223	0.006	93.19
GRD8	0.432	0.334	0.098	*	0.445	-0.013	86.22

*Treatment post mean is identical to treatment pre mean because no cases were discarded.
**The model decreased the balance across groups for this variable. However, post mean differences were close to zero indicating this may be of little concern.

Table A8. Covariate Balance Summary for SF4

Covariate	Treatment Pre Mean	Comparison Pre Mean	Pre Mean diff	Treatment Post Mean	Comparison Post Mean	Post Mean diff	% Improvement post-matching
Propensity Score	0.1395	0.1095	0.030	*	0.1395	0.000	99.93
2010–2011 RLA	466.68	476.72	-10.04	*	465.84	0.833	91.70
2010–2011 MATH	613.93	624.84	-10.91	*	615.65	-1.722	84.22
FRPL	0.583	0.448	0.134	*	0.555	0.027	79.35
SPED	0.090	0.081	0.009	*	0.097	-0.006	22.88
RACE	0.020	0.044	-0.023	*	0.006	0.013	40.48
GRD6	0.347	0.340	0.007	*	0.361	-0.013	-95.17**
GRD7	0.222	0.383	-0.161	*	0.215	0.006	95.69
GRD8	0.430	0.276	0.154	*	0.423	0.006	95.49

*Treatment post mean is identical to treatment pre mean because no cases were discarded.
**The model decreased the balance across groups for this variable. However, post mean differences were close to zero indicating this may be of little concern.

Appendix B. Power Analyses for Student-level Analyses

Table B1. Sample Size and Power by Sampling Frame (Hypothesis 2)			
Sample	Comparison	Treatment	Power to Detect Moderate Effect (α ≥ .95)
SF16			
All	18,875	435	.
Matched	435	435	.
Unmatched	18,440	0	.
Final N	870		YES
SF17			
All	18446	466	.
Matched	466	466	.
Unmatched	17980	0	.
Final N		932	YES
SF18			
All	18747	375	.
Matched	375	375	.
Unmatched	18372	0	.
Final N		750	YES

Table B2. Sample Size and Power by Sampling Frame (Hypothesis 3)			
Sample	Comparison	Treatment	Power to Detect Moderate Effect (α ≥ .95)
SF26			
All	18875	159	.
Matched	159	159	.
Unmatched	18716	0	.
Final N		318	NO (.76)
SF27			
All	18446	240	.

Table B2. Sample Size and Power by Sampling Frame (Hypothesis 3)			
Sample	Comparison	Treatment	Power to Detect Moderate Effect (α ≥ .95)
Matched	240	240	.
Unmatched	18,206	0	.
Final N		480	NO (.90)
SF28			
All	18,747	168	.
Matched	168	168	.
Unmatched	18,579	0	.
Final N		336	NO (.78)

Table B3. Sample Size and Power by Sampling Frame (Hypothesis 4)			
Sample	Comparison	Treatment	Power to Detect Moderate Effect (α ≥ .95)
SF3			
All	56068	148	.
Matched	148	148	.
Unmatched	55920	0	.
Final N		296	NO (.73)

Table B4. Sample Size and Power by Sampling Frame (Hypothesis 5)			
Sample	Comparison	Treatment	Power to Detect Moderate Effect (α ≥ .95)
SF4			
All	1132	144	.
Matched	144	144	.
Unmatched	988	0	.
Final N		288	NO (.72)

Appendix C. Detailed Statistics for Linear Models Used to Test the Impact of Treatment when Accounting for Measured Covariates

Table C1. Model Summaries for SF16 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.791[a]	.626	.623	.568752161645	.626	240.367	6	863	.000
2	**.793[b]**	**.628**	**.625**	**.567148612508**	**.003**	**5.887**	**1**	**862**	**.015**

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C2. ANOVA Statistics for SF16 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	466.522	6	77.754	240.367	.000[b]
	Residual	279.162	863	.323		
	Total	745.684	869			
2	Regression	468.415	7	66.916	208.036	.000[c]
	Residual	277.269	862	.322		
	Total	745.684	869			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C3. Coefficient Summaries for SF16 (2011–2012 Math Achievement)

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	-9.437	.318		-29.642	.000	-10.061	-8.812
SSM11	.011	.001	.508	17.171	.000	.009	.012
SSR11	.007	.001	.281	9.114	.000	.005	.008
WHITE11	.035	.092	.008	.384	.701	-.144	.215
LSES11	-.145	.042	-.078	-3.407	.001	-.228	-.061
SPED11	-.370	.077	-.104	-4.804	.000	-.522	-.219
SEX11	.076	.040	.041	1.892	.059	-.003	.156
(Constant)	-9.399	.318		-29.571	.000	-10.023	-8.775
SSM11	.011	.001	.508	17.239	.000	.009	.012
SSR11	.007	.001	.281	9.138	.000	.005	.008
WHITE11	.034	.091	.008	.377	.706	-.145	.214
LSES11	-.145	.042	-.078	-3.426	.001	-.228	-.062
SPED11	-.360	.077	-.101	-4.675	.000	-.511	-.209
SEX11	.079	.040	.043	1.964	.050	.000	.158
treat	**-.094**	**.039**	**-.051**	**-2.426**	**.015**	**-.169**	**-.018**

a. Dependent Variable: ZSSM12

Table C4. Model Summaries for SF16 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.477[a]	.228	.222	.568752161648	.228	42.388	6	863	.000
2	**.483[b]**	**.233**	**.227**	**.567148612511**	**.005**	**5.887**	**1**	**862**	**.015**

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C5. ANOVA Statistics for SF16 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	82.269	6	13.712	42.388	.000[b]
	Residual	279.162	863	.323		
	Total	361.432	869			
2	Regression	84.163	7	12.023	37.379	.000[c]
	Residual	277.269	862	.322		
	Total	361.432	869			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C6. Coefficient Summaries for SF16 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2.474	.318		7.771	.000	1.849	3.099
	SSM11	-.009	.001	-.633	-14.900	.000	-.010	-.008
	SSR11	.007	.001	.403	9.114	.000	.005	.008
	WHITE11	.035	.092	.012	.384	.701	-.144	.215
	LSES11	-.145	.042	-.111	-3.407	.001	-.228	-.061
	SPED11	-.370	.077	-.149	-4.804	.000	-.522	-.219
	SEX11	.076	.040	.059	1.892	.059	-.003	.156
2	(Constant)	2.512	.318		7.902	.000	1.888	3.135
	SSM11	-.009	.001	-.632	-14.921	.000	-.010	-.008
	SSR11	.007	.001	.403	9.138	.000	.005	.008
	WHITE11	.034	.091	.011	.377	.706	-.145	.214
	LSES11	-.145	.042	-.112	-3.426	.001	-.228	-.062
	SPED11	-.360	.077	-.145	-4.675	.000	-.511	-.209
	SEX11	.079	.040	.061	1.964	.050	.000	.158
	treat	**-.094**	**.039**	**-.073**	**-2.426**	**.015**	**-.169**	**-.018**

a. Dependent Variable: MATHGAIN

Table C7. Model Summaries for SF17 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.773[a]	.598	.595	.616209861140	.598	229.066	6	925	.000
2	**.775[b]**	**.601**	**.598**	**.614210903534**	**.003**	**7.031**	**1**	**924**	**.008**

a. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11, treat

Table C8. ANOVA Statistics for SF17 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	521.878	6	86.980	229.066	.000b
	Residual	351.236	925	.380		
	Total	873.114	931			
2	Regression	524.531	7	74.933	198.627	.000c
	Residual	348.584	924	.377		
	Total	873.114	931			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11, treat

Table C9. Coefficient Summaries for SF17 (Math Achievement)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-10.599	.367		-28.917	.000	-11.318	-9.879
	SSM11	.012	.001	.536	18.642	.000	.011	.013
	SSR11	.007	.001	.239	8.188	.000	.005	.008
	WHITE11	.224	.137	.034	1.641	.101	-.044	.492
	LSES11	-.184	.045	-.095	-4.110	.000	-.272	-.096
	SPED11	-.123	.102	-.026	-1.204	.229	-.323	.077
	SEX11	.144	.042	.074	3.455	.001	.062	.226
2	(Constant)	-10.530	.366		-28.750	.000	-11.249	-9.811
	SSM11	.012	.001	.536	18.674	.000	.011	.013
	SSR11	.007	.001	.239	8.213	.000	.005	.008
	WHITE11	.222	.136	.034	1.632	.103	-.045	.489
	LSES11	-.187	.045	-.096	-4.175	.000	-.274	-.099
	SPED11	-.114	.102	-.024	-1.120	.263	-.314	.086
	SEX11	.137	.042	.071	3.302	.001	.056	.219
	treat	**-.107**	**.040**	**-.055**	**-2.652**	**.008**	**-.186**	**-.028**

a. Dependent Variable: ZSSM12

Table C10. Model Summaries for SF17 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.456[a]	.208	.203	.616209861137	.208	40.577	6	925	.000
2	**.463[b]**	**.214**	**.208**	**.614210903531**	**.006**	**7.031**	**1**	**924**	**.008**

a. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11, treat

Table C11. ANOVA Statistics for SF17 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	92.447	6	15.408	40.577	.000[b]
	Residual	351.236	925	.380		
	Total	443.683	931			
2	Regression	95.100	7	13.586	36.012	.000[c]
	Residual	348.584	924	.377		
	Total	443.683	931			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, WHITE11, LSES11, SPED11, SSM11, SSR11, treat

Table C12. Coefficient Summaries for SF17 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	3.113	.367		8.494	.000	2.394	3.833
	SSM11	-.010	.001	-.622	-15.405	.000	-.011	-.009
	SSR11	.007	.001	.335	8.188	.000	.005	.008
	WHITE11	.224	.137	.048	1.641	.101	-.044	.492
	LSES11	-.184	.045	-.133	-4.110	.000	-.272	-.096
	SPED11	-.123	.102	-.036	-1.204	.229	-.323	.077
	SEX11	.144	.042	.104	3.455	.001	.062	.226
2	(Constant)	3.182	.366		8.688	.000	2.463	3.901
	SSM11	-.010	.001	-.623	-15.482	.000	-.011	-.009
	SSR11	.007	.001	.335	8.213	.000	.005	.008
	WHITE11	.222	.136	.048	1.632	.103	-.045	.489
	LSES11	-.187	.045	-.134	-4.175	.000	-.274	-.099
	SPED11	-.114	.102	-.033	-1.120	.263	-.314	.086
	SEX11	.137	.042	.100	3.302	.001	.056	.219
	treat	**-.107**	**.040**	**-.078**	**-2.652**	**.008**	**-.186**	**-.028**

a. Dependent Variable: MATHGAIN

Table C13. Model Summaries for SF18 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.663[a]	.440	.435	.718734366395	.440	97.220	6	743	.000
2	**.676[b]**	**.456**	**.451**	**.708537706902**	**.017**	**22.539**	**1**	**742**	**.000**

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C14. ANOVA Statistics for SF18 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	301.331	6	50.222	97.220	.000b
	Residual	383.818	743	.517		
	Total	685.149	749			
2	Regression	312.646	7	44.664	88.967	.000c
	Residual	372.503	742	.502		
	Total	685.149	749			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C15. Coefficient Summaries for SF18 (Math Achievement)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-8.328	.474		-17.581	.000	-9.258	-7.398
	SSM11	.011	.001	.524	14.842	.000	.010	.013
	SSR11	.003	.001	.114	3.211	.001	.001	.005
	WHITE11	.024	.114	.006	.208	.836	-.200	.247
	LSES11	-.143	.055	-.075	-2.585	.010	-.251	-.034
	SPED11	-.297	.099	-.091	-2.992	.003	-.492	-.102
	SEX11	-.009	.056	-.005	-.156	.876	-.118	.101
2	(Constant)	-8.269	.467		-17.700	.000	-9.186	-7.351
	SSM11	.011	.001	.524	15.074	.000	.010	.013
	SSR11	.003	.001	.118	3.373	.001	.001	.005
	WHITE11	.029	.112	.007	.259	.795	-.191	.249
	LSES11	-.140	.054	-.073	-2.573	.010	-.247	-.033
	SPED11	-.261	.098	-.080	-2.659	.008	-.454	-.068
	SEX11	-.004	.055	-.002	-.078	.938	-.112	.103
	treat	**-.247**	**.052**	**-.129**	**-4.748**	**.000**	**-.348**	**-.145**

a. Dependent Variable: ZSSM12

Table C16. Model Summaries for SF18 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.415[a]	.173	.166	.718734366375	.173	25.824	6	743	.000
2	**.444[b]**	**.197**	**.189**	**.708537706884**	**.024**	**22.539**	**1**	**742**	**.000**

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C17. ANOVA Statistics for SF18 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	80.042	6	13.340	25.824	.000[b]
	Residual	383.818	743	.517		
	Total	463.860	749			
2	Regression	91.357	7	13.051	25.997	.000[c]
	Residual	372.503	742	.502		
	Total	463.860	749			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C18. Coefficient Summaries for SF18 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	4.341	.474		9.165	.000	3.412	5.271
	SSM11	-.009	.001	-.514	-11.980	.000	-.010	-.007
	SSR11	.003	.001	.138	3.211	.001	.001	.005
	WHITE11	.024	.114	.007	.208	.836	-.200	.247
	LSES11	-.143	.055	-.091	-2.585	.010	-.251	-.034
	SPED11	-.297	.099	-.111	-2.992	.003	-.492	-.102
	SEX11	-.009	.056	-.006	-.156	.876	-.118	.101
2	(Constant)	4.401	.467		9.421	.000	3.484	5.318
	SSM11	-.009	.001	-.513	-12.134	.000	-.010	-.007
	SSR11	.003	.001	.143	3.373	.001	.001	.005
	WHITE11	.029	.112	.009	.259	.795	-.191	.249
	LSES11	-.140	.054	-.089	-2.573	.010	-.247	-.033
	SPED11	-.261	.098	-.097	-2.659	.008	-.454	-.068
	SEX11	-.004	.055	-.003	-.078	.938	-.112	.103
	treat	**-.247**	**.052**	**-.157**	**-4.748**	**.000**	**-.348**	**-.145**

a. Dependent Variable: MATHGAIN

Table C19. Model Summaries for SF26 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.790[a]	.625	.618	.526955678942	.625	86.318	6	311	.000
2	**.800[b]**	**.639**	**.631**	**.517467436514**	**.015**	**12.510**	**1**	**310**	**.000**

a. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11

b. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11, treat

Table C20. ANOVA Statistics for SF26 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	143.815	6	23.969	86.318	.000[b]
	Residual	86.359	311	.278		
	Total	230.174	317			
2	Regression	147.164	7	21.023	78.512	.000[c]
	Residual	83.009	310	.268		
	Total	230.174	317			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11

c. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11, treat

Table C21. Coefficient Summaries for SF26 (Math Achievement)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-9.658	.525		-18.387	.000	-10.691	-8.624
	SSM11	.011	.001	.503	10.818	.000	.009	.013
	SSR11	.007	.001	.286	5.979	.000	.005	.009
	WHITE11	-.134	.181	-.026	-.741	.459	-.489	.221
	LSES11	-.155	.063	-.091	-2.471	.014	-.278	-.032
	SPED11	-.468	.106	-.161	-4.405	.000	-.678	-.259
	SEX11	-.016	.063	-.009	-.251	.802	-.139	.108
2	(Constant)	-9.507	.518		-18.369	.000	-10.525	-8.489
	SSM11	.011	.001	.501	10.961	.000	.009	.013
	SSR11	.007	.001	.284	6.052	.000	.005	.009
	WHITE11	-.140	.177	-.027	-.789	.430	-.489	.209
	LSES11	-.159	.062	-.093	-2.577	.010	-.280	-.037
	SPED11	-.448	.105	-.154	-4.284	.000	-.654	-.242
	SEX11	-.008	.062	-.005	-.126	.899	-.129	.114
	treat	**-.206**	**.058**	**-.121**	**-3.537**	**.000**	**-.321**	**-.091**

a. Dependent Variable: ZSSM12

Table C22. Model Summaries for SF26 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.529[a]	.280	.266	.526955678968	.280	20.191	6	311	.000
2	**.555[b]**	**.308**	**.293**	**.517467436541**	**.028**	**12.510**	**1**	**310**	**.000**

a. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11

b. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11, treat

Table C23. ANOVA Statistics for SF26 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	33.641	6	5.607	20.191	.000[b]
	Residual	86.359	311	.278		
	Total	120.000	317			
2	Regression	36.991	7	5.284	19.735	.000[c]
	Residual	83.009	310	.268		
	Total	120.000	317			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11

c. Predictors: (Constant), SEX11, WHITE11, SPED11, LSES11, SSM11, SSR11, treat

Table C24. Coefficient Summaries for SF26 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2.253	.525		4.288	.000	1.219	3.286
	SSM11	-.009	.001	-.557	-8.646	.000	-.011	-.007
	SSR11	.007	.001	.396	5.979	.000	.005	.009
	WHITE11	-.134	.181	-.036	-.741	.459	-.489	.221
	LSES11	-.155	.063	-.126	-2.471	.014	-.278	-.032
	SPED11	-.468	.106	-.223	-4.405	.000	-.678	-.259
	SEX11	-.016	.063	-.013	-.251	.802	-.139	.108
2	(Constant)	2.403	.518		4.644	.000	1.385	3.422
	SSM11	-.009	.001	-.561	-8.857	.000	-.011	-.007
	SSR11	.007	.001	.394	6.052	.000	.005	.009
	WHITE11	-.140	.177	-.038	-.789	.430	-.489	.209
	LSES11	-.159	.062	-.129	-2.577	.010	-.280	-.037
	SPED11	-.448	.105	-.213	-4.284	.000	-.654	-.242
	SEX11	-.008	.062	-.006	-.126	.899	-.129	.114
	treat	**-.206**	**.058**	**-.168**	**-3.537**	**.000**	**-.321**	**-.091**

a. Dependent Variable: MATHGAIN

Table C25. Model Summaries for SF27 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.728[a]	.531	.525	.652136894472	.531	89.121	6	473	.000
2	.729[b]	.531	.524	.652747500812	.000	.115	1	472	.734

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C26. ANOVA Statistics for SF27 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	227.408	6	37.901	89.121	.000[b]
	Residual	201.159	473	.425		
	Total	428.567	479			
2	Regression	227.458	7	32.494	76.263	.000[c]
	Residual	201.109	472	.426		
	Total	428.567	479			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C27. Coefficient Summaries for SF27 (Math Achievement)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-9.782	.541		-18.089	.000	-10.845	-8.719
	SSM11	.012	.001	.535	12.377	.000	.010	.014
	SSR11	.005	.001	.196	4.479	.000	.003	.007
	WHITE11	.149	.250	.019	.598	.550	-.341	.640
	LSES11	-.171	.066	-.090	-2.600	.010	-.301	-.042
	SPED11	-.121	.130	-.030	-.926	.355	-.377	.135
	SEX11	.150	.061	.079	2.471	.014	.031	.269
2	(Constant)	-9.770	.542		-18.013	.000	-10.836	-8.705
	SSM11	.012	.001	.534	12.354	.000	.010	.014
	SSR11	.005	.001	.196	4.480	.000	.003	.007
	WHITE11	.151	.250	.019	.603	.547	-.340	.642
	LSES11	-.172	.066	-.091	-2.604	.010	-.301	-.042
	SPED11	-.120	.130	-.030	-.923	.357	-.377	.136
	SEX11	.150	.061	.079	2.474	.014	.031	.269
	treat	**-.020**	**.060**	**-.011**	**-.340**	**.734**	**-.137**	**.097**

a. Dependent Variable: ZSSM12

Table C28. Model Summaries for SF27 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.449[a]	.202	.191	.652136894458	.202	19.904	6	473	.000
2	.449[b]	.202	.190	.652747500798	.000	.115	1	472	.734

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C29. ANOVA Statistics for SF27 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	50.788	6	8.465	19.904	.000[b]
	Residual	201.159	473	.425		
	Total	251.947	479			
2	Regression	50.837	7	7.262	17.045	.000[c]
	Residual	201.109	472	.426		
	Total	251.947	479			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C30. Coefficient Summaries for SF27 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	3.930	.541		7.267	.000	2.867	4.993
	SSM11	-.010	.001	-.586	-10.402	.000	-.012	-.008
	SSR11	.005	.001	.256	4.479	.000	.003	.007
	WHITE11	.149	.250	.025	.598	.550	-.341	.640
	LSES11	-.171	.066	-.118	-2.600	.010	-.301	-.042
	SPED11	-.121	.130	-.040	-.926	.355	-.377	.135
	SEX11	.150	.061	.103	2.471	.014	.031	.269
2	(Constant)	3.942	.542		7.267	.000	2.876	5.008
	SSM11	-.010	.001	-.587	-10.398	.000	-.012	-.008
	SSR11	.005	.001	.256	4.480	.000	.003	.007
	WHITE11	.151	.250	.025	.603	.547	-.340	.642
	LSES11	-.172	.066	-.118	-2.604	.010	-.301	-.042
	SPED11	-.120	.130	-.040	-.923	.357	-.377	.136
	SEX11	.150	.061	.103	2.474	.014	.031	.269
	treat	**-.020**	**.060**	**-.014**	**-.340**	**.734**	**-.137**	**.097**

a. Dependent Variable: MATHGAIN

Table C31. Model Summaries for SF28 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.601[a]	.361	.349	.744333247367	.361	30.959	6	329	.000
2	**.630[b]**	**.397**	**.384**	**.724035146399**	**.036**	**19.705**	**1**	**328**	**.000**

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C32. ANOVA Statistics for SF28 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	102.914	6	17.152	30.959	.000[b]
	Residual	182.277	329	.554		
	Total	285.190	335			
2	Regression	113.244	7	16.178	30.860	.000[c]
	Residual	171.946	328	.524		
	Total	285.190	335			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C33. Coefficient Summaries for SF28 (Math Achievement)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-7.224	.742		-9.741	.000	-8.683	-5.765
	SSM11	.010	.001	.447	8.096	.000	.007	.012
	SSR11	.003	.001	.108	1.816	.070	.000	.005
	WHITE11	-.273	.194	-.063	-1.404	.161	-.654	.109
	LSES11	-.242	.086	-.131	-2.803	.005	-.411	-.072
	SPED11	-.283	.180	-.073	-1.574	.116	-.637	.071
	SEX11	.081	.089	.044	.911	.363	-.094	.257
2	(Constant)	-7.036	.723		-9.736	.000	-8.457	-5.614
	SSM11	.009	.001	.446	8.298	.000	.007	.012
	SSR11	.003	.001	.109	1.870	.062	.000	.005
	WHITE11	-.253	.189	-.058	-1.340	.181	-.625	.118
	LSES11	-.230	.084	-.125	-2.746	.006	-.395	-.065
	SPED11	-.266	.175	-.068	-1.521	.129	-.610	.078
	SEX11	.071	.087	.039	.816	.415	-.100	.242
	treat	**-.351**	**.079**	**-.191**	**-4.439**	**.000**	**-.507**	**-.196**

a. Dependent Variable: ZSSM12

Table C34. Model Summaries for SF28 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.466[a]	.217	.203	.744333247341	.217	15.229	6	329	.000
2	**.512[b]**	**.262**	**.246**	**.724035146373**	**.044**	**19.705**	**1**	**328**	**.000**

a. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C35. ANOVA Statistics for SF28 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	50.624	6	8.437	15.229	.000b
	Residual	182.277	329	.554		
	Total	232.901	335			
2	Regression	60.954	7	8.708	16.611	.000c
	Residual	171.946	328	.524		
	Total	232.901	335			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11

c. Predictors: (Constant), SEX11, LSES11, WHITE11, SPED11, SSM11, SSR11, treat

Table C36. Coefficient Summaries for SF28 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	5.445	.742		7.343	.000	3.987	6.904
	SSM11	-.011	.001	-.548	-8.967	.000	-.013	-.008
	SSR11	.003	.001	.120	1.816	.070	.000	.005
	WHITE11	-.273	.194	-.070	-1.404	.161	-.654	.109
	LSES11	-.242	.086	-.145	-2.803	.005	-.411	-.072
	SPED11	-.283	.180	-.080	-1.574	.116	-.637	.071
	SEX11	.081	.089	.049	.911	.363	-.094	.257
2	(Constant)	5.634	.723		7.797	.000	4.212	7.056
	SSM11	-.011	.001	-.550	-9.243	.000	-.013	-.008
	SSR11	.003	.001	.120	1.870	.062	.000	.005
	WHITE11	-.253	.189	-.065	-1.340	.181	-.625	.118
	LSES11	-.230	.084	-.138	-2.746	.006	-.395	-.065
	SPED11	-.266	.175	-.076	-1.521	.129	-.610	.078
	SEX11	.071	.087	.043	.816	.415	-.100	.242
	treat	**-.351**	**.079**	**-.211**	**-4.439**	**.000**	**-.507**	**-.196**

a. Dependent Variable: MATHGAIN

Table C37. Model Summaries for SF3 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.700[a]	.491	.476	.594238923336	.491	34.556	8	287	.000
2	**.719[b]**	**.516**	**.501**	**.580011587217**	**.026**	**15.253**	**1**	**286**	**.000**

a. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11

b. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11, treat

Table C38. ANOVA Statistics for SF3 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	97.619	8	12.202	34.556	.000[b]
	Residual	101.345	287	.353		
	Total	198.964	295			
2	Regression	102.750	9	11.417	33.936	.000[c]
	Residual	96.214	286	.336		
	Total	198.964	295			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11

c. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11, treat

Table C39. Coefficient Summaries for SF3 (Math Achievement)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-1.257	.129		-9.721	.000	-1.512	-1.003
	PLM11	.364	.043	.463	8.488	.000	.280	.449
	PLR11	.142	.039	.203	3.674	.000	.066	.219
	GRD6	.122	.086	.070	1.427	.155	-.046	.291
	GRD7	.177	.091	.090	1.938	.054	-.003	.356
	WHITE11	.238	.246	.041	.967	.334	-.246	.722
	LSES11	-.141	.074	-.085	-1.901	.058	-.287	.005
	SPED11	-.671	.134	-.219	-5.003	.000	-.935	-.407
	SEX11	.032	.075	.020	.436	.663	-.114	.179
2	(Constant)	-1.125	.131		-8.604	.000	-1.382	-.867
	PLM11	.358	.042	.454	8.535	.000	.275	.440
	PLR11	.145	.038	.207	3.833	.000	.071	.220
	GRD6	.128	.084	.074	1.531	.127	-.037	.292
	GRD7	.182	.089	.093	2.044	.042	.007	.357
	WHITE11	.238	.240	.041	.989	.323	-.235	.710
	LSES11	-.138	.072	-.083	-1.903	.058	-.280	.005
	SPED11	-.653	.131	-.213	-4.990	.000	-.911	-.396
	SEX11	.035	.073	.021	.480	.631	-.108	.178
	treat	**-.264**	**.068**	**-.161**	**-3.905**	**.000**	**-.397**	**-.131**

a. Dependent Variable: ZSSM12

Table C40. Model Summaries for SF3 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.379[a]	.143	.120	.660594616962	.143	6.008	8	287	.000
2	**.422[b]**	**.178**	**.152**	**.648318321312**	**.034**	**11.972**	**1**	**286**	**.001**

a. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11

b. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11, treat

Table C41. ANOVA Statistics for SF3 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20.974	8	2.622	6.008	.000[b]
	Residual	125.243	287	.436		
	Total	146.217	295			
2	Regression	26.006	9	2.890	6.875	.000[c]
	Residual	120.211	286	.420		
	Total	146.217	295			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11

c. Predictors: (Constant), SEX11, WHITE11, GRD6, LSES11, SPED11, PLM11, GRD7, PLR11, treat

Table C42. Coefficient Summaries for SF3 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.262	.144		1.822	.069	-.021	.545
	PLM11	-.249	.048	-.369	-5.214	.000	-.343	-.155
	PLR11	.097	.043	.162	2.255	.025	.012	.182
	GRD6	.066	.095	.044	.691	.490	-.122	.253
	GRD7	.283	.101	.169	2.795	.006	.084	.483
	WHITE11	.118	.273	.024	.433	.665	-.420	.657
	LSES11	-.038	.082	-.027	-.463	.644	-.200	.124
	SPED11	-.377	.149	-.144	-2.533	.012	-.671	-.084
	SEX11	-.023	.083	-.017	-.280	.779	-.186	.140
2	(Constant)	.393	.146		2.691	.008	.106	.681
	PLM11	-.255	.047	-.378	-5.445	.000	-.347	-.163
	PLR11	.100	.042	.166	2.359	.019	.017	.183
	GRD6	.071	.093	.048	.765	.445	-.112	.255
	GRD7	.289	.099	.172	2.901	.004	.093	.484
	WHITE11	.118	.268	.024	.440	.660	-.410	.646
	LSES11	-.035	.081	-.025	-.432	.666	-.194	.124
	SPED11	-.360	.146	-.137	-2.462	.014	-.648	-.072
	SEX11	-.021	.081	-.015	-.256	.798	-.181	.139
	treat	**-.261**	**.075**	**-.186**	**-3.460**	**.001**	**-.410**	**-.113**

a. Dependent Variable: MATHGAIN

Table C43. Model Summaries for SF4 (Math Achievement)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.691[a]	.477	.462	.647828147782	.477	31.857	8	279	.000
2	.692[b]	.479	.462	.648215096971	.001	.667	1	278	.415

a. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11

b. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11, treat

Table C44. ANOVA Statistics for SF4 (Math Achievement)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	106.957	8	13.370	31.857	.000[b]
	Residual	117.091	279	.420		
	Total	224.048	287			
2	Regression	107.237	9	11.915	28.357	.000[c]
	Residual	116.811	278	.420		
	Total	224.048	287			

a. Dependent Variable: ZSSM12

b. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11

c. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11, treat

Table C45. Coefficient Summaries for SF4 (Math Achievement)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-1.543	.144		-10.697	.000	-1.827	-1.259
	PLM11	.437	.048	.510	9.027	.000	.341	.532
	PLR11	.143	.042	.194	3.422	.001	.061	.225
	GRD6	.028	.094	.015	.297	.767	-.157	.213
	GRD7	.080	.102	.038	.790	.430	-.120	.281
	WHITE11	.586	.329	.078	1.780	.076	-.062	1.234
	LSES11	-.024	.081	-.014	-.299	.765	-.184	.135
	SPED11	-.468	.136	-.155	-3.429	.001	-.737	-.199
	SEX11	.022	.083	.013	.267	.790	-.142	.187
2	(Constant)	-1.512	.149		-10.132	.000	-1.806	-1.218
	PLM11	.435	.048	.508	8.977	.000	.340	.530
	PLR11	.143	.042	.194	3.431	.001	.061	.225
	GRD6	.028	.094	.015	.293	.770	-.158	.213
	GRD7	.081	.102	.038	.794	.428	-.120	.281
	WHITE11	.603	.330	.080	1.826	.069	-.047	1.253
	LSES11	-.023	.081	-.013	-.281	.779	-.182	.137
	SPED11	-.470	.137	-.155	-3.441	.001	-.739	-.201
	SEX11	.027	.084	.015	.326	.745	-.138	.192
	treat	-.063	.077	-.036	-.817	.415	-.214	.089

a. Dependent Variable: ZSSM12

Table C46. Model Summaries for SF4 (Math Gains)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.326[a]	.107	.081	.714297081012	.107	4.160	8	279	.000
2	.331[b]	.109	.080	.714478631858	.003	.858	1	278	.355

a. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11

b. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11, treat

Table 47. ANOVA Statistics for SF4 (Math Gains)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16.981	8	2.123	4.160	.000[b]
	Residual	142.351	279	.510		
	Total	159.332	287			
2	Regression	17.419	9	1.935	3.791	.000[c]
	Residual	141.913	278	.510		
	Total	159.332	287			

a. Dependent Variable: MATHGAIN

b. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11

c. Predictors: (Constant), SEX11, LSES11, SPED11, GRD7, WHITE11, PLM11, GRD6, PLR11, treat

Table 48. Coefficient Summaries for SF4 (Math Gains)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.168	.159		1.057	.291	-.145	.481
	PLM11	-.216	.053	-.299	-4.042	.000	-.321	-.111
	PLR11	.070	.046	.112	1.512	.132	-.021	.160
	GRD6	-.025	.104	-.016	-.237	.813	-.229	.180
	GRD7	.191	.112	.106	1.707	.089	-.029	.412
	WHITE11	.438	.363	.069	1.207	.228	-.276	1.153
	LSES11	.040	.089	.027	.452	.652	-.136	.216
	SPED11	-.236	.150	-.093	-1.570	.118	-.532	.060
	SEX11	-.031	.092	-.021	-.333	.739	-.212	.151
2	(Constant)	.207	.164		1.258	.210	-.117	.531
	PLM11	-.218	.053	-.302	-4.078	.000	-.323	-.113
	PLR11	.070	.046	.113	1.525	.128	-.020	.161
	GRD6	-.025	.104	-.016	-.240	.810	-.229	.179
	GRD7	.192	.112	.107	1.711	.088	-.029	.413
	WHITE11	.459	.364	.072	1.261	.208	-.257	1.175
	LSES11	.042	.089	.028	.472	.637	-.134	.218
	SPED11	-.239	.151	-.094	-1.586	.114	-.535	.058
	SEX11	-.024	.092	-.016	-.264	.792	-.206	.157
	treat	-.078	.085	-.053	-.926	.355	-.245	.088

a. Dependent Variable: MATHGAIN

GLOBAL21

Students deserve it • The world demands it

West Virginia Department of EDUCATION

Jorea M. Marple, Ed.D.

State Superintendent of Schools