Public Affairs
Research Council
of Alabama

High-Dose Tutoring in Birmingham City Schools

Academic Outcome Evaluation

Joe Adams
Research Coordinator

Rachel D. Cochran
Senior Research Associate

High-Dose Tutoring in Birmingham City Schools

Published by the Public Affairs Research Council of Alabama
Birmingham, Alabama
www.parcalabama.org

Contents

Introduction

High-dose tutoring is recognized as one of the most effective academic interventions for students in public schools. Through a grant managed by the United Way of Central Alabama, local colleges were engaged in identifying students to serve as tutors in Birmingham City Schools (BCS). The United Way and BCS collaborated with the Public Affairs Research Council of Alabama (PARCA) to evaluate the impact of the tutoring program on students' academic achievement. The program began in January of 2022 and was implemented for the first full academic year in 2022-23.

Tutors and Student Participants

Tutoring services were provided by students from a group of colleges in three geographic regions in the Birmingham area. Region West included tutors from Miles College and Lawson State Community College. Region Central included UAB, Samford University, and Birmingham-Southern College student tutors. Region East included student tutors from Jefferson State Community College. In total, 66 tutors provided 705.5 hours for 689 students in 44 BCS schools.

Although there is no widely accepted single definition of the term "high dose tutoring," there is agreement that for tutoring to be considered high dosage, the tutor-student ratio should be very low (preferably 1:1) and there should be multiple sessions per week over the course of the academic year, resulting in between 50 and 100 hours of tutoring exposure per student. No information was provided regarding the actual number of students in a tutoring cluster. If we assume that the tutor-student ratio was 1:1, on average, students received about one hour of tutoring (mean = 1.02 hours per student). If the tutor-student ratio was 1:3 or higher, then the average per-student time tutored would increase accordingly. Nevertheless, these averages do not meet the threshold for what is considered high-dosage tutoring, according to the literature.

Table 1. Tutoring Hours by Region and School

Tutoring Hours Provided in Each Area by Respective Colleges						
School	**Hours**	**School**	**Hours**	**School**	**Hours**	**Total**
Colleges: Miles, Lawson		**Colleges: UAB, Samford, BSC**		**Colleges: Jeff State**		**Hours**
Wylam (K-5)	27	Washington (K-8)	35	Huffman Academy (K-5)	24	**86.0**
Arrington (K-5)	26	Norwood (K-5)	31	Dupuy (6-12)	20	**77.0**
West End (K-5)	21	Glen Iris (K-5)	29	Ossie Ware Mitchell (6-8)	20	**70.0**
Jones Valley (6-8)	20	Inglenook (K-8)	24	Robinson (K-5)	20	**64.0**
South Hampton (K-8)	20	Ramsay (9-12)	18	Sun Valley (K-5)	20	**58.0**
Wenonah (9-12)	18	Wilkerson (6-8)	16.5	Huffman High (9-12)	19	**53.5**
Bush Hills STEAM (6-8)	14	Phillips (K-8)	16	Woodlawn (9-12)	19	**49.0**
Green Acres (6-8)	13	Hudson (K-8)	15	Hayes (K-8)	17	**45.0**
Oxmoor Valley (K-5)	12	Virtual	14	Oliver (K-5)	16	**42.0**
Brown (K-5)	12	Parker (9-12)	13	Putnam (6-8)	14	**39.0**
Jackson-Olin (9-12)	12	EPIC (K-5)	13	W.J. Christian (K-8)	11	**36.0**
Central Park (K-5)	11	Tuggle (K-5)	12	Martha Gaskins (K-5)	11	**34.0**
Minor (K-5)	5	Carver (9-12)	9	Smith (6-8)	9	**23.0**
		Hemphill (K-5)	8	Avondale (K-5)	8	**16.0**
		Princeton (K-5)	0	Barrett (K-5)	7	**7.0**
				Huffman Middle (6-8)	6	**6.0**
Western Area Schools	**211**	**Central Area Schools**	**253.5**	**Eastern Area Schools**	**241**	**705.5**

There were substantial inconsistencies between the data provided by the tutors and the data provided by the BCS schools. For example, BCS rosters indicated 12 schools where no students were tutored. However, those schools were represented in the tutoring dataset as having received tutoring hours.

Table 2. Sample Roster Inconsistencies

BCS School	N Students Tutored (BCS Student Roster)	N Hours Tutoring (United Way Tutor Roster)
Brown	0	12
Carver	0	9
Central Park	0	11
Green Acres	0	13
Huffman Academy	0	0
Martha Gaskins	0	0
Ossie Ware	0	20
Parker	0	13
Washington	0	35
Smith	0	9
Wenonah	0	18
Woodlawn	0	19

Other inconsistencies prevented matching all students to their respective tutors and their individual exposure to tutoring services (dosage). The student rosters provided by BCS included names of students

who received tutoring by grade level and school. In many cases, there was also an indication of the content area in which the students were tutored (math, ELA, reading, etc.), although in some cases, no subject matter was specified as can be seen in Table 3.

Table 3. Number of Students Tutored by Content Area

Number of Students Tutored per Rosters

Subject Area

Grade	Both	ELA	Math	NA	Grand Total
1		4			**4**
2		24	15		**35**
3	10	61	30		**101**
4	12	33	90	15	**150**
5	3	11	102	14	**130**
6		6	34		**40**
7		9	24		**33**
8		3	5		**8**
9	6	7			**13**
10	4	19	33		**56**
11	8	10			**18**
12		4			**4**
Not Listed		1		97	**98**
Grand Total	**43**	**192**	**333**	**126**	**689**

Over half of the students (55%) who participated in tutoring were in Grades 3-5, and tutoring in mathematics was more common than tutoring in ELA, according to the records provided. A total of 98 (14%) of the tutored students had no grade level indication, and 18% of the students (n=126) had no content area listed in the BCS roster. There was considerable overlap between the two, meaning the absence of a grade designation frequently coincided with the absence of a content area designation.

Matching Procedures

Although 689 students were identified by BCS as having participated in the HDT program, not all students in the roster were able to be matched to students with SSIDs in the iReady and ACAP datasets. Student identifiers were not present in the BCS student rosters, which posed a challenge in matching student roster names to SSIDs. PARCA conducted multiple series of automated matching attempts based on student names. These included various combinations of first, last, and conjoined dual first and last names, resulting in 431 matches in the first stage, and another 89 in the second stage. A total of 169

residual, unmatched cases were then searched individually and manually in a third matching stage. In that process, it was discovered that surnames were sometimes truncated or partially omitted, first names were sometimes abbreviated, nicknames were used in the rosters that weren't present in the iReady/ACAP data, and/or diacritical marks found in the iReady/ACAP database were not present in the rosters. Quite often, the school listed in the iReady/ACAP data did not match the school listed in the tutoring data (most often occurring with Dupuy Alternative School).

As a result of the three stages of matching procedures, a total of 552 students of the original 689 (80%) were matched to corresponding SSIDs. Of the 552 tutored students with SSIDs, differing numbers of students had iReady and/or ACAP scores available for both the 2021-22 academic year (baseline) and the 2022-23 academic year (HDT). Data for both years was necessary to calculate the gain/loss differences between tutored students and non-tutored students. The resulting datasets for both ACAP and iReady were dramatically smaller than the original number of students in the tutoring sample.

Methodology

For tutored students with iReady or ACAP scores in both 2022 (baseline-pre) and 2023 (HDT-post), gains/losses in scaled scores in ELA and Mathematics, were examined. The gains/losses in the tutored students' pre-post scaled scores were compared to gains/losses in non-tutored students' scores. A series of t-tests and ANOVAs were conducted to assess the statistical significance of differences between the performance of the tutored group and the non-tutored group. Measures were taken to control for baseline scores for each subject in 2021-2022 (pre-test year), grade level, and gender, which was observed to be an important factor in some cases.

Specifically for the iReady assessments, sub-analyses were conducted based on the tutoring content area. Tutoring treatments were categorized by PARCA as belonging to one of four groups: 1) math, including geometry, algebra, and other areas of mathematical focus; 2) ELA, including reading and writing; 3) both ELA and math; and 4) others that did not record the subject matter (NA). Various combinations of these categories were examined using regression analysis to further assess the impact of tutoring on student iReady performance. Six multiple regression sub-analyses were conducted using two streams of three combinations of tutoring designations for each subject:

<u>ELA Assessment Regression Analyses</u>

- *ELA* only, both *ELA and math*, and *NA*

- *ELA* only, and both *ELA and math*
- *ELA* only

Mathematics Assessment Regression Analyses

- *Math* only, both *ELA and math*, and *NA*
- *Math* only, both *ELA and math*
- *Math* only

Specifically for the ACAP assessments, tutored students were matched to a sample of non-tutored students based on equivalence of school, grade, baseline (2022) ACAP score, and gender. Grade-level sub-analyses (paired samples t-tests) were conducted when sample size allowed. All analyses for ELA included ELA only, ELA and math, and the no designation (NA) students. All analyses for Mathematics included math only, ELA and math, and the no designation (NA) students.

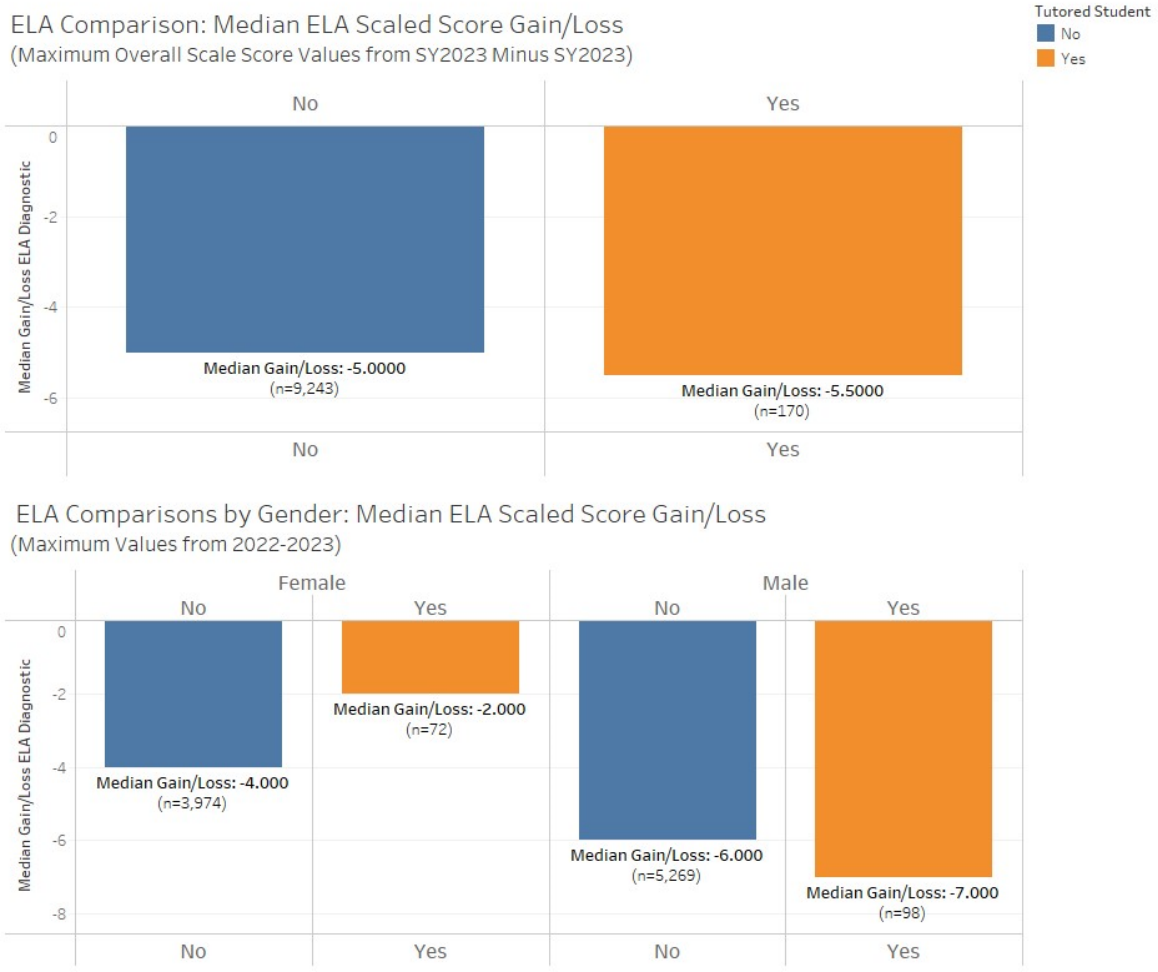
iReady Findings

In general, analyses of the 2022 to 2023 gains and losses on the iReady assessment did not yield statistically significant ($p<.05$) differences in performance between students who received tutoring and those who did not. There was one exception in a sub-analysis in mathematics, which will be described in the mathematics section of the report. Tableau summary dashboards for both ELA and Mathematics can be found at https://tinyurl.com/HDTOutcomes

iReady ELA Analysis

A total of 170 students' iReady ELA scores were included in the analysis. Overall, students had year-over-year declines. Both tutored and non-tutored students lost ground from 2022 to 2023. In fact, tutored students showed a slightly greater median loss (-5.5 points) than non-tutored students (-5.0 points). This difference was not statistically significant. In fact, it was a nearly identical decrease. Detailed t-test and ANOVA results can be found in Appendix I.

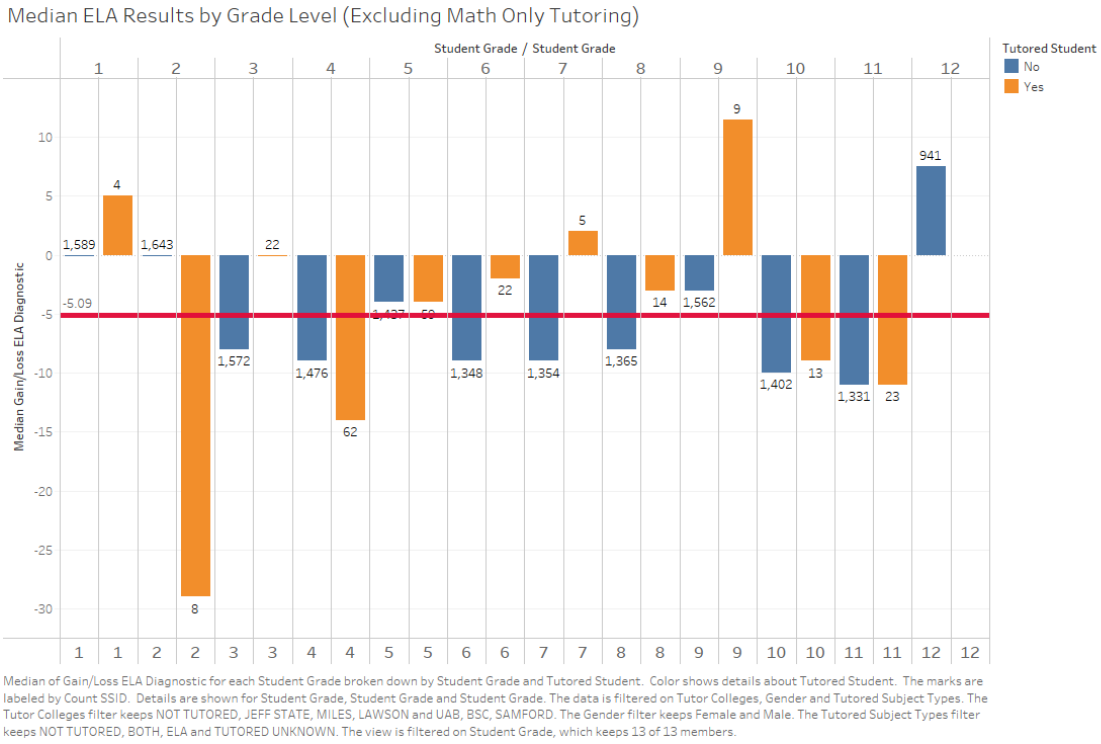
Figure 1. iReady ELA Gain-Loss in Scaled Scores



There were some differences in scaled score losses based on gender. Girls in the tutored group showed less of a decline than girls in the non-tutored group. The pattern was the opposite for boys. Tutored boys showed more of a decline from 2022 to 2023 than non-tutored boys. This finding is worth deeper exploration in subsequent years when more robust datasets are available.

A closer look at gains and losses by grade levels illustrates the variance in scores. There was one grade level (12^{th}) for which there were no ELA scores at both timepoints, despite the fact that were 80 tutored students with iReady mathematics scores in 12^{th} grade. The gains/losses across grade levels for which there were tutored and non-tutored students in the dataset showed extreme highs and lows

Figure 2. iReady ELA Gain-Loss in Scaled Scores by Grade Level



Very low performance in 2nd grade, 4th grade, and 11th grade appears to account for some of the differences between the overall performance of tutored and non-tutored students on the ELA iReady assessment. Most of the tutored students fell in the 3rd to 8th grade range, and in some of those grade levels, difference scores for tutored students were better than for non-tutored students.

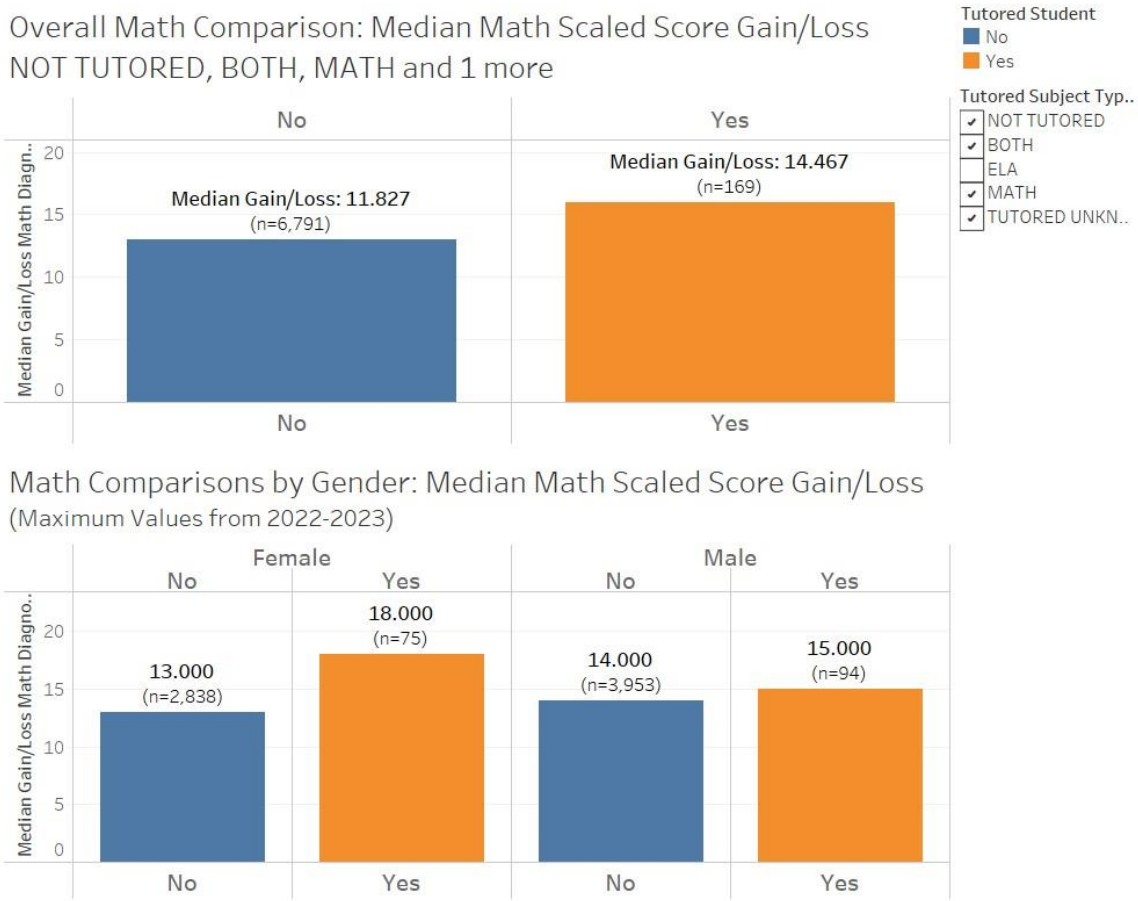
Although the medians were noticeably different at certain grade levels between tutored and non-tutored students, it is challenging to draw any conclusions about why given the absence of a consistent pattern of performance. PARCA does recommend following up on the 9th grade students (n=9) and 7th grade students (n=5), as their positive growth from 2022 to 2023 stood out. Grade 6 may also be worthy of further exploration to understand if there were particular schools where a critical mass of students received tutoring and whether the hours of tutoring or experience of the tutors might have played a role.

A total of 3 regression sub-analyses were conducted in ELA based on combinations of categories of tutoring by content area. The first combined students tutored in ELA only, both ELA and mathematics, and students with no content area designation. The results were not statistically significant. The second regression analysis included students tutored in ELA only and both ELA and math. The results were not statistically significant. The third regression included those students who were designated as tutored in ELA only. Those results were also not statistically significant. Details of the regression analyses can be found in Appendix I.

iReady Mathematics Analysis

Shifts from 2022 to 2023 in Mathematics were much better than in ELA. Both tutored and non-tutored students improved from 2022 to 2023. Tutored students showed a slightly greater average gain (14.5) points) than non-tutored students (11.8 points), although this difference was not statistically significant ($p<.05$). Detailed t-test and ANOVA results can be found in Appendix I.

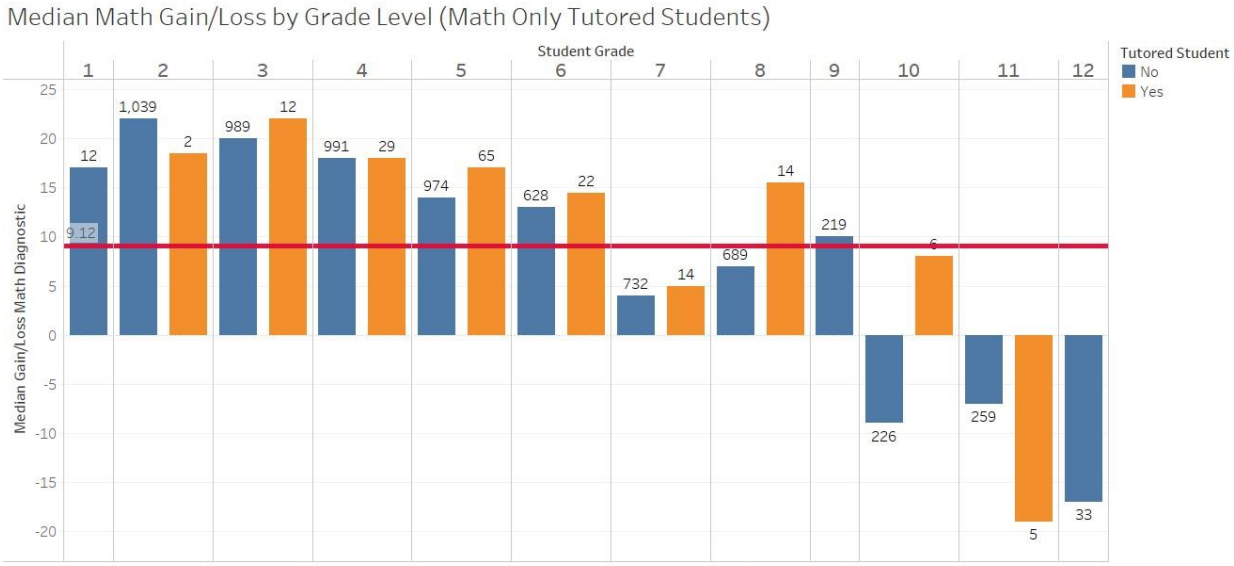
Figure 3. iReady Mathematics Gain-Loss in Scaled Scores



Similar in pattern to the ELA assessment data, tutored girls accounted for much of the gains in iReady mathematics scores. Tutored girls outperformed non-tutored girls by 5 points. Tutored boys only slightly outperformed non-tutored boys.

Similar to the results for ELA, results in Mathematics depicted by grade level revealed significant variance in student performance.

Figure 4. iReady Mathematics Gain-Loss in Scaled Scores by Grade Level



Tutored students in Grade 3 (n=12), Grade 5 (n=65), Grade 8 (n=14), and Grade 10 (n=6) showed substantially better growth in scores than non-tutored students. These 157 students represented over half (57%) of the sample of tutored students with mathematics (n=169). Medians were similar in the tutored and non-tutored groups at all grade levels except two. Medians were substantially higher in the tutored group in Grades 8 and 10. It may be beneficial to conduct some grade-level analyses in mathematics to glean a clearer picture of student performance.

A total of 3 regression sub-analyses were conducted in Mathematics based on combinations of categories of tutoring by content area. The first combined students tutored in mathematics only, both ELA and mathematics, and students with no content area designation. The results were not statistically significant. The second regression analysis included students tutored in math only and both ELA and

math. The results were not statistically significant. The third regression included those students who were designated as tutored in math only. Those results *were* statistically significant ($p<.05$).

Tutored students for whom *math alone* was the designated tutoring content area showed statistically better 2022-23 gains on the iReady math than non-tutored students. Regression analysis indicated a 4-point difference between groups that was statistically significant. These benefits were concentrated in the lower grades, as students in higher grades tended to have decreases in math scaled scores. This margin was not significant with respect to changes in percentiles, year over year. Nevertheless, it is worth attention in future tutoring years as it may indicate an opportunity for positively impacting younger students' mathematics scores with math-focused tutoring. Details of the regression analyses can be found in Appendix I.

ACAP Findings

The Alabama Comprehensive Assessment Program (ACAP) is the adopted high-stakes standardized assessment in the state. The test contains an ELA section (a portion of which is used to assess Reading proficiency) and a Mathematics section. It is administered to students in Grades 2 through 8 in the spring of each academic year. Because the nature and purpose of the ACAP are different from that of the iReady, PARCA used slightly different matching procedures, analyses, and illustrations of results.

Rather than comparing tutored students to *all* non-tutored students in the BCS system, PARCA selected a purposive matched comparison group of non-tutored students using a pre-determined set of criteria. This ensured that samples were equivalent in size and additional characteristics. A sample of non-tutored students was selected to ensure:

- Tutored and non-tutored students were in the same BCS school
- Tutored and non-tutored students were in the same grade level
- Tutored and non-tutored students had 2022 (baseline) ACAP scaled scores within 25 points of each other
- Tutored and non-tutored students were the same gender

From this group of potential matches, a random selection of "eligible" non-tutored comparison students was chosen. It is important to note that there were instances where there was only one potential match for a tutored student using the criteria above. There were also instances where tutored students could not be matched according to all the criteria above, which resulted in close-but-not-exactly-the-same sample sizes for tutored and non-tutored students, as can be seen in the table below.

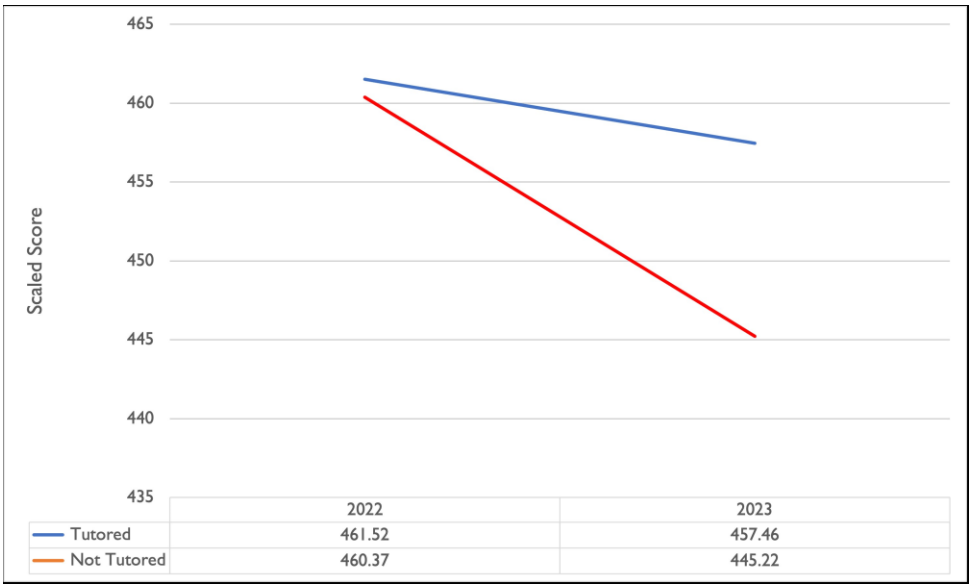
Table 4. ACAP Sample

ACAP Assessment	N Tutored Students	N Non-Tutored Matched Students
ELA	144	126
Mathematics	228	217

ACAP ELA Analysis

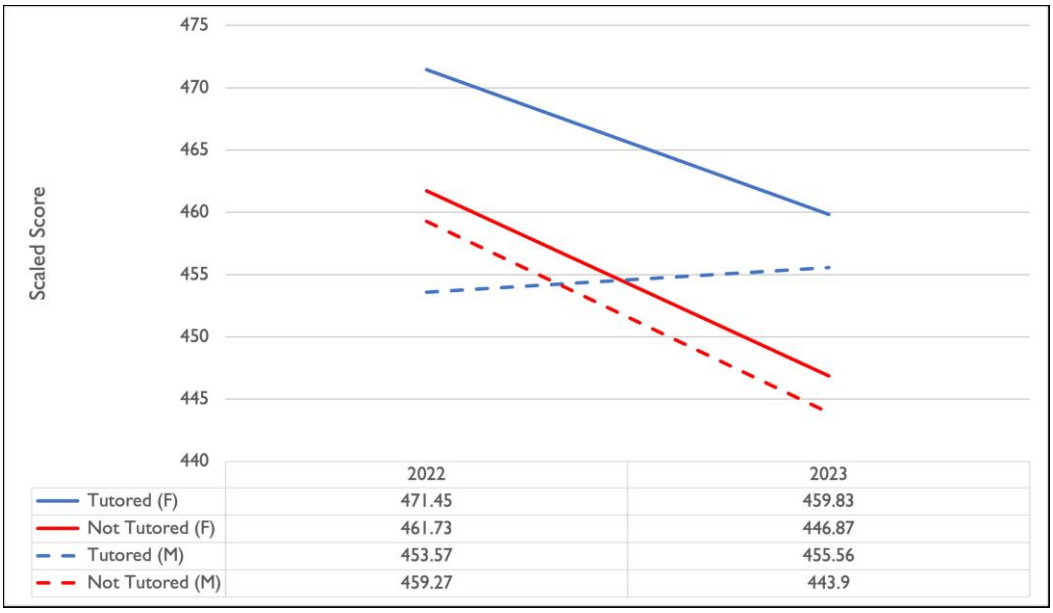
Tutored students included in the analysis were designated in the BCS roster as having been tutored in ELA, ELA and Mathematics, or NA. All math-only tutored students were excluded. Overall, ACAP scores in ELA declined for both tutored and non-tutored students from 2022 to 2023. Declines were smaller for tutored students—a mean of about -4 points for tutored students compared to -15 for non-tutored students. The differences were not statistically significant but are noteworthy, especially given that non-tutored students had very similar scores to tutored students at baseline. It is important to mention that the ELA portion of the ACAP assessment was modified for the 2023 administration to better reflect the 2021 Alabama Course of Study. It has been advised by the ALSDE that no 2022-23 comparisons be made. However, because the declines were consistent with declines in the iReady data, PARCA has presented them here but cautions against drawing definitive conclusions.

Figure 5. ACAP ELA Overall Scaled Scores



As with the iReady, there were differences in performance based on gender. However, on the ACAP, the pattern was reversed. Tutored boys showed slight gains from 2022 to 2023 (about +2 points), outperforming matched non-tutored boys who had about a 15-point decline. Tutored girls had baseline scores that were about 10 points higher than non-tutored girls and both groups declined at similar rates, although tutored girls had slightly less of a decrease than non-tutored girls.

Figure 6. ACAP ELA Scaled Scores by Gender



A series of grade-level paired t-tests were also conducted when the sample size was appropriate. The table below shows the breakdown of mean gains/losses by grade level. The grade level indicated is the grade level for the 2023 administration.

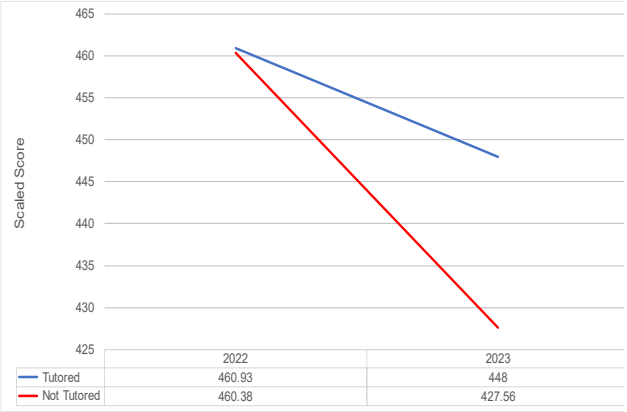
Table 6. ACAP ELA Gains/Losses by Grade Level

Grade Level	N Tutored	Mean Gain/Loss	N Not Tutored	Mean Gain/Loss
3rd	54	-12.93	45	-32.82
4th	51	+2.88	43	-7.30
5th	18	-5.78	18	-7.89
6th	5	-24.8	5	-41.8
7th	11	+27.64	11	+23.46
8th	5	-22.00	4	-6.40

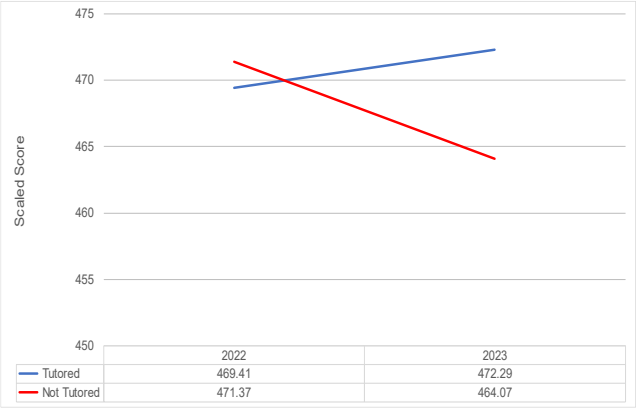
ACAP ELA scores for both tutored and non-tutored students showed substantial variance. Mean gains were as high as 27.64 points for tutored students in Grade 7 to -32.82 points for non-tutored students in Grade 3. With the exception Grade 8, tutored students outperformed non-tutored students, but because of the extreme variability across the grades and small sample sizes at specific grade levels, it is challenging to glean a true understanding of student performance in general, much less as it relates to tutoring. Where sample sizes were sufficient for further analysis (Grade 3 and Grade 4), PARCA conducted repeated measures ANOVAs to test the magnitude and significance of differences.

Figure 7. ACAP ELA Sample Grade Level Gains-Losses

Grade 3 ELA



Grade 4 ELA

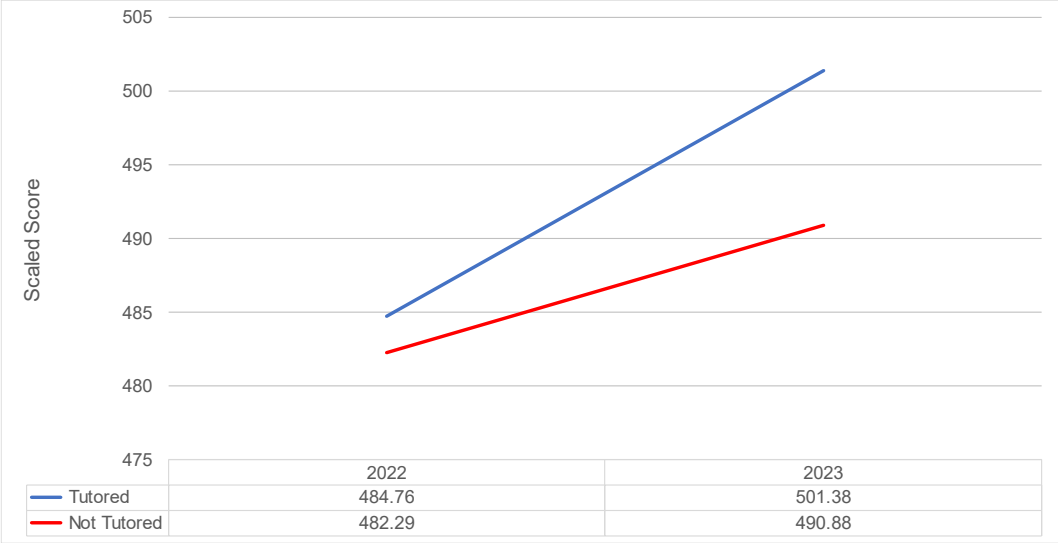


Although neither analysis yielded statistically significant ($p<.05$) results according to paired samples t-tests, there does appear to be an emerging pattern of better performance in those students who were tutored when controlling for 2022 baseline ELA score. The differences are not dramatic, but the consistency across grade levels indicates potential. Having larger sample sizes of students in other grade levels with scores at baseline and follow-up will allow for more rigorous testing of results.

ACAP Mathematics Analysis

Tutored students included in the analysis were designated in the BCS roster as having been tutored in Mathematics, ELA and Mathematics, or NA. All ELA-only tutored students were excluded. As far as PARCA understands it, the Mathematics portion of the ACAP was not changed in 2023, only the ELA portion. Therefore, year-to-year comparisons should be stable and valid. Overall, ACAP scores in Mathematics improved for both tutored and non-tutored students.

Figure 8. ACAP Mathematics Overall Scaled Score



The 228 tutored students showed a mean growth of close to 17 points. The 217 non-tutored students showed a mean growth of close to 9 points. Although tutored students outperformed non-tutored students, the difference in gains between the two groups was not statistically significant. There were some interesting gender differences in the baseline and post-test scores, although growth from 2022 to 2023 was similar for girls and boys in both groups.

Figure 9. ACAP Mathematics Scaled Scores by Gender

Girls in both the tutored and non-tutored groups started with a lower baseline score (2022 ACAP) than tutored and non-tutored boys. Tutored girls' mean growth in 2023 was close to 15 points compared to

almost 9 for non-tutored girls. Tutored boys' mean growth was 18 points, which very slightly widened the gap between tutored girls and boys. Non-tutored boys' mean growth was close to 9 points, which was the same as growth for non-tutored girls. These gender differences by tutored and non-tutored groups were not statistically significant.

A series of grade-level paired samples t-tests were also conducted. There were sufficient sample sizes at all grades in mathematics with the exception of Grade 8, for which PARCA had scores for only 5 tutored students matched to 6 non-tutored students. The table below shows the breakdown of mean gain/losses by grade level. The grade level indicated is the grade level for the 2023 administration.

Table 7. ACAP Mathematics Gains/Losses by Grade Level

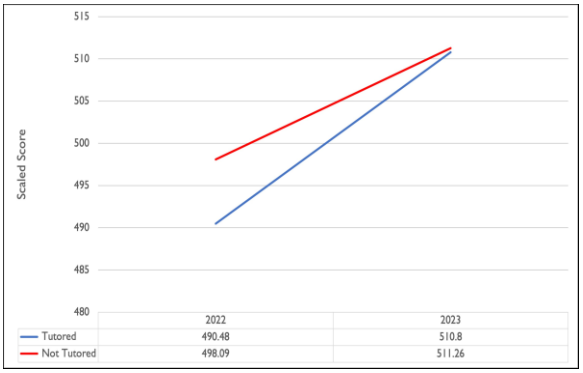
Grade Level	N Tutored	Mean Gain/Loss	N Not Tutored	Mean Gain/Loss
3rd	40	+20.32	35	+13.17
4th	88	+4.06	84	+5.84
5th	42	+24.69	41	+9.8
6th	31	+19.71	28	+1.78
7th	22	+34.41	23	+18.17
8th	5	+42.80	6	+6.84

There was one grade level (4th grade) for which growth in scores was slightly better for non-tutored students than for tutored students. Non-tutored students outgained tutored students by 1.78 points in 4th grade. The difference was not statistically significant, but it is important because such a large portion of the ACAP mathematics scores (39%) came from students in 4th grade 2023.

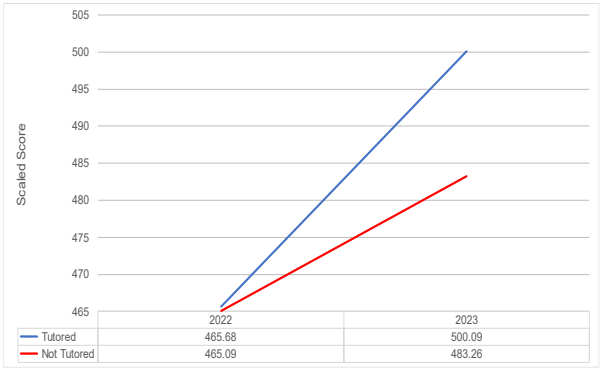
For all other grade levels, tutored students showed more growth on the Mathematics ACAP from 2022 to 2023 than did non-tutored students. In some instances, the results were notable, although not statistically significant ($p<.05$).

Figure 10. ACAP Mathematics Sample Grade Level Gains-Losses

Grade 3



Grade 7



In Grade 3, tutored students started with baseline (2022) mean scores that were about 8 points lower than non-tutored students' scores. However, the 20-point gain for tutored students by 2023 compared to the 13-point gain for non-tutored students closed the gap between them. In middle grades, students in both groups had similar mean baseline scores, but tutored students showed more growth than non-tutored students. In 7th grade, the mean growth for the n=22 tutored students was 34.41 points, whereas the mean growth for non-tutored students was only 18.17 points.

Conclusion and Recommendations

High-dose tutoring has been shown in the literature to have the potential to dramatically accelerate learning outcomes for students. According to Ed Week, "The research on high-dosage tutoring—generally defined as one-on-one tutoring or tutoring in very small groups at least three times a week, or for about 50 hours over a semester—is robust, and it is convincing. On average, the effect sizes are among the largest of all interventions seen in education."

There is not a clear consensus on the dosage needed to catalyze academic growth, but most studies point to the need for a significant number of hours before measurable effects are achieved. There is also not consistent guidance in the literature with regard to procedures. However, most research is in agreement about some key factors in successful implementation. Researchers at Duke University have summarized those critical components in their operationalization of high-dosage tutoring. High-dose tutoring should:

- Include at least three tutoring sessions per week with the student,
- Involve sessions lasting at least 30 minutes,
- Be capped at a maximum student-to-tutor ratio of 3:1,
- Build relationships between tutor and student that is “sustained and strong,”
- Integrate monitoring of the student’s “knowledge and skills,” and
- Provide coaching or oversight of tutors to ensure “high-quality interactions.”

The BCS high-dose tutoring program did not demonstrate having provided tutoring services at the level described. Despite the challenges at the program implementation level, there appears to be an emerging pattern, especially in the ACAP data, where tutored students show more growth (or less of a decline) than non-tutored students. Results were not statistically significant, but the consistency of the pattern across grade levels indicates the potential for the program to make modifications to processes and procedures that could yield better results moving forward. In addition, issues with matching and adequate sample sizes presented a problem in this evaluation. A large portion of the original roster of students who were tutored was not included in the iReady or ACAP analysis, some because of matching challenges and some because of missing scores. Small sample sizes could have contributed to a lack of statistical significance in some cases where differences seemed stark between groups.

Recommendations to bring the program into closer alignment with best practices and improve the data infrastructure that will allow for more precise tracking of progress include:

- Expand the program to include regional coordinators who will have considerable oversight of the tutoring process at each school and ensure records are thorough and accurate.
- Offer training and orientation for tutors and opportunities for structured tutor collaboration.
- Develop, maintain, and enforce clear expectations for tutors, and incorporate systems for ongoing productive feedback.
- Encourage tutors to complete a brief reflection sheet after each tutoring session to process thoughts about the successes and challenges of the session and plans for the next session.
- Consider targeting lower grades, especially in mathematics and reading.
- Consider allocating more resources to mathematics where focused tutoring seemed to have a positive effect.
- Work toward increasing tutoring dosages per student to meet the standards determined to be sufficient according to the research.
- Maintain more precise and accurate records for future evaluation and project management.
- Collect Student ID numbers for every student served.
- Record hours tutored for each student as well as for each tutor.

Potential implementation strategies based on published studies include:

- Selecting highly qualified tutors from the ranks of retired teachers or teachers not currently working full-time and desiring part-time work due to other commitments or circumstances.
- If necessary, rationing HDT to provide quality services for recipient students using a lottery or other mechanism for selection based on the most effective use among students in the lower grades.

References

https://www.edweek.org/leadership/high-dosage-tutoring-is-effective-but-expensive-ideas-for-making-it-work/2020/08

https://ssri.duke.edu/what-is-high-dosage-tutoring-and-why-does-the-states-investment-in-it-matter/

https://studentsupportaccelerator.org/sites/default/files/Higg_Impact_Tutoring_Definition.pdf

https://www.povertyactionlab.org/sites/default/files/publication/Evidence-Review_The-Transformative-Potential-of-Tutoring.pdf

Appendix I

iReady ANOVAs and Regression Analyses

Report for Linear Model ELA_ALL_TUTORED

Basic Summary

Call:

lm(formula = ELA.Gain.Loss ~ X2022_ELADiagnMax_Overall.Scale.Score + Sex + Student.Grade + Tutored.Student, data = the.data)

Residuals:

	Min	1Q	Median	3Q	Max
	-287.53	-17.07	1.66	20.16	185.96

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	12.22317	2.432925	5.0241	5.15e-07	***
X2022_ELADiagnMax_Overall.Scale.Score	-0.06015	0.005852	-10.2781	< 2.2e-16	***
Sexmale	-1.35762	0.703891	-1.9287	0.05379	.
Student.Grade	1.92417	0.186228	10.3324	< 2.2e-16	***
Tutored.StudentYes	0.91457	1.892457	0.4833	0.62892	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.971 on 9568 degrees of freedom

Multiple R-squared: 0.01286, Adjusted R-Squared: 0.01245

F-statistic: 31.17 on 4 and 9568 degrees of freedom (DF), p-value < 2.2e-16

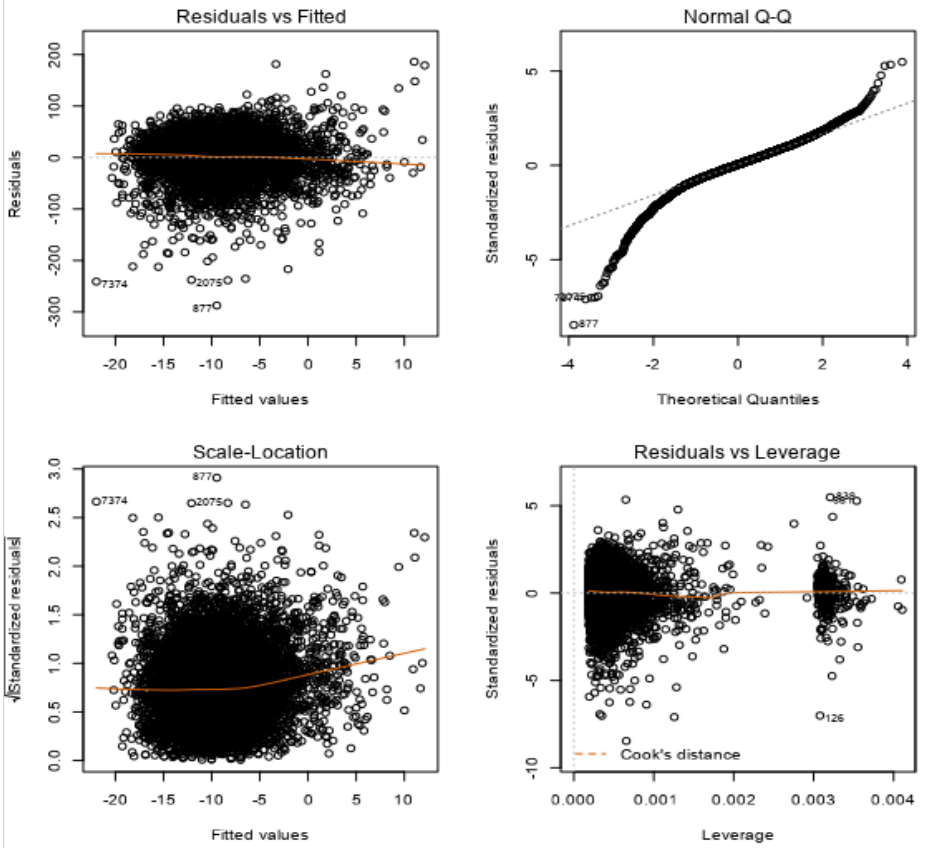
Type II ANOVA Analysis

Response: ELA.Gain.Loss

	Sum Sq	DF	F value	Pr(>F)	
X2022_ELADiagnMax_Overall.Scale.Score	121913.38	1	105.64	< 2.2e-16	***
Sex	4293.1	1	3.72	0.05379	.
Student.Grade	123204.11	1	106.76	< 2.2e-16	***
Tutored.Student	269.53	1	0.23	0.62892	
Residuals	11041989.74	9568			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Basic Diagnostic Plots



Report for Linear Model ELA_ALL_TUTORED

Basic Summary

Call:

lm(formula = ELA.Gain.Loss ~ X2022_ELADiagnMax_Overall.Scale.Score + Sex + Student.Grade + Tutored.Student, data = the.data)

Residuals:

	Min	1Q	Median	3Q	Max
	-287.69	-17.13	1.67	20.13	185.85

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	12.03116	2.449113	4.9125	9.14e-07	***
X2022_ELADiagnMax_Overall.Scale.Score	-0.06002	0.005889	-10.1912	< 2.2e-16	***
Sexmale	-1.30577	0.711555	-1.8351	0.06652	.
Student.Grade	1.94515	0.187485	10.3750	< 2.2e-16	***
Tutored.StudentYes	-1.89760	2.636580	-0.7197	0.47171	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34.032 on 9401 degrees of freedom

Multiple R-squared: 0.01316, Adjusted R-Squared: 0.01274

F-statistic: 31.33 on 4 and 9401 degrees of freedom (DF), p-value < 2.2e-16

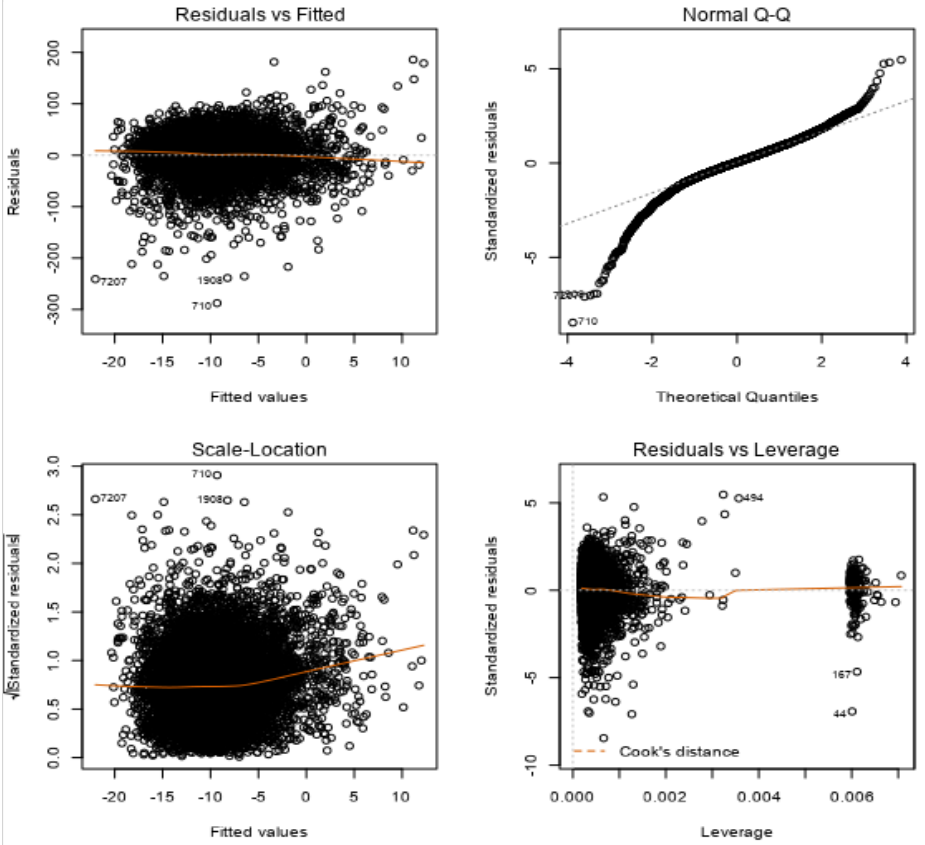
Type II ANOVA Analysis

Response: ELA.Gain.Loss

	Sum Sq	DF	F value	Pr(>F)	
X2022_ELADiagnMax_Overall.Scale.Score	120285.49	1	103.86	< 2.2e-16	***
Sex	3900.14	1	3.37	0.06652	.
Student.Grade	124663.21	1	107.64	< 2.2e-16	***
Tutored.Student	599.92	1	0.52	0.47171	
Residuals	10887774.33	9401			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Basic Diagnostic Plots



Report for Linear Model ELA_ALL_TUTORED

Basic Summary

Call:

lm(formula = ELA.Gain.Loss ~ X2022_ELADiagnMax_Overall.Scale.Score + Sex + Student.Grade + Tutored.Student, data = the.data)

Residuals:

	Min	1Q	Median	3Q	Max
	-287.70	-17.14	1.66	20.18	185.84

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	12.01584	2.452005	4.900	9.72e-07	***
X2022_ELADiagnMax_Overall.Scale.Score	-0.06002	0.005896	-10.180	< 2.2e-16	***
Sexmale	-1.27581	0.712870	-1.790	0.07354	.
Student.Grade	1.94489	0.187668	10.363	< 2.2e-16	***
Tutored.StudentYes	-3.50465	2.823132	-1.241	0.21449	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34.054 on 9379 degrees of freedom

Multiple R-squared: 0.01326, Adjusted R-Squared: 0.01284

F-statistic: 31.52 on 4 and 9379 degrees of freedom (DF), p-value < 2.2e-16

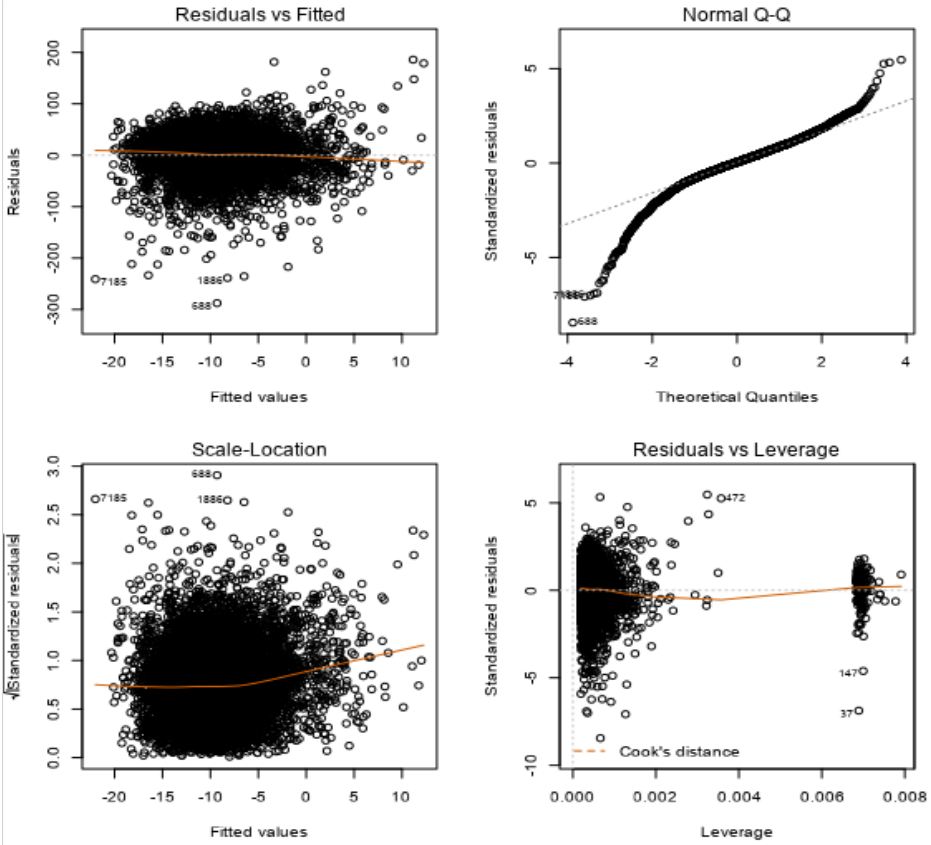
Type II ANOVA Analysis

Response: ELA.Gain.Loss

	Sum Sq	DF	F value	Pr(>F)	
X2022_ELADiagnMax_Overall.Scale.Score	120167.06	1	103.62	< 2.2e-16	***
Sex	3714.36	1	3.2	0.07354	.
Student.Grade	124549.75	1	107.4	< 2.2e-16	***
Tutored.Student	1787.14	1	1.54	0.21449	
Residuals	10876453.67	9379			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Basic Diagnostic Plots



Report for Linear Model MATH_ALL_TUTORED

Basic Summary

Call:

lm(formula = Math.Gain.Loss ~ X2022_ELADiagnMax_Overall.Scale.Score + Sex + Student.Grade + Tutored.Student + X2022_MathDiagnMax_Overall.Scale.Score, data = the.data)

Residuals:

	Min	1Q	Median	3Q	Max
	-128.051	-11.729	0.396	11.565	150.148

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	67.7610	2.69416	25.1511	< 2.2e-16	***
X2022_ELADiagnMax_Overall.Scale.Score	0.2386	0.00601	39.6982	< 2.2e-16	***
Sexmale	1.1801	0.53333	2.2127	0.02695	*
Student.Grade	-0.2504	0.16899	-1.4820	0.13838	
Tutored.StudentYes	0.6310	1.36206	0.4633	0.64318	
X2022_MathDiagnMax_Overall.Scale.Score	-0.4229	0.01077	-39.2519	< 2.2e-16	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.769 on 6939 degrees of freedom

Multiple R-squared: 0.2596, Adjusted R-Squared: 0.2591

F-statistic: 486.6 on 5 and 6939 degrees of freedom (DF), p-value < 2.2e-16

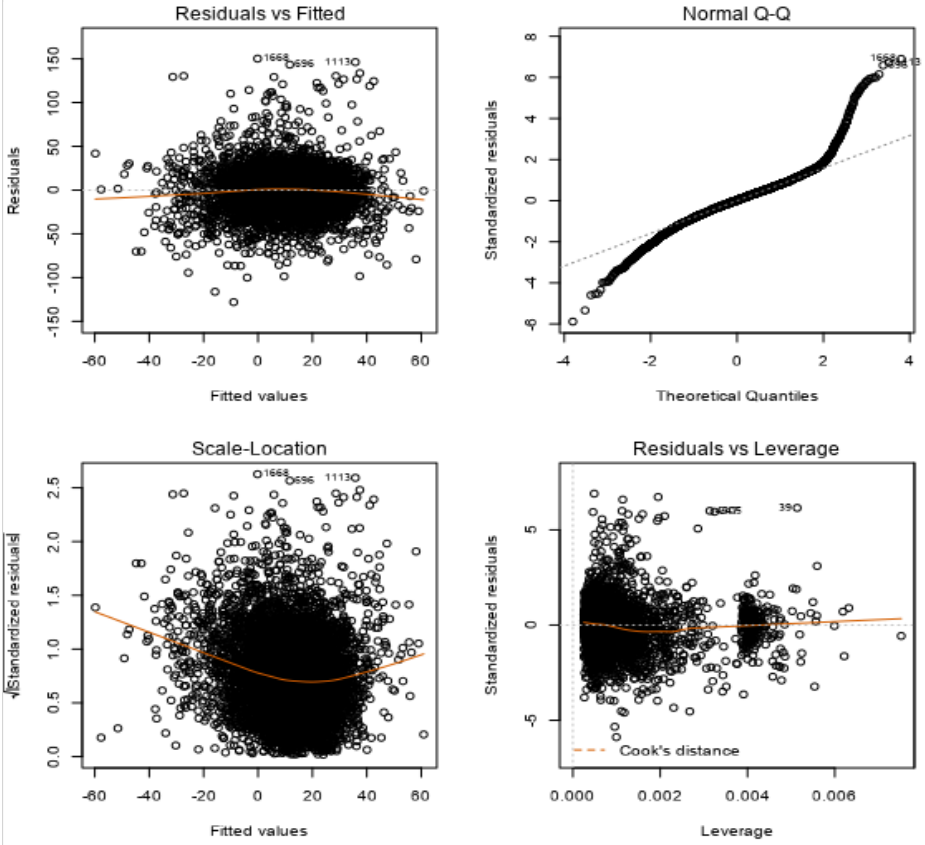
Type II ANOVA Analysis

Response: Math.Gain.Loss

	Sum Sq	DF	F value	Pr(>F)	
X2022_ELADiagnMax_Overall.Scale.Score	746792.86	1	1575.95	< 2.2e-16	***
Sex	2320.15	1	4.9	0.02695	*
Student.Grade	1040.82	1	2.2	0.13838	
Tutored.Student	101.71	1	0.21	0.64318	
X2022_MathDiagnMax_Overall.Scale.Score	730096.48	1	1540.71	< 2.2e-16	***
Residuals	3288182.13	6939			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Basic Diagnostic Plots



Report for Linear Model MATH_NOT_ELA

Basic Summary

Call:

lm(formula = Math.Gain.Loss ~ Sex + Student.Grade + Tutored.Student + X2022_MathDiagnMax_Overall.Scale.Score, data = the.data)

Residuals:

	Min	1Q	Median	3Q	Max
	-139.29	-11.91	1.12	13.30	172.83

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	67.4306	2.979702	22.6300	< 2.2e-16	***
Sexmale	-0.5658	0.587640	-0.9628	0.33567	
Student.Grade	-0.3701	0.184791	-2.0027	0.04525	*
Tutored.StudentYes	5.0855	1.879970	2.7051	0.00684	**
X2022_MathDiagnMax_Overall.Scale.Score	-0.1269	0.008616	-14.7228	< 2.2e-16	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.089 on 6955 degrees of freedom

Multiple R-squared: 0.09134, Adjusted R-Squared: 0.09081

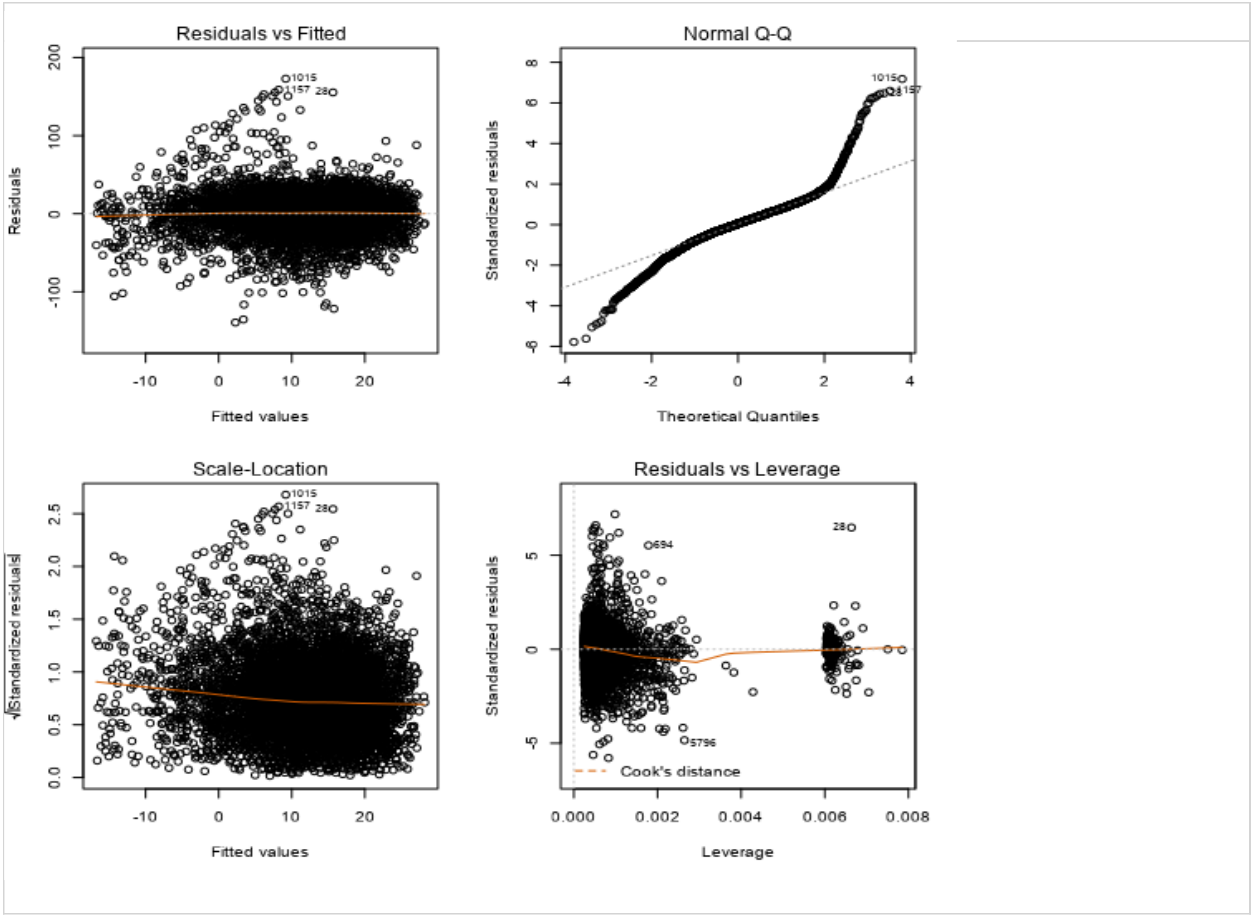
F-statistic: 174.8 on 4 and 6955 degrees of freedom (DF), p-value < 2.2e-16

Type II ANOVA Analysis

Response: Math.Gain.Loss

	Sum Sq	DF	F value	Pr(>F)	
Sex	537.95	1	0.93	0.33567	
Student.Grade	2327.39	1	4.01	0.04525	*
Tutored.Student	4246.36	1	7.32	0.00684	**
X2022_MathDiagnMax_Overall.Scale.Score	125785.65	1	216.76	< 2.2e-16	***
Residuals	4035949.7	6955			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residuals vs Fitted
Residuals
Fitted values
1015
157
280
Normal Q-Q
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
694
5796

Report for Linear Model MATH_MATH_ONLY

Basic Summary

Call:

lm(formula = Math.Gain.Loss ~ Sex + Student.Grade + Tutored.Student + X2022_MathDiagnMax_Overall.Scale.Score, data = the.data)

Residuals:

	Min	1Q	Median	3Q	Max
	-139.27	-11.96	1.13	13.29	172.83

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	67.4124	2.983204	22.5973	< 2.2e-16	***
Sexmale	-0.5366	0.588866	-0.9112	0.36223	
Student.Grade	-0.3670	0.185011	-1.9834	0.04736	*
Tutored.StudentYes	4.4040	1.974595	2.2303	0.02576	*
X2022_MathDiagnMax_Overall.Scale.Score	-0.1269	0.008626	-14.7100	< 2.2e-16	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.109 on 6939 degrees of freedom

Multiple R-squared: 0.09115, Adjusted R-Squared: 0.09063

F-statistic: 174 on 4 and 6939 degrees of freedom (DF), p-value < 2.2e-16

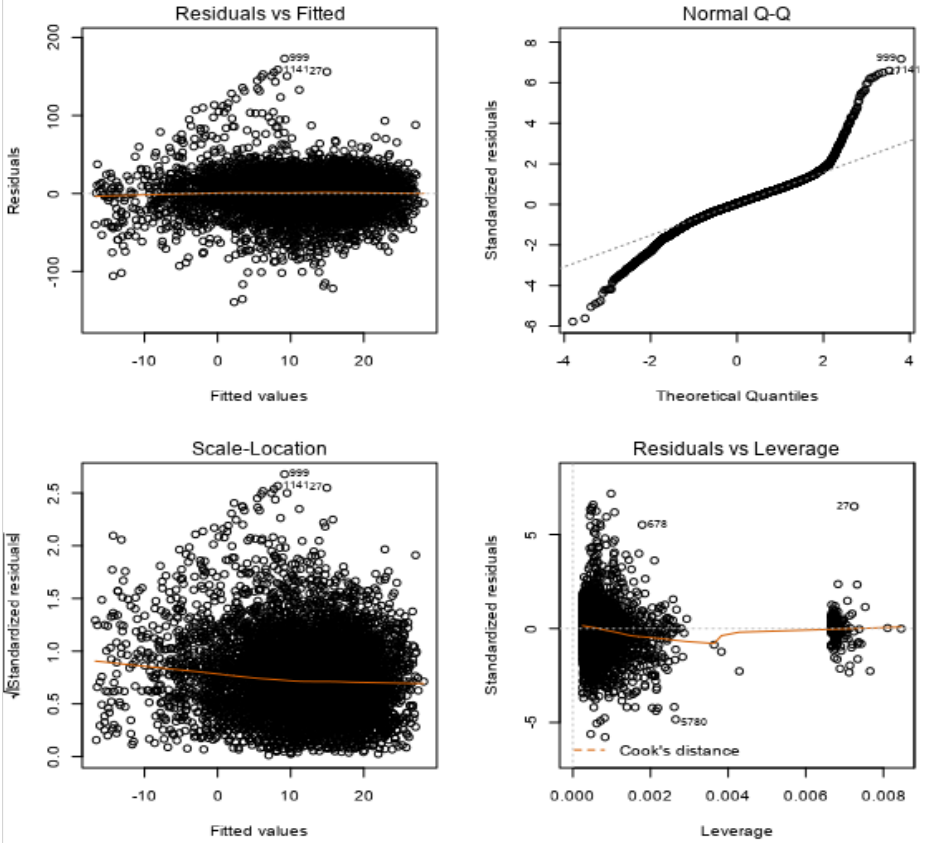
Type II ANOVA Analysis

Response: Math.Gain.Loss

	Sum Sq	DF	F value	Pr(>F)	
Sex	482.58	1	0.83	0.36223	
Student.Grade	2286.66	1	3.93	0.04736	*
Tutored.Student	2891.34	1	4.97	0.02576	*
X2022_MathDiagnMax_Overall.Scale.Score	125772.39	1	216.38	< 2.2e-16	***
Residuals	4033292.53	6939			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Basic Diagnostic Plots



Appendix II

ACAP ANOVAs and T-Tests

Overall Analysis of ELA (combining all grades):

Select Cases: HD_Tutor_Recode_ELA = 1 | HD_Tutor_Recode_ELA = 2

Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00 (Treatment)	461.52	48.395	144
	2.00 (Comparison)	460.37	45.919	126
	Total	460.98	47.172	270
ELA_Scale_Score_23	1.00 (Treatment)	457.46	54.211	144
	2.00 (Comparison)	445.22	64.897	126
	Total	451.75	59.639	270

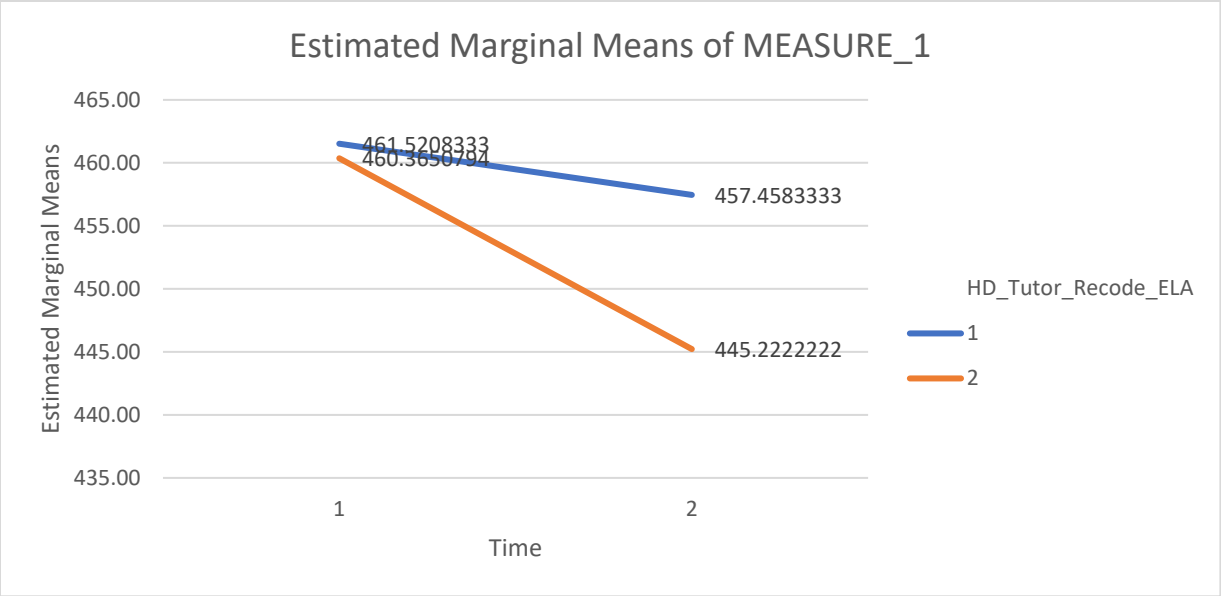
Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Time	Pillai's Trace	.040	11.136[b]	1.000	268.000	<.001
	Wilks' Lambda	.960	11.136[b]	1.000	268.000	<.001
	Hotelling's Trace	.042	11.136[b]	1.000	268.000	<.001
	Roy's Largest Root	.042	11.136[b]	1.000	268.000	<.001
Time * HD_Tutor_Recode_ELA	Pillai's Trace	.014	3.707[b]	1.000	268.000	.055
	Wilks' Lambda	.986	3.707[b]	1.000	268.000	.055
	Hotelling's Trace	.014	3.707[b]	1.000	268.000	.055
	Roy's Largest Root	.014	3.707[b]	1.000	268.000	.055

a. Design: Intercept + HD_Tutor_Recode_ELA

Within Subjects Design: Time

b. Exact statistic



Gender Female

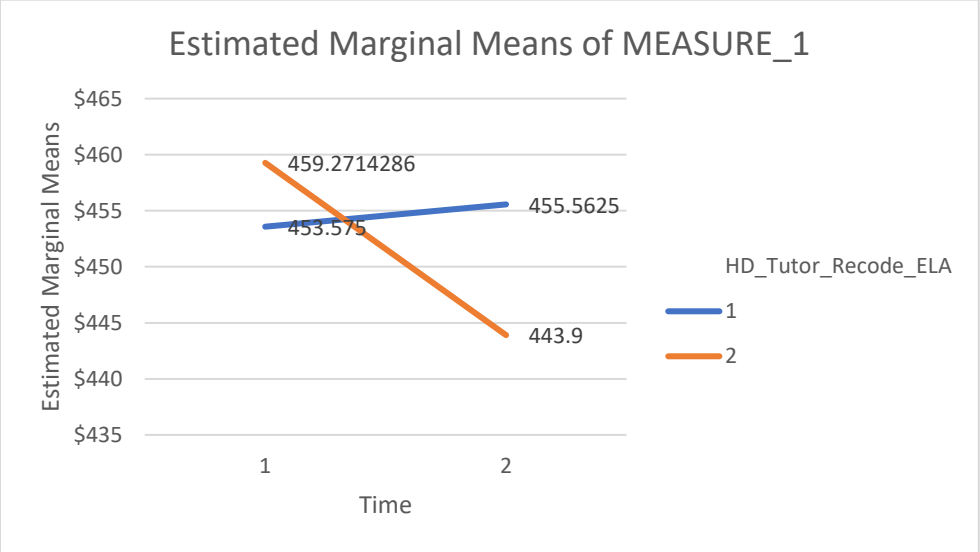
Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00 (Treatment)	471.45	42.069	64
	2.00 (Comparison)	461.73	40.337	56
	Total	466.92	41.385	120
ELA_Scale_Score_23	1.00	459.83	57.438	64
	2.00	446.87	61.032	56
	Total	453.78	59.248	120

Gender Male

Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00	453.57	51.810	80
	2.00	459.27	50.202	70
	Total	456.23	50.974	150
ELA_Scale_Score_23	1.00	455.56	51.773	80
	2.00	443.90	68.240	70
	Total	450.12	60.097	150



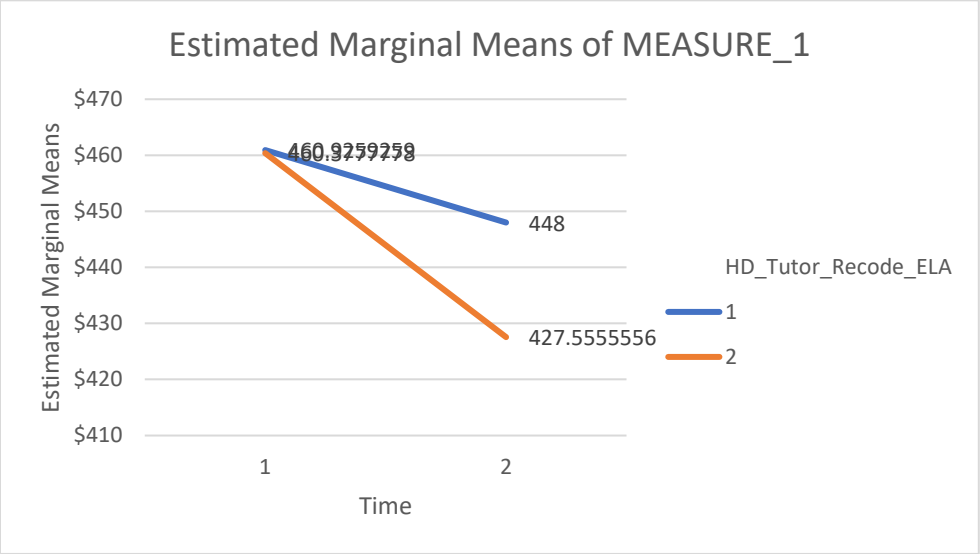
2nd to 3rd Grade

Not statistically significant

Descriptive Statistics

HD_Tutor_Recode_ELA	Mean	Std. Deviation	N

ELA_Scale_Score_22	1.00	460.93	49.670	54
	2.00	460.38	39.890	45
	Total	460.68	45.263	99
ELA_Scale_Score_23	1.00	448.00	57.069	54
	2.00	427.56	74.809	45
	Total	438.71	66.172	99



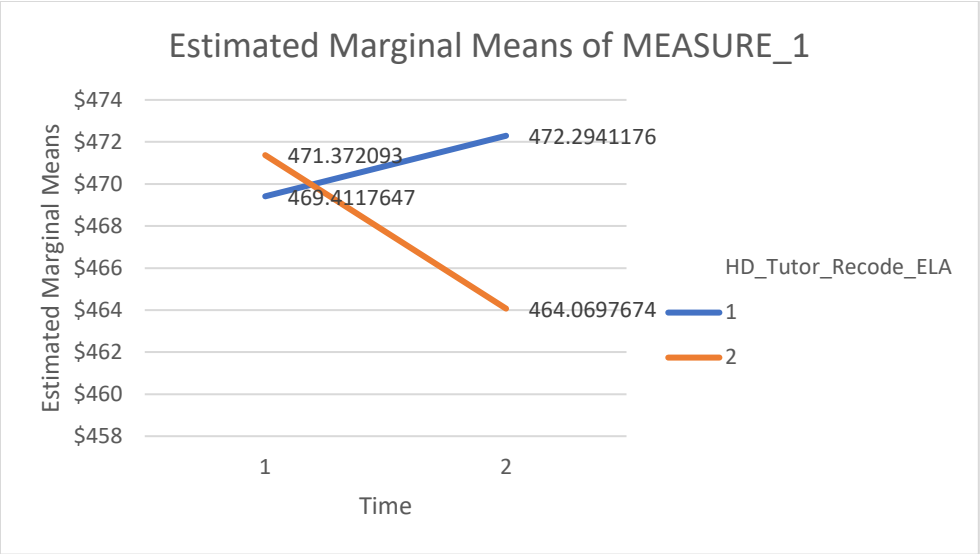
3rd to 4th Grade

Not statistically significant

Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00 (Treatment)	469.41	46.269	51
	2.00 (Comparison)	471.37	44.270	43
	Total	470.31	45.134	94
ELA_Scale_Score_23	1.00	472.29	43.981	51
	2.00	464.07	60.748	43

Total	468.53	52.187	94



4th to 5th Grade

** Small n's

Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00	465.06	45.743	18
	2.00	462.61	48.847	18
	Total	463.83	46.656	36
ELA_Scale_Score_23	1.00	459.28	49.906	18
	2.00	454.72	46.802	18
	Total	457.00	47.739	36

5th to 6th Grade

No comparison … n's are exceptionally small

Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00	436.40	28.632	5
	2.00	431.40	35.317	5
	Total	433.90	30.425	10
ELA_Scale_Score_23	1.00	411.60	30.566	5
	2.00	389.60	29.048	5
	Total	400.60	30.409	10

6th to 7th Grade

No comparison – small n's

Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00	443.00	59.768	11
	2.00	437.27	62.049	11
	Total	440.14	59.523	22
ELA_Scale_Score_23	1.00	470.64	57.181	11
	2.00	460.73	51.445	11
	Total	465.68	53.320	22

7th to 8th Grade

No comparison – small n's

Descriptive Statistics

	HD_Tutor_Recode_ELA	Mean	Std. Deviation	N
ELA_Scale_Score_22	1.00	440.60	50.023	5
	2.00	431.50	53.220	4
	Total	436.56	48.335	9
ELA_Scale_Score_23	1.00	418.60	94.944	5
	2.00	425.50	50.077	4
	Total	421.67	73.897	9

Overall Analysis of Math (combining all grades):

Select Cases: HD_Tutor_Recode_Math = 3 | HD_Tutor_Recode_Math = 4

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	484.76	43.838	228
	4.00	482.29	47.068	217
	Total	483.56	45.407	445
Math_Scale_Score_23	3.00	501.38	43.760	228
	4.00	490.88	60.404	217
	Total	496.26	52.741	445

Multivariate Tests[a]

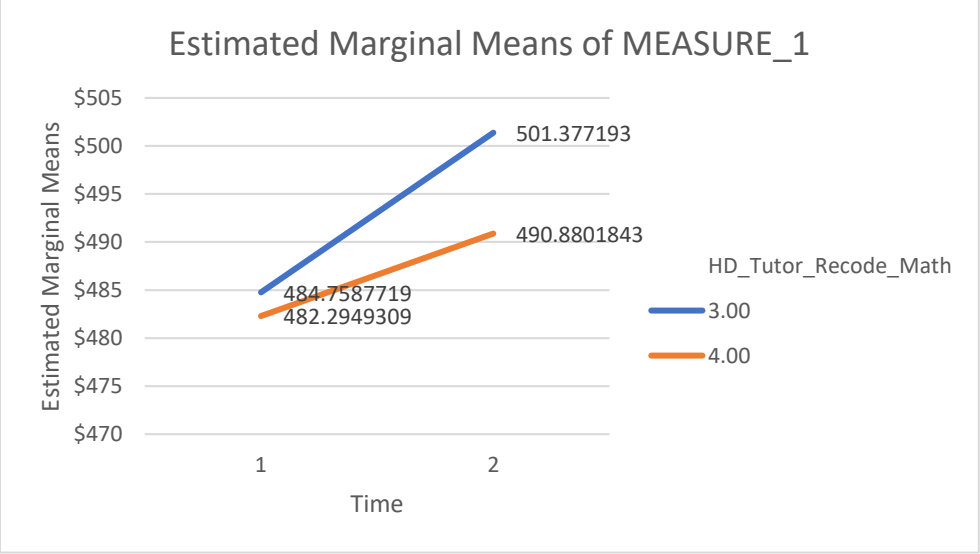
Effect	Value	F	Hypothesis df	Error df	Sig.

Time	Pillai's Trace	.065	30.552[b]	1.000	443.000	<.001
	Wilks' Lambda	.935	30.552[b]	1.000	443.000	<.001
	Hotelling's Trace	.069	30.552[b]	1.000	443.000	<.001
	Roy's Largest Root	.069	30.552[b]	1.000	443.000	<.001
Time * HD_Tutor_Recode_Math	Pillai's Trace	.007	3.104[b]	1.000	443.000	.079
	Wilks' Lambda	.993	3.104[b]	1.000	443.000	.079
	Hotelling's Trace	.007	3.104[b]	1.000	443.000	.079
	Roy's Largest Root	.007	3.104[b]	1.000	443.000	.079

a. Design: Intercept + HD_Tutor_Recode_Math

Within Subjects Design: Time

b. Exact statistic

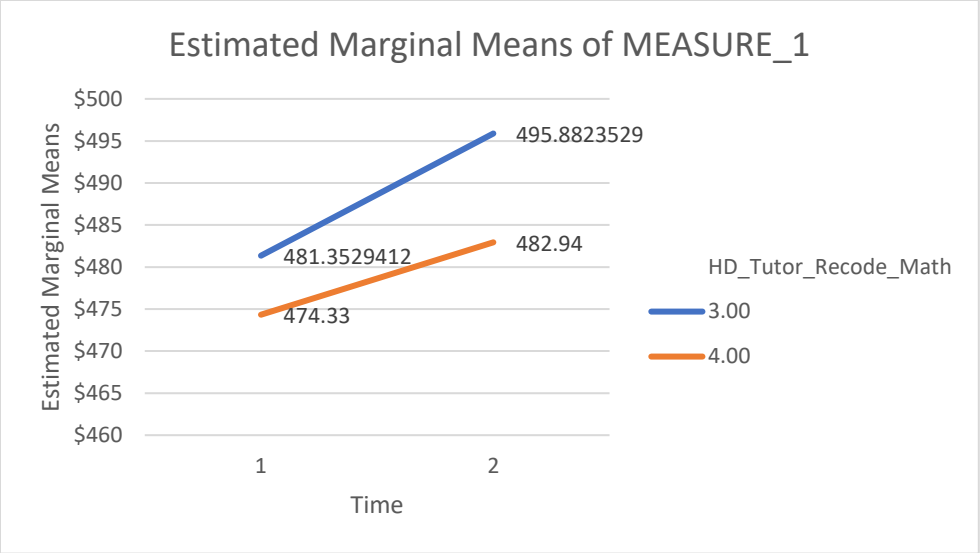


Gender Female

Statistically not different

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	481.35	43.773	102
	4.00	474.33	46.783	100
	Total	477.88	45.312	202
Math_Scale_Score_23	3.00	495.88	49.257	102
	4.00	482.94	63.774	100
	Total	489.48	57.135	202



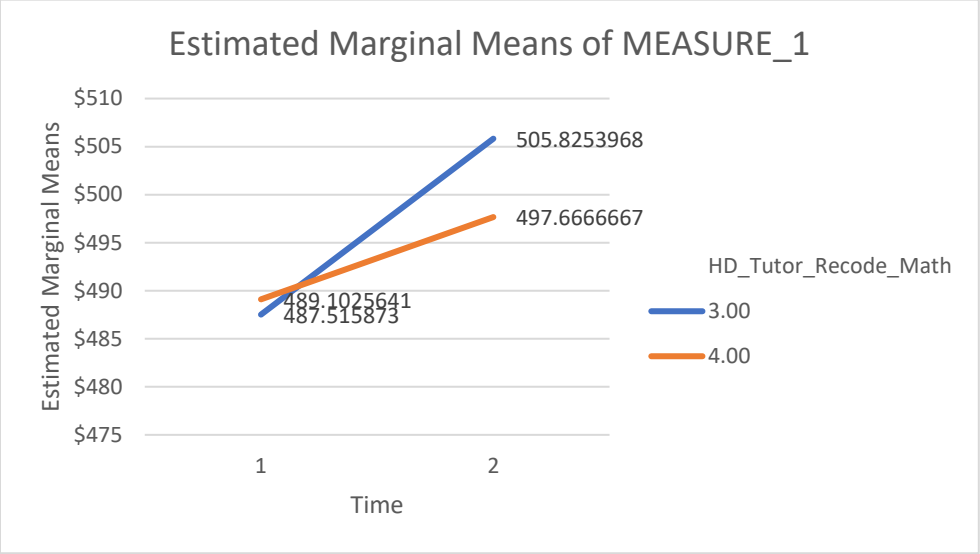
Gender Male

Statistically not different

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00 (Treatment)	487.52	43.870	126
	4.00 (Comparison)	489.10	46.432	117
	Total	488.28	45.035	243
Math_Scale_Score_23	3.00	505.83	38.374	126

	4.00	497.67	56.764	117
	Total	501.90	48.185	243

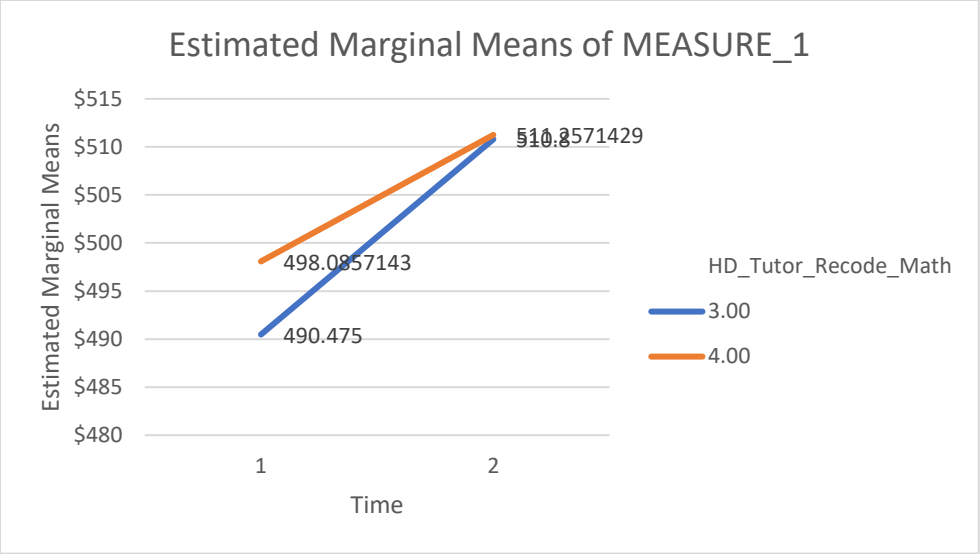


2nd to 3rd

Statistically not different

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	490.48	35.132	40
	4.00	498.09	43.956	35
	Total	494.03	39.406	75
Math_Scale_Score_23	3.00	510.80	37.841	40
	4.00	511.26	47.992	35
	Total	511.01	42.579	75



3rd to 4th

Statistically not different

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	489.17	43.148	88
	4.00	483.74	46.577	84
	Total	486.52	44.806	172
Math_Scale_Score_23	3.00	493.23	46.828	88
	4.00	489.58	66.366	84
	Total	491.45	57.069	172

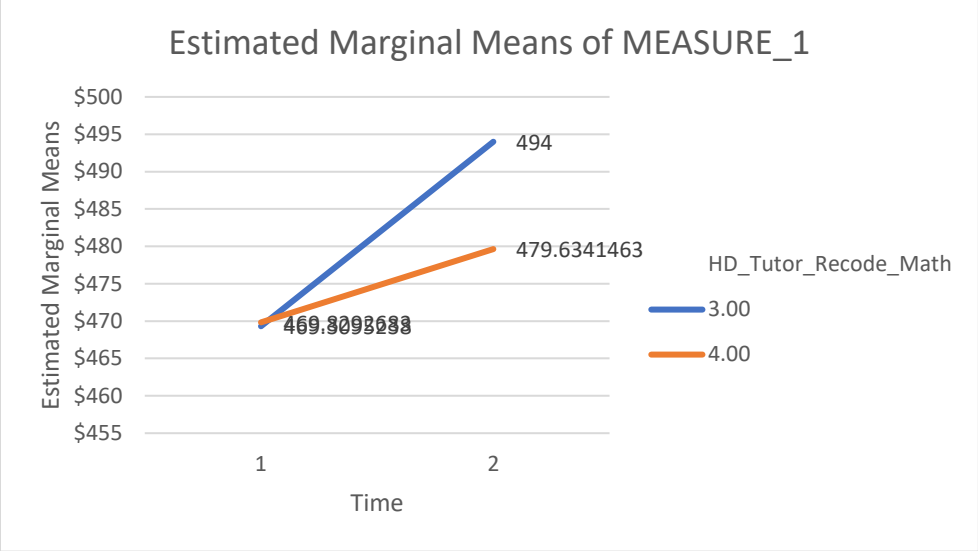
4th to 5th

Statistically not different

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	469.31	40.538	42

	4.00	469.83	48.958	41
	Total	469.57	44.620	83
Math_Scale_Score_23	3.00	494.00	30.254	42
	4.00	479.63	52.704	41
	Total	486.90	43.184	83



5th to 6th

Statistically not different

Descriptive Statistics

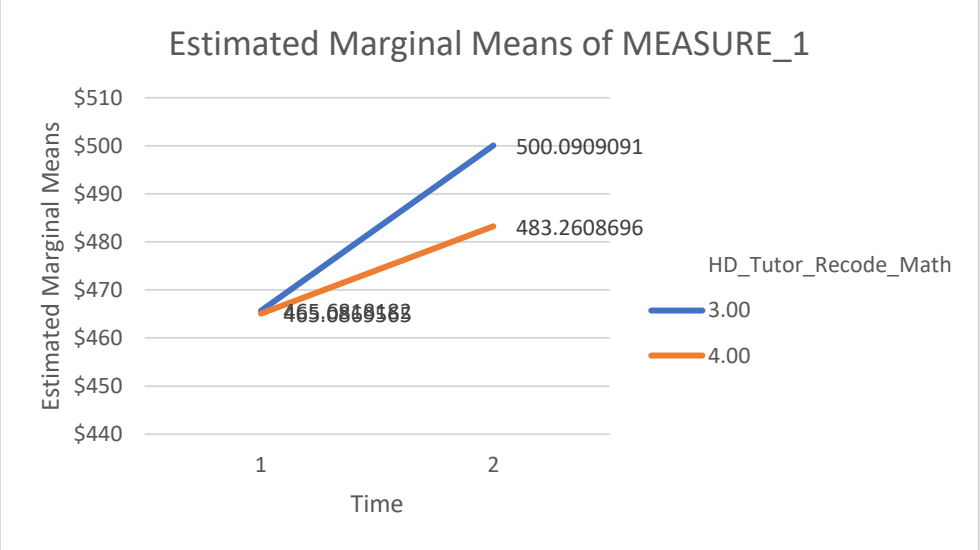
	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	500.39	33.878	31
	4.00	491.79	29.764	28
	Total	496.31	32.013	59
Math_Scale_Score_23	3.00	520.10	35.691	31
	4.00	493.57	56.842	28
	Total	507.51	48.388	59

6th to 7th

Statistically not different

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	465.68	65.438	22
	4.00	465.09	61.051	23
	Total	465.38	62.509	45
Math_Scale_Score_23	3.00	500.09	60.686	22
	4.00	483.26	54.399	23
	Total	491.49	57.530	45



7th to 8th

No comparison—small Ns

Descriptive Statistics

	HD_Tutor_Recode_Math	Mean	Std. Deviation	N
Math_Scale_Score_22	3.00	478.20	37.272	5

	4.00	476.83	35.227	6
	Total	477.45	34.303	11
Math_Scale_Score_23	3.00	521.00	51.166	5
	4.00	483.67	108.688	6
	Total	500.64	85.638	11