MobyMax Math Curriculum

A Strong, Evidence-Based Intervention for Improving Student Outcomes

William Koch PhD
Satyendra Pandey PhD
Valarmathy Selvakumar PhD

In the first quarter of 2018, MobyMax conducted an independent, large-scale experimental study with over 4,000 students in 230 classrooms across the United States to test the efficacy of MobyMax Math.

The "Gold Standard" study used a randomized control experimental design that qualifies as "Tier 1 – Strong Evidence" under the ESSA guidelines for evidence-based interventions. In addition, the study was designed to meet the evidence standards of the IES and WWC.

The study showed that MobyMax Math was very effective at producing positive student growth with an effect size of 0.698. This is the equivalent of one year of academic growth from just 100 minutes of use per week.

The students in the experimental group using MobyMax showed a 53% improvement over the students in the control group who did not use MobyMax. Both groups had the same basal instruction within their classroom.

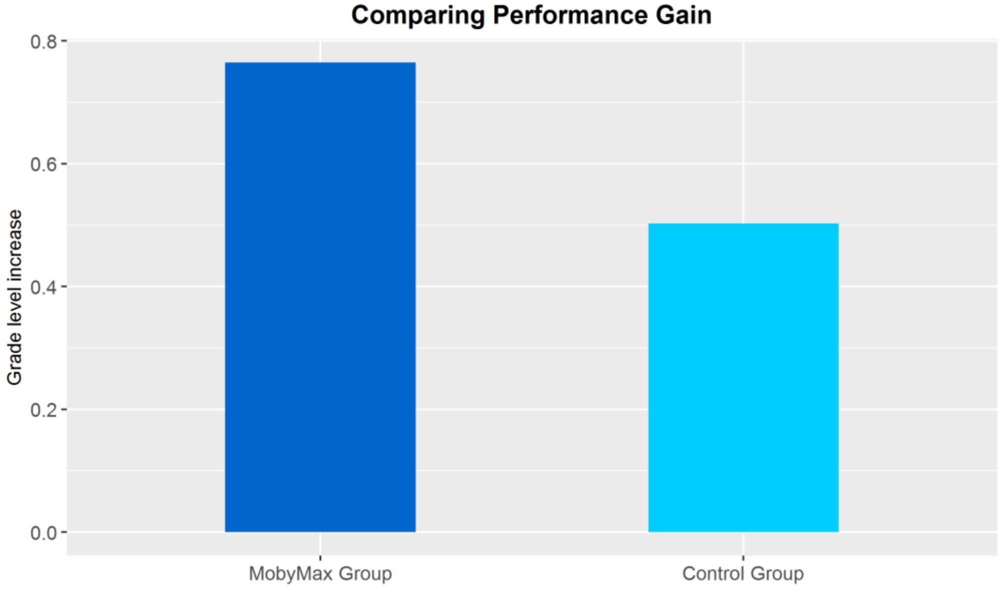


Table of Contents

Executive Summary

Background

MobyMax creates digital curriculum, assessments, and learning tools for teachers and students in grades K-8. MobyMax products are used in over 82% of all K-8 schools in the United States. This study focuses specifically on MobyMax's K-8 Math Curriculum (MobyMax Math).

Purpose

The main purpose of this study was to examine the efficacy of MobyMax Math. In addition, data was collected on factors that contributed to making MobyMax Math effective in producing positive student outcomes.

Setting

The study was conducted on over 4,000 students in 230 classrooms from a representative sample of K-8 schools across the United States during the first quarter of 2018.

Conclusion

The study showed that MobyMax Math has a statistically significant positive effect on improving student outcomes with an effect size of .698. This equates to roughly one school year of growth with just 100 minutes of use per week.

Given the size and design of the study, the results of the analysis indicate MobyMax Math is an effective intervention that reliably produces positive student outcomes for K-8 students.

Project Background

This research is an effort to provide evidence of the efficacy of MobyMax Math in improving student proficiency in math. This study was designed to address the ESSA requirements for evidence-based interventions and adhere to Tier 1 criteria for strong evidence of effectiveness.

Every Student Succeeds Act (ESSA) Requirements

The ESSA requires that all practices or programs be evidenced based. It defines four tiers of evidence based on strength:

- Tier 1 – strong evidence – a program or practice that is supported by at least one randomized control experimental study
- Tier 2 – moderate evidence – a program that is supported by at least one quasi-experimental study
- Tier 3 – promising evidence – a program or practice that is supported by at least one correlational study with statistical controls for selection bias
- Tier 4 – demonstrates a rationale – a program or practice that has a well-defined logic model or theory of action, is supported by research, and has some effort underway to determine its effectiveness

Study Qualifies as ESSA Tier 1 “Strong Evidence”

ESSA requirements to qualify as strong evidence are:

1. The study must be a well-designed and well-implemented randomized control experimental study.
The study employed a randomized control experimental design. In addition, the study was designed to meet the What Works Clearinghouse (WWC) evidence standards.

2. The statistical analysis must demonstrate a statistically significant positive effect on improving student outcomes.
The study showed that MobyMax Math has a statistically significant positive effect on improving student outcomes with an effect size of .698. This equates to roughly one school year of academic growth with just 100 minutes of use per week.

3. The study must use a large, multisite sample that overlaps with the populations proposed to receive the intervention.
The study was based on a representative sample of K-8 schools across the United States that included over 4,000 students in 230 classrooms. A large sample, as defined by ESSA, is a sample of 350 or more students or 50 or more classrooms.

Design and Methodology

Groups

- MobyMax randomly divided the students in each classroom into two equally sized groups: a math group and a language group
- The math group was the experimental group. Students in this group worked in MobyMax Math for 10 weeks. The language group was the control group. Students in this group worked in MobyMax Language for 10 weeks and were not permitted to work in MobyMax Math.
- Teachers continued to teach their standard curriculum throughout the course of the study. MobyMax was used as a supplement.
- Because the control and experimental groups were created within classrooms rather than between classrooms, factors related to the classrooms, teachers, or schools are controlled for experimentally. This methodological control is both important and uncommon for large-scale education experiments.
- Since the control and experimental groups in each classroom received the same basal curriculum, the effect of the basal curriculum is accounted for by subtracting the growth of the control group from the growth of the experimental group.

Student Time Worked

- Students were asked to work three hours per week in their respective subjects with the expectation that the actual time spent would vary considerably. The time variation allowed the study to analyze the relation between time spent and student improvement.
- The actual time spent by individual students varied from over 5 hours per week to 1 hour per week. The study did not include classrooms that averaged less than 1 hour per week in the analysis.

Pre- and Post-tests

- Students were tested at the onset and conclusion of the study with two tests: a math placement test that covered grades K through 8 and a math benchmark test that covered each student's current grade level.

Study Hypotheses

This study analyzed six questions concerning the effectiveness of MobyMax Math:

- **Study Question 1:** How effective is MobyMax Math at producing student growth when used as a supplement to basal curriculum?

- **Study Question 2a:** Within the experimental group, do students with a high Prior Year Learning Deficit (PYLD) experience more growth than students with a low PYLD, and if so, how much?

- **Study Question 2b:** Within the experimental group, which variables for the high PYLD quartile students are the most significant in terms of growth compared to students in the Low PYLD quartile?

- **Study Question 3:** Within the experimental group, is overall time spent using MobyMax Math correlated with student growth?

- **Study Question 4a:** Is the number of problems completed in MobyMax Math correlated with student growth?

- **Study Question 4b:** Is the number of problems *completed correctly* correlated with student growth?

- **Study Question 5a:** Do students in special education classrooms experience higher growth when using MobyMax Math than students in general education classrooms?

- **Study Question 5b:** Within the experimental group, are there differences in factor importance when comparing special education and general education students with respect to growth produced by MobyMax Math, and if so, which variables?

- **Study Question 6**: Does grade level have a significant effect on student growth?

Data Measures

ResearchGroup: Control or Experimental

ClassroomType: Special Education or General Education

MathBasalTime: From the survey, teacher's estimate of time students spent per week on basal curriculum plus time spent on other (non-Moby) supplemental curriculum

GradeLevel: The actual grade level of the student assigned by the teacher. Note that grade 0 = grade K.

ScorePlacementInitialOverall: The overall score from the initial placement test

ScorePlacementEndOverall: The overall score from the ending placement test

Gain: ScorePlacementEndOverall – ScorePlacementInitialOverall

TimeTotal: Total time spent in MobyMax in seconds. Note that this includes the time spent taking the placement tests. As a result, students in the control group also have time for TimeTotal, but it should not count as time spent in Moby supplemental curriculum. Similarly, this time coming from the placement tests should be subtracted out of the experimental group time, which could be done by taking the average time spent taking the placement test for the control group (i.e. TimeTotal for control students), and subtracting that value out of the TimeTotal of each individual experimental student.

ProblemTotal: Total number of problems done. Note that this does NOT include the placement test problems, so these are all problems that count as MobyMax supplemental curriculum (unlike the TotalTime field).

ProblemCorrect: Total number of problems done correctly. Note that this does NOT include the placement test problems, so these are all problems that count as MobyMax supplemental curriculum.

PYLD: Prior Year Learning Deficit = GradeLevel (the actual grade level such as 5.0 for 5th grade) - ScorePlacementPriorLevelInitial (the sum of all the prior grade levels before the 5th grade)

Sample Description

The study was conducted on over 4,000 students in 230 classrooms from a representative sample of K-8 schools across the United States. To select the classrooms for participation, teachers with MobyMax licenses and between 8 and 35 students in their classroom were emailed information about the study and about how to register their classroom if interested in participating. Participation was limited to one classroom per school.

Table 1: Descriptive Statistics

Variable	Control	Experimental
Special Ed: n	496	534
General Ed: n	1840	1948
Math: Mean	5.75	5.65
Math: sd	3.72	3.57
Grade Level: mean	3.38	3.40
Grade Level: sd	1.68	1.74
Score (initial): mean	3.03	3.03
Score (initial): sd	1.25	1.25
Score (end): mean	3.53	3.79
Score (end): sd	1.31	1.40
PYLD: mean	10.97	11.14
PYLD: sd	11.03	11.53

Table 1 presents baseline data on the final cohort to be analyzed.[1] Given the study design, the number of observations, and the randomization protocol, there were no significant statistical differences between the control and experimental groups on baseline measurements at the $p < .05$ level. The math baseline measurement has the lowest p-value ($p = .32$) indicating the Control group has a marginally higher baseline score which would tend to make estimates of the overall study effects weighted against finding results and thus more

[1] These data include students in the language group who answered more than 100 problems in MobyMax Math and students enrolled in the math group who did not answer any problems in MobyMax Math as indicated in the description for the *ProblemTotal* variable. In addition, there were several outliers in the experimental group which were included as described in Appendix A. This follows an intent-to-treat protocol as the most robust and conservative observation inclusion criteria for randomized control trials.

conservative.[2] This strengthens the statistical validity of any differences between the groups for gains in math proficiency being attributable to using MobyMax supplemental materials.

Also, following a more conservative inclusion/exclusion criteria for the analysis data, intention-to-treat (ITT) analysis was followed rather than per-protocol analysis. Randomized controlled trials often result in two major study problems: protocol non-compliance and missing outcomes data. For this study the first problem showed up in a number of controls not following protocol and using MobyMax material when they should not have. Intention-to-treat essentially includes all observations based upon randomization assignment. Problems with subjects not following protocol, or any non-compliance issues, are ignored after randomization. Consequently, ITT analysis estimates result in generally conservative treatment effects.[3]

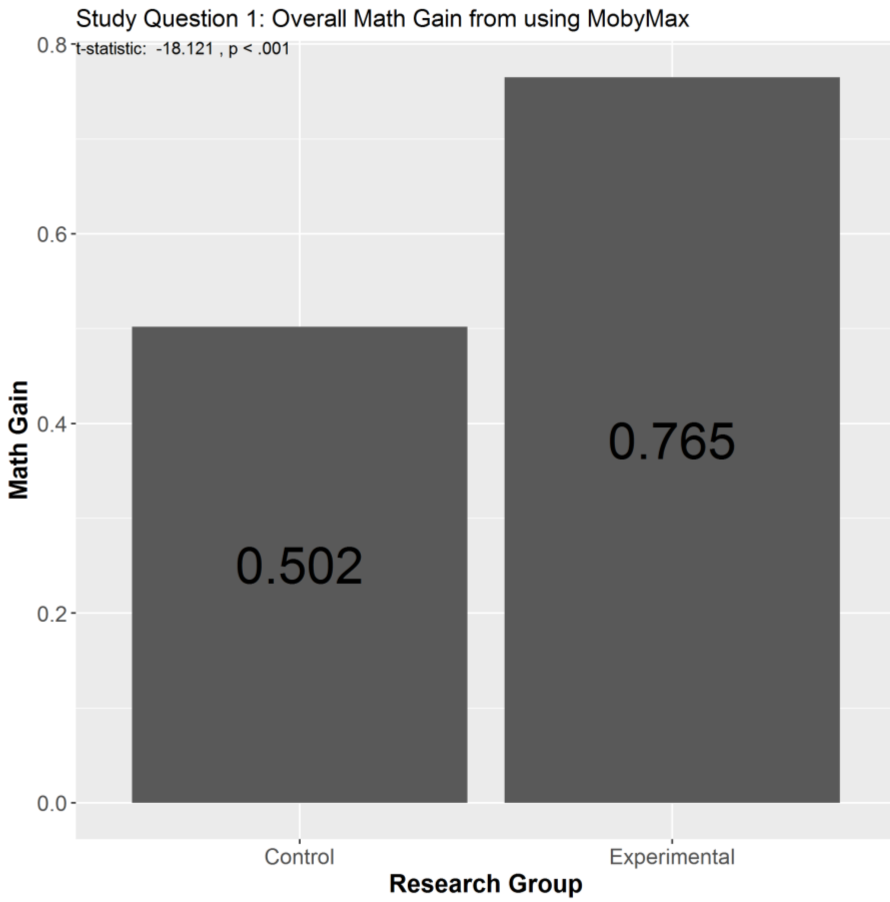
---

[2] Appendix B compares the two groups on baseline measurements.

[3] Gupta, Sandeep K. (2011): Intention-to-treat: A Review

Results

Study Question 1: How effective is MobyMax Math at producing student growth when used as a supplement to basal curriculum.

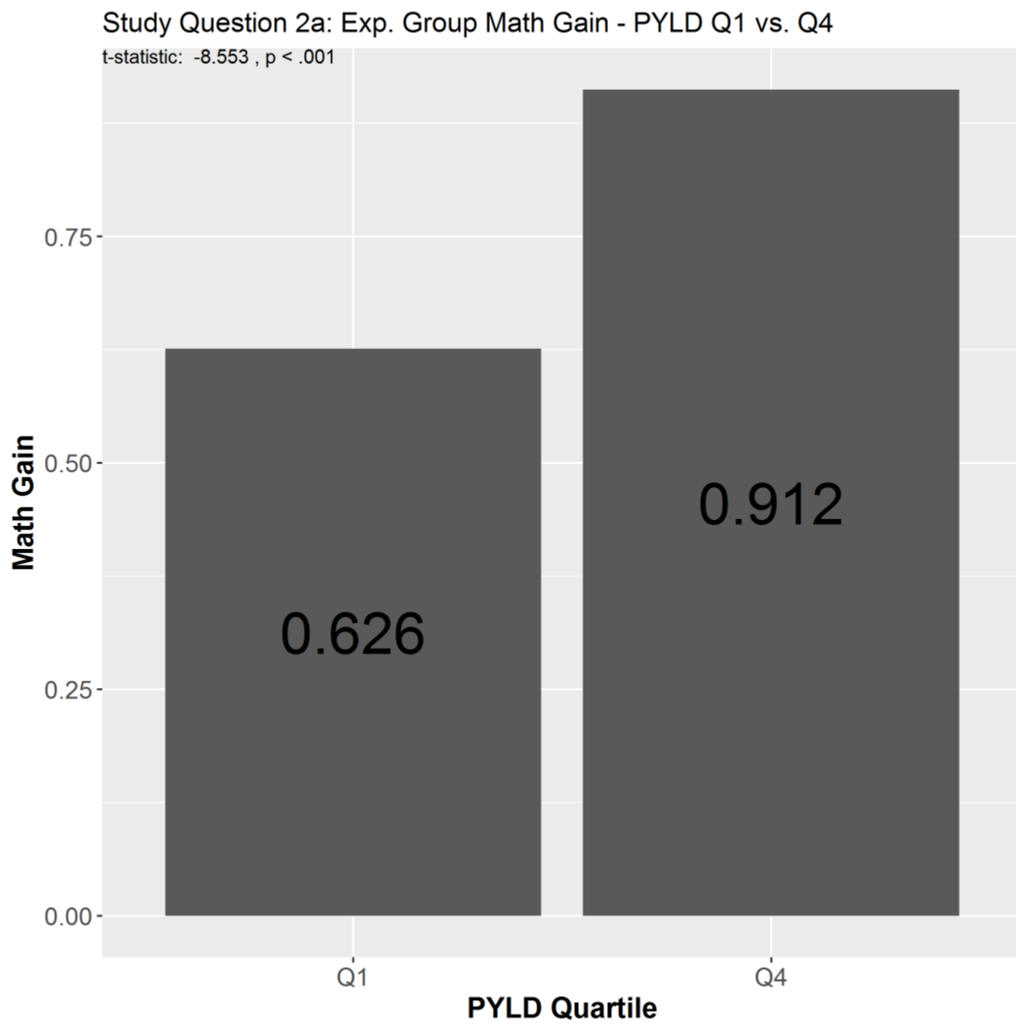


The experimental group exhibited .263 grade levels more growth than the control group (.765 – 0.502 = .263). This result was significant at the $p < .001$ level. Given the randomized design of the study and that baseline measures of performance showed no statistical difference between the control and experimental groups[4], this growth can be attributed to MobyMax Math.

The effect size for MobyMax Math was .698. An effect size of this magnitude qualifies MobyMax Math as a "highly effective" treatment for improving student outcomes. Effect size was calculated using Cohen's *d*, a widely used method for calculating the difference between two means, measured in standard deviations.

Study Question 2a: Within the experimental group, do students with a high PYLD experience more growth than students with a low PYLD, and if so, how much?

[4] As indicated in Table 1



PYLD (Prior Year Learning Deficit) is a measure of overall current grade level controlling for previous grade levels. Thus, a higher PYLD indicates a greater learning deficit (i.e. missing skills from previous grade levels). Within the experimental group, the PYLD was separated into quartiles. The lowest quartile (those students with the smallest learning deficit) were then compared to the highest quartile (those students with the greatest learning deficit). Students with the greatest deficit (high PYLD) showed increased growth (.912 - .626 = .288 grade levels) over those with the smallest deficit (low PYLD). This result was significant at the $p < .001$ level. Put another way, those with the highest learning deficit benefitted the most from MobyMax Math supplemental material, although both groups indicated substantial gain.

Study Question 2b: Within the experimental group, which variables for the high PYLD quartile students are the most significant in terms of growth compared to students in the Low PYLD quartile?[5]

[5] To control for time and number of problem effects, *ProblemTotal* was included in the regression. The variable was significant at the p < .001 level. However, the effect size was < .00001, so it not reported in the table.

Given the placement test score gain differences between the high and low PYLD quartiles in the experimental group, additional statistical models were run to delve into possible reasons for this result. Table 2b compares a regression for the full experimental group, PYLD Q1, and PYLD Q4[6]. The adj. R^2 for the full experimental group regression was .323. This indicates that variables included in the model explain ~32% of the variance in the outcome variable MobyMax score gain. Other than *MathBasalTime* (teacher estimate of non-MobyMax time spent on math), all other variables in the equation are significant at the $p < .01$ or $p < .001$ level. Specifically, controlling for *TimeTotal* and *ProblemTotal*, the more likely a student was in a general education classroom, the higher the grade level, the lower the initial placement test score, and the higher the PYLD, the higher the MobyMax score gain.

However, when running separate regressions for the top and bottom quartiles, a more nuanced picture emerges. For the low PYLD quartile group both classroom type and grade level remain significant indicators for improving the MobyMax score gain, but the initial placement test score (*Score_Initial*) is no longer a statistically significant factor.

Table 2b: Score Gain Determinants by PYLD Quartile

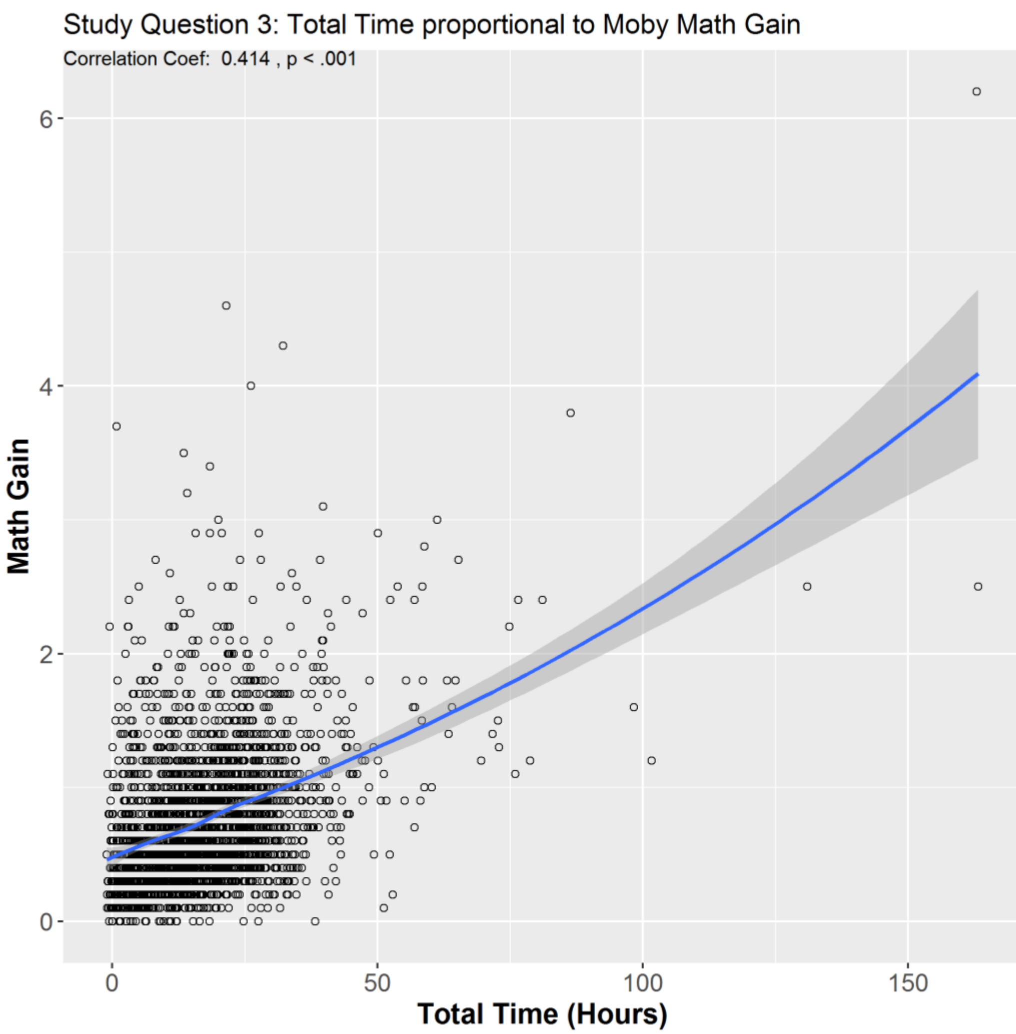
Coefficients	*Response*					
	Math Gain		Gain (PYLD - Q1)		Gain (PYLD - Q4)	
	Estimate	*std. Error*	*Estimate*	*std. Error*	*Estimate*	*std. Error*
Constant	0.171 **	0.057	0.012	0.092	0.099	0.142
ClassroomType	0.137 ***	0.024	0.100 *	0.042	0.279 ***	0.055
MathBasalTime	0.001	0.003	0.004	0.005	-0.005	0.006
GradeLevel	0.134 ***	0.020	0.078 **	0.026	0.024	0.021
Score_Initial	-0.128 ***	0.020	-0.047	0.026	-0.038	0.033
TimeTotal	0.004 ***	0.001	0.004 *	0.002	0.005	0.003
PYLD	-0.008 **	0.003				
Observations	2482		621		620	
R^2 / adj. R^2	.325 / .323		.352 / .346		.295 / .288	
Notes				** p<.05 ** p<.01 *** p<.001*		

For the High PYLD group (those students showing the greatest grade level deficit), controlling for other individual student characteristics, only *ClassroomType* remains a significant predictor, but with over 2.7 times the effect size compared to the experimental group overall. Specifically, a student in a general education classroom (*ClassroomType* = 2) has a score gain of 0.279 compared to a score gain of only .10 for a student in a special education classroom. This result

[6] Appendix D shows the results of a Variance Inflation Factor analysis. Any variable with a VIF > 10 is excluded. The model is then re-run without the excluded variable and the VIFs of the remaining variables is calculated. This process stops when all the remaining variables have a VIF < 5.

speaks to the overall effectiveness of MobyMax supplemental material in different classroom type environments and indicates possible tailoring of material for different types of classes going forward.

Study Question 3: Within the experimental group, is overall time spent[7] using MobyMax Math correlated with higher student growth?

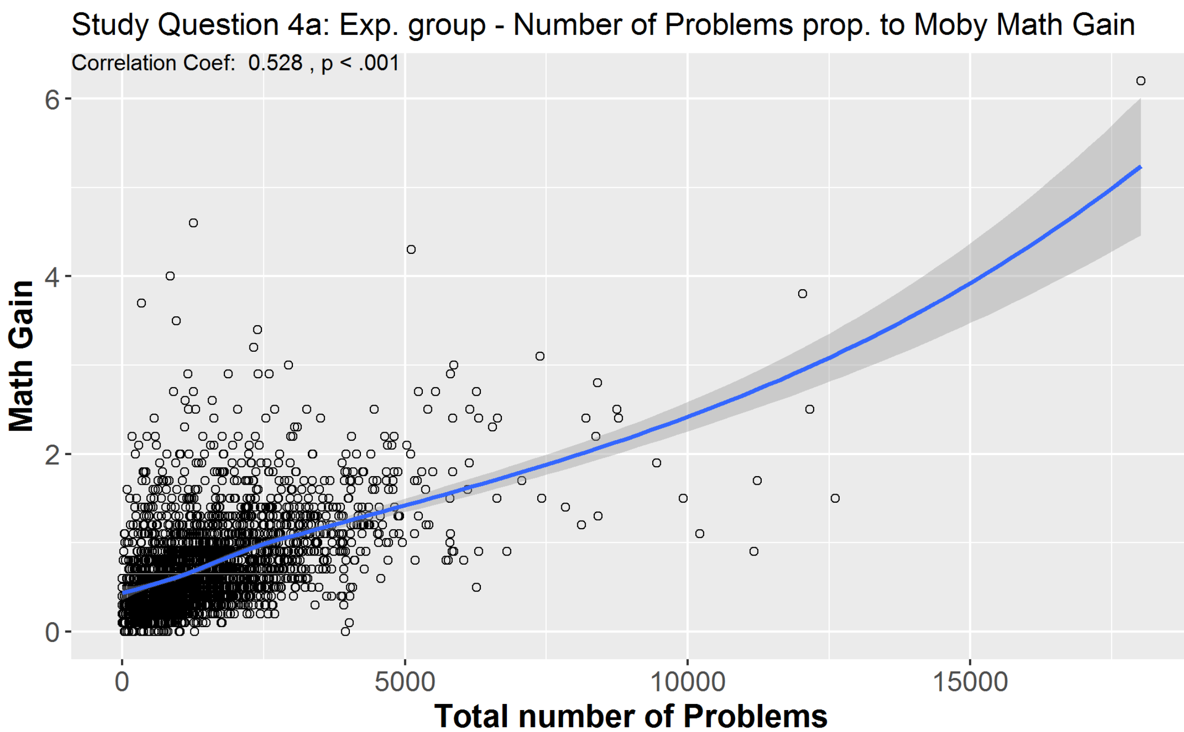


Within the experimental group the correlation between total time spent on MobyMax Math and score gain was calculated with an associated scatter plot and blue line with grey data dispersion region to highlight the relationship between the two measures.[8] A correlation of .414, significant at the $p < .001$ level, indicates a moderately strong relationship between total time spent using

[7] The TimeTotal variable is time spent only on MobyMax supplemental materials and does not include time spent in class on math-related lessons and material.
[8] For the correlational analyses and graphs, a LOWESS (Locally Weighted Scatterplot Smoothing) method was used. As the data become more sparse for results at the margins, the visual accuracy (given by the grey line) diminishes.

MobyMax Math and score gain. Put another way, ~17% of the variation in MobyMax score gain can be attributed to the *TimeTotal* variable.[9] While not statistically controlling for other variables factoring into MobyMax score gain, this is in line with basic intuition. The more time a student spends in MobyMax Math, the better the results.

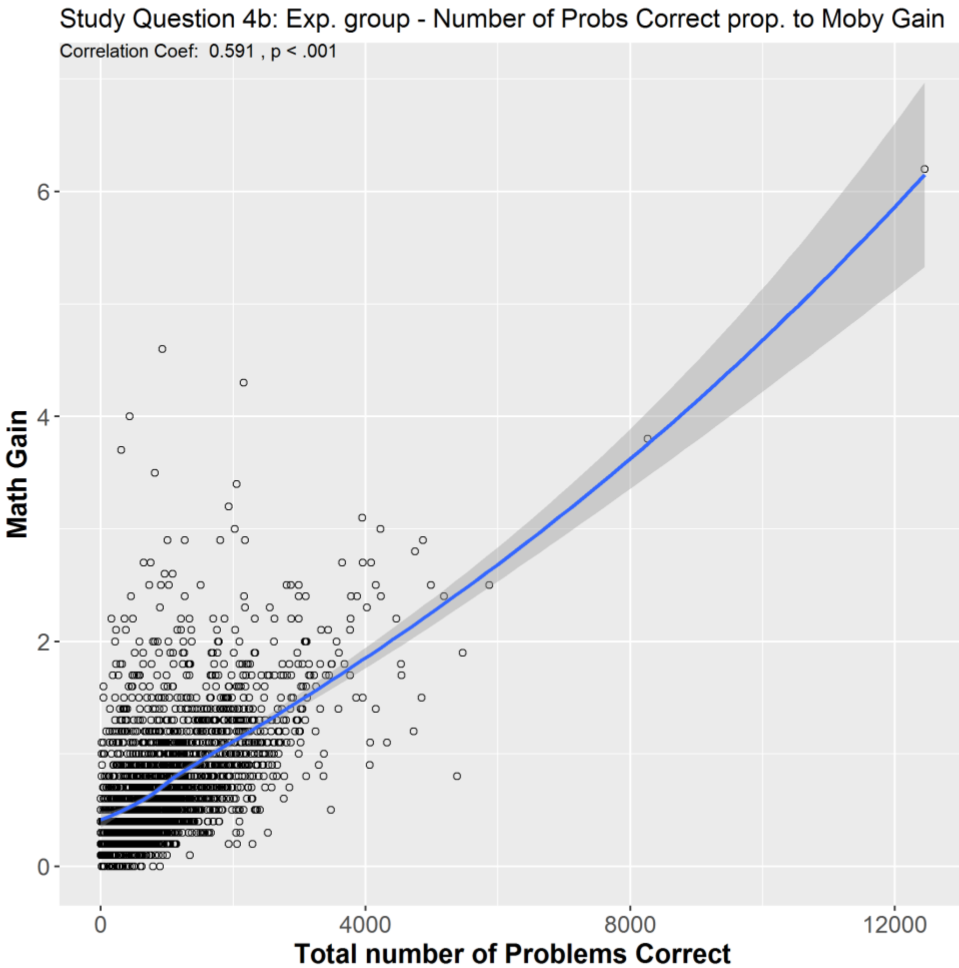
Study Question 4a: Is the number of problems completed in MobyMax Math correlated with higher student growth?



Similar to the examination of the relation between total time spent in MobyMax Math and score gain, within the experimental group, the correlation between the number of problems completed in MobyMax Math and score gain was calculated and illustrated with an associated scatter plot and line. While total time and number of problems completed are related, this question might be more associated with student focus on the supplemental material as opposed to just time spent on the material. A correlation of .528, significant at the $p < .001$ level, indicates a strong relationship between the number of problems completed and student growth. Put differently, ~28% of the variation in MobyMax score gain can be attributed to the *ProblemTotal* variable.[10] Since the correlation between the number of problems completed and score gain was stronger than the correlation between time spent and score gain, it is likely that effort and focus while using MobyMax Math are more important than simply time spent in the program.

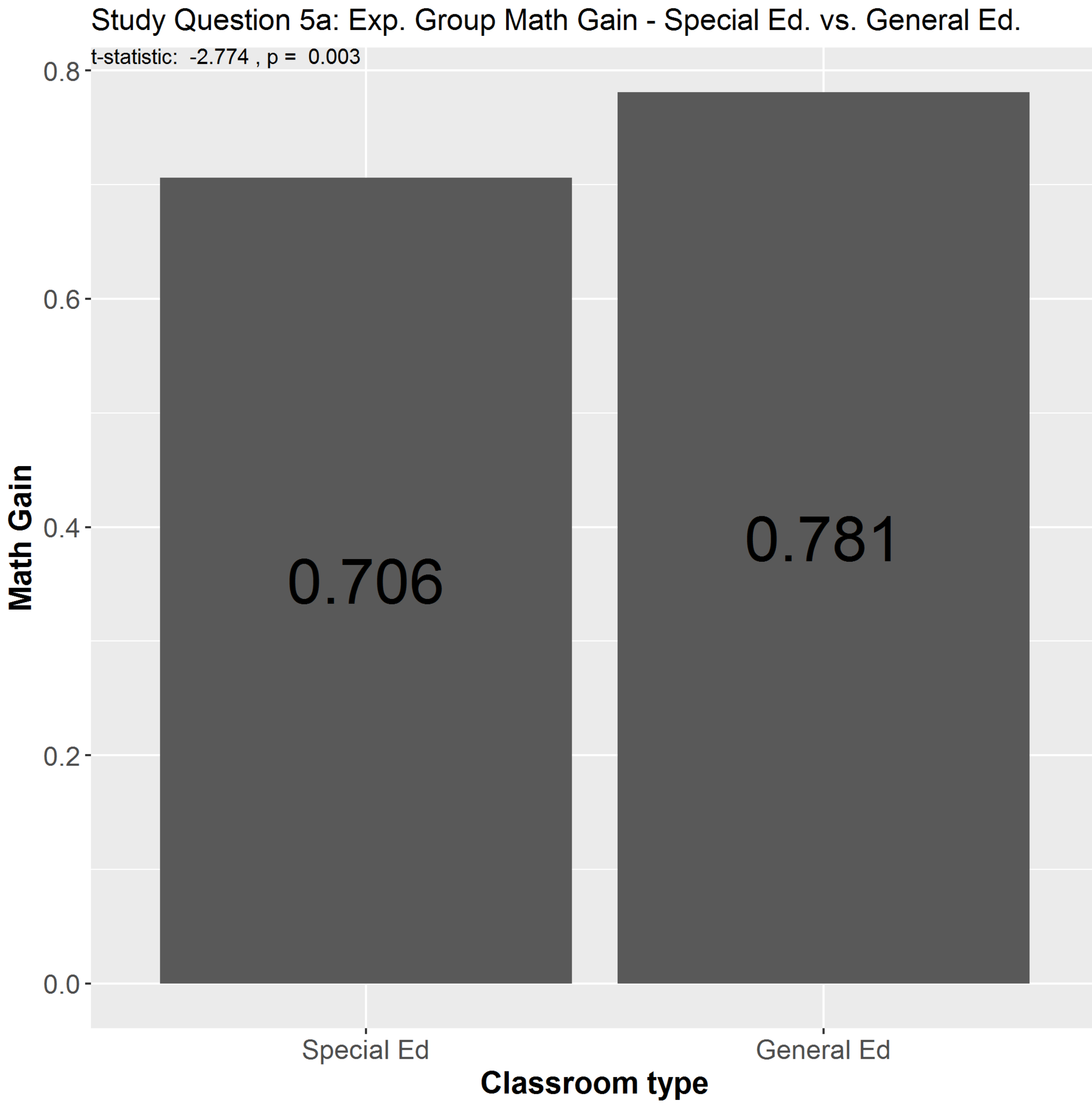
[9] Since this statistic doesn't control for other effects it should be seen as a general indicator of the importance of the statistical relationship.
[10] Again, this result should be interpreted as a general indicator of the importance of the statistical relationship.

Study Question 4b: Is the number of problems *completed correctly* correlated with higher student growth?



In this analysis, the relationship between number of problems *completed correctly* is correlated with MobyMax score gain. Similar to the analysis on total number of problems completed, a correlation of .591, significant at the $p < .001$ level, indicates a strong relationship between the number of problems completed correctly and score gain. With a somewhat higher correlation coefficient than in Study Question 4a, ~35% of the variation in MobyMax score gain can be attributed to the number of problems completed correctly. This indicates that both effort and success factor significantly into higher MobyMax score gains. Again, while not statistically controlling for other variables associated with MobyMax score gain, this is also in line with general intuition. The more problems a student worked on and the more success she had at solving those problems, the better the results.

Study Question 5a: Do students in special education classrooms experience higher student growth when using MobyMax than students in general education classrooms?



Within the experimental group, this analysis compares the MobyMax Math score gain of students in special education and general education classrooms. Given that an argument could be made for either classroom type benefitting more from the supplemental material, this comparison consisted of a two-sided t-test rather than one-sided. Using a two-sided test the overall result was significant at the $p < .01$ level. However, the effect size of the difference (.781 - .706 = .075 grade levels), while statistically significant, is relatively small. Being in a general education class indicates, on average, a .075 grade level increase. To further examine reasons for this difference, a regression analysis was performed on each classroom type within the experimental group. The results are presented in Table 5b.

Study Question 5b: Within the experimental group, are there differences in factor importance when comparing special education and general education students with respect to growth produced by MobyMax Math, and if so, which variables?[11]

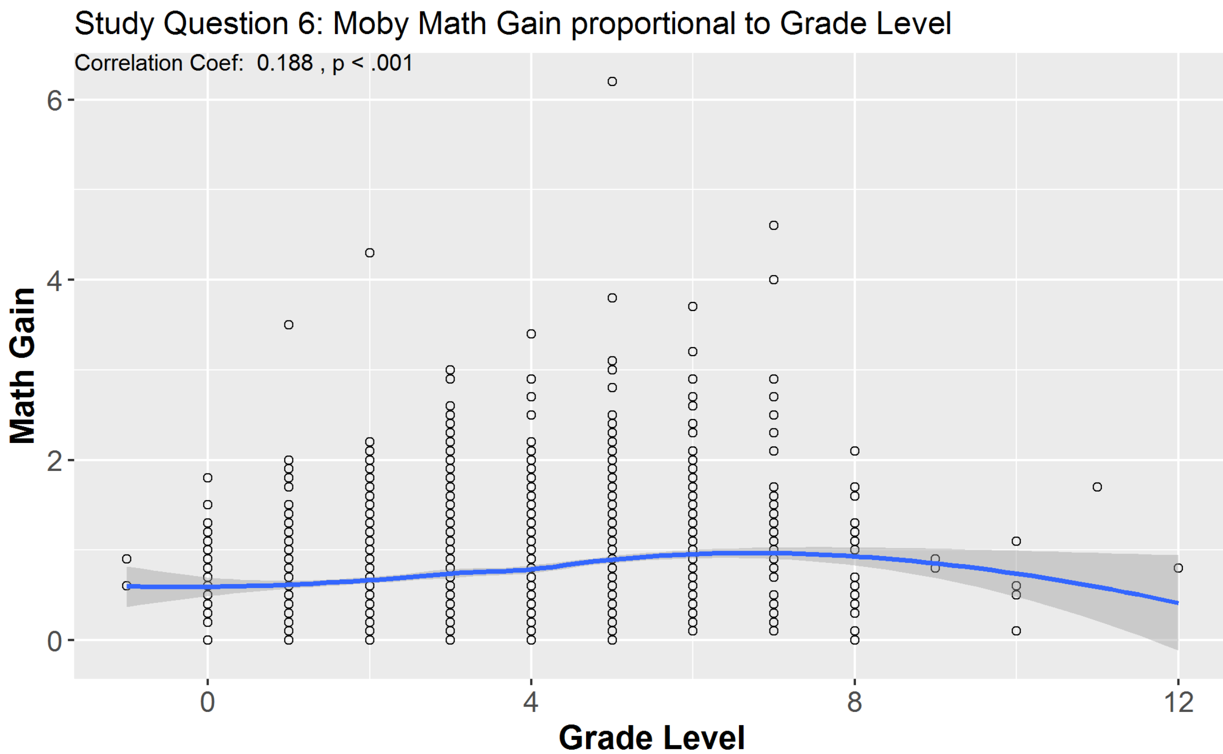
Table 5b: Score Gain Determinants by Classroom Type

Coefficients	*Response*					
	Math Gain		Gain (Special Educ.)		Gain (General Educ.)	
	Estimate	*std. Error*	*Estimate*	*std. Error*	*Estimate*	*std. Error*
Constant	0.171 **	0.057	0.499 ***	0.088	0.306 ***	0.038
ClassroomType	0.137 ***	0.024				
MathBasalTime	0.001	0.003	0.016 *	0.007	-0.003	0.003
GradeLevel	0.134 ***	0.020	0.074	0.043	0.122 ***	0.022
Score_Initial	-0.128 ***	0.020	-0.119 **	0.045	-0.102 ***	0.022
TimeTotal	0.004 ***	0.001	0.004	0.003	0.004 **	0.001
PYLD	-0.008 **	0.003	-0.005	0.005	0.002	0.003
Observations	2482		534		1948	
R^2 / adj. R^2	.325 / .323		.173 / .163		.410 / .408	
Notes				* *p<.05* ** *p<.01* *** *p<.001*		

The results of this analysis are interesting in that the model's statistical explanatory power, as indicated by the higher adj. R^2, increases when students in the general education classroom are considered separately from the complete experimental cohort. In particular, after controlling for other factors related to MobyMax score gain, an increase in the grade level for a student in a general education classroom is associated with a score gain of .119 at the $p < .001$ significance level. However, the change in grade level for students in special education classrooms is not statistically significant. For both classroom types initial placement test score is inversely related to MobyMax score gain and significant at the $p < .01$ level or better; the lower the initial test score, the greater the improvement. Further, for the special education cohort, the *MathBasalTime* variable factors positively into MobyMax score gain. The higher the *MathBasalTime*, the higher the overall score gain. This is not the case for general education students. Finally, for students in both types of classrooms, once they are divided into separate regression analyses, PYLD stops being a significant factor in MobyMax score gain. This may indicate that the initial placement score is a proxy for PYLD in this analysis.

[11] As with the regression comparing PYLD quartiles, to control for time and number of problem effects, *TimeTotal* and *ProblemTotal* were included in the regression. Both were significant at the p < .001 level but only *TimeTotal* is shown in the table. The effect size for *ProblemTotal* was < .00001 so not reported in the table.

Study Question 6: Does overall grade level have a significant effect on student growth?



While statistically significant at the *p* < .001 level, the correlation coefficient for MobyMax score gain ~ Grade Level of 0.188 indicates a weak statistical relationship with ~ 3.5% of the MobyMax score gain associated with increases in grade level. Further, for any student with a grade level > 8 the relationship is statistically tenuous as indicated by the downward direction of the trend line and widening grey dispersion region.[12] Given the significance of grade level in the two regression tables (baseline regression models for Table 2b and Table 5), for grade level to indicate any statistically significant relation to MobyMax score gain, other factors must be controlled to show the marginal effect. Grade Level, by itself, does not have a strong statistical relation to MobyMax score gain.

[12] Appendix C is a sensitivity analysis by looking at the same question excluding students with Grade Level > 8. The results reduce the correlation coefficient from .188 to .178 while remaining statistically significant.

Conclusion

The study showed that MobyMax Math has a statistically significant positive effect on improving student outcomes with an effect size of .698. This is equivalent to roughly one school year of additional growth with 100 minutes of use per week.

The study randomized control and experimental groups within classrooms. The randomization procedure within classrooms strengthens the validity of statistical assumptions behind the methods used in the analysis. In addition, the sample size was large enough to examine subsets of the data with more than sufficient observations to allow more nuanced analysis.

Given the size and design of the study, the results of the analysis indicate MobyMax Math is an effective intervention that reliably produces positive student outcomes for K-8 students.

References

Adrian A. Dragulescu and Cole Arendt (2018). xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files. R package version 0.6.1. https://CRAN.R-project.org/package=xlsx

Gupta, Sandeep H., “Intent to Treat Concept: A Review,” 2011, *Perspectives in Clinical Research*, Jul-Sep; 2(3): 109–112.
URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159210/

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

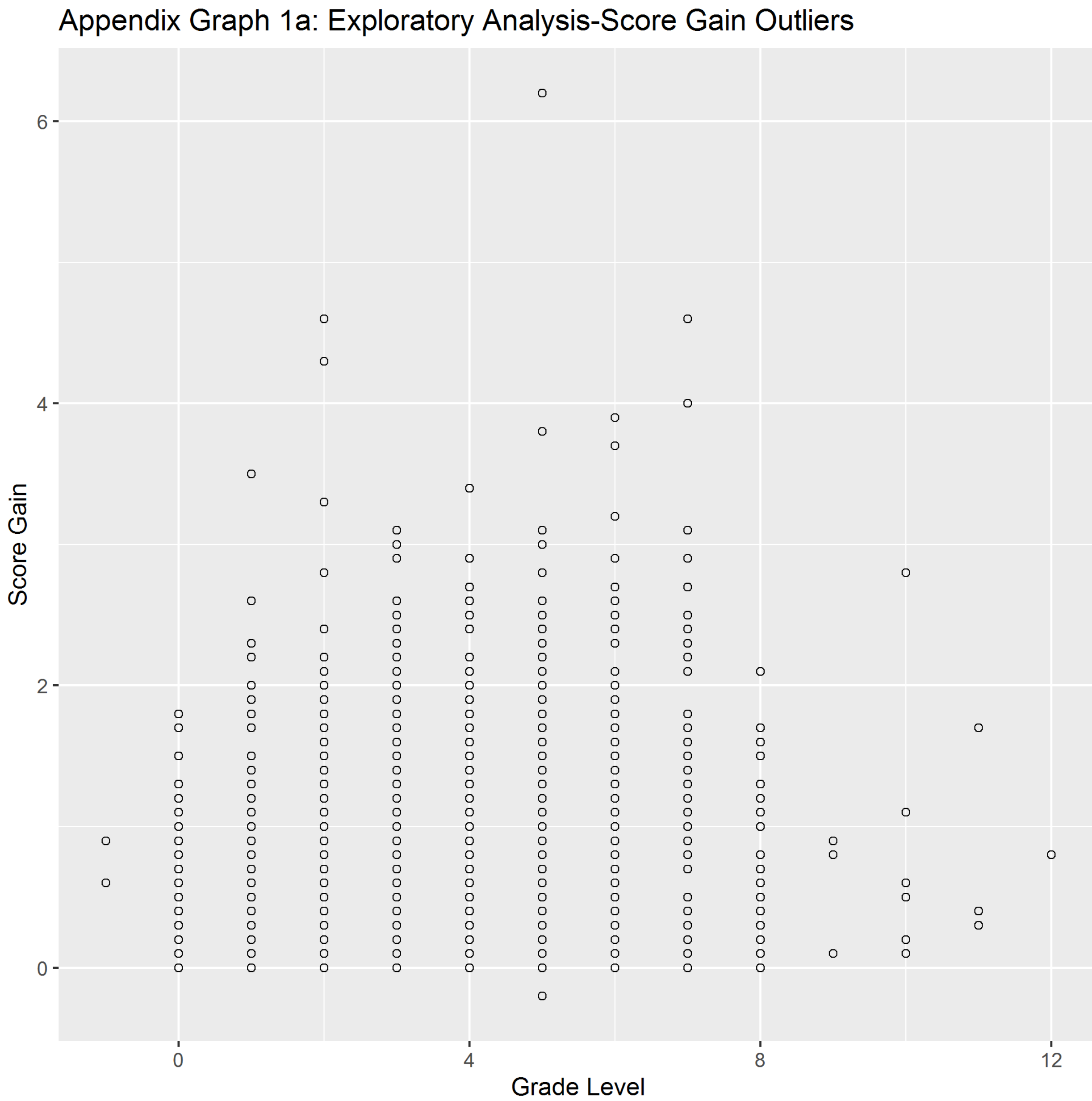
Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.5. https://CRAN.R-project.org/package=dplyr

Lüdecke D (2018). _sjPlot: Data Visualization for Statistics in Social Science_. R package version 2.4.1, <URL: https://CRAN.R-project.org/package=sjPlot>.

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

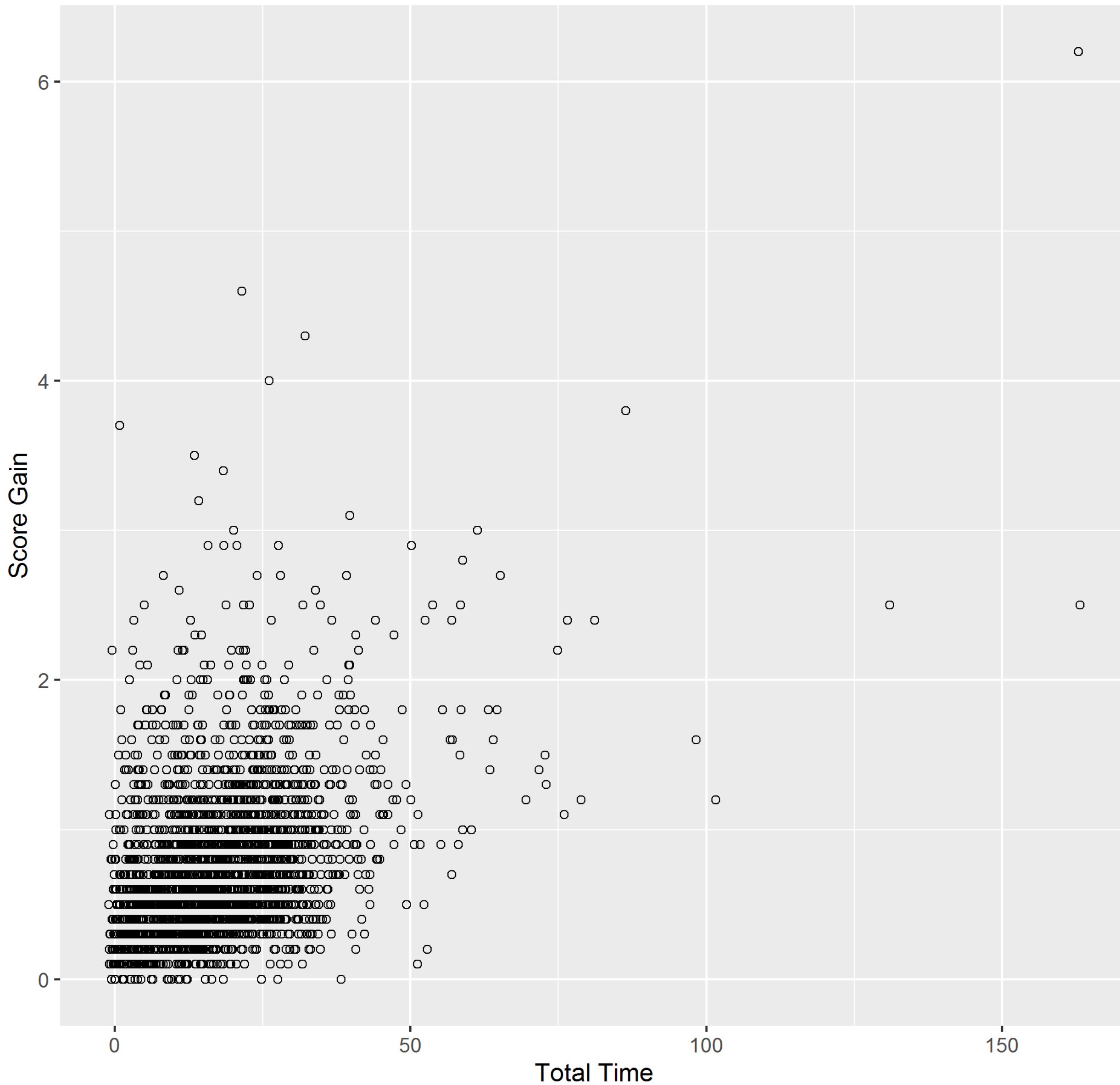
Appendix – Graphs

Appendix A: Data Outliers



A scatter plot of Score Gain vs Grade Level indicates one observation is a potential outlier with a Score Gain > 6. To maintain the robustness of the study randomization, and given that this observation is still within an acceptable score-gain range, it was kept in the data to be analyzed.

Appendix Graph 1b: Exploratory Analysis-Total Time Outliers



A scatter plot of Score Gain vs. Total Time (in hours) within the experimental group indicates two observations as potential outliers with Total Time > 125 hours. Since not necessarily extreme values given the 10-week time frame for the study, these observations were also included in the main analysis as part of the randomized data.

Appendix B: Control and Experimental Group baseline comparisons

Appendix Table 1a: Randomization Design Descriptive statistics - Categorical Vars

ClassroomType	*ResearchGroup*		***Total***
	Control	Experimental	
Special Ed	447	458	905
General Ed	1775	1803	3578
Total	2222	2261	4483

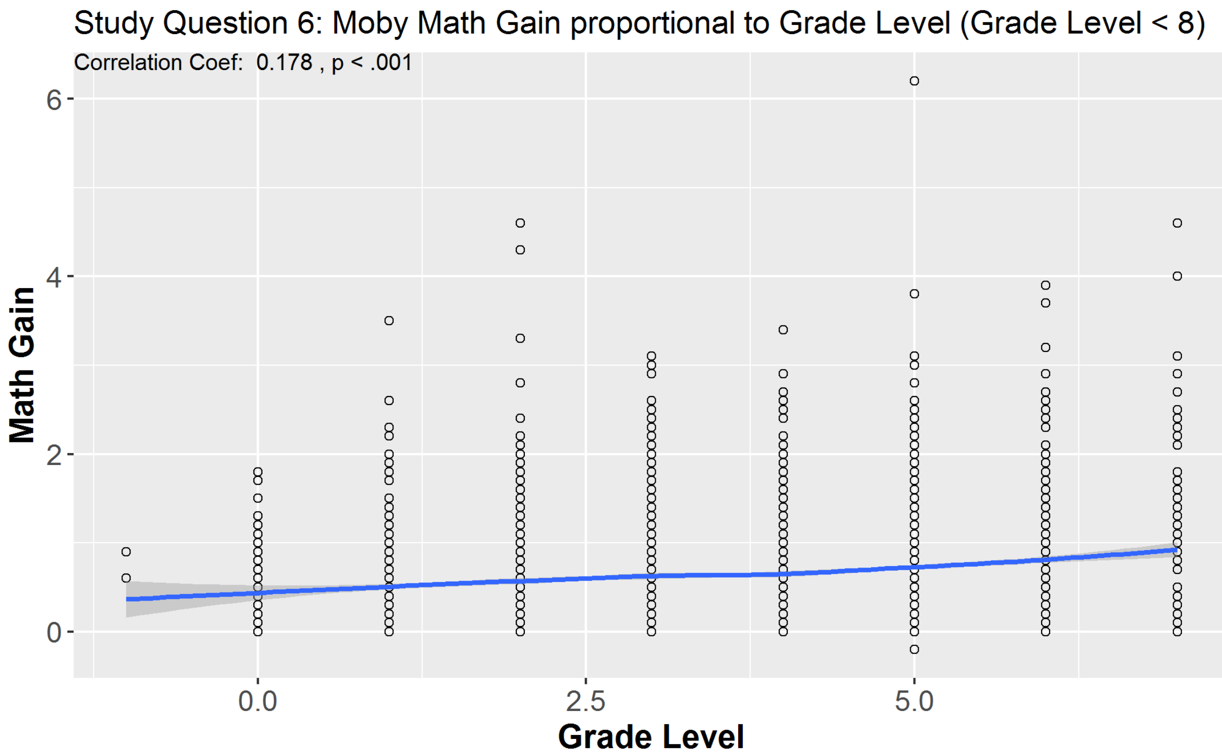
$\chi2=0.006 \cdot df=1 \cdot \varphi=0.002 \cdot p=0.937$

Appendix Table 1b: Randomization Design Descriptive statistics - Numerical Vars

Variable	*Control*	*Experimental*	*p-value*
MathBasal	5.75	5.65	0.32
Grade Level	3.38	3.40	0.67
Initial Score	3.03	3.03	0.97
PYLD	10.97	11.14	0.61

Baseline measurements between Experimental and Control groups indicates they are not statistically different at the $p < .1$ level

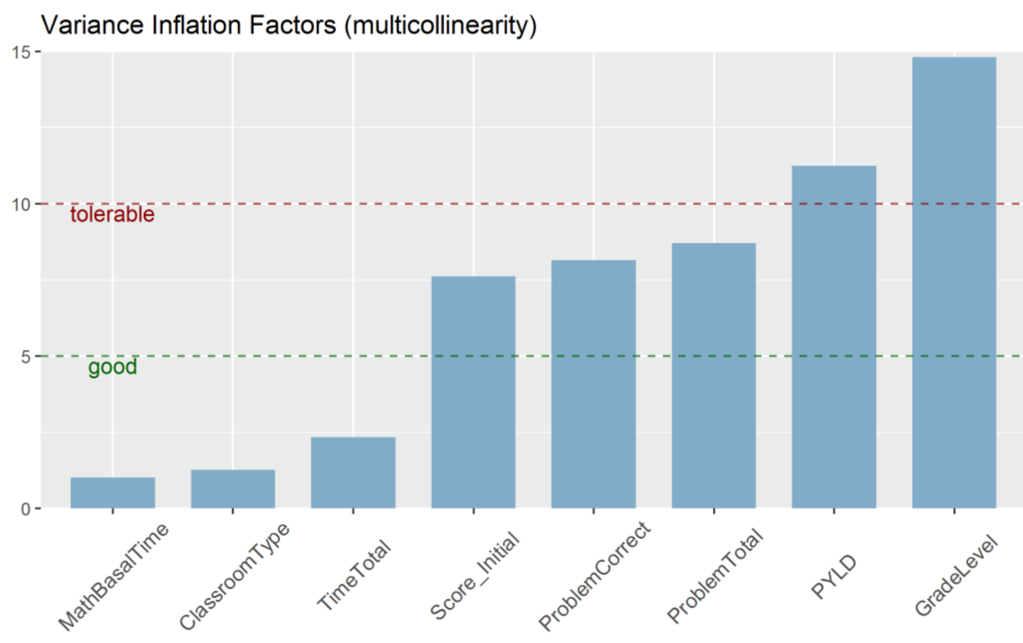
Appendix C: Study Question 6 with Grade Level < 8



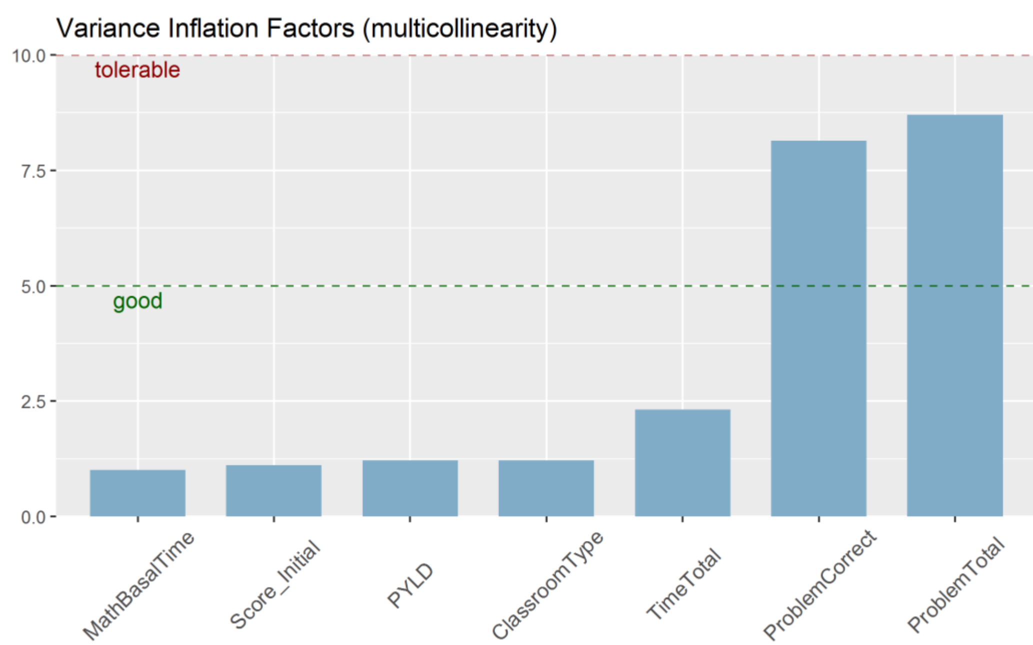
After removing students with a grade level greater or equal to 8 from the Study Question 6 analysis, the results change very little although the higher grade level loess-smoothing line exhibits much less dispersion. In terms of the decrease in the strength of the statistical association between Grade Level and Score gain this translates into going from an r^2 of ~3.5% $(.188)^2$ to < 3.2% $(.178)^2$.

Appendix D: Variance Inflation Factor Analysis

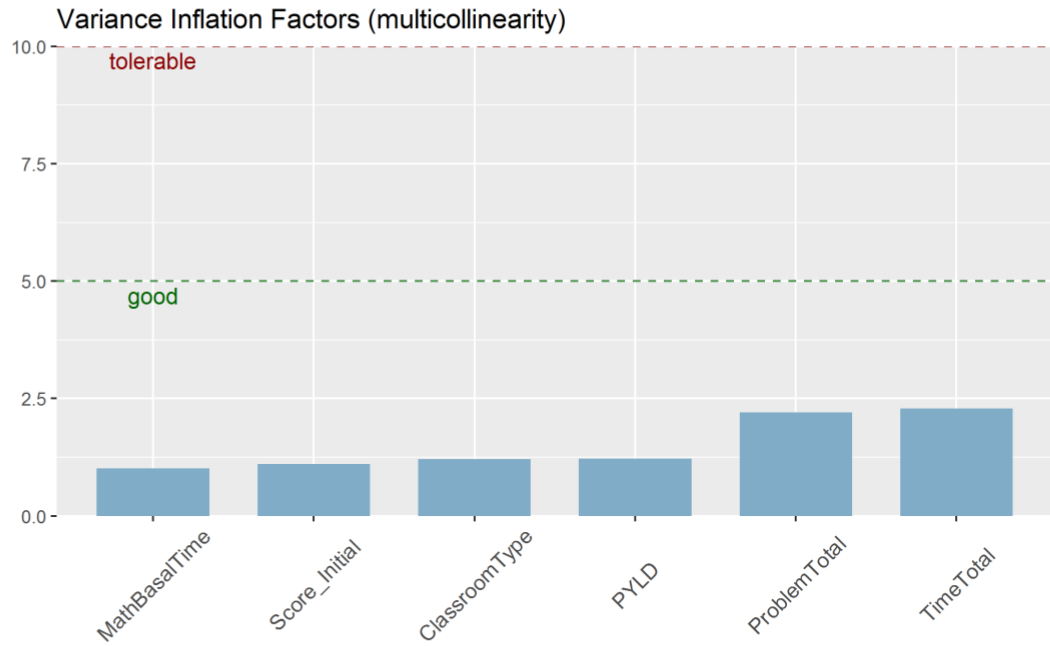
Initial VIF analysis with all variables included



VIF analysis with *GradeLevel* (variable with highest VIF) removed



VIF analysis with *ProblemCorrect* (variable with next highest VIF) removed[i]



[i] Although the graph indicates that *ProblemTotal* has the next highest VIF, other variations on the VIF test showed that, on average, *ProblemCorrect* had the higher VIF.