Reproducing *the* Results *of a* Published Trial

Research Design & Analysis Core
Version: 1.1 • 31 March 2022

Duke GLOBAL HEALTH INSTITUTE | Research Design & Analysis Core

Contents

Introduction

Reproducibility is crucial to ensuring the integrity of scientific research. Reproducibility means we will get the same results as the original study if we apply identical methods to the **same dataset.** For example, when independent researchers were unable to reproduce the results of research by Anil Potti, this led to retractions of papers and the halting of phase III cancer clinical trials which were exposing cancer patients to potentially harmful treatments that likely provided no additional benefits (Gewin, 2012; Baggerly, 2018). Partly due to such high profile cases of data fraud or error, it is becoming increasingly common for scientific journals to require analysis datasets (and, sometimes, the code from the statistical software) to be made publicly available as a supplement to published trial results, expressly for the purpose of allowing independent investigators to verify the authors' results.

Reproducibility is distinct from *replicability*. *Replication* is a fundamental principle in the accumulation of scientific evidence. If the results of a research study, such as a randomized controlled trial, are valid, then they should be able to be replicated by multiple independent investigators with independent samples. However, it can be expensive and time consuming to fully replicate a complex study.

The purpose of this guide is to walk through the reproduction of the published results from an individually randomized controlled trial aimed at evaluating the effects of a podcast intervention developed to help parents of primary school children in Uganda make beneficial health choices (Semakula et al. 2017). The raw data (in .xlsx format) for the study is available in the paper’s Supplementary Appendix 2. The “Informed Health Choices” paper was selected by virtue of being a recent global health trial, which made its data publicly available. This guide is not intended as an actual referendum on the scientific validity of the results published therein.

The article and accompanying data referenced and used throughout this guide can be accessed on the Lancet website by registering for free with an email address and password.

The current guide will follow the main results of the paper step-by-step, reproducing the results and explaining how they relate back to the primary scientific hypotheses that guided the study. Note that the paper contains a number of sub-group analyses (located in both the Results section and Supplementary Appendix 1); for simplicity, we restrict our reproduction to the main analyses for the two co-primary outcomes (mean score and passing score) and secondary outcome (mastery score), explained in the "Primary Objectives and Hypotheses" section below. In addition, we occasionally diverge from the published methods to present alternative ways of analyzing the data, which are explained in the **Boxes**. Code for reproducing these results in SAS, R, and Stata is given in **Core Guide Supplemental Material 1,** which can be accessed via the DGHI-RDAC GitHub account.

Accessing the Data

1. Create a free account on the Lancet website by registering with an email address and password
2. From the main podcast article, save the .xlsx file from the Supplementary appendix 2 and under the file name "mmc2_parents.xlsx". The pdf of the article and all supplemental material from the original publication can be found here.
3. From the corresponding primary school article, save the .xlsx file from the Supplementary appendix 2 under the file name "mmc2_children.xlsx"

Note: Both datasets are necessary for reproducing the analyses

Trial Background

The trial was designed to assess the effectiveness of the “Informed Health Choices” podcast, which was developed to help parents of primary school children in Uganda understand how to assess claims about the effectiveness of health treatments. See the Procedures sub-section of the paper (p. 391) for a detailed description of how the podcast was designed (in addition, the podcast is available online in the paper’s Supplementary Appendix 1). In short, the podcast consisted of 13 episodes, each 5-10 minutes in length, which were developed to address nine key health concepts identified a priori as important for the public to understand. Examples of health concepts include 1) effective treatments may have harmful side-effects and are not 100% safe and 2) personal experiences/anecdotes are an unreliable basis for assessing the actual effects of a treatment. A full description of the concepts is in Table 1 of the paper.

The eligible primary schools for this study were those enrolled in a separate but related cluster-randomized trial (Nsangi et al., 2017). The goal of that cluster trial was to evaluate the effects of an Informed Health Choices intervention (in the form of textbooks, exercise books, and other resources) on attitudes of children in the school. The individually randomized trial recruited parents whose children attended schools in the cluster trial (regardless of treatment assignment in that trial), and randomly assigned them to one of two groups: the podcast group, which delivered episodes of the podcast over a period of 7-10 weeks, or the control group, which delivered typical public service announcements covering the same basic concepts over the same time period. That is, the investigators were looking at the degree to which a dedicated teaching program, like the podcast, could improve people’s treatment appraisal skills above and beyond what would be expected if they had only listened to the widely available public service announcements.

The first step in reproducing an analysis is developing a Statistical Analysis Plan (SAP), if one does not already exist (e.g. provided by the original authors). In our case, an SAP was not provided in supplementary material; thus, an example SAP for the present analysis is given in Core Guide Supplemental Material 2. Note that this is just an example and should not be considered as a general template for a complete and sufficient SAP.

Primary Objectives and Hypotheses

The investigators were interested in evaluating the effects of the podcast on the ability of parents of primary school children to assess health related claims. To do so, they administered a test to the parents, and the resulting scores were used for the two co-primary outcome comparisons: the mean score (calculated as the percentage of correct answers) achieved on the test in each of the two study groups, and the proportion of participants in each study group with a passing score (defined as 11 or more correct test answers out of a total of 18). Each of these comparisons attempts to answer the scientific question: does listening to the podcast help prepare parents to make sound judgments about health related treatments, compared to parents who only listened to a series of typical public service announcements?

In addition, the investigators specified a secondary outcome comparison, the proportion of participants in each study arm with a score demonstrating that they achieved mastery of the key concepts evaluated by the test (defined as 15 or more correct test answers out of 18). This comparison also addresses whether or not parents who listened to the podcast developed a greater understanding of the key health concepts covered than parents who only listened to a series of typical public service announcements.

The statistical null hypothesis corresponding to each outcome is shown in the table provided on the next page.

Type of Outcome	Individual-Level Outcome	Group-Level Summary Statistic	Null Hypothesis
Co-Primary	Percentage: Each participant's score is the % of test questions answered correctly out of a total of 18 questions. (Note: this variable is treated as continuous in the analysis)	Mean score	The mean score among participants in the podcast group is equal to the mean score among participants in the control group.
Co-Primary	Binary: A participant has a passing score (1) if they answered 11 or more test questions correctly (0 otherwise).	Proportion of participants with a passing score	The proportion of participants who passed the test in the podcast group is equal to the proportion of participants who passed the test in the control group.
Secondary	Binary: A participant demonstrates master of key concepts (1) if they answered 15 or more test questions correctly (0 otherwise).	Proportion of participants demonstrating mastery of key concepts	The proportion of participants who demonstrated mastery of key concepts in the podcast group is equal to the proportion of participants who demonstrated mastery of key concepts in the control group.

Baseline & Demographic Characteristics

Before proceeding with any statistical analysis comparing our treatment groups, it is imperative that we scrutinize the distribution of baseline and demographic characteristics across these groups in order to ensure that they are comparable. If the groups have very different demographic profiles (e.g. the podcast group is mostly old men and the control group is mostly young women), it may call into question the *validity* of any comparisons between the groups. Even if there is no difference between the groups, it is important to understand the characteristics of the study population in order to properly interpret the results and, potentially, generalize them to a broader population.

In the paper, the baseline and demographic characteristics of the study sample are shown in Table 3 (p. 395). Since the data for individuals who were randomized but not included in the final sample were not made publicly available, we are only partially able to reproduce this table (i.e. only the “included” columns). This reproduction is shown in **Table R-1.1.**

The investigators not only tabulate the characteristics of the participants included in the final sample for each group, but also the characteristics of participants who were randomized to each group but dropped out of the study before any outcomes could be measured (see the study CONSORT diagram: Figure 1, p. 394 of the paper, not reproduced here). This allows them to look for potential bias in their results; that is, due to some details of the implementation of the intervention or the logistics of follow-up, were they disproportionately likely to measure outcomes on particular types of people?

Table R-1.1: Baseline and demographic characteristics by randomized group assignment. Data are presented as n (%).

	Control Group (n=273)	**Podcast Group (n=288)**
Took the test in Luganda[1]	237 (87%)	254 (88%)
Education (%)		
Primary Education/None	144 (53%)	145 (50%)
Secondary Education	68 (25%)	89 (31%)
Tertiary Education	61 (22%)	54 (19%)
Training in research* (%)	84 (31%)	96 (33%)
Prior participation in research† (%)	74 (27%)	72 (25%)
Sex (%)		
Female	208 (76%)	221 (77%)
Male	65 (24%)	67 (23%)
Sources of health care‡ (%)		
Government health facility	163 (60%)	177 (61%)
Private not-for-profit health facility	25 (9%)	32 (11%)
Private for-profit health facility	107 (39%)	93 (32%)
Alternative Medical practitioners	7 (3%)	8 (3%)
Advice about treatment[§] (%)		
Friends or relatives	77 (28%)	46 (16%)
Health workers	183 (67%)	236 (82%)
Community leaders	4 (1%)	6 (2%)
Radio or television programs	31 (11%)	19 (7%)
Alternative medicine practitioners	5 (2%)	8 (3%)
Internet	2 (1%)	3 (1%)

[1] Variable could not be found in the dataset – summary statistics were taken from the paper

***** "Have you ever had any training in scientific research (statistics, epidemiology, or randomized trials)?"

† "Have you ever been a participant in a scientific research study?"

‡ "If you or your family member are unwell, where do you commonly seek medical attention (select all that apply)?"

[§] "If you need to make a decision on what treatments to use, where do you usually get advice (select all that apply)?"

For example, Table 3 tells us that (in both groups) individuals included in the final sample were more likely to have training in scientific research compared to individuals who dropped out. This tells us our outcomes are all measured on individuals with a higher degree of education compared to the general population, and thus we are likely overestimating the degree to which the general population has beneficial health knowledge. However, this difference does not appear to be different by study group, which gives us more confidence that any treatment effect estimate comparing the groups is not confounded by education.

BOX A: Comparing “included” and “dropped out” participants

As we saw, in Table 3, the investigators tabulate the characteristics of “included” and “dropped out” participants separately within each study group. This is useful for looking for patterns of differential drop-out by study arm (e.g. are control group males more likely to drop out then podcast group males?). In **Table R-1.1,** we show only the characteristics of “included” individuals by study group. This is useful for assessing the degree to which the two study groups are comparable in our analysis. Alternatively, we can compare the characteristics of “included” and “dropped out” participants collapsed across study groups (see **Table R-1.2**). Once we’ve established that there are no differential patterns between groups, the collapsed table is useful for looking for baseline and demographic characteristics associated with drop out.

Results—Mean Differences

A simple way of comparing outcomes between groups is simply to compare the means. These are reported in the first two columns (for control and podcast groups, respectively) of Table 4 of the paper (p. 395), for both co-primary outcomes (i.e. mean score and passing score) as well as the secondary outcome (mastery score). In the third column, the authors also report an "Adjusted difference" between the means, with a 95% confidence interval. The footnote indicates that the "adjustment" is for parental education level and the child's school group in the corresponding primary school trial. We can see that the adjusted difference in mean scores (podcast group – control group) is 15.5%, with a confidence interval (CI) from 12.5% to 18.6%. This means, conditional on their level of education and the child's school group, parents who listened to the podcast achieved, on average, a score 15.5 percentage points higher than parents who only listened to the public service announcements.

Similarly, the adjusted difference in passing scores is 34%, with a confidence interval from 26% to 41%; that is, considering the same adjustments, the percentage of parents who passed the test after listening to the podcast was 34 percentage points higher than for parents who listened to the standard public service announcements. Finally, the adjusted difference in mastery scores is 26% (CI: 15%-39%), indicating that, adjusted for parental education level and child's school group, the percentage of parents who mastered the test after listening to the podcast was 26 percentage points higher than for parents who listened to the public service announcements. **Table R-2** on the next page reproduces the paper's Table 4, our focus in this section has been solely on the first three columns.

Table R-2: Summary of outcomes in each group, with adjusted mean differences & adjusted odds ratios (with 95% CIs) using parental education & child's school group as the adjustment factors.

Outcome	Control Group (N=273)	Podcast Group (N=273)	Adjusted Difference	Adjusted Odds Ratio	P-Value
Mean Score Mean (S.D.)	52.4% (17.6)	67.8% (19.6)	15.5% (12.5 - 18.6)	…	<0.0001
Passing Score N(%)	103 (38%)	203 (70%)	34% (26 - 41)	4.2 (2.9 - 6.0)	<0.0001
Mastery Score N (%)	17 (6%)	91 (32%)	26% (15 - 39)	7.2 (4.1 - 12.5)	<0.0001

BOX B: Unadjusted Mean Differences

To fully understand the "adjusted" differences, however, we should take a step back. What exactly are the "adjusted" differences, and how do they compare to the "unadjusted" differences? **Table R-3** reproduces the first three columns of the paper's Table 4, but adds in a column for the unadjusted difference in the means for each group. For mean scores, we see the unadjusted difference in means (podcast group-control group) is 15.4%, with a confidence interval from 12.3% to 18.5%. This is very similar to the adjusted difference. Similarly, for both other outcomes, we see that the unadjusted differences and the adjusted differences look almost identical to one another. This implies that we do not have evidence to show that parents with lower or higher education nor children of the different school groups were differentially impacted by listening to the podcast.

How were the adjusted mean differences calculated? For mean score, they were calculated by fitting a multiple *linear regression* with mean score as the outcome and, as predictors, an indicator variable for assignment to the podcast group, an indicator variable for whether the parent's child was attending an intervention school from the associated cluster-randomized trial (see "Trial Background" section, above), and a three-level categorical variable for parental education level (primary, secondary, tertiary); the regression coefficient estimate for the podcast group indicator variable from this model is the adjusted mean difference between podcast and control groups.

For passing and mastery score, the outcomes are binary. Therefore, the investigators used a *logistic regression* model, similar to the *linear regression* model described above. As before, the model included a single binary indicator variable for membership in the podcast group as well as a categorical variable for parental education and an indicator variable for school group membership . Now, instead of mean score, the outcomes are passing and mastery score. The Statistical Analysis section of the paper (p. 393) describes the method used to derive the adjusted differences from the *logistic regression* model*:*

> "We converted ORs from the logistic regression analyses to risk differences using the control group odds as the reference, multiplying these odds by the OR to estimate the intervention group odds, and converting the control and intervention group odds to proportions to calculate difference. We calculated the adjusted standardized mean difference (Hedges' g) for comparison to effect sizes reported in meta-analyses."

We follow this method to fully reproduce their results. However, we note that an adjusted risk difference derived from a generalized linear model with a binomial error structure and an identity link could be an alternative method here.

BOX C: Hypothesis Test for the Difference in Means

In the paper, the analysis for mean score was based on the adjusted mean differences derived from a *linear regression* model, as discussed in the text. The analyses for passing and mastery scores, meanwhile, were based on adjusted odds ratios derived from a *logistic regression* model. We may instead be interested in unadjusted comparisons on these outcomes; that is, in applying hypothesis tests to the unadjusted differences between groups.

For mean score, a natural method for such a comparison would be a *two-sample t-test*. This tests the null hypothesis that the mean score in the podcast group is the same as the mean score in the control group (equivalently, it tests the null hypothesis that the difference in means between the groups is equal to 0). The results of this t-test are shown in **Table R-4.**

For passing and mastery scores, we could use the analogue of the t-test for proportion data, the *two-sample Z-test of proportions*. This tests the null hypothesis that the proportion in the podcast group is the same as the proportion in the control group (equivalently, that the difference in proportions between groups is equal to 0). Alternatively, these data could be analyzed as a contingency table, with the counts of individuals with/without a passing/mastery score in each group, using *Pearson's chi-square test of independence.* This tests the null hypothesis that there is no association between the classification variables (i.e. achieving a passing score and being in the podcast group). These results are summarized in **Table R-4.**

In this case, you can see that all of the tests reject the null hypothesis at the 0.05 level.

Results—Odds Ratios

For the mean score outcome, the adjusted mean differences constituted the final reported analysis. However, for the other two outcomes (passing and mastery score), the final analysis was based on an adjusted odds ratio, calculated from the *logistic regression* model described above. Instead of the transformation of the results used to calculate the adjusted mean differences, the odds ratios are based on the regression coefficient estimates directly. Where before (in the *linear regression*) the regression coefficient estimate for our indicator variable was interpreted as the adjusted mean difference between groups, now the regression coefficient estimate for our indicator variable is interpreted as the difference in the adjusted *log-odds* of achieving a passing/mastery score for individuals in the podcast group compared to the control group. We exponentiate this coefficient to get our adjusted *odds ratio*.

The results from this model are presented as the "Adjusted odds ratios" (with corresponding 95% CIs) in the second-to-last column of Table 4 in the paper (p. 395). The reproduced adjusted odds ratios are presented in **Table R-2** in the previous section. For comparison, as a parallel to our treatment of the mean differences in the previous section, we also present the unadjusted odds ratios for each outcome in **Table R-5**. These can be calculated either by fitting another *logistic regression* model where we exclude the adjustment factors (parental education, child's school group allocation) and exponentiating the resulting coefficient, or by analyzing the data as a contingency table and calculating the odds ratio directly. See **Core Guide Supplemental Material 3** for an example contingency table with the data from the study.

The adjusted odds ratios comparing the podcast to control group are 4.2 (with 95% CI: 2.9-6.0) for passing scores and 7.2 (with 95% CI: 4.1-12.5) for mastery scores. This tells us that the odds of a parent in the podcast group achieving a passing score on the test are estimated to be 4.2 times higher than for a parent in the control group. Similarly, the odds of a parent in the podcast group achieving a mastery score on the test are estimated to be 7.2 times higher than for a parent in the control group. As we saw with our mean differences, there is very little difference between the adjusted and unadjusted results, leading us to believe that parental education and child's school group does not significantly impact the association between group assignment and the odds of achieving a passing/mastery score.

BOX D: Alternatives to the Odds Ratio

Odds ratios are one of the most common effect size measures used when analyzing binary data, in part due to it being the natural interpretation of the exponentiated co-efficients from the *logistic regression model.* However, a major drawback in using the odds ratio is the difficulty in interpreting what it means. The ratio of two odds is not intuitive, with odds themselves already being defined as the ratio of the probability of an event occurring to its complement (i.e. the probability of the event not occurring).

BOX D: Alternatives to the Odds Ratio (Continued)

More intuitively, we may like to know "how many times more likely is this event to occur than to not occur?" This effect measure is known as the *relative risk*, and is the ratio of the probabilities of the event occurring in each group. Like odds ratios, an unadjusted relative risk can be calculated from a contingency table (see **Core Guide Supplemental Material 3**). In order to derive adjusted *relative risks*, instead of using a *logistic regression* model we fit a *log-binomial regression* model (i.e. we use a log link function instead of a logit link). In this model, the exponentiated regression coefficient is interpreted as a *relative risk.*

A drawback to odds ratios and relative risks is that they are only *relative* measures of effect. They only give us an idea of how well one group is doing relative to another. We may be more interested in an *absolute* measure of effect, analogous to the mean differences we calculated before. For binary outcomes, an absolute effect measure is the *risk difference.* It is simply the difference in the probabilities of the event occurring in each group. One way to calculate the adjusted *risk difference* is through a generalized linear model with a binomial error structure and an identity link. Note that the adjusted *risk difference* is not equivalent to our adjusted difference in **Table R-3.** Unadjusted and adjusted *odds ratios*, *relative risks*, and *risk differences* are shown in **Table R-5.**

See Gallis & Turner (2019) for a more detailed treatment of this subject.

Summary

We have shown how to reproduce the primary results of this trial by using the publicly available data. And we have highlighted issues that can arise in reproducing results when insufficient detail regarding statistical methods are reported in the manuscript. We have also examined several alternative analysis methods that could have been used by the investigators to answer their scientific question. In all cases, the results were not sensitive to the choice of analytic method. Whether we choose to examine the unadjusted mean differences between groups or to fit a regression model adjusting these differences by potentially pertinent demographic characteristics, we reach the same general conclusion: the parents that listened to the podcast, on average, performed better than parents who only listened to the typical public service announcements. This finding is consistent across both co-primary outcomes and the secondary outcome, and persists across several measures of absolute and relative effect. This is an important step in demonstrating the reproducibility of published results.

References

Baggerly K. What ‘data thugs’ really need. *Nature.* 2018. doi:10.1038/d41586-018-06903-2

Gallis JA and Turner EL. Relative measures of association for binary outcomes: challenges and recommendations for the global health researcher. *Annals of Global Health*. 2019;85(1):137. doi:10.5334/aogh.2581

Gewin, V. Research: Uncovering misconduct. *Nature.* 2012; 485: 137-139. doi:10.1038/nj7396-137a

Nsangi A, Semakula D, Oxman AD, et al. Effects of the Informed Health Choices primary school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects: a cluster-randomised controlled trial. *Lancet*. 2017;390(10092):374-388. doi:10.1016/S0140-6736(17)31226-6

Semakula D, Nsangi A, Oxman AD, et al. Effects of the Informed Health Choices podcast on the ability of parents of primary school children in Uganda to assess claims about treatment effects: a randomised controlled trial. *Lancet*. 2017;390(10092): 389-398. doi:10.1016/S0140-6736(17)31225-4

Prepared by: Rachel D’Anna, Ryan Simmons, and John Gallis
Reviewed by: Duke Global Health Institute | Research Design & Analysis Core
Photo by: David Brooke Martin @dbmartin00 via Unsplash

Referenced Tables

Table R-1.2 : Baseline & demographic characteristics of the included participants compared to those who dropped-out, collapsed across randomized group assignment. Data are presented as n (%). The numbers in the "Dropped out" column are taken directly from Table 3 of the pa-

	Included (n=561)	**Dropped out[1] (n=114)**
Took the test in Luganda[2]	491 (87.5%)	101 (88.6%)
Education (%)		
Primary Education/None	289 (51.5%)	55 (48.2%)
Secondary Education	157 (28.0%)	38 (33.3%)
Tertiary Education	115 (20.5%)	21 (18.4%)
Training in research* (%)	180 (32.1%)	18 (15.8%)
Prior participation in research† (%)	146 (26.0%)	19 (16.7%)
Sex (%)		
Female	429 (76.5%)	90 (78.9%)
Male	132 (23.5%)	24 (21.1%)
Sources of health care‡ (%)		
Government health facility	340 (60.6%)	76 (66.7%)
Private not-for-profit health facility	57 (10.1%)	19 (16.7%)
Private for-profit health facility	200 (35.7%)	54 (47.3%)
Alternative Medical practitioners	15 (2.7%)	3 (2.6%)
Advice about treatment[§] (%)		
Friends or relatives	123 (21.9%)	58 (50.9%)
Health workers	419 (74.7%)	99 (86.8%)
Community leaders	10 (1.8%)	5 (4.4%)
Radio or television programs	50 (8.9%)	34 (29.8%)
Alternative medicine practitioners	13 (2.3%)	3 (2.6%)
Internet	5 (0.9%)	3 (2.6%)

[1] Public data does not contain dropped-out participants, summary statistics taken from paper.

[2] Variable could not be found in the dataset – summary statistics taken from the paper.

* "Have you ever had any training in scientific research (statistics, epidemiology, or randomized trials)?"

† "Have you ever been a participant in a scientific research study?"

‡ "If you or your family member are unwell, where do you commonly seek medical attention (select all that apply)?"

[§] "If you need to make a decision on what treatments to use, where do you usually get advice (select all that apply)?"

Table R-3: Summary of outcomes in each group, with unadjusted and adjusted differences. Differences for mean score derived from linear regression model, differences for passing score and mastery score derived from converted ORs from the logistic regression. Adjusted differences use parental education and child's school group as the adjustment factors.

Outcome	Control Group (N=273)	Podcast Group (N=273)	Unadjusted Difference (95% CI)	Adjusted Difference (95% CI)
Mean Score Mean (S.D.)	52.4% (17.6)	67.8% (19.6)	15.4% (12.3 - 18.5)	15.5% (12.5 - 18.6)
Passing Score N(%)	103 (38%)	203 (70%)	33% (25 - 40)	34% (26 - 41)
Mastery Score N (%)	17 (6%)	91 (32%)	25% (15 - 38)	26% (15 - 39)

Table R-4: Results of hypothesis tests for the mean differences between groups

Outcome	Unadjusted Mean Difference (95% CI)	Hypothesis Test	Test Statistic (DF)	P Value
Mean Score	15.4% (12.3 - 18.5)	Two-sample t-test	9.77 (559)	<0.0001
Passing Score	33% (25 - 40)	Two-sample Z-test	7.88 (559)	<0.0001
		Chi-square test	60.6 (1)	<0.0001
Mastery Score	25% (15 - 38)	Two-sample Z-test	7.62 (559)	<0.0001
		Chi-square test	58.0 (1)	<0.0001

Table R-5: Summary of different unadjusted and adjusted effect measures for passing and mastery score outcomes. Unadjusted odds ratios, relative risks, and risk differences are all calculated from contingency tables. Adjusted odds ratios, relative risks, and risk differences are all calculated from the models described in **Box D**, using parental education and child's school group as the adjustment factors.

	Unadjusted			**Adjusted**		
Outcome	**Odds Ratio**	**Relative Risk**	**Risk Difference**	**Odds Ratio**	**Relative Risk**	**Risk Difference**
Passing Score	3.9 (2.8 – 5.6)	1.9 (1.6 – 2.2)	32.8% (25.0 – 40.6)	4.2 (2.9 – 6.0)	1.9 (1.5 – 2.4)	32.8% (25.1 – 40.6)
Mastery Score	7.0 (4.0 – 12.1)	5.1 (3.1 – 8.3)	25.4% (19.3 – 31.5)	7.2 (4.1 – 12.5)	5.1 (3.1 –8.6)	25.4% (19.4 – 31.4)