Comparability of Online and Paper/Pencil Mathematics Performance Measures

John L. Moon
University of Nebraska-Lincoln, john.moon@nebraska.gov

Follow this and additional works at: http://digitalcommons.unl.edu/cehsdiss

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Psychology Commons, and the Science and Mathematics Education Commons
Comparability of Online and Paper/Pencil Mathematics Performance Measures

By

John L. Moon

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Psychological Studies in Education
(Quantitative, Qualitative, & Psychometric Methods)

Under the Supervision of Professor Delwyn L. Harnisch

Lincoln, Nebraska

May, 2013
Comparability of Online and Paper/Pencil Mathematics Performance Measures

John Lee Moon, Ph.D.
University of Nebraska, 2013

Advisor: Delwyn L. Harnisch

The purpose of this study was to investigate the relationship between student mathematics performance of 4th, 8th, and 11th grade students in Nebraska and the mode of test administration, online and paper-pencil. Schools were allowed to select the mode of test administration for their school with some exceptions for students needing accommodations. This resulted in four test groups, namely students taking the online tests in schools selecting paper or online assessments along with students taking the paper-pencil tests in schools selecting paper or online assessments.

Since the students in the study were clustered within schools, the data from the Nebraska State Accountability (NeSA) mathematics assessment (2010) were analyzed using hierarchical linear modeling (HLM) for the three grade level groups. The use of HLM allowed the researcher to adjust for and model the dependence of students clustered within schools. Both school level and student level variables were included in the model to control for sample differences between test modes. Student variables such as gender, students eligible for free and reduced lunch (FRL), students receiving special education services (SPED), English language learners eligible for support (ELL), and the seven ethnicities were incorporated in all three grade level models along with school level percents of FRL, SPED, ELL, and ethnicity.
In each of the three grades, the results failed to indicate a significant effect on mathematics performance between students taking paper-pencil tests in schools selecting paper-pencil assessments (P/P) and students taking online tests in schools selecting online assessments (O/O) ($p = 0.0518$). Significant differences were noted between results for P/P test takers and the results for those students taking paper-pencil tests in schools selecting online assessments (P/O) ($p < 0.0001$). Likewise significant differences were found in results of students taking P/P tests and the results of students taking online tests in schools selecting paper assessments (O/P) ($p < 0.0001$).

The state policy makers have considered the expansion of online testing throughout the state. The advantages of computer testing with the assurance of comparable performance on both test modes need to be considered before moving forward with online assessments in Nebraska schools.
Acknowledgements

I would like to express my appreciation to my advisor and committee chair, Dr. Delwyn L. Harnisch, for his support and guidance through this very long process. Without his encouragement, I would not have finished. I thank him for the many hours he put into this process.

To my committee members, Dr. Charles Ansorge, Dr. Jody Isernhagen, and Dr. David Fowler, I wish to express my thanks for their time, insights, and helpful comments that they made on this project and on the final document.

To the Nebraska Department of Education assessment team, I would like to thank them for their moral support and encouragement. Finally, I thank my wife, Patricia, for her patience during the many weekend hours and her confidence in me. Without the support from each of you, the completion of this dissertation would not have been possible.
# Table of Contents

Chapter 1—Introduction ............................................................................................................. 1  
  Context ................................................................................................................................. 1  
  Purpose of Study .................................................................................................................. 2  
  Research Questions ............................................................................................................ 3  
  Theoretical Perspective ..................................................................................................... 4  
  Assumptions ....................................................................................................................... 4  
  Definition of Terms .......................................................................................................... 5  
  Delimitations ..................................................................................................................... 7  
  Limitations ......................................................................................................................... 8  
  Significance of Study ....................................................................................................... 8  

Chapter 2—Literature Review ............................................................................................... 9  
  Meta-analysis—K-12 Comparability Studies ..................................................................... 9  
  Specific Grade Level Comparability Studies ................................................................ 11  
  Test Comparability for Nebraska Assessments ............................................................. 16  

Chapter 3—Methods ............................................................................................................ 18  
  Research Design .............................................................................................................. 18  
  General Design Issues .................................................................................................... 21  
  Instrumentation ............................................................................................................... 22  
  Sampling .......................................................................................................................... 24  
  Administration ................................................................................................................ 25  
  Scoring ............................................................................................................................ 26  

Chapter 4—Results ............................................................................................................ 27  
  Introduction Hierarchical Linear Models .................................................................... 27  
  Variables and Coding for Analysis .............................................................................. 28
HLM Analysis Process ........................................................................................................ 36

HLM Analysis—Step One—Empty Model ........................................................................ 36

HLM Analysis—Step Two—Add Student and School Administrative Mode Variables ........................................... 37

HLM Analysis—Step Three—Add Level 1 Explanatory Variables .................................. 39

HLM Analysis—Step Four—Add Level 2 Explanatory Variables and Level 1 Interactions with Student Administrative Mode ................................................................. 40

Analysis of Model Fit for Grade 4 .................................................................................. 43

Step 1: Grade 4—Empty model—Calculation of ICC ...................................................... 43

Step 2: Grade 4 Model—School and Student Administrative Mode Effects Only .......... 45

Step 3: Grade 4 Model—Addition of Student and School Explanatory Variables—Control for Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race ................................................................. 48

Step 4: Grade 4 Model—Addition of Interaction of Student Administrative Mode and Student Level Variables—Effect of Mode on Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race ................................................................. 54

Analysis of Model Fit for Grade 8 .................................................................................. 58

Step 1: Grade 8—Empty Model—Calculation of ICC ...................................................... 58

Step 2: Grade 8 Model—School and Student Administrative Mode Effects Only .......... 60

Step 3: Grade 8 Model—Addition of Student and School Explanatory Variables—Control for Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race ................................................................. 63

Step 4: Grade 8 Model—Addition of Interaction of Student Administrative Mode and Student Level Variables—Effect of Mode on Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race ................................................................. 70

Analysis of Model Fit for Grade 11 ................................................................................. 74
Step 1: Grade 11—Empty Model—Calculation of ICC ........................................ 74

Step 2: Grade 11 Model—School and Student Administrative Mode Effects Only ................................................................................. 75

Step 3: Grade 11 Model—Addition of Student and School Explanatory Variables—Control for Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race .......................................................... 78

Step 4: Grade 11 Model—Addition of Interaction of Student Administrative Mode and Student Level Variables—Effect of Mode on Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race .......................................................... 85

Chapter 5—Discussion ........................................................................................................ 91

Introduction ......................................................................................................................... 91

Model Review ..................................................................................................................... 91

Proportion of Variance Explained ..................................................................................... 97

Effects of School Level Variables ..................................................................................... 101

Subgroup Performance by Mode ....................................................................................... 103

Limitations of Study .......................................................................................................... 107

Further Research .............................................................................................................. 108

Conclusions and Implications .......................................................................................... 110

References ......................................................................................................................... 113

Appendices ........................................................................................................................ 117
List of Tables

Table 3.1  Number of Students Tested on NeSA-M in 2011 by Grade ............ 23
Table 4.1  Four Testing Groups by Student/School Administrative Mode ................................................................. 29
Table 4.2  Dummy Codes for Ethnicity/Race Variables.................................. 30
Table 4.3  Descriptive Statistics for the Variables in Both Levels—Grade 4 ................................................................. 32
Table 4.4  Descriptive Statistics for the Variables in Both Levels—Grade 8 ................................................................. 33
Table 4.5  Descriptive Statistics for the Variables in Both Levels—Grade 11 ................................................................. 35
Table 4.6  Fit Statistics Fixed Effects and Random Effects of Empty Model for Grade 4 Mathematics Achievement ................. 44
Table 4.7  Fixed Effects and Random Effects of Empty Model for Grade 4 Mathematics Achievement .................................. 45
Table 4.8  Summary of Fit Statistics for Fixed Effects and Random Effects of Admin Mode Models for Grade 4 Mathematics Achievement ................................................................. 46
Table 4.9  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Explanatory Variables for Administrative Mode in Grade 4 Mathematics Achievement ................................................................. 47
Table 4.10 Summary of Fit Statistics for Fixed Effects and random Effects of Level 1 and 2 Explanatory Variables for Grade 4 Mathematics Achievement ................................................................. 49
Table 4.11 Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Effect of Student Explanatory Variables for FRL, Gender, SPED, ELL and Ethnicity as well as School Level Explanatory Variables for Percent FRL, and Percent White on Grade 4 Mathematics Achievement ................................................................. 51
Table 4.12 Summary of Fit Statistics for Addition of Interactions with Administrative Mode for FRL, Gender, SPED, and ELL for Grade 4 Mathematics Achievement ................................................................. 55
Table 4.13  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Previous Model with Effect of interactions Between Student Administrative Mode and Student Explanatory Variables for FRL, Gender, SPED, and ELL on Grade 4 Mathematics Achievement ........................................... 56

Table 4.14  Fit Statistics Fixed Effects and Random Effects of Empty Model for Grade 8 Mathematics Achievement ........................................... 59

Table 4.15  Fixed Effects and Random Effects of Empty Model for Grade 8 Mathematics Achievement .......................................................... 59

Table 4.16  Summary of Fit Statistics for Fixed effects and Random Effects for Admin Mode Models for Grade 8 Mathematics Achievement .......................................................... 61

Table 4.17  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Effect of Student Explanatory Variables for Administrative Mode in Grade 8 Mathematics Achievement .......................................................... 62

Table 4.18  Summary of Fit Statistics for Fixed Effects and Random Effects of Level 1 and 2 Explanatory Variables for Grade 8 Mathematic Achievement .......................................................... 64

Table 4.19  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Effect of student Explanatory Variables for FRL, Gender, SPED, ELL and Ethnicity as well as School Level Explanatory Variables for Percent FRL, Percent SPED, and Percent White on Grade 8 Mathematics Achievement .......................................................... 66

Table 4.20  Summary of Fit Statistics for Addition of Interactions with Administrative Mode with FRL, Gender, SPED, and ELL for Grade 8 Mathematics Achievement .......................................................... 72

Table 4.21  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Previous Model with Effect of Interactions Between Student Administrative Mode and Student Explanatory Variables for FRL, Gender, SPED, and ELL on Grade 8 Mathematics Achievement .......................................................... 73

Table 4.22  Fit Statistics Fixed Effects and Random Effects of Empty Model for Grade 11 Mathematics Achievement .......................................................... 74

Table 4.23  Fixed Effects and Random Effects of Empty Model for Grade 11 Mathematics Achievement .......................................................... 75
Table 4.24  Summary of Fit Statistics for Fixed Effects and Random Effects of Admin Mode Models for Grade 11 Mathematics Achievement ................................................................. 76

Table 4.25  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Explanatory Variables for Administrative Mode in Grade 11 Mathematics Achievement................................................................. 77

Table 4.26  Summary of Fit Statistics for Fixed Effects and Random Effects of Level 1 and 2 Explanatory Variables for Grade 11 Mathematics Achievement ................................................................. 79

Table 4.27  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Effect of Student Explanatory Variables for FRL, Gender, SPED, ELL and Ethnicity as well as School Level Explanatory Variables for Percent FRL, Percent SPED, and Percent White on Grade 11 Mathematics Achievement................................................................. 81

Table 4.28  Summary of Fit Statistics for Addition of Interactions with Administrative Mode with FRL, Gender, SPED, and ELL for Grade 11 Mathematics Achievement ................................................................. 86

Table 4.29  Estimates of Mathematics Achievement for the Four Administrative Groups Using the Previous Model with Effect of Interactions Between Student Administrative Mode and Student Explanatory Variables for FRL, Gender, SPED, and ELL on Grade 11 Mathematics Achievement................................................................. 86

Table 5.1  Summary of Grade 4 Mathematics Performance Estimates for Student/School Administrative Mode ................................................................. 94

Table 5.2  Summary of Grade 8 Mathematics Performance Estimates for Student/School Administrative Mode ................................................................. 94

Table 5.3  Summary of Grade 11 Mathematics Performance Estimates for Student/School Administrative Mode ................................................................. 95

Table 5.4  Proportion Variance Reduction comparison (Pseudo R²) of Two-Level Empty Model and Final Model for Grade 4................................. 98

Table 5.5  Proportion Variance Reduction comparison (Pseudo R²) of Two-Level Empty Model and Final Model for Grade 8................................. 98

Table 5.6  Proportion Variance Reduction comparison (Pseudo R²) of Two-Level Empty Model and Final Model for Grade 11................................. 99
<table>
<thead>
<tr>
<th>Table 5.7</th>
<th>Proportion Variance Reduction Comparison (Pseudo R$^2$) of Model with Interaction and Final Model for Grade 4</th>
<th>104</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 5.8</td>
<td>Proportion Variance Reduction Comparison (Pseudo R$^2$) of Model with Interaction and Final Model for Grade 8</td>
<td>105</td>
</tr>
<tr>
<td>Table 5.9</td>
<td>Proportion Variance Reduction Comparison (Pseudo R$^2$) of Model with Interaction and Final Model for Grade 11</td>
<td>106</td>
</tr>
</tbody>
</table>
List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>Table Showing Mathematics Performance of Grades 4, 8, and 11</td>
<td>117</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Table Showing Deviance Comparisons and Pseudo R²</td>
<td>121</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Final Programs for Grade 4, 8, and 11 Models</td>
<td>138</td>
</tr>
<tr>
<td>Appendix D</td>
<td>IRB Approval Letter</td>
<td>142</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Context

As required in the amended Quality Education Accountability Act, 79-760, the Nebraska Department of Education (NDE) has built Nebraska State Accountability (NeSA) assessments to measure reading and mathematics achievement in grades 3, 4, 5, 6, 7, 8, and 11 for the purposes of accountability (Standards, Assessment, and Accountability Update (SAA-6), 2010). For mathematics, a transition from the district-wide criterion-referenced assessment system to a balanced statewide approach began in 2009-2010 with the development and pilot of the first statewide assessment. With the help of their partners, Data Recognition Corporation (DRC) out of Maple Grove, Minnesota, and Computer Assisted Learning (CAL) from Lawrence, Kansas, NDE has involved Nebraska teachers in the design and development of the NeSA mathematics test. In 2011, the NeSA mathematics assessments were administered in both online and paper/pencil modes to approximately 148,000 students including students receiving English Language Learner (ELL) and Special Education (SPED) support. NDE reported the results of the NeSA mathematics assessments to Nebraska educators and the public through the State of the Schools Report (SOSR) for all seven grades.

For eight years Nebraska educators perfected a local assessment system aligned to standards that used an array of assessments administered either through computer or with paper-pencil. Those wanting to be able to compare scores on tests used by different districts were faced with relatively unique assessments in each district. The local systems usually did not mix the modes of administration. This system was entitled STARS:
School-based Teacher-led Assessment and Reporting System. In the eighth annual report of the Nebraska Comprehensive Evaluation Project (CEP) the authors interviewed teachers and administrators about the new statewide tests (Isernhagen & Mills, 2009). In the interviews, Isernhagen and Mills found that “educators shared both hope and apprehension about this new assessment” (Isernhagen & Mills, 2009, p. 40). The new Nebraska assessment system offered both modes of administration, online and paper-pencil. It is evident that there is a need to determine whether student performance on the NeSA Mathematics assessment is affected by mode of administration, online versus paper-pencil. Pommerich (2004) offered that it is important for testing programs to conduct comparability studies based on their own tests and technology provided for online measures since findings from previous studies cannot be generalized to similar situations. This study was approved by the Institutional Review Board (IRB) at the University of Nebraska-Lincoln (UNL) indicating adequate safeguards for the rights and welfare of the participants in this study (Appendix D).

**Purpose of Study**

The comparability study investigated the difference in performance on the Nebraska State Accountability (NeSA) mathematics assessment administered through an online web-based process or one utilizing paper-pencil materials. For NeSA assessments, schools leaders selected the assessment mode for students within their respective schools. Purpose of the study was to establish the equity of the two administrative assessment modes and assure schools that student performance results are equivalent whether online or paper-pencil. Equivalence of online and paper-pencil administrative modes must be
established before scores from either mode can be used interchangeably with scores from the other mode.

**Research Questions**

The current study of comparability between assessment administrative modes addressed four research questions. Each question examined how student mathematics performance differed for students taking paper based assessments perform when compared to similar students taking computer based assessments in either schools selecting paper testing or online testing. The study controlled for significant effects by student level variables such as gender, free and reduced lunch (FRL), English language learners (ELL), special education (SPED), and race/ethnicity in estimating mathematics performance on the statewide mathematics assessment. In addition significant school level variables such as percent of students receiving free and reduced lunch or percent of white students were controlled in estimating mathematics performance on the statewide mathematics assessment. The purpose of this study is to answer the following questions:

- After controlling for significant student level and school level variables, to what extent do the mathematics performance scores for students taking paper assessments in schools selecting paper testing differ from scores for students taking online assessments in schools selecting online testing?
- After controlling for significant student level and school level variables, to what extent do the mathematics performance scores for students taking paper assessments in schools selecting paper testing differ from scores for students taking paper assessments in schools selecting online testing?
• After controlling for significant student level and school level variables, to what extent do the mathematics performance scores for students taking paper assessments in schools selecting paper testing differ from scores for students taking online assessments in schools selecting paper testing?

• After controlling for significant student level and school level variables, to what extent do mathematics performance scores on paper assessments differ from scores on online assessments for students in different demographic subgroups (gender, FRL, ELL, SPED, or race/ethnicity)?

**Theoretical Perspective**

Whenever online and paper-pencil assessments of the same content are both administered, professional testing standards specify the need to conduct a comparability study across the assessments to determine that mode does not affect student performance and to provide evidence of measurement equivalence.

**Assumptions**

Comparability of administrative mode effects for the NeSA Mathematics online and paper/pencil tests cannot be assumed. Since findings on previous studies cannot be generalized, the NeSA tests and the technology utilized in the online assessments need to be compared between test takers. The online and paper/pencil forms of NeSA Mathematics assessment were constructed to be equivalent using a common test blue print. The items for the online form were a literal transfer of a fixed number of items from a paper/pencil form to a computer screen in a static manner.
Definition of Terms

*Between-school variance*—the variance between the schools’ true means, also called the variance for random intercept.

*Cluster*—refers to the people being in naturally occurring organizational unit such as students in schools, classrooms in schools, or schools in districts. The responses of people from the same cluster are likely to exhibit some degree of relatedness with each other (McCoach, 2010). The responses may not be independent and need to be considered dependent for analysis.

*Computer Based or Online test (CBT)*—tests presented using computers through an online connection to deliver questions and receive student responses.

*Cut Score*—a score set to differentiate between levels of student performance on an assessment such as meeting the standard or not meeting the standard. For NeSA assessments the cut scores were set using the Bookmark Method and the Contrasting Group method.

*Explained Variance*—“proportional reductions in the estimated variance components, $\sigma^2$ and $\tau^2$ in the random-intercept model for two levels, as analogues of $R^2$ values” (Snijders & Bosker, 1999, p. 99). To compute this proportional reduction in the components variance statistic, the difference in variances between the larger model and simpler model are divided by the variance of the baseline model. “The proportional reduction in variance provides a rough estimate of the proportion of variance that is explained using the set of independent variables” (McCoach, 2010, p. 254).

*Hierarchical Linear Modeling (HLM)*—a methodology for conducting a study that allows one to simultaneously model the impact of both individual (level 1) and
contextual (level 2) variables on the dependent variable. The model may include cross-level interactions between higher level and lower level variables. Using HLM, researchers can adjust for and model variables that are dependent on each other.

*Intraclass Correlation Coefficient (ICC)*—a coefficient that captures the degree of relationship among units from the same cluster. It is defined as the ratio of the between group variance divided by the total variance. It is the proportion of variance that is accounted for by the group level (Snijders & Bosker, 1999). In this study, the within-group variance is the variance within the schools about their true means, while the between-group variance is the variance between the schools’ true means. The total variance is then equal to the sum of the within-group and between-group variances (Snijders & Bosker, 1999).

*Nebraska State Accountability—Mathematics (NeSA-M)* assessment - test given to all students in grades 3 through 8 and 11 to measure mathematics achievement on the Nebraska State Mathematics Standards as indicated in the Table of Specifications.

*Paper-pencil test (PPT)*—tests using paper test booklets to present the questions and a paper form to collect student responses (NeSA answer sheets).

*Power of a statistical test*—the probability that the test will correctly reject a false null hypothesis or will identify a treatment effect if one really exists. Factors such as the sample size, the size of the treatment effect, and the value chosen for the alpha level can all influence the power of a hypothesis test (Gravetter & Wallnau, 2007).

*Quasi-experimental*—a non-experimental study that compares groups that are not created by manipulating an independent variable. In this study the groups were 
determined by a participant variable, namely the student/school administrative assessment modes.

*Raw Score*—the number of items the student answered correctly out of the total questions on the assessment. If each question is counted equally, then the raw score for a multiple choice test is simply the number correct responses.

*Statistically Significant*—a result that is very unlikely to occur when the null hypothesis is true. The result is sufficient to reject the null hypothesis (Gravetter & Wallnau, 2007).

*Table of Specifications (TOS)*—a matrix showing the content assessed on the test and the number of items for each content area. The NeSA TOS includes alignment of assessment items with Webb’s Depth of Knowledge (Webb, 1997) with a level of cognitive expectation required to complete the task.

*Within-school variance*—the variance within the schools about their true means, also called the residual variability.

**Delimitations**

This study is delimited to students in Nebraska schools who completed the NeSA mathematics assessment in 2011 in one of two ways. Schools determined the administrative mode for the majority of their students by selecting one administrative mode for the assessment either online or paper-pencil. Only students needing support as identified in 504 plans, special education individual education plans (IEPs), or ELL accommodations were allowed to take the assessment in a different mode. Since students are clustered within schools, hierarchical linear models were investigated to explain the between- and within-cluster variability of the outcome variable, mathematics
achievement. Predictors at both the individual level (level 1) and the school level (level 2) were included in the model to explain the variance in mathematics achievement.

**Limitations**

Because a strictly controlled experiment to study comparability within the context of a statewide assessment program is difficult, a quasi-experimental study with a posttest only design was selected where the impact of both individual and school level variables on mathematics achievement were modeled. Pretest observations were not available on the posttest, the NeSA Mathematics assessment (NeSA-M). Without pretest results to incorporate in the design, the effects of student level variables (level 1) such as gender, social economic status, race/ethnicity, ELL identification, and SPED status were considered. In addition, the impact of school level variables, the percent of FRL students, percent of SPED students, and the percent ELL students along with the percent of the seven race/ethnicity groups in the school, were examined.

**Significance of Study**

If the results of this study show that mathematics performance on the NeSA assessment do not show a difference across administrative modes of assessment, then the scores from the assessment modes are interchangeable. There is no advantage or disadvantage to students testing in a specific testing mode. If the results of this study do show a difference, then policymakers can use the information from the study to decide whether to make adjustments to test scores due the administrative modes.
Chapter 2

Literature Review

First the literature review will discuss three comparability summaries of earlier studies by Paek (2005), Wang, Jiao, Young, Brooks, & Olson (2007), and finally Kingston (2009). The summaries failed to present a consistent picture of the relationship between computer-based and paper-based testing. After discussing the summaries, several specific mathematics comparability studies were presented and reviewed in more detail.

Meta-analysis—K-12 Comparability Studies

Paek (2005) concludes from a review of “K-12 comparability studies to date show that, in general, computer and paper versions of traditional multiple-choice tests are comparable across grades and academic subjects” (p. 17). She further states that as we identify mode differences by grade or subject, we “may be able to recommend improvements in technology to minimize or remove the disparity across modes” (Paek, 2005, p. 18). Paek (2005) lists several reasons to support the continued movement toward computer-based assessments, namely reduced lag time in score reporting, decreased paper use, and reduced mailing costs as well as analysis of student performance not possible from paper tests alone. Paek (2005) recommends that future steps include states continuing “to conduct comparability studies of their high-stakes, large scale assessments across grades and subjects” (p. 18).

In a synthesis of available research on comparability for mathematics assessments, Wang et al. (2007) concluded “The results based on the final selected studies with homogeneous effect sizes show that administrative mode had no statistically
significant effect on K-12 student mathematics tests” (Wang et al., 2007, p. 219). The meta-analysis examined data from 14 primary studies that contained 44 independent experiments. This agrees with a similar finding for reading achievement scores reported by the same researchers (Wang, Jiao, Young, Brooks, & Olson, 2008). The researchers stated that “findings indicate that the administrative mode had no statistically significant effect on K-12 student reading achievement scores” (Wang et al., 2008, p. 5). The researchers further stated that certain variables such as study design, grade level, sample size, type of test, computer delivery method, and whether practice was provided did not lead to differences in student mathematics mean scores between CBT (computer-based test) and PPT (paper-and-pencil test) test administration modes. The investigators found that mean effect size (ES) for linear fixed form is relatively lower than those for the computerized adaptive test algorithm. The researchers mentioned one limitation to the study was the limited sample of the CBT and PPT comparability studies for K-12 mathematics assessments.

In a meta-analysis of comparability studies of computer-administered and paper-administered multiple-choice tests between 1997 and 2007, Neal Kingston (2009) reported that the estimated effect size was a very small. He reasoned that changes in computer-based test administration systems and student’s level of computer experiences has affected the consistency of the results with some favoring computer-based and others favoring paper-based. His grade level analysis found no affect for elementary, middle, or high school. For subject analysis he found small effects favoring computer-based for English Language Arts and Social Studies with effect size of .11 and .15 respectively, but just the opposite for Mathematics with a small effect of negative .06 favoring paper-
based. He recommended “greater attention needs to be paid to possible reasons for
differences observed for mathematics tests” (Kingston, 2009, p. 32). Kingston suggests
that the use of scratch paper by computer students may be different for paper test students
where the work is completed on the test booklet. The change of focus for computer
tested students may present a small cognitive difference from the paper tested students,
resulting in a small effect size favoring paper-based testing.

**Specific Grade Level Comparability Studies**

As stated by Bennett, Braswell, Oranje, Sandene, Kaplan, and Yan (2008) in an
investigation of computerized testing for National Assessment of Education Progress
(NAEP), “comparability is important because if delivery mode affects scores” (2008,
p. 4), then the overall scores are suspect and any conclusions based on the scores are not
valid. If not comparable, student scores as well as aggregated scores for schools and
districts would need adjustments to compensate for the mode effect. The researchers
state three additional concerns when comparing assessment results between assessment
modes, namely results:

- that are to be compared over time after delivery mode has changed from paper
to computer;
- that are aggregated across individuals taking assessments in different modes
  especially when the mode selection is not voluntary; or
- that are to be compared for different groups when mode effects are different
  across the groups.

The investigation examined eighth-grade results from both computer and paper-and-
pencil testing conditions on the same 26 items with 16 multiple choice, 8 short
constructed response, and 2 extended constructed response. Besides both testing conditions being timed, one section of the test allowed students to use a scientific calculator. The study reported a statistically significant difference between testing conditions ($p < 05$) with a small effect size (0.14). The authors stated that “constructed-response items appeared to shift in difficulty more than multiple-choice items when presented on computer as compared to paper” (Bennett et al., 2008, p. 26). This finding was consistent with a similar result found by Russell (1999) where skills in computer keyboarding lessened the negative difference between online and paper-pencil performance on an assessment with open-ended items. The constructed response items on the NAEP assessment needed more adaptation than the multiple-choice items. The researchers suggested that difficulty in “the translation of constructed-response items to electronic delivery” (Bennett et al., 2008, p. 26) may change the nature of what is being measured to some computer skill instead of mathematic performance. The date of data collection (2001) presents one limitation to interpreting the results in the Bennett et al. (2008) study. From a technology point of view, many changes have occurred since 2001 such as student familiarity with computers, changes to computer assessment design, hardware ease of use, and dependability of the technology overall. As a result of current changes in the use of technology for assessing student achievement in mathematics, the following two administrative conditions may be more comparable at the present time.

The study of 7th grade student mathematics achievement by Poggio, Glasnapp, Yang, and Poggio (2005) found “no meaningful statistical differences in the computer test scores attained by the same students on a computerized fixed form assessment and an equated form of that assessment when taken in a traditional paper and pencil format”
During 2003, the study involved students from schools who volunteered to participate in having their students take the Kansas large scale assessment twice once in the computerized assessment system (CBT) and once in the paper-and-pencil format (P&P). There were four forms that were constructed to be parallel using common test specifications and equated based on analyses using linear and IRT methods following a random/equivalent groups design. The CBT items were presented one at a time in the center of the basic template with supporting features for:

- navigation to specific items;
- response-choice striker;
- item-review marker for later review; and
- access to tools such as formulas, a calculator, ruler, and highlighter.

All students and staff were required to sit through an instructional tutorial for the CBT assessment. A follow-up survey found that students were comfortable within the testing CBT environment. Most of the students completed the CBT first with a smaller number of students taking the P & P first. The researchers reported that the main effect for both test order and test mode were not significant. The researchers concluded that no significant interactions of test mode with gender, socio-economic status, or academic placement were detected. They suggested that equivalent forms were the case “if the computer-based test is constructed in such a way that it reflects the paper-and-pencil version on the computer screen” (Poggio et al., 2005, p. 26).

A paper published in 2007 by Society for the Advancement of Excellence in Education (SAEE), a Canadian independent non-profit education research agency, reports the results from the Grade 7 Foundation Skills Assessment (FSA) administered in paper
and electronic modes. Gaskell and Marshall found a significant difference in Numeracy multiple-choice assessment with students doing significantly better in the paper mode than the electronic (online) mode of the assessment. In this study the researchers reported “that the difference between paper and electronic modes was greater for males than females” (Gaskill & Marshall, 2007, p. 38), but much of the gender difference is “attributed to some larger school having considerable gender differences” (Gaskill & Marshall, 2007, p. 29). The research reviewed evidence to discern whether the student’s level of ability, based on FSA achievement in grade 4, affected the difference in scores between modes. The mode difference was smaller than the total difference for above-average and average achieving students, but disappeared for below-average students.

The researchers reported no discernible differences in navigation between paper and electronic versions of the assessment since the items in both formats were virtually identical except the items on the computer were presented one at a time. One difference noted in the study between modes was the required use of paper-and-pencil work to find the answer on both modes of the assessment. The researchers felt that this difference introduced more opportunities for mistakes in that the students in the online version had to transfer the necessary information to paper, work the problem on screech paper and then select the answer on the screen. The researchers recommend that students practice working problems on the electronic administration along with adequate space adjacent to the computer for calculating the solutions.

Johnson and Green (2006) conducted a project with 11-year-olds from United Kingdom where the students were given two sets of matched mathematics items, one set on-line and the other on paper. The researchers found no statistically significant
differences between overall performance on paper and computer, but did identify individual question-level differences. Contrary to the study by Poggio et al. (2005), the researchers raised important questions about the merits of transferring questions between modes. Several issues where the “medium of assessment might matter” (Johnson & Green, 2006, p. 23) are the way students:

1. navigate through the test questions,
2. transfer information from screen to paper and vice versa, and
3. select a problem solving strategy.

Additionally the researchers found students preferred paper-and-pencil over online for questions on shape, space, and measurement. The researchers suggest that “as computer-based testing becomes more widespread it is important that students have the opportunity to be as familiar as possible with the experience of test taking on computers so that valid inferences can be made about their ability” (Johnson & Green, 2006, p. 30).

The Texas Education Agency in Technical Report (2008) stated that each time a new test form is offered in both computer and pencil/paper modes a comparability study is conducted. A quasi-experimental study design using a matched samples comparability analysis (MSCA) allows districts the flexibility of choosing the mode of test administration. Mode affects have consistently been reported on the Texas Assessment of Knowledge and Skills (TAKS) assessments in Mathematics. Based on TAKS online versus paper comparability studies from 2005 through 2008, the students taking the TAKS online tend to score one point lower than students taking the test on paper. For high stakes testing such as exit level tests for graduation even a difference of one point between modes can affect a large number of students. As result, the Texas Education
Agency (TEA) (2008) has recommended the following for statewide assessment systems:

“comparability studies continue to be conducted to help ensure the defensibility of the testing programs” (Texas Education Agency in Technical Report Series, 2008, p. 35).

TEA suggested that mode effects may be related to factors that differ across test administrations such as test questions, test scoring, testing conditions, and examinee groups.

**Test Comparability for Nebraska Assessments**

To maintain the transparency of the Nebraska assessment process, an analysis of mode difference for statewide mathematics assessments needs to be conducted and reported for all grades. Because the research has not been conclusive in establishing the lack of a mode difference in student mathematics achievement scores between administrative modes, the Nebraska Department of Education (NDE) is obligated to investigate potential mode affects with the NeSA mathematics assessment. In the future the NeSA assessments will continue to be offered in both modes with districts/schools still deciding which testing mode to utilize for students. Although, NeSA tests began with a strong movement toward online assessments with more than 50% of the schools selecting online mathematics assessment, the assessment office believes that the percent of schools using online assessments will continue to increase over the years to the point where fewer and fewer students are being assessed using the paper-pencil mode. Still there will always be a requirement to offer the paper-based assessments for students requiring specific accommodations requiring paper-based assessment. Since the computer-based and paper-based forms of the same assessment can be considered two different forms of the NeSA mathematics assessment, the Standards for Educational and
Psychological Testing (AERA, APA, & NCME, 1999) has suggested in Standard 4.10 that “A clear rationale and supporting evidence should be provided for any claim that scores earned on different forms of a test may be used interchangeably.” Therefore the following study has been designed to measure comparability of administrative modes for the 2011 NeSA mathematics assessment.
Chapter 3

Methods

Research Design

A quasi-experimental design was selected for the comparability study due to the difficulty in carrying out a controlled experiment such as a counterbalanced test-retest design or a random assignment in the context of a statewide assessment program. Rather than assigning administrative mode, the selection of assessment mode was determined by each school based on school characteristics rather than student characteristics. When selecting the mode of testing, school administrators/teachers looked at instructional issues such as school use of computers and overall computer literacy and the technological issues such as computer availability and internet connectivity. Because the students could not be randomly assigned to test-retest groups or control/experimental groups, the researcher cannot assume that the computer-based testing group is equivalent to the paper-based testing group. The researcher is forced to work with intact groups rather than groups randomly assigned to different conditions. A posttest-only nonequivalent control group design was investigated where the paper/pencil assessed group was compared to the online assessed group. With this design there is no basis for “assuming group equivalency or assuming that posttest differences might have occurred anyway” (Girden & Kabacoff, 2010, p. 263). Although initial nonequivalence cannot be controlled experimentally, it is possible to achieve some degree of equivalence statistically. Explanatory variables at the student and school level used as covariates were included in the model data analysis process to limit the possibility of alternative explanations for the findings about administrative mode effect.
For comparability studies, Lottridge, Nicewander, Schulz, and Mitzel (2008) point out that the two basic designs are within-subjects design and the between-subjects design. Since the students took mathematics assessment once on either paper or computer, but not both, the research study is a between-subjects design. The study involved two units of analysis, one at the school level and one at the student level. Because students are nested in schools, the investigation of the effect of test administrative mode required the use of multilevel methods and considered the mode effects at both the school level and the student level. Lee (2000) identifies four steps in conducting an analysis using Hierarchical Linear Modeling (HLM). They are:

1. Establish whether multilevel methods are necessary by partitioning the variance in mathematics achievement into two parts – proportion of variance between students in the same school and the proportion of variance that lies systematically between schools. The intraclass correlation (ICC) for a 2-level empty model is the proportion of total variance that lies systematically between schools. When ICC is greater than 10% of the total variance, the researcher needs to consider multilevel methods.

2. Estimate the within-school or Level 1 model by investigating the characteristics of individual students that are associated with the mathematics achievement. Student characteristics considered at this level were student administrative mode, gender, race/ethnicity, SES (eligible for free or reduced lunch), LEP eligibility (ELL), and student receiving special services (SPED).

3. Finally the school effects for student characteristics included at level 2 are estimated as function of school characteristics. The school characteristics
investigated were school administrative mode, proportion of females, proportion of each ethnicity (seven race/ethnicity groups), average SES (eligible for free or reduced lunch), proportion of ELLs, and proportion of SPED.

4. The investigation explored not only the average outcomes for the student characteristics but cross-level intersections of student characteristics with administrative mode, as well.

First the schools voluntarily selected the mode of test administration, computer-based and paper-based mode, and then the students within each mode group completed the assessment using paper-based assessments or online assessments. In this study, all students from each of the grades were divided into the online and paper/pencil assessments within a school selecting paper testing or online testing. The participants were divided into four groups with each being administered the same mathematics assessment items via a computer or on paper. Since the school administration selected administrative mode for each group and both groups were given only a posttest, the design can be represented as:

- Students paper assessment in school paper testing (P/P) NR X₁ O
- Students paper assessment in school online testing (P/O) NR X₂ O
- Students online assessment in school online testing (O/O) NR X₃ O
- Students online assessment in school paper testing (O/P) NR X₄ O

where NR represents nonrandom assignment, X₁ represents the paper-based mathematics assessment in school selecting paper testing, X₂ represents the paper-based mathematics assessment in school selecting online testing, X₃ represents the online-based mathematics assessment,
assessment in school selecting online testing, \( X_4 \) represents the online-based mathematics assessment in school selecting paper testing, and \( O \) represents the performance (test score) (Shadish, Cook, & Campbell, 2002). The assessment results of students using the paper-based assessment in schools selecting paper testing \( (X_1) \) were compared to the assessment results of students completing the assessment in the other three situations \( (X_2, X_3, \text{ and } X_4) \) using HLM. Other research has shown that alternative explanations for differences in mathematics achievement can involve student level factors such as free and reduced lunch (FRL) or gender and school level factors such as school average FRL or school minority concentrations. To limit the alternate explanations for the findings on administrative test mode, the HLM analysis included student factors at level 1 and school factors at level 2.

**General Design Issues**

Based on CCSSO technical paper submitted by Lottridge et al. (2008), four general design issues for comparability studies need to be identified and addressed for the following dimensions

- instrumentation
- sampling,
- administration, and
- scoring.

Each dimension as it relates to the Nebraska State Accountability assessment for mathematics is considered in the following sections.
Instrumentation

To build quality assessments, more multiple choice items were fielded tested in 2010, than needed to build the 2011 mathematics assessments. From the field tested items, items were reviewed by NDE/DRC team, and the best items were selected for the operational test form based on their item statistics. The following item statistics along with the criteria for each was used to evaluate the items. Test items were selected by evaluating

- the item mean value \((p\text{-value})\) – greater than .30 but less than .95,
- the corrected item to total correlation – greater than .1, and
- the frequency distribution of item responses – correct answer selected by more than 50% of the respondents.

Each assessment has a grade level Table of Specifications (TOS) which defines the total number of items on the respective assessment and the range of items per standard/indicator. The NeSA Table of Specifications (2010) for each test is located on the Statewide Assessment Website. The number of multiple choice items selected for each test measuring a student’s overall mathematics achievement varies across the grades with 50 items for grade 3 and more items for grades 4 through 11. See Table 3.1 for number of items on the NeSA-M assessment for grades 4, 8, and 11. The items on each assessment were evaluated for alignment to standards, appropriateness to grade level, and lack of bias due to race/ethnicity or gender. The 2011 operational NeSA-M assessment was administered in both modes to students in grades 3, 4, 5, 6, 7, 8, and 11 between March 26 and May 6, 2011. For this study only the results from three grades will be analyzed, grade 4 in the elementary level, grade 8 in the middle level, and grade 11 in the
Table 3.1

*Number of Students Tested on NeSA-M in 2011 by Grade*

<table>
<thead>
<tr>
<th>Grade</th>
<th>No. of Operational Items</th>
<th>Total Students</th>
<th>Total Tested on Paper</th>
<th>Total Tested on Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>55</td>
<td>21596</td>
<td>8518</td>
<td>13078</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>20543</td>
<td>8500</td>
<td>12043</td>
</tr>
<tr>
<td>11</td>
<td>60</td>
<td>20814</td>
<td>9686</td>
<td>11128</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>62953</td>
<td>26704</td>
<td>36249</td>
</tr>
</tbody>
</table>

The raw score for each student’s grade level assessment was calculated by counting the number of correct answers. Although the raw scores are converted to scale scores, the raw scores were used in the analysis to measure student achievement.

Table 3.1 presents information on the number of students tested on the NeSA-M assessment for each grade by test mode.

Additional information showing the average raw scores for the different subgroups is presented in Appendix B. Tables B1, B2, and B3 show the performance results for 4th, 8th, and 11th grade respectively. Each table shows the average mathematics performance for all students along with results for students in the following subgroups - male, female, SPED, Non-SPED, FRL, Non-FRL, ELL, and Non-ELL. The results are divided into the four administrative modes, Paper/Paper, Online/Paper, Online/Online, and Paper/Online. Initially the tables suggest some differences. The following analysis examined the data for possible significance using HLM.
Sampling

All Nebraska students in grades 3, 4, 5, 6, 7, 8, and 11 participated in the NeSA Mathematics assessment in either the computer-based or paper-based mode of assessment. The only exception was those students taking the NeSA Alternate mathematics assessment, designed for significantly cognitively disabled students. School administrators were permitted to select the mode of assessment that best matches the instructional practice in their respective schools. With the exception of ELL or SPED students with accommodations, all students within the same school took the assessment in the same mode. For 2011, the percentage of students taking the mathematics assessment online was 58% while the percentage taking the paper/pencil assessment was 42%. For mathematics, in grades 4, 8, and 11 about 27,000 students took the assessment using paper and pencil and over 36,000 students took the assessment online. See Table 3.1 for number of students taking the online mathematics assessment versus paper/pencil mathematics assessment in grades 4, 8, and 11. For each grade level, between 8,000 and 10,000 students took the paper assessments while approximately 11,000 to 13,000 took the online assessments.

With a statewide assessment that required for all students to participate and allowed schools to select the administrative mode, random assignment of students to assessment mode was not possible. Because student data were available for all students and HLM required a large sample of student data, the researcher decided to include all students from both modes in the analysis. By using the entire research population, arguments can be made that the investigated research population is supposed to be representative of a wider population – namely pupils in later years or other grades
To account for school differences and potential student differences, covariates for school factors and student factors were included in the HLM analysis. Student level factors such as gender, free and reduced lunch status, student receiving special education services, race/ethnicity, and student with English Language Learner support were included in the model analysis to control for student level variation in the four assessment groups. School level factors such as the percent female, the percent of each race/ethnicity, the percent ELL, the percent free and reduced lunches, and the percent SPED were included in the model to account for variation due to school differences.

**Administration**

For both modes, the tests were administered during the same test period, March 28 through May 6. District test coordinators received equivalent administrative training for each administrative mode. The administrators at the local schools were responsible for training all test administrators at their respective schools, selecting and scheduling testing locations, and promoting security of the assessments, test booklets for the paper-based assessment or computer passwords for the online assessment, within their school building. Test administrative procedures were standardized for each grade through online and the paper/pencil test manuals that included assessment procedures for the test administrators. Each manual included specific directions for administering the NeSA mathematics assessments. The supporting materials such as grade level practice tests were available for both the computer-based or paper-based modes. Practice tutorials to familiarize students with the online tools and format were provided for students taking the online assessment. The Nebraska Department of Education encouraged all
administers to provide students with an opportunity to practice the assessment mode that the school selected. For both modes, the tests were low-stakes, untimed, and consist of the same item sets for all students. Every effort was made to make the test administration for both modes equivalent.

**Scoring**

Scoring procedures for multiple choice items were the same for both modes of the assessment. During the online assessment, the student answers were collected electronically for each item, while answers for the paper-based assessment were collected on answer sheets. The answer sheets were scanned and the answers recorded electronically in the DRC data system along with the electronic entries from the online assessment. Each set of answers was compared to the respective answer key for that grade and scored one point for each correct answer. The raw score for the mathematics assessment was the total number of correct answers. The student’s raw score was used as an indicator of mathematics achievement.
Chapter 4

Results

Introduction Hierarchical Linear Models

Since schools selected the mode of assessment, an analysis of administrative mode effects, by their very nature, are hierarchical because students are nested in schools. Students in the same school have similar experiences and therefore their educational outcomes are dependent rather than independent. Prior to the availability of Hierarchical Linear Models (HLM), research on multilevel data would analyze data at either individual or the school level but not both. HLM technique allows researchers to consider more than one unit of analysis when analyzing multilevel data (Lee, 2000). Along with student and school level data, the 2011 NeSA-Mathematics assessment results from grades 4, 8, and 11 were coded and analyzed using a Hierarchical Linear Model (HLM – Multilevel) SAS 9.2 (32) (English) software. The SAS mixed procedure (proc mixed) used Maximum Likelihood (ML) as the estimation method in fitting the model. The hierarchical linear model is a type of regression model that is particularly suitable for multilevel data (Snijders & Bosker, 1999). Since the data involves more than 10 groups with group size ranging from small ($n$ ranging from one to greater than 100), a random coefficient model was selected where the study involves two populations, a population of schools and a population of individuals. In a random coefficient model, the random residuals and coefficients can be regarded as representing the effects of unmeasured variables and the approximate nature of the linear model (Snijders & Bosker, 1999). The SAS program for each grade level model is listed in Appendix C.
Variables and Coding for Analysis

The HLM analysis utilized a two-level hierarchical linear model for mathematics achievement as measured by the NeSA-Mathematics assessments. The dependent variable in the analysis was the student’s raw score on the NeSA-M assessment. Administrative mode for student and school were added to the model to define the four comparison groups:

- students taking paper assessments in schools selecting paper testing (paper/paper),
- students taking paper assessments in schools selecting online testing (paper/online),
- students taking online assessments in schools selecting paper testing (online/paper), and
- students taking online assessments in schools selecting online testing (online/online).

By adding the cross-level interaction between school and student administrative modes, all four testing groups can be defined in the analysis. With the cross-level interaction effects added to the model, the score differential between students taking paper or online assessments in schools selecting paper or online testing can be analyzed. See Table 4.1 showing the four testing groups where zero represents paper mode and one represents online mode for both student and school administrative codes. The mathematics achievement score for students taking paper assessments in schools selecting paper testing represented the comparison student score in the model analysis where both the Student_Online_Paper and the School_Online Paper were equal to zero. Whereas when
Table 4.1

Four Testing Groups by Student/School Administrative Mode

<table>
<thead>
<tr>
<th>Student Administrative Mode</th>
<th>Paper Testing = 0</th>
<th>Online Testing = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper Assessment = 0</td>
<td>0 / 0</td>
<td>1 / 0</td>
</tr>
<tr>
<td>Online Assessment = 1</td>
<td>0 / 1</td>
<td>1 / 1</td>
</tr>
</tbody>
</table>

Student_Online_Paper and School_Online Paper were both equal to one, the differential score for students taking online assessments in schools selecting online testing was represented. Likewise the differential score for students taking online assessments in schools selecting paper testing was represented by Student_Online_Paper equal to one and School_Online Paper equal to zero while the differential score for students taking paper assessments in schools selecting online testing was represented by the reverse Student_Online_Paper equal to zero and School_Online Paper equal to one.

Each of the three grade level models started with the same explanatory variables at student level (level 1) and the school level (level 2). For Level 1, variables consisted of the student gender indicator, the seven ethnicity/race classifications, the student’s eligibility status for free and reduced lunch, the student’s indicator for support of special education services, and English Language Learner indicator for English language support. To represent the seven ethnicity/race classifications, six dummy variables were used, contrasting the last six against the first classification. This means that all six dummy variables were 0 for white with the first code (e1d) being 1 for Hispanic, the second one (e2d) being 1 for American Indian or Alaska Native, the third one (e3d) being 1 for Asian, the fourth one (e4e) being 1 for black or African American, the fifth one...
(e5d) being 1 for Native Hawaiian or other Pacific Islander, and the sixth one (e6d) being 1 for two or more races. The dummy codes for ethnicity/race were coded as shown in Table 4.2. To analyze the individual ethnicity/race variables the six dummy variables (e1d, e2d, e3d, e4d, e5d, and e6d) were added to the model as a group.

Table 4.2

Dummy Codes for Ethnicity/Race Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>e1d</th>
<th>e2d</th>
<th>e3d</th>
<th>e4d</th>
<th>e5d</th>
<th>e6d</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hispanic</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asian</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Two or More Races</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

At Level 2, school variables for percent of female students, percent of students in each ethnicity/race subgroup, percent of students eligible for free and reduced lunch, percent of students receiving special education support, and percent of English Language Learners were included in the model. To make the interpretation more meaningful, each Level 2 variable was centered at the grand mean. For example, the school percent of students eligible for free/reduced lunches was centered at .47 for grade 4, .43 for grade 8, and .35 for grade 11. When centering the data on the mean, the interpretations can be simplified by referring any changes in the school average relative to the mean. The
interpretation would reference a change in mathematics achievement relative to a change in percent above or below the average, for example, for a 10% increase in the school’s percent free/reduced lunch above the state average, the estimated student mathematics score on average decreased by two points.

The descriptive statistics for the level 1 and level 2 variables for grades 4, 8 and 11 are presented for 4th, 8th, and 11th grades in Tables 4.3, 4.4, and 4.5 respectively. The Tables 4.3 through 4.5 present the mean, standard deviation, minimum value, and maximum value for the dependent variable, independent student and school variables. For the 4th grade, 60.6% of the students tested online and 62.7% of the schools selected online testing, while only 58.6% of 8th grade students tested online and 59.8% of the schools for 8th graders selected online testing. Fifty-three percent of the 11th grade students were tested online, and 54% of the schools for 11th graders selected online testing.

In Appendix A, Tables A1, A2, and A3 show the mathematics performance of students in grades 4, 8, and 11 by demographic subgroup in each of the test modes before controlling for student and school variables. As shown in the tables most of the students were tested by either online in a school selecting online testing or on paper in a school selecting paper testing. Differences in the four groups cannot be tested since the groups were not randomly selected. HLM analysis was used to control for student (level 1) and school (level 2) variables. Using a similar HLM process, Liu and Koirala (2010) investigated the relationship between mathematics self-efficacy and mathematics achievement with student-level variables and school-level variables.
Table 4.3

*Descriptive Statistics for the Variables in Both Levels – Grade 4*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1 Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw Score – NeSA-M</td>
<td>21596</td>
<td>40.9</td>
<td>9.6</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Gender_Code – (0 = male, 1 = female)</td>
<td>21596</td>
<td>.490</td>
<td>.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic_Code – recoded with 6 dummy codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (0 = white)</td>
<td>21596</td>
<td>.696</td>
<td>.460</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic (e1d = 1)</td>
<td>21596</td>
<td>.168</td>
<td>.374</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>American Indian/Alaska Native (e2d = 1)</td>
<td>21596</td>
<td>.016</td>
<td>.124</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Asian (e3d = 1)</td>
<td>21596</td>
<td>.021</td>
<td>.144</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Black (e4d = 1)</td>
<td>21596</td>
<td>.068</td>
<td>.252</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander (e5d = 1)</td>
<td>21596</td>
<td>.001</td>
<td>.036</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Two or More Races (e6d = 1)</td>
<td>21596</td>
<td>.030</td>
<td>.171</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ELLcode2 (0 = Not LEP Eligible, 1 = LEP Eligible)</td>
<td>21596</td>
<td>.083</td>
<td>.276</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Food_Program_Code2 (0 = Not Eligible, 1 = Eligible)</td>
<td>21596</td>
<td>.466</td>
<td>.499</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Special_Education_Code (0 = Not SPED, 1 = SPED)</td>
<td>21596</td>
<td>.180</td>
<td>.384</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stud_Online_Paper (0 = Paper, 1 = Online)</td>
<td>21596</td>
<td>.606</td>
<td>.489</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Level 2 Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Female (Centered at .50)</td>
<td>21596</td>
<td>.490</td>
<td>.080</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent Free/Reduced Lunch (Centered at .47)</td>
<td>21596</td>
<td>.466</td>
<td>.265</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent ELL (Centered at .08)</td>
<td>21596</td>
<td>.083</td>
<td>.136</td>
<td>0</td>
<td>.763</td>
</tr>
<tr>
<td>Percent SPED (Centered at .18)</td>
<td>21596</td>
<td>.180</td>
<td>.078</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.3 continues
Table 4.4

**Descriptive Statistics for the Variables in Both Levels – Grade 8**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1 Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw Score – NeSA-M</td>
<td>20543</td>
<td>43.1</td>
<td>12.1</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>Gender_Code (0 = male, 1 = female)</td>
<td>20543</td>
<td>.488</td>
<td>.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic_Code – recoded with 6 dummy codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (0 = white)</td>
<td>20543</td>
<td>.720</td>
<td>.449</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic (e1d = 1)</td>
<td>20543</td>
<td>.154</td>
<td>.361</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>American Indian/Alaska Native (e2d = 1)</td>
<td>20543</td>
<td>.014</td>
<td>.116</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Asian (e3d = 1)</td>
<td>20543</td>
<td>.020</td>
<td>.140</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Black (e4d = 1)</td>
<td>20543</td>
<td>.062</td>
<td>.241</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander (e5d = 1)</td>
<td>20543</td>
<td>.001</td>
<td>.033</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Two or More Races (e6d = 1)</td>
<td>20543</td>
<td>.029</td>
<td>1.68</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.4 continues
<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELLcode2 (0 = Not LEP Eligible, 1 = LEP Eligible)</td>
<td>20543</td>
<td>.033</td>
<td>.179</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Food_Program_Code2 (0 = Not Eligible, 1 = Eligible)</td>
<td>20543</td>
<td>.427</td>
<td>.494</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Special_Education_Code (0 = Not SPED, 1 = SPED)</td>
<td>20543</td>
<td>.139</td>
<td>.346</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stud_Online_Paper (0 = Paper, 1 = Online)</td>
<td>20543</td>
<td>.586</td>
<td>.493</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Level 2 Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent Female (Centered at .50)</td>
<td>20543</td>
<td>.488</td>
<td>.063</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent Free/Reduced Lunch (Centered at .47)</td>
<td>20543</td>
<td>.427</td>
<td>.226</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent ELL (Centered at .03)</td>
<td>20543</td>
<td>.033</td>
<td>.179</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent SPED (Centered at .13)</td>
<td>20543</td>
<td>.139</td>
<td>.53</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent White (Centered at .72)</td>
<td>20543</td>
<td>.720</td>
<td>.247</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent Hispanic (Centered at .15)</td>
<td>20543</td>
<td>.154</td>
<td>.187</td>
<td>0</td>
<td>.808</td>
</tr>
<tr>
<td>Percent American Indian/Alaska Native (Centered at .01)</td>
<td>20543</td>
<td>.014</td>
<td>.059</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent Asian (Centered at .02)</td>
<td>20543</td>
<td>.020</td>
<td>.023</td>
<td>0</td>
<td>.114</td>
</tr>
<tr>
<td>Percent Black or African American (Centered at .06)</td>
<td>20543</td>
<td>.062</td>
<td>.108</td>
<td>0</td>
<td>.550</td>
</tr>
<tr>
<td>Percent Native Hawaiian or Pacific Islander (Centered at .001)</td>
<td>20543</td>
<td>.001</td>
<td>.005</td>
<td>0</td>
<td>.071</td>
</tr>
<tr>
<td>Percent Two or more races (Centered at .03)</td>
<td>20543</td>
<td>.029</td>
<td>.031</td>
<td>0</td>
<td>.500</td>
</tr>
<tr>
<td>School_Online_Paper (0 = Paper, 1 = Online)</td>
<td>20543</td>
<td>.598</td>
<td>.490</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 4.5

Descriptive Statistics for the Variables in Both Levels – Grade 11

<table>
<thead>
<tr>
<th>Level 1 Variables</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Score – NeSA-M</td>
<td>20814</td>
<td>38.4</td>
<td>13.3</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>Gender_Code – (0 = male, 1 = female)</td>
<td>20814</td>
<td>.490</td>
<td>.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ethnic_Code – recoded with 6 dummy codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (0 = white)</td>
<td>20814</td>
<td>.750</td>
<td>.433</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic (e1d = 1)</td>
<td>20814</td>
<td>.126</td>
<td>.331</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>American Indian/Alaska Native (e2d = 1)</td>
<td>20814</td>
<td>.014</td>
<td>.119</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Asian (e3d = 1)</td>
<td>20814</td>
<td>.020</td>
<td>.141</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Black (e4d = 1)</td>
<td>20814</td>
<td>.061</td>
<td>.240</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander (e5d = 1)</td>
<td>20814</td>
<td>.001</td>
<td>.039</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Two or More Races (e6d = 1)</td>
<td>20814</td>
<td>.026</td>
<td>.160</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ELLcode2 (0 = Not LEP Eligible, 1 = LEP Eligible)</td>
<td>20814</td>
<td>.022</td>
<td>.147</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Food_Program_Code2 (0 = Not Eligible, 1 = Eligible)</td>
<td>20814</td>
<td>.352</td>
<td>.477</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Special_Education_Code (0 = Not SPED, 1 = SPED)</td>
<td>20814</td>
<td>.116</td>
<td>.319</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stud_Online_Paper (0 = Paper, 1 = Online)</td>
<td>20814</td>
<td>.535</td>
<td>.499</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 2 Variables</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Female (Centered at .50)</td>
<td>20814</td>
<td>.488</td>
<td>.067</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent Free/Reduced Lunch (Centered at .35)</td>
<td>20814</td>
<td>.351</td>
<td>.192</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent ELL (Centered at .02)</td>
<td>20814</td>
<td>.022</td>
<td>.036</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent SPED (Centered at .12)</td>
<td>20814</td>
<td>.115</td>
<td>.056</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.5 continues
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent White (Centered at .75)</td>
<td>20814</td>
<td>.750</td>
<td>.229</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent Hispanic (Centered at .13)</td>
<td>20814</td>
<td>.126</td>
<td>.155</td>
<td>0</td>
<td>.764</td>
</tr>
<tr>
<td>Percent American Indian/Alaska Native (Centered at .01)</td>
<td>20814</td>
<td>.014</td>
<td>.063</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Percent Asian (Centered at .02)</td>
<td>20814</td>
<td>.020</td>
<td>.022</td>
<td>0</td>
<td>.090</td>
</tr>
<tr>
<td>Percent Black or African American (Centered at .06)</td>
<td>20814</td>
<td>.061</td>
<td>.105</td>
<td>0</td>
<td>.539</td>
</tr>
<tr>
<td>Percent Native Hawaiian or Pacific Islander (Centered at .001)</td>
<td>20814</td>
<td>.001</td>
<td>.005</td>
<td>0</td>
<td>.067</td>
</tr>
<tr>
<td>Percent Two or more races (Centered at .03)</td>
<td>20814</td>
<td>.026</td>
<td>.032</td>
<td>0</td>
<td>.143</td>
</tr>
<tr>
<td>School_Online_Paper (0 = Paper, 1 = Online)</td>
<td>20814</td>
<td>.544</td>
<td>.498</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**HLM Analysis Process**

In the following sections the steps to the HLM model are explained. Following the explanation the HLM analysis for mathematics achievement in three grade levels, namely grade 4, 8, and 11 are discussed. The results from the HLM analysis were examined for each grade level through the following steps:

- **Step One – Empty Model**
- **Step Two – Add Student and School Administrative Mode Variables**
- **Step Three – Add Level 1 and Level 2 Explanatory Variables**
- **Step Four – Level 1 Interactions With Student Administrative Mode**

**HLM Analysis—Step One—Empty Model.** Using an Empty Model, the first step in conducting an HLM analysis is to partition the variance in the dependent variable (individual mathematics scores) into two parts. The empty model can be expressed like this (Snijders & Bosker, 1999):
Level 1 model: \( Y_{ij} = \beta_{0j} + R_{ij} \)
where \( Y_{ij} \) represents the mathematics achievement of \( i \)th student in the \( j \)th school, 
\( \beta_{0j} \) is the intercept, the mean mathematics achievement in the \( j \)th school, 
\( R_{ij} \) is the random error of \( i \)th student in the \( j \)th school.

Level 2 model: \( \beta_{0j} = \gamma_{00} + U_{0j} \)
where \( \gamma_{00} \) is a level 2 coefficient, 
and \( U_{0j} \) is level 2 random effect.

One part represents the proportion of variance that lies between students in the same school (within-school – \( \text{var}(R_{ij}) \)) and the other part is the proportion of variance that lies systemically between schools (\( \text{var}(U_{0j}) \)) The total variance in the model would be represented by the sum of the within-school variance and the between-school variance, \([\text{var}(R_{ij}) + \text{var}(U_{0j})]\). The intraclass correlation (ICC) can be calculated by dividing between-schools variance by the total variance in the model, ICC = \( \frac{\text{var}(U_{0j})}{\text{var}(R_{ij}) + \text{var}(U_{0j})} \). When the ICC is greater than 10%, this parameter indicates that multilevel analysis (HLM) should be considered.

**HLM Analysis—Step Two—Add Student and School Administrative Mode**

**Variables.** The second step adds the variables for student administrative mode and school administrative mode to set up four comparison groups. The inclusion of the cross-level interaction between student and school administrative mode permitted an examination of the effect of student administrative mode in the different school testing settings. There are two online assessment groups – students taking the online assessment in either a school that selects online testing (online/online) or paper/pencil testing (online/paper). Likewise, there are two paper/pencil assessment groups – students taking paper assessments in a school that selects online testing (paper/online) or paper/pencil testing (paper/paper). The coding for student administrative mode variable was paper equal zero, while the coding for online was equal one. Likewise, the coding for school
administrative mode variable was also paper equal zero and online was equal one. With this coding the estimated intercept represented the students taking paper assessments in a school that selected paper/pencil testing (paper/paper or zero/zero). With the other three codes paper/online (zero/one), online/paper (one/zero), and online/online (one/one), the estimated coefficients represented the difference between respective assessment group and the paper/paper group. The intercept and differences between assessment groups were estimated using the following model (Snijders & Bosker, 1999).

Level 1:
$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + R_{ij}$$
where $Y_{ij}$ represents the mathematics achievement of $i$th student in the $j$th school,
$\beta_{0j}$ is the intercept, the mean mathematics achievement in the $j$th school,
$\beta_{1j}$ is the slope for student online/paper mode,
$X_{ij}$ represents the values of student online/paper mode of $i$th student in the $j$th school, and
$R_{ij}$ is the random error of $i$th student in the $j$th school.

Level 2:
$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{12} (W_{1j} * X_{ij}) + U_{0j}$$
$$\beta_{1j} = \gamma_{10} \text{ (online/paper admin mode)}$$
where $\gamma_{00}$ is a level 2 coefficient,
where $\gamma_{01}$ is a level 2 coefficient,
where $\gamma_{12}$ is a level 2 coefficient for the cross-level interaction,
$W_{1j}$ is level 2 value for SCH online/paper mode in the $j$th school, and
$U_{0j}$ is level 2 random effect.

The assessment groups were estimated by $\beta_{0j} + \beta_{1j}X_{ij}$ or substituting for $\beta_{0j}$ and $\beta_{1j}$ the formula became $[\gamma_{00} + \gamma_{01}W_{1j} + \gamma_{12} (W_{1j} * X_{ij}) + \gamma_{10}X_{ij}]$. The paper/paper assessment group where $X_{ij}$ and $W_{1j}$ were both zero was estimated by the intercept $\gamma_{00}$, while the paper/online assessment group where $X_{ij}$ was zero and $W_{1j}$ was one was estimated by the intercept $\gamma_{00}$ plus the coefficient $\gamma_{01}$ ($\gamma_{00} + \gamma_{01}$). The online/paper assessment group where $X_{ij}$ was one and $W_{1j}$ was zero was estimated by the intercept $\gamma_{00}$ plus the coefficient $\gamma_{10}$ ($\gamma_{00} + \gamma_{10}$), while the online/online assessment group where $X_{ij}$ and $W_{1j}$ were both one
was estimated by the intercept $\gamma_{00}$ plus the coefficients $\gamma_{01}$, $\gamma_{02}$, and $\gamma_{10}$ ($\gamma_{00} + \gamma_{012} + \gamma_{10}$).

These estimates were examined for differences after adding each student level variable and school level variable to the model. With the addition of fixed effects for two explanatory variables and one interaction along with the random effect of student mode to the empty model, the model now has eight parameters – four regression coefficients and four variance components.

**HLM Analysis—Step Three—Add Level 1 Explanatory Variables.** The third step adds explanatory variables at the student level to estimate the within-school or Level 1 model (Snijders & Bosker, 1999). Individual students characteristics associated with Mathematics achievement are investigated. Student variables are added to the model explain the remaining variability in the outcome by examining the significance of parameter estimates. With additional explanatory variables added to model, the model increases the number of regression coefficients and variance components. The addition of random effects increased the variance components, while addition of fixed effects increased the regression coefficients.

**Level 1:**

\[ Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \beta_{2j}X_{ij} + \beta_{3j}X_{ij} + \beta_{4j}X_{ij} + \beta_{5j}X_{ij} + \beta_{6j}X_{ij} + \beta_{7j}X_{ij} + \beta_{8j}X_{ij} + \beta_{9j}X_{ij} + \beta_{10j}X_{ij} + \beta_{11j}X_{ij} + \epsilon_{ij} \]

where $Y_{ij}$ represents the mathematics achievement of $i$th student in the $j$th school, $\beta_{0j}$ is the intercept, the mean mathematics achievement in the $j$th school, $\beta_{1j}$, $\beta_{2j}$, $\beta_{3j}$, $\beta_{4j}$, $\beta_{5j}$, $\beta_{6j}$, $\beta_{7j}$, $\beta_{8j}$, $\beta_{9j}$, $\beta_{10j}$, and $\beta_{11j}$ are the slopes for student online/paper mode, Free and Reduced Lunch, gender, special education, ELL, and ethnicity (six dummy variables) in the $j$th school, $X_{ij}$ represents the values of student online/paper mode, Free and Reduced Lunch, gender, special education, ELL, and ethnicity (six dummy variables) of $i$th student in the $j$th school, and $\epsilon_{ij}$ is the random error of $i$th student in the $j$th school.
Level 2:
\[ \beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \gamma_{02} (W_{1j} \times X_{ij}) + U_{0j} \]
\[ \beta_{1j} = \gamma_{10} \text{ (online/paper admin mode)} \]
\[ \beta_{2j} = \gamma_{20} \text{ (FRL)} \]
\[ \beta_{3j} = \gamma_{30} \text{ (Gender)} \]
\[ \beta_{4j} = \gamma_{40} \text{ (SPED)} \]
\[ \beta_{5j} = \gamma_{50} \text{ (ELL)} \]
\[ \beta_{6j} = \gamma_{60} \text{ (HISP)} \]
\[ \beta_{7j} = \gamma_{70} \text{ (AMIN)} \]
\[ \beta_{8j} = \gamma_{80} \text{ (ASIA)} \]
\[ \beta_{9j} = \gamma_{90} \text{ (BLAC)} \]
\[ \beta_{10j} = \gamma_{100} \text{ (NHPI)} \]
\[ \beta_{11j} = \gamma_{110} \text{ (Two or More)} \]
where \( \gamma_{00} \) is a level 2 coefficient,
where \( \gamma_{01} \) is a level 2 coefficients for SCH online/paper mode,
where \( \gamma_{02} \) is a level 2 coefficient for the cross-level interaction of SCH online/paper mode with student online/paper mode,
\( W_{1j} \) is level 2 value for SCH online/paper mode in the jth school,
and \( U_{0j} \) is level 2 random effect.

When parameters of a statistical model are estimated by maximum likelihood (ML) method, the estimation also provides a likelihood which can be transformed into the deviance defined as minus twice the natural logarithm of the likelihood (Snijders & Bosker, 1999). The difference in deviance between two models fitted to the same data set can be used along with the difference in degrees of freedom as a test statistic having a chi-squared distribution to determine whether the model fits better. For models that fit better, the added parameter was kept in the model, while for models that do not fit better, the added parameter was removed from the model. This procedure was followed for each of the fourteen parameters. This final model served as the starting point for the next step, adding school variables (level 2).

**HLM Analysis—Step Four—Add Level 2 Explanatory Variables and Level 1 Interactions with Student Administrative Mode.** The fourth step adds explanatory variables at the school level to estimate the between-school or Level 2 model (Snijders &
Bosker, 1999). School characteristics associated with Mathematics achievement are investigated at this step. The purpose of adding school-level variables to the model is still reduction in outcome variability. School variables such as school administrative mode (SCH online/paper mode), school percent of free and reduced lunch (PCT_FRL), school percent of male students (PCT_MALE), school percent of SPED students (PCT_SPED), school percent of ELL students (PCT_ELL), school percent of white students (PCT_WH), school percent of Hispanic students (PCT_HIS), school percent of American Indian/Alaska native students (PCT_AMI), school percent of Asian students (PCT_ASI), school percent of black students (PCT_BL), school percent of native Hawaiian/other Pacific students (PCT_NHPI), and school percent of students with two or more races (PCT_TWO) were added to the model.

Both the residual variance as well as the random intercept variance can be reduced by adding school level variables. Interactions between level-1 variables and student administrative mode were added to investigate any joint effects of the level-1 variables and the student administrative mode. When no effects are observed over and above the separate effects of the level-1 variables, the variables are said to not interact. Any significant interaction effect between a level 1 variable and the student administrative mode would indicate a joint effect as a consequence of the combined variables beyond the separate effect of each variable. A joint effect with administrative mode and a specific demographic characteristic indicated that a student’s mathematics achievement score was affected by the online versus paper mode. A significant positive value would indicate that the combination enhances the effect of each other, whereas a significant negative value would indicate that the combination operate at cross purposes.
Level 1:

\[ Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{ij} + \beta_3 X_{ij} + \beta_4 X_{ij} + \beta_5 X_{ij} + \beta_6 X_{ij} + \beta_7 X_{ij} + \beta_8 X_{ij} + \beta_9 X_{ij} + \beta_{10} X_{ij} + \beta_{11} X_{ij} + \beta_{12} (X_{ij} \times X_{ij}) + \beta_{13} (X_{ij} + X_{ij}) + \beta_{14} (X_{ij} + X_{ij}) + \beta_{15} (X_{ij} + X_{ij}) + \beta_{16} (X_{ij} + X_{ij}) + \beta_{17} (X_{ij} + X_{ij}) + \beta_{18} (X_{ij} + X_{ij}) + \beta_{19} (X_{ij} + X_{ij}) + \beta_{20} (X_{ij} + X_{ij}) + \beta_{21} (X_{ij} + X_{ij}) + R_{ij} \]

where \( Y_{ij} \) represents the mathematics achievement of \( i \)th student in the \( j \)th school,
\( \beta_{0j} \) is the intercept, the mean mathematics achievement in the \( j \)th school,
\( \beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j}, \beta_{5j}, \beta_{6j}, \beta_{7j}, \beta_{8j}, \beta_{9j}, \beta_{10j}, \) and \( \beta_{11j} \) are the slopes for student online/paper mode, Free and Reduced Lunch, gender, special education, ELL, and ethnicity (six dummy variables) in the \( j \)th school,
\( \beta_{12j}, \beta_{13j}, \beta_{14j}, \beta_{15j}, \beta_{16j}, \beta_{17j}, \beta_{18j}, \beta_{19j}, \beta_{20j}, \) and \( \beta_{21j} \) are the slopes for student online/paper mode interactions with Free and Reduced Lunch, gender, special education, ELL, and ethnicity (six dummy variables) in the \( j \)th school,
\( X_{ij} \) represents the values of student online/paper mode, Free and Reduced Lunch, gender, special education, ELL, and ethnicity (six dummy variables) of \( i \)th student in the \( j \)th school,
\( (X_{ij} \times X_{ij}) \) represents the values of student online/paper mode with Free and Reduced Lunch, gender, special education, ELL, and ethnicity (six dummy variables) of \( i \)th student in the \( j \)th school, and
\( R_{ij} \) is the random error of \( i \)th student in the \( j \)th school.

Level 2:

\[ \beta_0 = \gamma_{00} + \gamma_{01} W_{1j} + \gamma_{02} W_{2j} + \gamma_{03} W_{3j} + \gamma_{04} W_{4j} + \gamma_{05} W_{5j} + \gamma_{06} W_{6j} + \gamma_{07} W_{7j} + \gamma_{08} W_{8j} + \gamma_{09} W_{9j} + \gamma_{10} W_{10j} + \gamma_{11} W_{11j} + \gamma_{12} (W_{1j} \times X_{ij}) + U_{0j} \]
\[ \beta_{1j} = \gamma_{10} \text{ (online/paper admin mode)} \]
\[ \beta_{2j} = \gamma_{20} \text{ (FRL)} \]
\[ \beta_{3j} = \gamma_{30} \text{ (Gender)} \]
\[ \beta_{4j} = \gamma_{40} \text{ (SPED)} \]
\[ \beta_{5j} = \gamma_{50} \text{ (ELL)} \]
\[ \beta_{6j} = \gamma_{60} \text{ (HISP)} \]
\[ \beta_{7j} = \gamma_{70} \text{ (AMIN)} \]
\[ \beta_{8j} = \gamma_{80} \text{ (ASIA)} \]
\[ \beta_{9j} = \gamma_{90} \text{ (BLAC)} \]
\[ \beta_{10j} = \gamma_{100} \text{ (NHPI)} \]
\[ \beta_{11j} = \gamma_{110} \text{ (Two or More)} \]
\[ \beta_{12j} = \gamma_{120} \text{ (Stud_mode * FRL)} \]
\[ \beta_{13j} = \gamma_{130} \text{ (Stud_mode * Gender)} \]
\[ \beta_{14j} = \gamma_{140} \text{ (Stud_mode * SPED)} \]
\[ \beta_{15j} = \gamma_{150} \text{ (Stud_mode * ELL)} \]
\[ \beta_{16j} = \gamma_{160} \text{ (Stud_mode * HISP)} \]
\[ \beta_{17j} = \gamma_{170} \text{ (Stud_mode * AMIN)} \]
\[ \beta_{18j} = \gamma_{180} \text{ (Stud_mode * ASIA)} \]
\[ \beta_{19j} = \gamma_{190} \text{ (Stud\_mode \times BLAC)} \]
\[ \beta_{20j} = \gamma_{200} \text{ (Stud\_mode \times NHPI)} \]
\[ \beta_{21j} = \gamma_{210} \text{ (Stud\_mode \times Two or More)} \]

where \( \gamma_{00}, \ldots, \gamma_{012} \) are level 2 coefficients for SCH online/paper mode,

PCT\_FRL, PCT\_MALE, PCT\_SPED, PCT\_ELL, PCT\_WH, PCT\_HIS,
PCT\_AMI, PCT\_ASI, PCT\_BL, PCT\_NHPI, PCT\_TWO, and the cross-
level interaction of stud\_online/paper mode with sch\_online/paper mode,

where \( \gamma_{100}, \ldots, \gamma_{212} \) are level 2 coefficients for SCH online/paper mode
interaction with Free and Reduced Lunch, gender, special education, ELL,
and ethnicity (six dummy variables) in the jth school,

\( W_{1j}, W_{2j}, W_{3j}, W_{4j} \ldots W_{11j} \) are level 2 values such as SCH online/paper mode,
PCT\_FRL, PCT\_MALE, PCT\_SPED, PCT\_ELL, PCT\_WH, PCT\_HIS,
PCT\_AMI, PCT\_ASI, PCT\_BL, PCT\_NHPI, PCT\_TWO (note: percents
centered on grand mean)

and \( U_{0j} \) is level 2 random effect.

Like the level-1 models, model fit with level-2 variables and level-1 interactions can be
tested using the difference in deviance between two models fitted to the same data set.

Attention to reduction in residual error and random intercept error guides the model
selection as well. Interactions between all level 1 variables and student administrative
mode were investigated. Any significant interaction between the level 1 variable and
student administrative mode indicated some joint effect in addition to the main effect of
each level-1 variable.

**Analysis of Model Fit for Grade 4**

**Step 1: Grade 4—Empty Model—Calculation of ICC.** Table 4.6 shows the
initial fit statistics for the Empty Model. Table 4.7 presents the parameters for the empty
model. The average school mean mathematics achievement was statistically different
from zero (\( \gamma_{00} = 40.82, t = 185.38, df = 534, p < .0001 \)). The intraclass correlation (ICC)
can be calculated by dividing between-schools variance by the total variance in the
model, \( ICC = \frac{\text{var}(U_{0j})}{\text{var}(R_{ij}) + \text{var}(U_{0j})} \). The table shows the following parameters.
\[ \text{Var}(U_{ij}) = \tau^2 = 69.25 \quad \text{var}(R_{ij}) = \sigma^2 = 22.73 \]

\[ \text{ICC} = \frac{\sigma^2}{(\tau^2 + \sigma^2)} = \frac{22.73/(69.25 + 22.73)}{22.73 / 91.98} = .247 \]

Using these parameters, the intraclass correlation coefficient is equal to .247, indicating that 25\% of the variability in mathematics achievement was between schools. The rest of the variability or 75\% of the total variability was within the school. This ICC is rather high, compared to other results of educational research (values between .05 and .20 are common). This indicates that the grouping according to schools leads to an important similarity between the results of different students in the same school, although the within-school differences are far larger than the between-schools differences. The total variability was 91.98. For model analysis, student level explanatory variables were selected to reduce the variance within schools and additional school level explanatory variables were added to explain between-school variance in the subsequent models.

Table 4.6

*Fit Statistics Fixed Effects and Random Effects of Empty Model for Grade 4 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.2 Log Likelihood</td>
<td>154098.7</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>154104.7</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>154117.6</td>
</tr>
</tbody>
</table>
Table 4.7

**Fixed Effects and Random Effects of Empty Model for Grade 4 Mathematics Achievement**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>40.8236</td>
<td>0.2199</td>
<td>534</td>
<td>185.38</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means ($\tau_{00}$)</td>
<td>22.7376</td>
<td>1.5634</td>
<td>14.54</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Var. within School ($\tau^2$)</td>
<td>69.2591</td>
<td>0.6747</td>
<td>102.64</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

**Step 2: Grade 4 Model—School and Student Administrative Mode Effects**

**Only.** Table 4.8 shows the results of the model fit statistics model improvements with the addition of the level-1 student administrative mode variable (paper/pencil = 0 and online = 1), the level-2 school administrative mode variable (paper/pencil = 0 and online = 1), the random effect of student administrative mode, and the cross-level interaction of student by school administrative mode. The addition of fixed effect of student admin mode to the empty model improved the model as shown by a deviance difference of 362.3 with $df = 1, p < .0001$. Likewise the addition of fixed effect of school admin mode to the fixed student admin mode model was also significantly better with a deviance difference of 192.3 with $df = 1, p < .0001$. The model showed improvement with the addition of random effect of student mode effect and the cross-level interaction of student by school admin mode effect with a deviance difference of 64.1 with $df = 1, p < .0001$ and a deviance difference of 8.6 with $df = 1, p = .003$, respectively.
Table 4.8

**Summary of Fit Statistics for Fixed Effects and Random Effects of Admin Mode Models for Grade 4 Mathematics Achievement**

<table>
<thead>
<tr>
<th>Models</th>
<th>Two Level Unconditional Model</th>
<th>Fixed Student Admin Mode</th>
<th>Fixed School Admin Mode</th>
<th>Random Student Admin Mode</th>
<th>Interaction Student with School Admin Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Statistics</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>df</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>.2 Log Likelihood</td>
<td>154098.7</td>
<td>153736.4</td>
<td>153544.1</td>
<td>153480.0</td>
<td>153471.4</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>154104.7</td>
<td>153744.4</td>
<td>153554.1</td>
<td>153494.0</td>
<td>153487.4</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>154117.6</td>
<td>153761.6</td>
<td>153575.6</td>
<td>153524.1</td>
<td>1253521.8</td>
</tr>
</tbody>
</table>

Tables 4.9 show the results of model improvement. The overall mean mathematics achievement score across schools was still significantly different from zero ($\gamma_{00} = 40.89, t = 100.32, df = 205, p < .0001$). After adding student and school online/paper-pencil administrative mode as variables for mathematics achievement, the random intercept variance was increased by 10.78% ($\frac{22.74 - 25.18}{22.74} = -10.78\%$), relative to the empty model. The residual variance was reduced by 3.25% ($\frac{69.26 - 67.01}{69.26} = 3.25\%$), relative to the empty model.

By including the cross-level interaction, the difference between four combinations of school and student administrative modes could be tested. The four combinations of administrative mode were student taking assessment on paper in school that selected paper assessments (stud mode = 0 and sch mode = 0), student taking assessment online
Table 4.9

*Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Explanatory Variables for Administrative Mode in Grade 4 Mathematics*

Achievement

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ( \gamma_{00} )</td>
<td>40.8917</td>
<td>0.4076</td>
<td>205</td>
<td>100.32</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>1.7375</td>
<td>2.5908</td>
<td>137</td>
<td>0.67</td>
<td>.5036</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>.1597</td>
<td>.04840</td>
<td>364</td>
<td>.33</td>
<td>.7416</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-9.5020</td>
<td>0.7603</td>
<td>260</td>
<td>-12.50</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online Paper * Sch Online Paper</td>
<td>7.9242</td>
<td>2.6684</td>
<td>141</td>
<td>2.97</td>
<td>&lt; .0035</td>
</tr>
</tbody>
</table>

Random Effects

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means ( \tau_{00} )</td>
<td>25.1829</td>
<td>2.7166</td>
<td>9.27</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>-12.7892</td>
<td>3.2804</td>
<td>-3.90</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>22.1321</td>
<td>5.1810</td>
<td>4.27</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Var. within School ( \tau^2 )</td>
<td>67.0058</td>
<td>0.6542</td>
<td>102.43</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

in school that selected paper assessments (stud mode = 1 and sch mode = 0), student taking assessment on paper in school that selected online assessments (stud mode = 0 and sch mode = 1), and student taking assessment online in school that selected online assessments (stud mode = 1 and sch mode = 1). Table 4.9 shows the estimates for the four different test admin modes. The overall mean mathematics achievement score across schools was still significantly different from zero \( \gamma_{00} = 40.89, t = 100.32, \)
There was no significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing compared to students taking the online assessment in schools selecting online testing (difference = 0.1597, \( t = .33, df = 364, p = .7416 \)) or students taking online assessment in schools selecting paper testing (difference = 1.7375, \( t = .67, df = 137, p = .5036 \)). There was a significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper assessment in schools selecting online testing (difference = -9.5002, \( t = -12.50, df = 260, p < .0001 \)).

**Step 3: Grade 4 Model—Addition of Student and School Explanatory Variables—Control for Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race.** Table 4.10 shows the results of the model fit statistics for model improvements with the addition of the level-1 variables for free and reduced lunch, gender, English language learner, students receiving special services, and ethnicity/race and level-2 variables for percent of students receiving free and reduced lunch along with percent of white students. Other level-2 variables such as percent of female students, percent of students receiving special services, percent of ELL students, and percent of other ethnicities other than white were not significant.

The model was improved with the addition of fixed student FRL effect (deviance \( \Delta = 1156.8, df = 4, p < .0001 \)), fixed school percent of FRL centered at the grand mean of .47 (deviance \( \Delta = 57.8, df = 1, p < .0001 \)), and random student FRL effect (deviance \( \Delta = 32.9, df = 3, p < .0001 \)). Adding the cross-level interaction of student and school FRL effects did not improve the model. Adding demographics for gender (deviance \( \Delta = 38.2, df = 4, p < .0001 \)), English language learners (deviance \( \Delta = 179.3, df = 1, p < .0001 \)), and
Table 4.10

*Summary of Fit Statistics for Fixed Effects and Random Effects of Level 1 and 2 Explanatory Variables for Grade 4 Mathematics*

*Achievement*

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Add FRL Fixed Student Effect (L1)</th>
<th>Add FRL Fixed School Effect (L2)</th>
<th>Add FRL Random Student Effect (L1)</th>
<th>Add Gender Fixed Student Effect (L1)</th>
<th>Add ELL Fixed Student Effect (L1)</th>
<th>Add SPED Fixed Student Effect (L1)</th>
<th>Add Seven Ethnicities Fixed Student Effect (L1)</th>
<th>Add Percent White Fixed School Effect (L2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>.2 Log Likelihood</td>
<td>153471.4</td>
<td>152387.3</td>
<td>152329.5</td>
<td>152296.6</td>
<td>152291.3</td>
<td>152111.7</td>
<td>150591.9</td>
<td>150194.8</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>153487.4</td>
<td>152405.3</td>
<td>152349.5</td>
<td>152322.6</td>
<td>152319.3</td>
<td>152141.7</td>
<td>150625.9</td>
<td>150238.8</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>153521.8</td>
<td>152444.0</td>
<td>152392.5</td>
<td>152378.5</td>
<td>152382.8</td>
<td>152206.2</td>
<td>150699.0</td>
<td>150333.4</td>
</tr>
</tbody>
</table>
Special Education (deviance $\Delta = 1515.2$, $df = 1$, $p < .0001$) improved model fit. With a deviance change of 396.5 ($df = 1$, $p < .0001$), the addition of all seven student ethnicity/race variables improved model fit as well. One more school effect, namely the percent of white students in the school, improved the model with deviance change of 5.3 ($df = 1$, $p < .0001$).

Table 4.11 shows the parameters of final model with level-1 variables and level-2 variables that improved the model along with student and school admin mode estimates. All parameters in the table are for students in schools with 47% FRL (FRL_47) and 70% white students (PCT_WH70). After including student variables for free and reduced lunch (FRL), gender, English language learner (ELL), special education (SPED), and ethnicity/race in the model along with school variable for percent free and reduced lunch, and percent white students, the residual variability was reduced by 16.33% $((69.26 - 57.93)/69.26 = 16.33\%)$, relative to the empty model. Likewise the variance for random intercept was reduced by 34.71% $((22.73 – 14.84)/22.73 = 34.71\%)$, relative to the empty model. The overall mean mathematics achievement score across schools with 47% FRL and 70% white students was significantly different from zero ($\gamma_{00} = 44.85$, $t = 132.35$, $df = 189$, $p < .0001$). There was no significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing (P/P) compared to students taking the online assessment in schools selecting online testing (O/O) (difference = -0.6509, $t = -1.66$, $df = 309$, $p = .0982$) or students taking online assessment in schools selecting paper testing (O/P) (difference = 3.0611, $t = 1.49$, $df = 104$, $p = .1405$). There was significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper
Table 4.11

Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Effect of Student Explanatory Variables for FRL, Gender, SPED, ELL and Ethnicity as well as School Level Explanatory Variables for Percent FRL, and Percent White on Grade 4 Mathematics Achievement

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>44.8530</td>
<td>0.3389</td>
<td>189</td>
<td>132.35</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>3.0611</td>
<td>2.0608</td>
<td>104</td>
<td>1.49</td>
<td>.1405</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>-0.6509</td>
<td>0.3924</td>
<td>309</td>
<td>-1.66</td>
<td>.0982</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-5.3587</td>
<td>0.6449</td>
<td>264</td>
<td>-8.31</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online Paper*Sch Online Paper</td>
<td>1.6467</td>
<td>2.1337</td>
<td>108</td>
<td>0.77</td>
<td>.4419</td>
</tr>
<tr>
<td>FRL_Program</td>
<td>-2.9683</td>
<td>0.1350</td>
<td>484</td>
<td>-21.99</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Percent FRL (center at .47)</td>
<td>-0.2623</td>
<td>0.0957</td>
<td>615</td>
<td>-2.74</td>
<td>.0063</td>
</tr>
<tr>
<td>Gender_Code</td>
<td>-0.8086</td>
<td>0.1059</td>
<td>21000</td>
<td>-7.64</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Special_Education_Code</td>
<td>-5.7123</td>
<td>0.1434</td>
<td>21000</td>
<td>-39.83</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL Code</td>
<td>-3.5094</td>
<td>0.2386</td>
<td>21000</td>
<td>-14.71</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E1d—Hispanic</td>
<td>-1.4159</td>
<td>0.1907</td>
<td>21000</td>
<td>-7.43</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E2d—Am Ind or Al Na</td>
<td>-3.5559</td>
<td>0.4989</td>
<td>21000</td>
<td>-7.13</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E3d—Asian</td>
<td>0.9710</td>
<td>0.3822</td>
<td>21000</td>
<td>2.54</td>
<td>0.0111</td>
</tr>
<tr>
<td>E4d—Black</td>
<td>-4.6207</td>
<td>0.2571</td>
<td>20000</td>
<td>-17.97</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E5d—Nat Haw or Pac Is</td>
<td>-1.8411</td>
<td>1.4594</td>
<td>21000</td>
<td>-1.26</td>
<td>0.2071</td>
</tr>
<tr>
<td>E6d—Two or More</td>
<td>-1.1097</td>
<td>0.3161</td>
<td>21000</td>
<td>-3.51</td>
<td>0.0004</td>
</tr>
<tr>
<td>Percent White (center at .70)</td>
<td>0.2170</td>
<td>0.0927</td>
<td>632</td>
<td>2.34</td>
<td>0.0196</td>
</tr>
</tbody>
</table>

Table 4.11 continues
<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means ($\tau_{00}$)</td>
<td>14.8374</td>
<td>1.9496</td>
<td>7.61</td>
<td>$&lt; .0001$</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>-8.0316</td>
<td>2.1824</td>
<td>-3.68</td>
<td>0.0002</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>12.6614</td>
<td>3.2933</td>
<td>3.84</td>
<td>$&lt; .0001$</td>
</tr>
<tr>
<td>Covar. (3,1)</td>
<td>1.6004</td>
<td>0.7935</td>
<td>2.02</td>
<td>0.0437</td>
</tr>
<tr>
<td>Covar. (3,2)</td>
<td>0.1419</td>
<td>0.9364</td>
<td>0.15</td>
<td>0.8795</td>
</tr>
<tr>
<td>Rand FRL Slope Var. (3)</td>
<td>0.6922</td>
<td>0.4208</td>
<td>1.65</td>
<td>0.0500</td>
</tr>
<tr>
<td>Var. within School ($\tau^2$)</td>
<td>57.9319</td>
<td>0.5701</td>
<td>101.62</td>
<td>$&lt; .0001$</td>
</tr>
</tbody>
</table>

assessment in schools selecting online testing (P/O) (difference = -5.3587, $t = -8.31$, $df = 264$, $p < .0001$).

In schools where 47% of the students are eligible for FDL and 70% of the students are white, white male students not receiving SPED service and not qualified for ELL support, but qualified for Free or Reduced Lunch (FRL) performed 2.97 points lower on mathematics achievement than students not qualified for FRL ($\gamma_{30} = -2.9683$, $t = -21.99$, $df = 484$, $p < .0001$). Since the model included a random effect of student FRL, an examination of the variability of the difference between kids who get FRL and kids who do not get FRL indicated that the difference does vary significantly over schools (estimate = .6922, $z = 1.65$, $p = .0500$). On average the gap related to lunch status is 2.97 lower than the overall average, but across 95% of the schools, that gap is predicted to be anywhere from 4.62 lower to 1.36 lower (-2.9683 ± 1.96*SQRT(.6922) = -2.9863 ± 1.6307 = -4.62 to -1.36).

Likewise in these same schools with 47% FDL and 70% white students, the white male students not qualified for FRL and not receiving ELL support, but receiving for
SPED Services performed 5.71 points lower on mathematics achievement than students not receiving Special Education Services ($\gamma_{40} = -5.7123, t = -39.83, df = 21000, p < .0001$). In these same average schools, white male students not qualified for FRL and not receiving Special Education Services, but qualified for ELL support performed 3.51 points lower on mathematics achievement than student not identified for English language support ($\gamma_{70} = -3.5322, t = -14.71, df = 21000, p < .0001$). Likewise, the female students who were not qualified for FRL, not receiving ELL support, and not receiving Special Education Services performed .81 points lower on mathematics achievement than male students ($\gamma_{90} = -0.8086, t = -7.64, df = 21000, p < .0001$).

E1d through E6d were dummy coding variables for the ethnicity compared to the baseline variable, white students. Compared to the white Americans in the schools where 70% of the students are white and 47% of the students are eligible for FRL, the Hispanic American students ($\gamma_{60} = -1.4159, t = -7.43, df = 21000, p < .0001$), the American Indian/Alaska Native students ($\gamma_{70} = -3.5559, t = -7.13, df = 21000, p < .0001$), the Black/African American students ($\gamma_{90} = -4.6207, t = -17.97, df = 21000, p < .0001$), and the students with two or more races ($\gamma_{110} = -1.1097, t = -3.51, df = 21000, p = .0004$) performed significantly below White American students. The Asian American students’ performance was significantly above the performance by White American students ($\gamma_{80} = .9710, t = 2.54, df = 21000, p = .0111$) while no significant difference was found between the performance of Native Hawaiian or other Pacific Islander students and White Americans ($\gamma_{100} = -1.8411, t = -1.26, df = 21000, p = .2071$).

In addition to student characteristics, some school characteristics namely, the percent of students in the school who receive FRL and the percent of white students in
the school showed some significant improvements in the model. The school’s mean mathematics achievement score is estimated to be .2623 points lower for each 10% increase in the percentage of students who receive free/reduced lunch above 47% ($\gamma_{100} = -0.2623, t = -2.74, df = 21000, p = .0063$). Likewise for a 10% decrease in percent of students receiving FRL, the school’s mean mathematics achievement score is estimated to be higher by .2623 points for each decrease. The school’s mean mathematics achievement score is estimated to be .2170 points higher for each 10% increase in the percentage of white students above 70% ($\gamma_{100} = 0.2170, t = 2.34, df = 632, p = 0.0196$). Likewise for a 10% decrease in percent of white students below 70%, the school’s mean mathematics achievement score is estimated to be lower by .2170 points for each decrease. In Appendix B, Tables B1and B2 show the sequences of models along with model fit statistics and Pseudo R for each 4th grade model. The SAS program for the final 4th grade model is listed in Program C.1 in Appendix C.

**Step 4: Grade 4 Model—Addition of Interaction of Student Administrative Mode and Student Level Variables—Effects of Mode on Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity Race.** To ascertain the effect of administrative mode on five student level variables – free and reduced lunch, gender, ELL, SPED, and ethnicity/race (seven groups), interactions with each student level variable were added to the model. Of the 12 interactions only interactions between SPED, gender, free and reduced lunch, and ELL with administrative mode were significant. The model with these significant interactions was a better fit based on a deviance difference between the model with the interactions and the previous model with school level variables of percent white (deviance $\Delta = 42.1, df = 4, p < .0001$).
See Table 4.12 for model fit statistics for the addition of each of the significant interactions.

Table 4.12

Summary of Fit Statistics for Addition of Interactions with Administrative Mode with FRL, Gender, SPED, and ELL for Grade 4 Mathematics Achievement

<table>
<thead>
<tr>
<th>Models</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>.2 Log Likelihood</td>
<td>150172.6</td>
<td>150165.3</td>
<td>150156.2</td>
<td>150147.4</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>150220.6</td>
<td>150215.3</td>
<td>150208.2</td>
<td>150201.4</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>150323.7</td>
<td>150322.8</td>
<td>150319.9</td>
<td>150317.5</td>
</tr>
</tbody>
</table>

Table 4.13 shows the parameters of model with the four significant interactions of administrative mode and demographic variables. All parameters in the table are for students in schools with 47% (FRL (FRL_47) and 70% white students (PCT_WH70).

After including four significant interactions in the model along with other significant student and school variables, the residual variability was reduced by 16.53% ((69.26 – 57.81)/69.26 = 16.53%), relative to the empty model. Likewise the variance for random intercept was reduced by 34.89% ((22.73 – 14.80)/22.73 = 34.89%), relative to the empty model.
Table 4.13

*Estimates of Mathematics Achievement for the Four Administrative Groups Using the Previous Model with Effect of Interactions Between Student Administrative Mode and Student Explanatory Variables for FRL, Gender, SPED, and ELL on Grade 4*

**Mathematics Achievement**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ( (γ_{00}) )</td>
<td>44.6474</td>
<td>0.3481</td>
<td>204</td>
<td>128.24</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>3.7985</td>
<td>2.0606</td>
<td>102</td>
<td>1.84</td>
<td>0.0682</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>-0.3228</td>
<td>0.4142</td>
<td>363</td>
<td>-0.78</td>
<td>0.4363</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-5.3555</td>
<td>0.6541</td>
<td>279</td>
<td>-8.19</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online Paper*Sch Online Paper</td>
<td>1.2342</td>
<td>2.1226</td>
<td>103</td>
<td>0.58</td>
<td>0.5622</td>
</tr>
<tr>
<td>FRL_Program</td>
<td>-2.5352</td>
<td>0.2150</td>
<td>425</td>
<td>-11.84</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Percent FRL (center at .47)</td>
<td>-0.2551</td>
<td>0.0957</td>
<td>616</td>
<td>-2.67</td>
<td>0.0079</td>
</tr>
<tr>
<td>Gender_Code</td>
<td>-0.4661</td>
<td>0.1683</td>
<td>21000</td>
<td>-2.77</td>
<td>0.0056</td>
</tr>
<tr>
<td>Special_Education_Code</td>
<td>-6.3346</td>
<td>0.2247</td>
<td>21000</td>
<td>-28.19</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL Code</td>
<td>-2.9624</td>
<td>0.3166</td>
<td>20000</td>
<td>-9.36</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E1d—Hispanic</td>
<td>-1.4069</td>
<td>0.1905</td>
<td>21000</td>
<td>-7.38</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E2d—Am Ind or Al Na</td>
<td>-3.5593</td>
<td>0.4985</td>
<td>21000</td>
<td>-7.14</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E3d—Asian</td>
<td>0.9160</td>
<td>0.3821</td>
<td>21000</td>
<td>2.40</td>
<td>0.0165</td>
</tr>
<tr>
<td>E4d—Black</td>
<td>-4.5958</td>
<td>0.2569</td>
<td>20000</td>
<td>-17.89</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E5d—Nat Haw or Pac Is</td>
<td>-1.8318</td>
<td>1.4584</td>
<td>21000</td>
<td>-1.26</td>
<td>0.2091</td>
</tr>
<tr>
<td>E6d—Two or More</td>
<td>-1.0835</td>
<td>0.3158</td>
<td>21000</td>
<td>-3.43</td>
<td>0.0006</td>
</tr>
<tr>
<td>Percent White (center at .70)</td>
<td>0.2233</td>
<td>0.0928</td>
<td>633</td>
<td>2.41</td>
<td>0.0164</td>
</tr>
</tbody>
</table>

Table 4.13 continues
### Fixed Effects

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stud Online * Special Ed</td>
<td>1.0725</td>
<td>0.2917</td>
<td>21000</td>
<td>3.68</td>
<td>0.0002</td>
</tr>
<tr>
<td>Stud Online * Gender</td>
<td>-0.5649</td>
<td>0.2163</td>
<td>21000</td>
<td>-2.61</td>
<td>0.0099</td>
</tr>
<tr>
<td>Stud Online * FRL</td>
<td>-0.6882</td>
<td>0.2665</td>
<td>499</td>
<td>-2.58</td>
<td>0.0101</td>
</tr>
<tr>
<td>Stud Online * ELL</td>
<td>-1.3215</td>
<td>0.4428</td>
<td>20000</td>
<td>-2.96</td>
<td>0.0031</td>
</tr>
</tbody>
</table>

### Random Effects

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means (τ₀₀)</td>
<td>14.8043</td>
<td>1.9307</td>
<td>7.67</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>-7.8745</td>
<td>2.1806</td>
<td>-3.61</td>
<td>0.0003</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>12.3803</td>
<td>3.2909</td>
<td>3.76</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (3,1)</td>
<td>1.3795</td>
<td>0.7854</td>
<td>1.76</td>
<td>0.0790</td>
</tr>
<tr>
<td>Covar. (3,2)</td>
<td>0.4851</td>
<td>0.9328</td>
<td>0.52</td>
<td>0.6031</td>
</tr>
<tr>
<td>Rand FRL Slope Var. (3,3)</td>
<td>0.7307</td>
<td>0.4243</td>
<td>1.72</td>
<td>0.0425</td>
</tr>
<tr>
<td>Var. within School (τ²)</td>
<td>57.8137</td>
<td>0.5689</td>
<td>101.63</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

The overall mean mathematics achievement score across schools with 47% FRL and 70% white students was significantly different from zero ($\gamma_{00} = 44.65$, $t = 128.24$, $df = 204$, $p < .0001$). There was no significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing (P/P) compared to students taking the online assessment in schools selecting online testing (O/O) (difference = -0.3228, $t = -0.78$, $df = 363$, $p = .4363$) or students taking online assessment in schools selecting paper testing (O/P) (difference = 3.7985, $t = 1.84$, $df = 102$, $p = .0682$). There was a significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper assessment in
schools selecting online testing (P/O) (difference = -5.3555, \( t = -8.19, df = 279, p < .0001 \)).

In this revised model, the significant interaction between special education and administrative mode indicated that the effect of receiving special education services (-6.3346) on mathematics achievement was less negative by 1.0725 \( (df = 21000, t = 3.68, p = .0002) \) for students taking the assessment online. On the other hand, the significant interaction between gender and administrative mode indicated that the effect for female students (-0.4661) on their mathematics achievement was more negative by 0.5649 \( (df = 21000, t = -2.61, p = .0099) \) for female students taking the assessment online. In addition, the significant interaction of students eligible free or reduced lunch with administrative mode indicated that the effect of being eligible for FRL (-2.5352) on their mathematics achievement was more negative by 0.6882 \( (df = 499, t = -2.58, p = .0101) \) for FRL students taking the assessment online. Finally, the significant interaction between students receiving English language learner services and administrative mode indicated that the effect of receiving ELL services (-2.9624) on mathematics achievement was more negative by 1.3215 \( (df = 20000, t = -2.96, p = .0031) \) for students taking the assessment online.

**Analysis of Model Fit for Grade 8**

**Step 1: Grade 8—Empty Model—Calculation of ICC.** Table 4.14 shows the initial fit statistics for the Empty Model. Table 4.15 presents the parameters for the empty model. Average school mean mathematics achievement was statistically different from zero \( (\gamma_{00} = 44.17, t = 143.78, df = 295, p < .0001) \). The intraclass correlation (ICC)
Table 4.14

*Fit Statistics Fixed Effects and Random Effects of Empty Model for Grade 8 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2 Log Likelihood</td>
<td>156069.9</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>156075.9</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>156087.0</td>
</tr>
</tbody>
</table>

Table 4.15

*Fixed Effects and Random Effects of Empty Model for Grade 8 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>44.1738</td>
<td>0.3072</td>
<td>295</td>
<td>143.78</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means ($\tau_{00}$)</td>
<td>23.2255</td>
<td>2.2526</td>
<td>10.31</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Var. within School ($\tau^2$)</td>
<td>113.10</td>
<td>1.1238</td>
<td>100.64</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

can be calculated by dividing between-schools variance by the total variance in the model, $\text{ICC} = \frac{\text{var}(U_{0j})}{\text{var}(R_{ij}) + \text{var}(U_{0j})}$. The table shows the following parameters.

\[
\text{Var}(U_{0j}) = \tau^2 = 113.10 \quad \text{var}(R_{ij}) = \sigma^2 = 23.23
\]

\[
\text{ICC} = \sigma^2 / (\tau^2 + \sigma^2) = 23.23 / (113.10 + 23.23) = 23.23 / 136.33 = .170
\]

Using these parameters, the intraclass correlation coefficient is equal to .170, indicating that 17% of the variability in mathematics achievement was between schools. The 83% of the total variability was within the school. This ICC is about average,
compared to other results of educational research (values between .05 and .20 are common). This indicates that the grouping according to schools leads to an important similarity between the results of different students in the same school, although the within-school differences are far larger than the between-schools differences. The total variability was 136.33. For model analysis, student level explanatory variables were selected to reduce the variance within schools and additional school level explanatory variables were added to explain between-school variance in the subsequent models.

**Step 2: Grade 8 Model—School and Student Administrative Mode Effects Only.** Table 4.16 shows the model fit statistics for improvements made by the addition of the level-1 student administrative mode variable (paper/pencil = 0 and online = 1), the level-2 school administrative mode variable (paper/pencil = 0 and online = 1), the random effect of student administrative mode, and the cross-level interaction of student by school administrative mode. The addition of fixed effect of student admin mode to the empty model improved the model as shown by a deviance difference of 234.5 with \( df = 1 \), \( p < .0001 \). Likewise the addition of fixed effect of school admin mode to the fixed student admin mode model was also significantly better with a deviance difference of 119.2 with \( df = 1 \), \( p < .0001 \). The model showed improvement with the addition of random effect of student mode effect and the cross-level interaction of student by school admin mode effect with a deviance difference of 84.1 with \( df = 1 \), \( p < .0001 \) and a deviance difference of 106.1 with \( df = 1 \), \( p < .0001 \), respectively.

Table 4.17 shows the parameters for the improved model. The overall mean mathematics achievement score across schools was still significantly different from zero \( (\gamma_{00} = 43.29, t = 77.86, df = 99.4, p < .0001) \). After adding student and school
Table 4.16

*Summary of Fit Statistics for Fixed Effects and Random Effects of Admin Mode Models for Grade 8 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Models</th>
<th>Two Level Unconditional Model</th>
<th>Fixed Student Admin Mode</th>
<th>Fixed School Admin Mode</th>
<th>Random Student Admin Mode</th>
<th>Cross-level Student with School Admin Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>.2 Log Likelihood</td>
<td>AIC (smaller is better)</td>
<td>BIC (smaller is better)</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>-------------------------</td>
<td>---------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>156069.9</td>
<td>156075.9</td>
<td>156087.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>155835.4</td>
<td>155843.4</td>
<td>155858.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>155716.2</td>
<td>155726.2</td>
<td>155744.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>155632.1</td>
<td>155646.1</td>
<td>155672.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>155526.0</td>
<td>155542.0</td>
<td>155571.6</td>
<td></td>
</tr>
</tbody>
</table>

online/paper-pencil administrative mode as variables for mathematics achievement, the random intercept variance was decreased by 3.92% \(((23.23 – 22.32)/23.23 = 3.92%)\), relative to the empty model. The residual variance was reduced by 2.72% \(((113.10 – 110.02)/113.10 = 2.72\%)\), relative to the empty model.

By including the cross-level interaction, the difference between four combinations of school and student administrative modes could be tested. The four combinations of administrative mode were student taking assessment on paper in school that selected paper assessments (stud mode = 0 and sch mode = 0), student taking assessment online in school that selected paper assessments (stud mode = 1 and sch mode = 0), student taking assessment on paper in school that selected online assessments (stud mode = 0 and sch mode = 1), and student taking assessment online in school that selected online assessments (stud mode = 1 and sch mode = 1). Table 4.17 shows the estimates for the
Table 4.17

*Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Explanatory Variables for Administrative Mode in Grade 8 Mathematics*

**Achievement**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>43.2899</td>
<td>0.5560</td>
<td>99.4</td>
<td>77.86</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-29.5140</td>
<td>2.8159</td>
<td>98.2</td>
<td>-10.48</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>1.5227</td>
<td>0.6664</td>
<td>181</td>
<td>2.28</td>
<td>0.235</td>
</tr>
<tr>
<td>Stud Online Paper * Sch Online Paper</td>
<td>31.0367</td>
<td>2.7237</td>
<td>95.1</td>
<td>11.39</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means ($\tau_{00}$)</td>
<td>22.3228</td>
<td>3.872</td>
<td>6.22</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>-5.3407</td>
<td>3.6574</td>
<td>-1.46</td>
<td>0.1442</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>11.9403</td>
<td>5.2459</td>
<td>2.28</td>
<td>0.0114</td>
</tr>
<tr>
<td>Var. within School ($r^2$)</td>
<td>110.02</td>
<td>1.0940</td>
<td>100.56</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

four different test admin modes. The overall mean mathematics achievement score across schools was still significantly different from zero ($\gamma_{00} = 43.29$, $t = 77.86$, $df = 99.4$, $p < .0001$). There was a significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing compared to students taking the online assessment in schools selecting online testing (difference = 1.5227, $t = 2.28$, $df = 182$, $p = .0235$) or students taking online assessment
in schools selecting paper testing (difference = -13.4834, \( t = -12.48, df = 142, p < .0001 \)). Likewise, there was a significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper assessment in schools selecting online testing (difference = -29.5140, \( t = -10.48, df = 98.2, p < .0001 \)).

**Step 3: Grade 8 Model—Addition of Student and School Explanatory Variables—Control for Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race.** Table 4.18 shows the results of the model fit statistics for model improvements with the addition of the level-1 variables for free and reduced lunch, gender, English language learner, students receiving special services, and ethnicity/race and level-2 variables for percent of students receiving free and reduced lunch, percent of students receiving special services, and percent of white students. Other level-2 variables such as percent of female students, percent of ELL students, and percent of ethnicities other than white were not significant.

The model was improved with the addition of fixed student FRL effect (deviance \( \Delta = 1623.6, df = 4, p < .0001 \)), fixed school percent of FRL centered at the grand mean of .43 (deviance \( \Delta = 39.1, df = 1, p < .0001 \)), and random student FRL effect (deviance \( \Delta = 84.8, df = 3, p < .0001 \)). Adding the cross-level interaction of student and school FRL effects improved the model as well (deviance \( \Delta = 10.5, df = 1, p = .0012 \)). Adding demographics for gender (deviance \( \Delta = 101.8, df = 4, p < .0001 \)), English language learners (deviance \( \Delta = 282.2, df = 1, p < .0001 \)), and Special Education (deviance \( \Delta = 2840.6, df = 1, p < .0001 \)) improved model fit. In addition to student level SPED variable, the effect of the fixed school percent of SPED students centered at the grand
Table 4.18

Summary of Fit Statistics for Fixed Effects and Random Effects of Level 1 and 2 Explanatory Variables for Grade 8 Mathematics

Achievement

<table>
<thead>
<tr>
<th>Interaction Student with School Admin Mode</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add FRL Fixed Student Effect (L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add FRL Fixed School Effect (L2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add FRL Random Student Effect (L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add FRL School Effect (L2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Interaction FRL Student with FRL School Effect (L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Gender Fixed Student Effect (L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Gender Fixed School Effect (L2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add ELL Fixed Student Effect (L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add ELL Fixed School Effect (L2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add SPED Fixed Student Effect (L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add SPED Fixed School Effect (L2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Seven Ethnicities Fixed Student Effect (L1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Percent White Fixed School Effect (L2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>df</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2 Log Likelihood</td>
<td>155526.0</td>
<td>154092.6</td>
<td>154053.5</td>
<td>153968.7</td>
<td>153958.2</td>
<td>153951.7</td>
<td>153669.5</td>
<td>150828.9</td>
<td>150818.9</td>
<td>150393.6</td>
<td>150375.0</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>155542.0</td>
<td>154110.6</td>
<td>154073.5</td>
<td>153994.7</td>
<td>153986.2</td>
<td>153981.7</td>
<td>153701.5</td>
<td>150862.9</td>
<td>150854.9</td>
<td>150441.6</td>
<td>150426.1</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>155571.6</td>
<td>154143.9</td>
<td>154110.6</td>
<td>154042.9</td>
<td>154038.1</td>
<td>154037.3</td>
<td>153760.9</td>
<td>150925.9</td>
<td>150921.6</td>
<td>150530.5</td>
<td>150517.7</td>
</tr>
</tbody>
</table>
mean of .14 also improved the model fit (deviance $\Delta = 10$, $df = 1$, $p = .0016$). With a deviance change of 425.3 ($df = 6$, $p < .0001$), the addition of the student ethnicity/race improved model fit as well. One more school effect, namely the percent of white students in the school centered at the grand mean of .72, improved the model with deviance change of 18.6 ($df = 1$, $p < .0001$).

Table 4.19 shows the parameters of final model with level-1 variables and level-2 variables that improved the model along with student and school admin mode estimates. All parameters in the table are for students in schools with 43% FRL (FRL_43), 14% SPED (PCT_SPED14), and 72% white students (PCT_WH72). After including student variables for free and reduced lunch (FRL), gender, English language learner (ELL), special education (SPED), and ethnicity/race in the model along with school variable for percent free and reduced lunch, percent receiving SPED services, and percent white students, the residual variability was reduced by 24.02% ($(113.10 - 85.93)/113.10 = 24.02\%$), relative to the empty model. Likewise the variance for random intercept was reduced by 45.85% ($(22.23 - 12.58)/22.23 = 45.85\%$), relative to the empty model. The overall mean mathematics achievement score across schools with 47% FRL, 14% SPED, and 70% white students was significantly different from zero ($\gamma_{00} = 47.0221$, $t = 106.16$, $df = 84.9$, $p < .0001$). There was no significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing (P/P) compared to students taking the online assessment in schools selecting online testing (O/O) (difference = 0.9738, $t = 1.96$, $df = 133$, $p = .0518$). There was significant difference between students taking the paper assessment in schools selecting paper testing versus students taking online assessment in schools selecting paper testing (O/P)
Table 4.19

*Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Effect of Student Explanatory Variables for FRL, Gender, SPED, ELL and Ethnicity as well as School Level Explanatory Variables for Percent FRL, Percent SPED, and Percent White on Grade 8 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (γ₀₀)</td>
<td>47.0221</td>
<td>0.4429</td>
<td>84.9</td>
<td>106.16</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-11.5164</td>
<td>2.3618</td>
<td>449</td>
<td>-4.88</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>0.9738</td>
<td>0.4962</td>
<td>133</td>
<td>1.96</td>
<td>0.518</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-4.2705</td>
<td>0.9552</td>
<td>201</td>
<td>-4.47</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online Paper*Sch Online Paper</td>
<td>16.7308</td>
<td>2.5294</td>
<td>434</td>
<td>6.63</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>FRL_Program</td>
<td>-3.8417</td>
<td>0.1891</td>
<td>221</td>
<td>-20.32</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Percent FRL (center at .43)</td>
<td>-0.0518</td>
<td>0.1386</td>
<td>341</td>
<td>-0.37</td>
<td>.7090</td>
</tr>
<tr>
<td>Stud FRL * Percent FRL</td>
<td>-0.2558</td>
<td>0.0916</td>
<td>175</td>
<td>-2.79</td>
<td>.0058</td>
</tr>
<tr>
<td>Gender_Code</td>
<td>-0.4976</td>
<td>0.1314</td>
<td>20000</td>
<td>-3.79</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Special_Education_Code</td>
<td>-11.0383</td>
<td>0.1990</td>
<td>20000</td>
<td>-55.47</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Percent SPED (center at .14)</td>
<td>-0.8021</td>
<td>0.2819</td>
<td>449</td>
<td>-2.85</td>
<td>.0046</td>
</tr>
<tr>
<td>ELL_Code</td>
<td>-7.1452</td>
<td>0.3991</td>
<td>20000</td>
<td>-17.90</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E1d—Hispanic</td>
<td>-2.6843</td>
<td>0.2296</td>
<td>20000</td>
<td>-11.69</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E2d—Am Ind or Al Na</td>
<td>-3.4878</td>
<td>0.6369</td>
<td>18000</td>
<td>-5.48</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E3d—Asian</td>
<td>2.5263</td>
<td>0.4817</td>
<td>20000</td>
<td>5.25</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E4d—Black</td>
<td>-5.1846</td>
<td>0.3173</td>
<td>19000</td>
<td>-16.34</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E5d—Nat Haw or Pac Is</td>
<td>1.9907</td>
<td>1.9553</td>
<td>20000</td>
<td>1.02</td>
<td>0.3087</td>
</tr>
</tbody>
</table>

Table 4.19 continues
<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>E6d—Two or More</td>
<td>-1.7102</td>
<td>0.3997</td>
<td>20000</td>
<td>-4.28</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Percent White (center at .72)</td>
<td>0.5498</td>
<td>0.1252</td>
<td>316</td>
<td>4.39</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means (τ₀₀)</td>
<td>12.5816</td>
<td>2.5349</td>
<td>4.96</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>-11.2106</td>
<td>2.9690</td>
<td>-3.78</td>
<td>0.0002</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>17.8009</td>
<td>4.3046</td>
<td>4.14</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (3,1)</td>
<td>-0.9114</td>
<td>0.9643</td>
<td>-0.95</td>
<td>0.3446</td>
</tr>
<tr>
<td>Covar. (3,2)</td>
<td>2.7783</td>
<td>1.0629</td>
<td>2.61</td>
<td>0.0090</td>
</tr>
<tr>
<td>Rand FRL Slope Var. (3,3)</td>
<td>1.7199</td>
<td>0.5569</td>
<td>3.09</td>
<td>0.0010</td>
</tr>
<tr>
<td>Var. within School (τ²)</td>
<td>85.9295</td>
<td>0.8574</td>
<td>100.22</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

(difference = -11.5164, \( t = -4.88 \), \( df = 449 \), \( p < .0001 \)) and also students taking paper assessment in schools selecting online testing (P/O) (difference = -4.2705, \( t = -4.47 \), \( df = 201 \), \( p < .0001 \)).

In schools with 47% FRL, 14% SPED, and 70% white students, white male student not receiving SPED service or not qualified for ELL support, but qualified for Free or Reduced Lunch (FRL) performed 3.8417 points lower on mathematics achievement on average than a student not qualified for FRL (\( \gamma_{30} = -3.8417 \), \( t = -20.32 \), \( df = 221 \), \( p < .0001 \)). Since the model included a random effect of student FRL, the difference between kids who get free/reduced lunch and kids who do not varies significantly over schools. On the average, the gap related to lunch status is 3.8417 points lower, but across 95% of the schools, that gap is predicted to be anywhere from
1.27 lower to 6.41 lower \((-3.8417 \pm 1.96*\text{SQRT}(1.7199)) = -3.8417 \pm 2.5704 = -1.27 \) to \(-6.41\).

Likewise in these same schools with 47% FDL, 14% SPED, and 70% white students, the white male student not qualified for FRL, not receiving ELL support, but receiving for SPED Services performed 11.04 points lower on mathematics achievement on average than student not receiving Special Education Services (\(\gamma_{40} = -11.0383\), \(t = -55.47\), \(df = 20000\), \(p < .0001\)). In these same average schools, white male student not qualified for FRL and not receiving Special Education Services, but qualified for ELL support performed 7.1452 points lower on mathematics achievement than a student not identified for English language support (\(\gamma_{70} = -7.1452\), \(t = -17.90\), \(df = 20000\), \(p < .0001\)). In these same average schools, the female students who were not qualified for FRL, not receiving ELL support, and not receiving Special Education Services performed .50 points lower on mathematics achievement than male students (\(\gamma_{0} = -0.4976\), \(t = -3.79\), \(df = 20000\), \(p < .0002\)).

E1d through E6d were dummy coding variables for the ethnicity compared to the baseline variable, white students. Compared to the white Americans in the schools with 47% FRL, and 70% white, and 14% SPED, the Hispanic American students (\(\gamma_{60} = -2.6843\), \(t = -11.69\), \(df = 20000\), \(p < .0001\)), the American Indian/Alaska Native students (\(\gamma_{70} = -3.4878\), \(t = -5.48\), \(df = 18000\), \(p < .0001\)), the Black/African American students (\(\gamma_{90} = -5.1846\), \(t = -16.34\), \(df = 19000\), \(p < .0001\)), and the students with two or more races (\(\gamma_{110} = -1.7102\), \(t = -4.28\), \(df = 20000\), \(p < .0001\)) performed significantly below White American students. The Asian American students’ performance was significantly above the performance by White American students (\(\gamma_{80} = 2.5263\), \(t = 5.25\), \(df = 20000\),
while no significant difference was found between the performance of Native Hawaiian or other Pacific Islander students and White Americans ($\gamma_{100} = 1.9907, t = 1.02, df = 20000, p = .3087$).

In addition to comparison between groups based on student characteristics, school characteristics also showed some significant differences based on percent of students in the school who are qualified for free and reduced lunches, percent of student receiving special services, and the percent of white students in the school. For schools with 72% white students and 14% SPED students, the school’s mean mathematics achievement score is estimated to be 0.0518 points lower for each 10% increase in free/reduced lunch percent above 43% (not significant). For schools with 43% FRL and 14% SPED students, the school’s mean mathematics achievement score is estimated to be 0.5498 points higher for each 10% increase in the percentage of white students above 72%. Likewise for a 10% decrease in percent of white students below 72%, the school’s mean mathematics achievement score is estimated to be lower by .5498 points for each decrease of 10%. In schools with 72% white and 43% FRL, the schools average mathematics achievement is decreased by 0.0802 points for each 1% increase in the percent of SPED students, whereas the school’s average is increased by 0.0802 points for each 1% decrease in the percent of SPED students. In Appendix B, Tables B3 and B4 show the sequence of models along with model fit statistics and Pseudo R for each 8th grade model. The SAS program for the final 8th grade model is listed in Program C.2 in Appendix C.
Step 4: Grade 8 Model—Addition of Interaction of Student Administrative Mode and Student Level Variables—Effect of Mode on Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity Race. To ascertain the effect of administrative mode on five student level variables – free and reduced lunch, gender, ELL, SPED, and ethnicity/race (seven groups), interactions with each student level variable was added to the model. Of the 12 interactions only interactions between SPED, ELL, and black students with administrative mode were significant. The model with these significant interactions was a better fit based on a deviance difference between the model with the interactions and the previous model with school level variables of percent FRL, percent Sped, and percent white (deviance Δ = 18.7, df = 3, p = .0003). See Table 4.20 for model fit statistics for the addition of each of the significant interactions.

Table 4.21 shows the parameters of model with the three significant interactions of administrative mode and demographic variables. All parameters in the table are for students in schools with 43% FRL (FRL_43), 14% SPED students (PCT_SPED14), and 72% white students (PCT_WH72). After including the three significant interactions in the model along with other significant student and school variables, the residual variability was reduced by 24.08% ((113.10 – 85.86)/113.10 = 24.08%), relative to the empty model. Likewise the variance for random intercept was reduced by 46.66% ((22.73 – 12.38)/22.73 = 46.66%), relative to the empty model.

The overall mean mathematics achievement score across schools with 47% FRL, 14% SPED, and 72% white students was significantly different from zero (γ₀₀ = 47.0404, t = 106.37, df = 85.8, p < .0001) for the model including the significant interactions.
There was no significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing (P/P) compared to students taking the online assessment in schools selecting online testing (O/O) (difference = 0.9167, $t = 1.84$, $df = 136$, $p = .0673$). There was a significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper assessment in schools selecting online testing (P/O) (difference = -4.1192, $t = -4.27$, $df = 214$, $p < .0001$) or students taking online assessment in schools selecting paper testing (O/P) (difference = -11.0559, $t = -4.67$, $df = 447$, $p < .0001$).

In this revised model, the significant interaction between special education and administrative mode indicated that the effect of receiving special education services (-11.5313) on mathematics achievement was more positive by 0.8982 ($df = 20000$, $t = 2.27$, $p = .0234$) for students taking the assessment online. In addition, the significant interaction of black students with administrative mode indicated that the effect of being black (-4.0788) on their mathematics achievement was more negative by 1.7920 ($df = 18000$, $t = -2.89$, $p = .0039$) for black students taking the assessment online.

Finally, the significant interaction between students qualified for FRL and administrative mode indicated that the effect of FRL (-3.8499) on mathematics achievement was more negative by 1.7584 ($df = 19000$, $t = -2.27$, $p = .0229$) for students taking the assessment online.
Table 20

Summary of Fit Statistics for Addition of Interactions with Administrative Mode with FRL, Gender, SPED, and ELL for Grade 8 Mathematics Achievement

<table>
<thead>
<tr>
<th>Models</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Admin Mode * SPED Effect (L1)</td>
<td>26</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Add Admin Mode * ELL Effect (L1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Admin Mode * Black Students Effect (L1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Df</td>
<td>26</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>.2 Log Likelihood</td>
<td>150369.8</td>
<td>150364.6</td>
<td>150356.3</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>150421.8</td>
<td>150418.6</td>
<td>150412.3</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>150518.2</td>
<td>150518.7</td>
<td>150516.1</td>
</tr>
</tbody>
</table>

Table 4.21

Estimates of Mathematics Achievement for the Four Administrative Groups Using the Previous Model with Effect of Interactions Between Student Administrative Mode and Student Explanatory Variables for FRL, Gender, SPED, and ELL on Grade 8 Mathematics Achievement

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>47.0404</td>
<td>0.4422</td>
<td>85.8</td>
<td>106.37</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-11.0559</td>
<td>2.3692</td>
<td>447</td>
<td>-4.67</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>0.9167</td>
<td>0.4969</td>
<td>136</td>
<td>1.84</td>
<td>0.0673</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-4.1192</td>
<td>0.9651</td>
<td>214</td>
<td>-4.27</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online Paper*Sch Online Paper</td>
<td>16.0919</td>
<td>2.5301</td>
<td>424</td>
<td>6.36</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

Table 4.21 continues
<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRL_Program</td>
<td>-3.8499</td>
<td>0.1883</td>
<td>217</td>
<td>-20.45</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Sch Percent FRL (center at .43)</td>
<td>-0.0549</td>
<td>0.1384</td>
<td>341</td>
<td>-0.40</td>
<td>0.6916</td>
</tr>
<tr>
<td>Stud FRL * Sch FRL</td>
<td>-0.2613</td>
<td>0.0913</td>
<td>173</td>
<td>-2.86</td>
<td>0.0047</td>
</tr>
<tr>
<td>Gender_Code</td>
<td>-0.4949</td>
<td>0.1313</td>
<td>20000</td>
<td>-3.77</td>
<td>0.0002</td>
</tr>
<tr>
<td>Special_Education_Code</td>
<td>-11.5313</td>
<td>0.2990</td>
<td>20000</td>
<td>-38.56</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Sch Percent SPED (center at .14)</td>
<td>-0.8144</td>
<td>0.2816</td>
<td>450</td>
<td>-2.89</td>
<td>0.0040</td>
</tr>
<tr>
<td>ELL_Code</td>
<td>-6.2141</td>
<td>0.5697</td>
<td>20000</td>
<td>-10.91</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E1d—Hispanic</td>
<td>-2.6788</td>
<td>0.2296</td>
<td>20000</td>
<td>-11.67</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E2d—Am Ind or Al Na</td>
<td>-3.4715</td>
<td>0.6366</td>
<td>18000</td>
<td>-5.45</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E3d—Asian</td>
<td>2.4391</td>
<td>0.4838</td>
<td>20000</td>
<td>5.04</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E4d—Black</td>
<td>-4.0788</td>
<td>0.4947</td>
<td>18000</td>
<td>-8.24</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E5d—Nat Haw or Pac Is</td>
<td>2.0076</td>
<td>1.9545</td>
<td>20000</td>
<td>1.03</td>
<td>0.3044</td>
</tr>
<tr>
<td>E6d—Two or More</td>
<td>-1.6819</td>
<td>0.3996</td>
<td>20000</td>
<td>-4.21</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Percent White (center at .72)</td>
<td>0.5534</td>
<td>0.1251</td>
<td>317</td>
<td>4.42</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online * Special Ed</td>
<td>0.8982</td>
<td>0.3962</td>
<td>20000</td>
<td>2.27</td>
<td>0.0234</td>
</tr>
<tr>
<td>Stud Online * FRL</td>
<td>-1.7584</td>
<td>0.7729</td>
<td>19000</td>
<td>-2.27</td>
<td>0.0229</td>
</tr>
<tr>
<td>Stud Online * Black</td>
<td>-1.7920</td>
<td>0.6209</td>
<td>18000</td>
<td>-2.89</td>
<td>0.0039</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means ($\tau_{00}$)</td>
<td>12.3850</td>
<td>2.5011</td>
<td>4.95</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>-10.9182</td>
<td>2.9472</td>
<td>-3.70</td>
<td>0.0002</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>17.4120</td>
<td>4.3056</td>
<td>4.04</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (3,1)</td>
<td>-0.8110</td>
<td>0.9588</td>
<td>-0.85</td>
<td>0.3976</td>
</tr>
<tr>
<td>Covar. (3,2)</td>
<td>2.6587</td>
<td>1.0555</td>
<td>2.52</td>
<td>0.0118</td>
</tr>
<tr>
<td>Rand FRL Slope Var. (3,3)</td>
<td>1.6711</td>
<td>0.5537</td>
<td>3.02</td>
<td>0.0013</td>
</tr>
<tr>
<td>Var. within School ($\tau^2$)</td>
<td>85.8605</td>
<td>0.8567</td>
<td>100.22</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
Analysis of Model Fit for Grade 11

**Step 1: Grade 11—Empty Model—Calculation of ICC.** Table 4.22 shows the initial fit statistics for the Empty Model. Table 4.23 presents the parameters for the empty model. Average school mean mathematics achievement was statistically different from zero ($\gamma_{00} = 39.88$, $t = 121.04$, $df = 266$, $p < .0001$). The intraclass correlation (ICC) can be calculated by dividing between-schools variance by the total variance in the model, ICC = $\text{var}(U_{0j}) / [\text{var}(R_{ij}) + \text{var}(U_{0j})]$. The table shows the following parameters.

\[
\text{Var}(U_{0j}) = \tau_0^2 = 145.09 \quad \text{var}(R_{ij}) = \sigma^2 = 24.26
\]

\[
\text{ICC} = \sigma^2 / (\tau_0^2 + \sigma^2) = 24.26 / (145.09 + 24.26) = 24.26 / 169.35 = .143
\]

Table 4.22

*Fit Statistics Fixed Effects and Random Effects of Empty Model for Grade 11 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>.2 Log Likelihood</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
</tr>
</tbody>
</table>

Table 4.23

*Fixed Effects and Random Effects of Empty Model for Grade 11 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>$df$</th>
<th>$t$ value</th>
<th>Exact $p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>39.8763</td>
<td>0.3295</td>
<td>266</td>
<td>121.04</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>$z$ value</th>
<th>Exact $p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means ($\tau_{00}$)</td>
<td>24.26</td>
<td>2.5234</td>
<td>9.61</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Var. within School ($\tau^2$)</td>
<td>145.09</td>
<td>1.4315</td>
<td>101.35</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
The intraclass correlation coefficient is equal to .143, indicating that 14% of the variability in mathematics achievement was between schools. The rest of the total variability, 86%, was within the school. This ICC is about average, compared to other results of educational research (values between .05 and .20 are common). This indicates that the grouping according to schools leads to an important similarity between the results of different students in the same school, although the within-school differences are far larger than the between-schools differences. The total variability was 169.35. For model analysis, student level explanatory variables were selected to reduce the variance within schools and additional school level explanatory variables were added to explain between-school variance in the subsequent models.

**Step 2: Grade 11 Model—School and Student Administrative Mode Effects Only.** Table 4.24 shows the results of the model fit statistics model improvements with the addition of the level-1 student administrative mode variable (paper/pencil = 0 and online = 1), the level-2 school administrative mode variable (paper/pencil = 0 and online = 1), the random effect of student administrative mode, and the cross-level interaction of student by school administrative mode. The addition of fixed effect of student admin mode to the empty model improved the model as shown by a deviance difference of 132.5 with $df = 1, p < .0001$. Likewise the addition of fixed effect of school admin mode to the fixed student admin mode model was also significantly better with a deviance difference of 40.8 with $df = 1, p < .0001$. The model showed improvement with the addition of random effect of student mode effect and the cross-level interaction of student by school admin mode effect with a deviance difference of 119.9 with $df = 1, p < .0001$ and a deviance difference of 119.2 with $df = 1, p < .0001$, respectively.
Table 4.24

Summary of Fit Statistics for Fixed Effects and Random Effects of Admin Mode Models
for Grade 11 Mathematics Achievement

<table>
<thead>
<tr>
<th></th>
<th>Two Level Unconditional Model</th>
<th>Fixed Student Admin Mode</th>
<th>Fixed School Admin Mode</th>
<th>Random Student Admin Mode</th>
<th>Cross-level Student with School Admin Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fit Statistics</td>
<td>Fit Statistics</td>
<td>Fit Statistics</td>
<td>Fit Statistics</td>
<td>Fit Statistics</td>
</tr>
<tr>
<td>df</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>.2 Log Likelihood</td>
<td>163220.3</td>
<td>163087.8</td>
<td>163047.0</td>
<td>162927.1</td>
<td>162807.9</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>163226.3</td>
<td>163095.8</td>
<td>163057.0</td>
<td>162941.1</td>
<td>162823.9</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>163237.2</td>
<td>163110.3</td>
<td>163075.1</td>
<td>162966.4</td>
<td>162852.9</td>
</tr>
</tbody>
</table>

Table 4.25 shows the parameters for the improved model. The overall mean mathematics achievement score across schools was still significantly different from zero ($\gamma_{00} = 38.4227$, $t = 62.83$, $df = 98.3$, $p < .0001$). After adding student and school online/paper-pencil administrative mode as variables for mathematics achievement, the random intercept variance was increased by 7.71% ($\frac{26.13 - 24.26}{24.26} = -7.71\%$), relative to the empty model. The residual variance was reduced by 1.96% ($\frac{145.09 - 142.24}{145.09} = 1.96\%$), relative to the empty model.

By including the cross-level interaction, the difference between four combinations of school and student administrative modes could be tested. The four combinations of administrative mode were student taking assessment on paper in school that selected paper assessments (stud mode = 0 and sch mode = 0), student taking assessment online in...
Table 4.25

Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Explanatory Variables for Administrative Mode in Grade 11 Mathematics Achievement

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (γ00)</td>
<td>38.4227</td>
<td>0.6116</td>
<td>98.8</td>
<td>62.83</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-12.8523</td>
<td>1.8097</td>
<td>26.6</td>
<td>-7.10</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>2.2758</td>
<td>0.7155</td>
<td>170</td>
<td>3.18</td>
<td>.0017</td>
</tr>
<tr>
<td>Stud Online Paper * Sch Online Paper</td>
<td>26.8452</td>
<td>2.1057</td>
<td>25.6</td>
<td>12.75</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means (τ00)</td>
<td>26.1259</td>
<td>4.2378</td>
<td>6.17</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>-7.6473</td>
<td>5.6253</td>
<td>-1.36</td>
<td>0.1740</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>9.8903</td>
<td>7.8225</td>
<td>1.26</td>
<td>0.1031</td>
</tr>
<tr>
<td>Var. within School (r²)</td>
<td>142.24</td>
<td>1.4041</td>
<td>101.30</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

school that selected paper assessments (stud mode = 1 and sch mode = 0), student taking assessment on paper in school that selected online assessments (stud mode = 0 and sch mode = 1), and student taking assessment online in school that selected online assessments (stud mode = 1 and sch mode = 1). Table 4.25 shows the estimates for the four different test admin modes. The overall mean mathematics achievement score across schools was still significantly different from zero (γ00 = 38.4227, t = 62.83,
df = 98.3, \( p < .0001 \)). There was a significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing compared to students taking the online assessment in schools selecting online testing (difference = 2.2758, \( t = 3.18, df = 170, p = .0017 \)) or students taking online assessment in schools selecting paper testing (difference = -12.8523, \( t = -7.10, df = 26.6, p < .0001 \)). Likewise, there was a significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper assessment in schools selecting online testing (difference = -11.7170, \( t = -9.19, df = 45, p < .0001 \)).

**Step 3: Grade 11 Model—Addition of Student and School Explanatory Variables—Control for Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race.** Table 4.26 shows the model fit statistics for the addition of the level-1 variables for free and reduced lunch, gender, English language learner, students receiving special services, and ethnicity/race and level-2 variable for percent of students receiving FRL and white students. Other level-2
Table 4.26

Summary of Fit Statistics for Fixed Effects and Random Effects of Level 1 and 2 Explanatory Variables for Grade 11 Mathematics Achievement

<table>
<thead>
<tr>
<th>Interaction with School Admin Mode</th>
<th>Add FRL Fixed Student Effect (L1)</th>
<th>Add FRL Fixed School Effect (L2)</th>
<th>Add FRL Random Student Effect (L1)</th>
<th>Add FRL Random School Effect (L2) – Eliminate</th>
<th>Add Interaction FRL Student with FRL School Effect (L2)</th>
<th>Add Gender Fixed Student Effect (L1)</th>
<th>Add ELL Fixed Student Effect (L1)</th>
<th>Add ELL Fixed School Effect (L2)</th>
<th>Add SPED Fixed Student Effect (L1)</th>
<th>Add Seven Ethnicities Fixed Student Effect (L1)</th>
<th>Add Percent White Fixed School Effect (L2)</th>
<th>Dropped ELL School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Statistics</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>.2 Log Likelihood</td>
<td>162807.9</td>
<td>161497.3</td>
<td>161436.8</td>
<td>161397.8</td>
<td>161396.9</td>
<td>161397.5</td>
<td>161255.9</td>
<td>161248.9</td>
<td>158941.0</td>
<td>158367.9</td>
<td>158341.0</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>162823.9</td>
<td>161515.3</td>
<td>161456.8</td>
<td>161423.8</td>
<td>161424.9</td>
<td>161425.5</td>
<td>161285.9</td>
<td>161280.9</td>
<td>158975.0</td>
<td>158413.9</td>
<td>158387.0</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>162852.9</td>
<td>161547.9</td>
<td>161492.9</td>
<td>161470.8</td>
<td>161475.5</td>
<td>161476.2</td>
<td>161340.1</td>
<td>161338.8</td>
<td>159036.5</td>
<td>158497.1</td>
<td>158470.2</td>
<td></td>
</tr>
</tbody>
</table>
variables such as percent of female students, percent of SPED students, and percent of ethnicities were not significant.

The model was improved with the addition of fixed student FRL effect (deviance $\Delta = 1549.7$, $df = 4$, $p < .0001$), fixed school percent of FRL centered at the grand mean of .35 (deviance $\Delta = 60.5$, $df = 1$, $p < .0001$), and random student FRL effect (deviance $\Delta = 39.0$, $df = 3$, $p < .0001$). Adding the cross-level interaction of student and school FRL effects did not improve the model (deviance $\Delta = 0.9$, $df = 1$, $p = .3428$). The cross-level interaction effect was not included in the following models. Adding demographics for gender (deviance $\Delta = 39.3$, $df = 4$, $p < .0001$), English language learners (deviance $\Delta = 141.6$, $df = 1$, $p < .0001$), percent students receiving ELL services (deviance $\Delta = 7.0$, $df = 1$, $p < .0001$), and Special Education (deviance $\Delta = 2314.9$, $df = 2$, $p < .0001$) improved model fit. With a deviance change of 573.1 ($df = 6$, $p < .0001$), the addition of the student ethnicity/race improved model fit as well. One additional school effects, namely the percent of white students in the school centered at the grand mean of .75 improved the model with deviance change of 27.8 ($df = 1$, $p < .0001$). The factor percent students receiving ELL services was dropped from the model since it was no longer significant. There was no change in deviance between the model with the percent ELL students and one without the factor. In addition the fit statistics AIC and BIC were both smaller for the model without the percent ELL students factor. The more parsimonious model with fewer factors was selected.

Table 4.27 shows the parameters of final model with level-1 variables and level-2 variables along with student and school admin mode estimates. All parameters in the table are for students in schools with 35% of the student qualified for free and reduced
Table 4.27

*Estimates of Mathematics Achievement for the Four Administrative Groups Using the Model with Effect of Student Explanatory Variables for FRL, Gender, SPED, ELL and Ethnicity as well as School Level Explanatory Variables for Percent FRL, Percent SPED, and Percent White on Grade 11 Mathematics Achievement*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept (γ00)</td>
<td>42.4799</td>
<td>0.3626</td>
<td>77.8</td>
<td>117.15</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-5.7038</td>
<td>1.4392</td>
<td>26</td>
<td>-3.96</td>
<td>0.0005</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>0.8039</td>
<td>0.4365</td>
<td>162</td>
<td>1.84</td>
<td>0.0673</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-4.8188</td>
<td>0.8638</td>
<td>44.2</td>
<td>-5.58</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online Paper*Sch Online Paper</td>
<td>11.3265</td>
<td>1.6457</td>
<td>26</td>
<td>6.88</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>FRL_Program</td>
<td>-3.9349</td>
<td>0.2235</td>
<td>183</td>
<td>-17.61</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Percent FRL (center at .35)</td>
<td>-0.4565</td>
<td>0.1439</td>
<td>280</td>
<td>-3.17</td>
<td>.0017</td>
</tr>
<tr>
<td>Gender_Code</td>
<td>-0.7951</td>
<td>0.1511</td>
<td>21000</td>
<td>-5.26</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Special_Education_Code</td>
<td>-12.1887</td>
<td>0.2455</td>
<td>19000</td>
<td>-49.66</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL_Code</td>
<td>-7.3811</td>
<td>0.5396</td>
<td>20000</td>
<td>-13.68</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E1d—Hispanic</td>
<td>-4.1678</td>
<td>0.2743</td>
<td>19000</td>
<td>-15.20</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E2d—Am Ind or Al Na</td>
<td>-3.7876</td>
<td>0.7085</td>
<td>12000</td>
<td>-5.35</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E3d—Asian</td>
<td>3.3133</td>
<td>0.5499</td>
<td>21000</td>
<td>6.03</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E4d—Black</td>
<td>-6.6367</td>
<td>0.3647</td>
<td>18000</td>
<td>-18.20</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E5d—Nat Haw or Pac Is</td>
<td>-1.8476</td>
<td>1.9485</td>
<td>21000</td>
<td>-0.95</td>
<td>0.3430</td>
</tr>
<tr>
<td>E6d—Two or More</td>
<td>-1.7466</td>
<td>0.4843</td>
<td>21000</td>
<td>-3.61</td>
<td>0.0003</td>
</tr>
<tr>
<td>Percent White (center at .75)</td>
<td>0.7933</td>
<td>0.1320</td>
<td>223</td>
<td>6.01</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

Table 4.27 continues
lunch (PDT_FRL35) and 75% white students (PCT_WH75). After including student variables for free and reduced lunch (FRL), gender, English language learner (ELL), special education (SPED), and ethnicity/race in the model along with school variable for percent qualified for FRL and percent white students, the residual variability was reduced by 20.44% \( \frac{(145.09 - 115.43)}{145.09} = 20.44\% \), relative to the empty model. Likewise the variance for random intercept was reduced by 73.74% \( \frac{(24.26 - 6.37)}{24.26} = 73.74\% \), relative to the empty model. The overall mean mathematics achievement score across schools with 35% qualified for FRL and 75% white students was significantly different from zero \( \gamma_{00} = 42.4799, t = 117.15, df = 77.8, p < .0001 \). There was no significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing (P/P) compared to students taking the online assessment in schools selecting online testing (O/O) \( \text{difference} = 0.8039, t = 1.84, df = 162, p = .0673 \). There was significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper assessment in schools selecting online testing (P/O) \( \text{difference} = -4.8188, t = -5.58, df = 44.2, \)
$p < .0001$) and also students taking online assessment in schools selecting paper testing (O/P) (difference = -5.7038, $t = -3.96$, $df = 26$, $p = 0.0005$).

In schools with 35% FDL and 75% white students, the white male student not receiving SPED service or not qualified for ELL support, but qualified for Free or Reduced Lunch (FRL) performed 3.9349 points lower on mathematics achievement on average than a student not qualified for FRL ($\gamma_{30} = -3.9349$, $t = -17.61$, $df = 183$, $p < .0001$). Since the model included a random effect of student FRL, the difference between kids who get free/reduced lunch and kids who do not varies significantly over schools. On the average, the gap related to lunch status is 3.9349 points, but across 95% of the schools, that gap is predicted to be anywhere from -6.62 lower to -1.25 lower ($-3.9349 \pm 1.96*\text{SQRT}(1.8760) = -3.9349 \pm 2.6846 = -6.62$ to -1.25).

Likewise in these same schools with 35% FDL and 75% white students, the white male student not qualified for FRL, not receiving ELL support, but receiving for SPED Services performed 12.19 points lower on mathematics achievement on average than student not receiving Special Education Services ($\gamma_{40} = -12.1887$, $t = -49.66$, $df = 19000$, $p < .0001$). In these same average schools, white male student not qualified for FRL and not receiving Special Education Services, but qualified for ELL support performed 7.3811 points lower on mathematics achievement than a student not identified for English language support ($\gamma_{70} = -7.3811$, $t = -13.68$, $df = 19000$, $p < .0001$). In these same average schools, the female students who were not qualified for FRL, not receiving ELL support, and not receiving Special Education Services performed .7951 points lower on mathematics achievement than male students ($\gamma_{0} = -0.7951$, $t = -5.26$, $df = 21000$, $p < .0001$).
E1d through E6d were dummy coding variables for the ethnicity compared to the baseline variable, white students. Compared to the white Americans in the schools with 35% FRL and 75% white, the Hispanic American students ($\gamma_{60} = -4.1678$, $t = -15.20$, $df = 19000$, $p < .0001$), the American Indian/Alaska Native students ($\gamma_{70} = -3.7876$, $t = -5.35$, $df = 12000$, $p < .0001$), the Black/African American students ($\gamma_{90} = -6.6367$, $t = -18.20$, $df = 18000$, $p < .0001$), and the students with two or more races ($\gamma_{110} = -1.7466$, $t = -3.61$, $df = 21000$, $p < .0003$) performed significantly below White American students. The Asian American students’ performance was significantly above the performance by White American students ($\gamma_{80} = 3.3133$, $t = 6.03$, $df = 21000$, $p < .0001$) while no significant difference was found between the performance of Native Hawaiian or other Pacific Islander students and White Americans ($\gamma_{100} = -1.8476$, $t = -0.95$, $df = 21000$, $p = .3430$).

In addition to comparison between groups based on student characteristics, school characteristics also showed some significant differences based on percent of students in the school who received free and reduced lunches and the percent of white students in the school. For schools with 75% white students, the school’s mean mathematics achievement score is estimated to be .4565 points lower for each 10% increase in free/reduced lunch percent above 35%. For schools with 35% FRL, the school’s mean mathematics achievement score is estimated to be .7933 points higher for each 10% increase in the percentage of white students above 75%. Likewise for a 10% decrease in percent of white students below 75%, the school’s mean mathematics achievement score is estimated to be lower by .7933 points for each decrease of 10%. In Appendix B, Tables B5 and B6 show the sequence of models along with model fit statistics and Pseudo
R for each model. The SAS program for the final 11th grade model is listed as Program C.3 in Appendix C.

**Step 4: Grade 11 Model—Addition of Interaction of Student Administrative Mode and Student Level Variables—Effect of Mode on Free and Reduced Lunch, Gender, English Language Learner, Special Education, and Ethnicity/Race.** To ascertain the effect of administrative mode on five student level variables – free and reduced lunch, gender, ELL, SPED, and ethnicity/race (seven groups), interactions with each student level variable was added to the model. Of the 12 interactions only the interaction between students receiving ELL services and administrative mode were significant. The model with this significant interaction was a better fit based on a deviance difference between the model with the interaction and the previous model with school level variable of percent white students and students eligible for FRL (deviance $\Delta = 5.8$, $df = 1$, $p = .0160$). See Table 4.28 for model fit statistics for the addition the significant interactions. Adding the interactions of the other variables did not improve the model (deviance $\Delta = 1.3$, $df = 1$, $p = .7291$).

Table 4.29 shows the parameters of model with one significant interaction of administrative mode, level-1 variables and level-2 variables. All parameters in the table are for students in schools with 35% (FRL (FRL_35) and 75% white students (PCT_WH75). After including the one significant interaction in the model along with other significant student and school variables, the residual variability was reduced by 20.46% ($((145.09 - 115.40)/145.09 = 20.46\%)$, relative to the empty model. Likewise the variance for random intercept was reduced by 73.58% ($((24.26 - 6.41)/24.26 = 73.58\%)$, relative to the empty model.
Table 4.28

**Summary of Fit Statistics for Addition of Interactions with Administrative Mode with FRL, Gender, SPED, and ELL for Grade 11 Mathematics Achievement**

<table>
<thead>
<tr>
<th>Models</th>
<th>Fit Statistics</th>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Admin Mode * ELL, SPED, Gender, &amp; FRL Effect (L1)</td>
<td>df</td>
<td>27</td>
</tr>
<tr>
<td>2 Log Likelihood</td>
<td>158333.9</td>
<td>158335.2</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>158387.9</td>
<td>158383.2</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>158485.6</td>
<td>158470.0</td>
</tr>
</tbody>
</table>

Table 4.29

**Estimates of Mathematics Achievement for the Four Administrative Groups Using the Previous Model with Effect of Interactions Between Student Administrative Mode and Student Explanatory Variables for FRL, Gender, SPED, and ELL on Grade 11 Mathematics Achievement**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ((\beta_{00}))</td>
<td>42.4610</td>
<td>0.3634</td>
<td>74.5</td>
<td>116.84</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-5.6980</td>
<td>1.4511</td>
<td>27.4</td>
<td>-3.93</td>
<td>0.0005</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>0.8328</td>
<td>0.4373</td>
<td>161</td>
<td>1.90</td>
<td>0.0587</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-4.9135</td>
<td>0.8697</td>
<td>47.3</td>
<td>-5.65</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

Table 4.29 continues
### Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stud Online Paper*Sch Online Paper</td>
<td>11.4443</td>
<td>1.6595</td>
<td>27.6</td>
<td>6.90</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>FRL_Program</td>
<td>-3.9336</td>
<td>0.2235</td>
<td>183</td>
<td>-17.60</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Sch Percent FRL (center at .35)</td>
<td>-0.4573</td>
<td>0.1441</td>
<td>281</td>
<td>-3.17</td>
<td>0.0017</td>
</tr>
<tr>
<td>Gender_CODE</td>
<td>-0.7947</td>
<td>0.1510</td>
<td>21000</td>
<td>-5.26</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Special_Education_CODE</td>
<td>-12.1791</td>
<td>0.2455</td>
<td>19000</td>
<td>-49.61</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>ELL_CODE</td>
<td>-6.4685</td>
<td>0.6584</td>
<td>20000</td>
<td>-9.82</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E1d—Hispanic</td>
<td>-4.1524</td>
<td>0.2743</td>
<td>19000</td>
<td>-15.14</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E2d—Am Ind or Al Na</td>
<td>-3.7838</td>
<td>0.7085</td>
<td>12000</td>
<td>-5.34</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E3d—Asian</td>
<td>3.3204</td>
<td>0.5498</td>
<td>21000</td>
<td>6.04</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E4d—Black</td>
<td>-6.6314</td>
<td>0.3647</td>
<td>18000</td>
<td>-18.18</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>E5d—Nat Haw or Pac Is</td>
<td>-1.7548</td>
<td>1.9486</td>
<td>21000</td>
<td>-0.90</td>
<td>0.3678</td>
</tr>
<tr>
<td>E6d—Two or More</td>
<td>-1.7378</td>
<td>0.4842</td>
<td>21000</td>
<td>-3.59</td>
<td>0.0003</td>
</tr>
<tr>
<td>Percent White (center at .75)</td>
<td>0.7930</td>
<td>0.1322</td>
<td>223</td>
<td>6.00</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online * ELL</td>
<td>-2.6609</td>
<td>1.1019</td>
<td>17000</td>
<td>-2.41</td>
<td>0.0158</td>
</tr>
</tbody>
</table>

### Random Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>z value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var. in School means (τ₀)</td>
<td>6.4105</td>
<td>1.5983</td>
<td>4.01</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Covar. (2,1)</td>
<td>0.3453</td>
<td>2.2001</td>
<td>0.16</td>
<td>0.853</td>
</tr>
<tr>
<td>Rand Online Slope Var. (2,2)</td>
<td>3.1083</td>
<td>3.1472</td>
<td>0.99</td>
<td>0.1617</td>
</tr>
<tr>
<td>Covar. (3,1)</td>
<td>-0.7362</td>
<td>0.8908</td>
<td>-0.83</td>
<td>0.4085</td>
</tr>
<tr>
<td>Covar. (3,2)</td>
<td>-0.613</td>
<td>1.1202</td>
<td>-0.55</td>
<td>0.5939</td>
</tr>
<tr>
<td>Rand FRL Slope Var. (3,3)</td>
<td>1.8736</td>
<td>0.6919</td>
<td>2.71</td>
<td>0.0034</td>
</tr>
<tr>
<td>Var. within School (τ²)</td>
<td>115.40</td>
<td>1.1428</td>
<td>100.98</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
The overall mean mathematics achievement score across schools with 35% FRL and 75% white students, was significantly different from zero ($\gamma_{00} = 42.4610, t = 116.84, df = 74.5, p < .0001$). There was no significant difference in mathematics achievement between students taking the paper assessment in schools selecting paper testing (P/P) compared to students taking the online assessment in schools selecting online testing (O/O) (difference = 0.8328, $t = 1.90, df = 161, p = .0587$). There was a significant difference between students taking the paper assessment in schools selecting paper testing versus students taking paper assessment in schools selecting online testing (P/O) (difference = -4.9135, $t = -5.65, df = 47.3, p < .0001$) or students taking online assessment in schools selecting paper testing (O/P) (difference = -5.6980, $t = -3.93, df = 27.4, p = 0.0005$).

In schools with 35% FDL and 75% white students, the white male student not receiving SPED service or not qualified for ELL support, but qualified for Free or Reduced Lunch (FRL) performed 3.9336 points lower on mathematics achievement on average than a student not qualified for FRL ($\gamma_{30} = -3.9336, t = -17.60, df = 183, p < .0001$). Since the model included a random effect of student FRL, the difference between kids who get free/reduced lunch and kids who don’t varies significantly over schools. On the average, the gap related to lunch status is 3.9336 points, but across 95% of the schools, that gap is predicted to be anywhere from 6.23 lower to 1.64 lower ($-3.9336 \pm 1.96*\text{SQRT}(1.8736) = -3.9336 \pm 2.2971 = -6.23$ to -1.64).

Likewise in these same schools with 35% FDL and 75% white students, the white male student not qualified for FRL, not receiving ELL support, but receiving SPED Services performed 12.18 points lower on mathematics achievement on average than
student not receiving Special Education Services ($\gamma_{40} = -12.1791, t = -49.61, df = 19000, p < .0001$). In these same average schools, white male student not qualified for FRL and not receiving Special Education Services, but qualified for ELL support performed 6.4685 points lower on mathematics achievement than a student not identified for English language support ($\gamma_{70} = -6.4685, t = -9.82, df = 20000, p < .0001$). In these same average schools, the female students who were not qualified for FRL, not receiving ELL support, and not receiving Special Education Services performed .7947 points lower on mathematics achievement than male students ($\gamma_{0} = -0.7947, t = -5.26, df = 21000, p < .0001$).

$E_{1d}$ through $E_{6d}$ were dummy coding variables for the ethnicity compared to the baseline variable, white students. Compared to the white Americans in the schools with 35% FRL and 75% white, the Hispanic American students ($\gamma_{60} = -4.1512, t = -15.14, df = 19000, p < .0001$), the American Indian/Alaska Native students ($\gamma_{70} = -3.7838, t = -5.34, df = 12000, p < .0001$), the Black/African American students ($\gamma_{90} = -6.6314, t = -18.18, df = 18000, p < .0001$), and the students with two or more races ($\gamma_{110} = -1.7378, t = -3.59, df = 21000, p < .0003$) performed significantly below White American students. The Asian American students’ performance was significantly above the performance by White American students ($\gamma_{80} = 3.3204, t = 6.04, df = 21000, p < .0001$) while no significant difference was found between the performance of Native Hawaiian or other Pacific Islander students and White Americans ($\gamma_{100} = -1.7548, t = -0.90, df = 21000, p = .3678$).

In addition to changes to groups based on student characteristics, school characteristics showed changes as well, but still had significant differences based on
percent of students in the school who received free and reduced lunches and the percent of white students in the school. For schools with 75% white students, the school’s mean mathematics achievement score is estimated to be .4573 points lower for each 10% increase in free/reduced lunch percent above 35%. For schools with 35% FRL, the school’s mean mathematics achievement score is estimated to be .7930 points higher for each 10% increase in the percentage of white students above 75%. Likewise for a 10% decrease in percent of white students below 75%, the school’s mean mathematics achievement score is estimated to be lower by .7930 points for each decrease of 10%.

In this revised model, the significant interaction between students receiving English language learner services and administrative mode indicated that the effect of receiving ELL services (-6.4685) on mathematics achievement was more negative by 2.6609 (df = 17000, t = -2.41, p = 0.0158) for students taking the assessment online.
Chapter 5

Discussion

Introduction

In this study, multilevel (hierarchical) modeling was used to investigate the effects of administrative assessment mode on mathematics achievement on the Nebraska State Accountability mathematics assessment (NeSA-M) after controlling for student-level characteristics and school-level variables on mathematics performance. The investigation examined data from three grade levels to determine the effect of different test administrative modes on mathematics performance for all students and different demographic subgroups (gender, FRL, ELL, SPED, or race/ethnicity).

The intraclass correlations (ICC) for all three grade levels indicated that grouping according to schools leads to an important similarity between the mathematics performances of different pupils in the same school, although within-school differences were far larger than between-school differences. Since the ICC was 0.25, 0.17, and 0.14 for grade 4, 8, and 11 respectively, 25%, 17%, and 14% of the variability in mathematics performance was accounted for by the school and thus it would be incorrect to treat the student data as independent of school variables. Though the use hierarchical linear modeling, this school dependence was adjusted for and modeled. HLM allowed for the simultaneously modeling of the impact of both student (level 1) and school (level 2) variables on mathematics performance.

Model Review

In each grade level model, school and student administrative assessment mode was added to the empty model. Now, the predicted mathematics achievement score,
(γ₀₀), represented the score for a student taking the paper/pencil test in a school that selected paper/pencil testing. After adding school/student mode effect to the model, the average mathematics performance level for 4th graders taking a paper/pencil test in a school selecting paper/pencil assessments increased slightly, 40.82 to 40.89, from the average performance for all students in the empty model. The average mathematics performance level for 8th and 11th graders for the pencil/paper students in paper/pencil schools showed a slight decrease of 44.17 to 43.29 and 39.88 to 38.42, respectively. The initial comparison across testing modes for the three grade levels showed significant difference between testing modes for grades 8 and 11 and showed mixed effects for grade 4 with some significant differences and some not. Since there was no reason to believe the testing groups were equivalent, student-level and school-level variables were added to the model to control for potential score differences due to any school/student variables. The final model included significant variables for student demographics of FRL, gender, SPED, ELL, and race/ethnicity as well as any significant school variables for FRL, gender, SPED, ELL, and race/ethnicity.

In the final model, the predicted mathematics achievement score, (γ₀₀), for a student taking the paper/pencil test in a school that selected paper/pencil testing now represents the score of a white male student who does not qualify for SPED services, ELL support, or FRL eligibility in a school with an average rate of FRL and average percentage of white students. For grades 4 and 11, the final model included only two school level variables, percent of FRL students and percent of white students, while the model for grade 8 included percent of special education students as well. After controlling for the effects of student demographics and school characteristics in grade 4,
8, and 11, there was no significant effect on mathematics performance between students taking paper assessments in schools selecting paper testing (P/P) and those taking online assessments in schools selecting online testing (O/O). Tables 5.1, 5.2, and 5.3 present the summary of mathematics performance estimates for student/school administration modes for grades 4, 8, and 11, respectively. The average mathematics performance for 4th graders was non-significantly 0.65 items higher for students who took P/P assessments than for students who took O/O assessments. Like the 4th grade results, the average mathematics performance for 8th graders showed a non-significant difference with the O/O students scored higher than P/P students by 0.97 items. Similar to the 8th grade results, the average 11th grade mathematics performance revealed a non-significant difference where students taking the assessment O/O scored higher than P/P students by 0.80 items. Most of the students in grades 4, 8, and 11 were tested in one of these two configurations, paper assessments in schools selecting paper testing or online assessments in schools selecting online testing.

The final model also addresses the question of whether students taking paper assessments in schools selecting paper testing scored differently than students taking paper assessments in a schools selecting online testing. After controlling for the effects of student demographics and school characteristics in all three grades, there was a significant effect on mathematics performance between students taking paper assessments in schools selecting online testing and those taking paper assessments in schools selecting paper testing. For Grades 4, 8, and 11, students taking paper assessments in online schools scored significantly lower on average than students taking paper assessments in paper schools by 5.36, 4.27, and 4.82 items respectively. These
Table 5.1

**Summary of Grade 4 Mathematics Performance Estimates for Student/School Administrative Mode**

<table>
<thead>
<tr>
<th>Estimates in Final Model for Grade 4 (Table 4.11)</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>44.853</td>
<td>0.3389</td>
<td>189</td>
<td>132.35</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>3.0611</td>
<td>2.0608</td>
<td>104</td>
<td>1.49</td>
<td>0.1405</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>-0.6509</td>
<td>0.3924</td>
<td>309</td>
<td>-1.66</td>
<td>0.0982</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-5.3587</td>
<td>0.6449</td>
<td>264</td>
<td>-8.31</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

Table 5.2

**Summary of Grade 8 Mathematics Performance Estimates for Student/School Administrative Mode**

<table>
<thead>
<tr>
<th>Estimates in Final Model for Grade 8 (Table 4.19)</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>47.0221</td>
<td>0.4429</td>
<td>84.9</td>
<td>106.16</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-11.5164</td>
<td>2.3618</td>
<td>449</td>
<td>-4.88</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>0.9738</td>
<td>0.4962</td>
<td>133</td>
<td>1.96</td>
<td>0.0518</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-4.2705</td>
<td>0.9552</td>
<td>201</td>
<td>-4.47</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
Table 5.3

*Summary of Grade 11 Mathematics Performance Estimates for Student/School Administrative Mode*

<table>
<thead>
<tr>
<th>Estimates in Final Model for Grade 11 (Table 4.27)</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>df</th>
<th>t value</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept ($\gamma_{00}$)</td>
<td>42.4799</td>
<td>0.3626</td>
<td>77.8</td>
<td>117.15</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Stud Paper / School Paper</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stud Online / School Paper</td>
<td>-5.7038</td>
<td>1.4392</td>
<td>26</td>
<td>-3.96</td>
<td>0.0005</td>
</tr>
<tr>
<td>Stud Online / School Online</td>
<td>0.8039</td>
<td>0.4365</td>
<td>162</td>
<td>1.84</td>
<td>0.0673</td>
</tr>
<tr>
<td>Stud Paper / School Online</td>
<td>-4.8188</td>
<td>0.8638</td>
<td>44.2</td>
<td>-5.58</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

findings are initially troubling, until further investigation revealed that a student in an online testing school can only take paper/pencil assessment if the student qualified for a SPED or ELL accommodation.

As reported in the Nations Report Card, both SPED and ELL supported students in Nebraska have lower scores on the NEAP mathematics assessment than non-accommodated students (NCES, 2011). Further analysis shows that for all three grades the students taking paper tests in schools selecting online testing (P/O) had a larger percentage of the students receiving SPED and ELL support than the P/P and O/O modes. Table B1 in Appendix B shows that the 4th grade students taking P/O assessments were identified as 64% SPED qualified compared to 17% for P/P and 16% for O/O, while 32% received ELL support compared to 14% for P/P and 9% for O/O. In Table B2 for P/O testing, 86% of the 8th graders were SPED qualified compared to 13% for P/P and 12% for O/O, whereas only 7% of the students received ELL support compared to 10% for P/P.
and 9% for O/O. In Table B3 for P/O testing, 35% of the 11th grade students were labeled as SPED compared to 12% for P/P and 10% for O/O, and 17% of the students received ELL support compared to 10% for P/P and 4% for O/O. All the percents in the P/O group were much larger than the percents for students taking the P/P assessments or the O/O assessments. Additional study is recommended to determine whether there are mode differences in performance for students taking P/O assessments.

In addressing the question of whether students taking paper assessments in schools selecting paper testing (P/P) scored differently than students taking online assessments in schools selecting paper testing (O/P), the final model showed inconsistency between the three grades. After controlling for the effects of student demographics and school characteristics in all three grades, there was a significant effect on mathematics performance between students taking paper assessments in schools selecting paper testing (P/P) and those taking online assessments in schools selecting paper testing (O/P) for students in Grades 8 and 11, but no significant effect in Grade 4. For Grades 8 and 11, students taking online assessments in schools selecting paper testing (O/P) scored respectively 11.52 and 5.70 significantly lower than students taking paper assessment in paper schools (P/P). On the other hand, scores for grade 4 students were not significantly different between the two groups. Further analysis is recommended to determine the basis for the mode differences. Were the differences due to a selection bias for online students in schools selecting paper testing? Few students took online assessments in schools selecting paper/pencil testing (O/P). Only 83 fourth graders, 29 eighth graders, and 78 eleventh graders took the mathematics assessments online in schools selecting paper/pencil testing. This represents less than 0.4% of the students in
all three grades. No data were collected on reasons for selecting online instead of paper/pencil. HLM requires larger sample sizes than the small number of students taking the online assessments in schools selecting paper testing (O/P). To test for significant differences additional data would be needed to control for group differences and their potential effect on mathematics performance.

**Proportion of Variance Explained**

Relative to the two-level empty model, the final model explained a proportion of the between-school and within-school variance for each grade level. To determine the proportion of within school variance explained subtract the two within school variances and divide the difference by the empty model within variance, \[\frac{(\tau_1^2) - (\tau_2^2)}{(\tau_1^2)}\]. The proportion of between school variance is calculated using the same process with the between school variances, \[\frac{(\tau_{00})_1 - (\tau_{00})_2}{(\tau_{00})_1}\]. Table 5.4 presents estimates for the 4th grade models used in the proportion calculations. For the 4th grade, almost 41% of the between-school variance \[(25.18 - 14.84) / 25.18 = .4108\] and more than 13% of within-school variance \[(67.01 - 57.93) / 67.01 = .1354\] was explained by adding student/school level variables to the model. These values represent a large reduction in the variation from the two-level empty model. Table 5.4 presents estimates for the 4th grade models used in the proportion calculations.

Using a similar process for the 8th grade model, almost 44% of the between-school variance \[(22.32 - 12.58) / 22.32 = .4364\] and more than 21% of within-school variance \[(110.02 - 85.93) / 110.02 = .2190\] was explained by adding student/school level variables to the model. These values represent a large reduction in the variation
Table 5.4

*Proportion Variance Reduction Comparison (Pseudo $R^2$) of Two-Level Empty Model and Final Model for Grade 4*

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimates Random School Intercept</th>
<th>Estimates Two-Level Empty Model</th>
<th>Estimates Final Model with Variables</th>
<th>Proportion Reduction in Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance between Schools ($\tau_{00}$)</td>
<td>22.73</td>
<td>25.1829</td>
<td>14.8374</td>
<td>41.08%</td>
</tr>
<tr>
<td>Variance Within School ($\tau^2$)</td>
<td>69.26</td>
<td>67.0058</td>
<td>57.9319</td>
<td>13.54%</td>
</tr>
</tbody>
</table>

from the two-level empty model. Table 5.5 presents estimates for the 8th grade models used in the proportion calculations.

Table 5.5

*Proportion Variance Reduction Comparison (Pseudo $R^2$) of Two-Level Empty Model and Final Model for Grade 8*

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimates Random School Intercept</th>
<th>Estimates Two-Level Empty Model</th>
<th>Estimates Final Model with Variables</th>
<th>Proportion Reduction in Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance between Schools ($\tau_{00}$)</td>
<td>23.2255</td>
<td>22.3228</td>
<td>12.5816</td>
<td>43.64%</td>
</tr>
<tr>
<td>Variance Within School ($\tau^2$)</td>
<td>113.10</td>
<td>110.02</td>
<td>85.9295</td>
<td>21.90%</td>
</tr>
</tbody>
</table>

For the 11th grade, almost 76% of the between-school variance [(26.13 – 6.37) / 26.13 = .7560] and more than 18% of within-school variance [(142.24-115.43) / 142.24 = .1885] was explained by adding student/school level variables to the model. These
values represent a large reduction in the variation from the two-level empty model. Table 5.6 presents estimates for the 11th grade models used in the proportion calculations.

Table 5.6

Proportion Variance Reduction Comparison (Pseudo R²) of Two-Level Empty Model and Final Model for Grade 11

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimates Random School Intercept</th>
<th>Estimates Two-Level Empty Model</th>
<th>Estimates Final Model with Variables</th>
<th>Proportion Reduction in Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance between Schools (τ₀₀)</td>
<td>24.26</td>
<td>26.1259</td>
<td>6.374</td>
<td>75.60%</td>
</tr>
<tr>
<td>Variance Within School (τ²)</td>
<td>145.09</td>
<td>142.24</td>
<td>115.43</td>
<td>18.85%</td>
</tr>
</tbody>
</table>

Snijders and Bosker (1999) define “explained proportion of variance in the hierarchical linear model” (p. 101) in two parts. Level one, (R²)₁, becomes “the proportional reduction of error for predicting an individual outcome” (Snijders & Bosker, 1999, p. 101), the within school variance. Whereas level two, (R²)₂, is “the proportional reduction of error for predicting a group mean” (Snijders & Bosker, 1999, p. 101), the between school variance. The formulas are defined by:

\[
(R²)₁ = 1 - \frac{((τ^2)₂ + (τ₀₀)₂) / ((τ^2)₁ + (τ₀₀)₁)}{(τ^2)₁ / n + (τ₀₀)₁}
\]

\[
(R²)₂ = 1 - \frac{((τ^2)₂ / n + (τ₀₀)₂) / ((τ^2)₁ / n + (τ₀₀)₁)}{(τ^2)₁ + (τ₀₀)₁}
\]

where \(n\) represents a usual group size.

At level one, a ratio is computed by dividing the value of \(τ^2 + τ₀₀\) for the second model by the value of \(τ^2 + τ₀₀\) for the first model. The value for \((R²)₁\) is then computed by subtracting this ratio of \(τ^2 + τ₀₀\) values from one. This value represents the explained proportion of the variance within schools, predicting the individual outcome. For level
two, $\tau^2$ is divided by $n$ and added to $\tau_{00}$ for each model. The ratio of the sum for model two over sum for model one \( \left( \frac{\tau^2}{n} + \tau_{00} \right) / \left( \frac{\tau^2}{n} + (\tau_{00})_1 \right) \) is subtracted from one to get the level two proportion of variance representing the between school variance. This value represents the explained proportion of the variance between schools, predicting the school mean.

Using Snijders and Bosker (1999) process for the 4th grade results, the proportional reduction in the total variance \( (R^2)_1 \) is \( 1 - \frac{(57.93 + 14.84)}{(69.26 + 22.73)} = 1 - \frac{72.77}{91.99} = 0.209 \). A little over 20% of the variance was the proportional reduction of error in predicting an individual outcome (within school variance). For the level-two explained proportion of variance, \( (R^2)_2 \) is \( 1 - \frac{(57.93 + 30 + 14.84)}{(69.26 + 30 + 22.73)} = 1 - \frac{16.77}{25.03} = .330 \), where $n$ is equal to 30, the average class size. More than 30% of the between school proportion of variance (predicting the school mean) was explained by including the variables in the model. These represent a large reduction in the variance and support the final model. See Table 5.4 for the variance estimates used in the proportional reduction calculations.

Using Snijders and Bosker (1999) process, the 8th grade proportional reduction in the total variance \( (R^2)_1 \) is \( 1 - \frac{(85.93 + 12.58)}{(113.1 + 23.23)} = 1 - \frac{98.51}{136.33} = 0.277 \). A little over 27% of the variance was the proportional reduction of error in predicting an individual outcome (within school variance). For the level-two explained proportion of variance, \( (R^2)_2 \) is \( 1 - \frac{(85.93 + 30 + 12.58)}{(113.1 + 30 + 23.23)} = 1 - \frac{15.44}{27.00} = .428 \), where $n$ is equal to 30, the average class size. Over 42% of the between school variance (predicting the group mean) was explained due to the inclusion of the variables in the model. These represent a large reduction in the variance and
support the final model. Table 5.5 presents estimates for the models used in the proportion calculations.

The 11th grade proportional reduction in the total variance \((R^2)_1\) using Snijders and Bosker (1999) process is:

\[
1 - \frac{115.43 + 6.37}{145.09 + 24.26} = 1 - \frac{121.8}{169.35} = 0.281.
\]

A little over 28% of the variance was the proportional reduction of error in the predicting an individual outcome (within school variance). For the level-two explained proportion of variance, \((R^2)_2\) is:

\[
1 - \frac{(115.43 / 30 + 6.37) / (145.09 / 30 + 24.26)}{10.22 / 29.10} = .649.
\]

By including the variables in the model, more than 64% of the proportion of variance for predicting a group mean (between school variance) was explained. These represent a large reduction in the variance and support the final model. Table 5.6 presents estimates for the models used in the proportion calculations.

Based on this evidence, there is still some additional between-school variability to be explained by including other school-level variables, such as school average reading performance or school size. Adding student demographics and school level characteristics explained some of the within-school variance in mathematics performance, but by inclusion of more student-levels variables in the model, such as previous mathematics test scores or gifted identification, could explain more within-school variance.

**Effects of School Level Variables**

In the final model for three grades, the school free/reduced lunch percentage and the school percentage of white students showed significant effects on mathematics performance. Since the school level variables were centered at the state mean, the estimates were relative to the state mean. As the FRL rate increased above the state
means, the average mathematics score for each grade decreased. For each increase of
10% in the school FRL rate above state average of 47%, the grade 4 average mathematics
performance for the school decreased by 2.97. The 8th grade school average mathematics
performance showed a similar pattern with a decrease of 3.84 for each 10% increase in
the school’s FRL rate above state average of 43%. Likewise for the 11th grade, student
mathematics performance decreased by 3.93 for a 10% increase in school FRL rate above
the state average of 35%.

As the percentage of white students increased, the average mathematics score
increased. The school level variables for percentage of white students were centered at
grade level state means, so these estimates were relative to the state means as well. For a
10% increase in the percent of white students above the state mean of 70%, the average
4th grade mathematics school performance increased by .22 for each 10% increase. The
average 8th grade mathematics school performance was increased by .55 for each 10%
increase in the percent of white students above the state mean of 72%. The largest
effect of percent of white student was estimated for 11th grade students with an increase
of .79 in mathematics school performance for each 10% increase in the percent of white
students above the state mean of 75%.

For Grade 8 only, the percent of SPED students showed significance. For each
1% increase in the percent of SPED students above the state average of 14%, the average
mathematics school performance decreased by .081 for Grade 8. Other school level
variables such as percent ELL students and percent black students did not show a
significant effect on the average mathematics school performance for all three grades.
Subgroup Performance by Mode

The NeSA mathematics results did indicate mathematics achievement differences similar to the findings of the 2011 NAEP scores for Mathematics for Nebraska students at Grades 4 and 8 (NCES, 2011). There were mathematics achievement subgroup differences based on gender, racial/ethnicity, and free/reduced lunch eligibility. Male students performed significantly better than female students. White students scored higher than Hispanic and black students. Students who were eligible for free/reduced-price school lunch scored lower than students who were not eligible. Similar results for these groups were noted for Grade 11. Other differences noted in the NeSA analysis included performance differences between whites and other ethnicities. Differences between students receiving SPED services and students not receiving SPED services were identified. Students qualified for ELL support scored differently than students not qualified for ELL support. For this study, the question to be answered was not the observed differences, but whether the group differences were confounded by the testing mode. To determine whether the group differences were affected by testing mode, the interaction between student demographic variables and the students’ assessment mode were investigated.

For Grade 4 SPED, ELL, gender, and FRL had significant interaction effects with student assessment mode. Grade 4 SPED students improved their mathematics score by 1.07 when taking the assessment online, whereas ELL, FRL, and female students scored lower by 1.32, .69, and .56 respectively when taking the assessment online. Although the interaction effects were significant, the proportion reduction in random intercept variance (between schools) relative to the previous model was less than 1% or (14.83 -14.80)/
14.83 = 0.22%. The proportion reduction in residual variance (within school) relative to the previous model was also less than 1% or \((57.93 - 57.81) / 57.93 = 0.20\%\). Since the proportions of variance reduction in both the within school and between school variance were less than 1%, the effects of the additional variables do not represent a large effect on the overall model. Although the random online slope variance was reduced by a larger amount, 2.22%, this represents a small reduction as well. These interaction variables were not practically significant and were most likely due to the large sample size. The interaction variables were deleted from the final model. Table 5.7 presents estimates for the models used in the proportion variance reduction calculations.

Table 5.7

*Proportion Variance Reduction Comparison (Pseudo $R^2$) of Model with Interaction and Final Model for Grade 4*

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate Final Model</th>
<th>Estimate Model with Interactions</th>
<th>Proportion Reduction in Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance in Schools Means ($\tau_{00}$)</td>
<td>14.8374</td>
<td>14.8043</td>
<td>0.22%</td>
</tr>
<tr>
<td>Random Online Slope Var. (2,2)</td>
<td>12.6614</td>
<td>12.3803</td>
<td>2.22%</td>
</tr>
<tr>
<td>Random FRL Slope Var. (3,3)</td>
<td>0.6922</td>
<td>0.7307</td>
<td>-5.56%</td>
</tr>
<tr>
<td>Variance within School ($\tau^2$)</td>
<td>57.9319</td>
<td>57.8137</td>
<td>0.20%</td>
</tr>
</tbody>
</table>

For Grade 8 the interaction of SPED, FRL, and black students with assessment mode were significant. Similar to Grade 4, grade 8 SPED students taking the assessment online improved their mathematics performance by .90, whereas FRL and black students scored lower by 1.76 and 1.79 respectively. Although the interaction effects were significant, the proportion reduction in random intercept variance (between schools)
relative to the previous model was less than 2% or \((12.58 - 12.39) / 12.58 = 1.56\%\). The proportion reduction in residual variance (within school variance) relative to the previous model was also less than 1% or \((85.93 - 85.86) / 85.93 = 0.08\%\). Since the proportions of variance reduction for both the within school and between school variance based were less than 2%, the effects of the additional interaction variables do not represent a large effect on the overall model. Although the random online slope variance was reduced by a larger amount, 2.18\%, this represents a small reduction as well. These interaction variables were not practically significant and were most likely due to the large sample size. The interaction variables were deleted from the final model. Table 5.8 presents estimates for the models used in the proportion variance reduction calculations.

Table 5.8

*Proportion Variance Reduction Comparison (Pseudo R²) of Model with Interaction and Final Model for Grade 8*

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate Final Model</th>
<th>Estimate Model with Interactions</th>
<th>Proportion Reduction in Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance in Schools Means ((\tau_{00}))</td>
<td>12.5816</td>
<td>12.3850</td>
<td>1.56%</td>
</tr>
<tr>
<td>Random Online Slope Var. (2,2)</td>
<td>17.8009</td>
<td>17.4120</td>
<td>2.18%</td>
</tr>
<tr>
<td>Random FRL Slope Var. (3,3)</td>
<td>1.7199</td>
<td>1.6711</td>
<td>2.84%</td>
</tr>
<tr>
<td>Variance within School ((\tau^2))</td>
<td>85.9295</td>
<td>85.8605</td>
<td>0.08%</td>
</tr>
</tbody>
</table>

Grade 11 had only one subgroup that showed a significant interaction with assessment mode, namely ELL. ELL students in the 11\textsuperscript{th} grade scored lower by 2.66. Although the interaction effect was significant, the proportion reduction in random
intercept variance (between schools) relative to the previous model was less than 1% or 
\((6.37 - 6.41) / 6.37 = -0.57\%\). This negative value represents an increase in variance 
rather than a decrease in variance, a lack of model fit. The proportion reduction in 
residual variance (within school variance) relative to the model without the interactions 
was also less than 1% or \((115.43 - 115.40) / 115.43 = 0.03\%\). Since the proportions 
reduction in both within school and between school variance were less than 1%, the 
effect of the additional variable does not represent a large effect on the overall model. 
Although the random online slope variance was changed by -6.62%, this represents an 
increase in variance rather than a decrease in variance. The interaction variable was not 
practically significant and was most likely due to the large sample size. The interaction 
variable was deleted from the final model. Table 5.9 presents estimates for the models 
used in the proportion variance reduction calculations.

Table 5.9

*Proportion Variance Reduction Comparison (Pseudo R²) of Model with Interaction and 
Final Model for Grade 11*

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimate Final Model</th>
<th>Estimate Model with Interactions</th>
<th>Proportion Reduction in Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance in Schools Means (τ₀₀)</td>
<td>6.3740</td>
<td>64.105</td>
<td>-0.57%</td>
</tr>
<tr>
<td>Random Online Slope Var. (2,2)</td>
<td>2.9154</td>
<td>3.1083</td>
<td>-6.62%</td>
</tr>
<tr>
<td>Random FRL Slope Var. (3,3)</td>
<td>1.8760</td>
<td>1.8736</td>
<td>0.13%</td>
</tr>
<tr>
<td>Variance within School (τ²)</td>
<td>115.43</td>
<td>115.40</td>
<td>0.03%</td>
</tr>
</tbody>
</table>
Since all the interactions between student test mode and the demographic variables were not practically significant, the mathematics performance of Grade 4, 8, or 11 students in these subgroups were not affected by the mode of the assessment administration.

**Limitations of Study**

As with any research, generalization of the findings is limited by the test being examined and the Nebraska population of students. Results from this study cannot be generalized to other measures of mathematics performance administered in two modes, online and paper/pencil. The results are localized. Similar investigations must be conducted for each test and different populations. The methodology described in this study can be used in determining equivalence between online and paper/pencil administration of the same assessments.

As noted in the discussion some variability for within schools and between schools was explained in the final model, but additional variance still remains. The study was limited to the data collected in the state data system. Analysis using hierarchical linear modeling is dependent on the available explanatory variables in the data set that were incorporated into the model. Snijders and Bosker (1999) state that “the basic idea of multilevel analysis is that data sets with a nesting structure that includes unexplained variability at each level of nesting . . . are often adequately represented by the hierarchical linear model” (Snijders & Bosker, 1999, p. 3). In other words more variables could improve the model fit by explaining additional variability in mathematics performance, but maybe not. For this study, the available data were added to the model and tested for
improvement. Data not available such as previous NeSA grade level scores, or pretest mathematics scores, or other mathematics assessment scores could improve the model.

As stated earlier, the experimental design was limited to a posttest only quasi-experimental design where the online and paper/pencil test groups may differ in systematic ways other than test mode. To better control for systematic differences, random assignment to test mode groups would be ideal. Random sampling was not feasible, so alternative explanations were considered by controlling for student characteristics and school characteristics. Shadish, Cook, and Campbell (2002) stated “in many cases these alternative explanations are never completely enumerable in advance” (Shadish et al., 2002, p. 14). More explanatory variables would address this issue as well.

**Further Research**

This study examined mean difference between administrative modes, but did not analyze differential item functioning (DIF) where group differences are identified at the item level. Since the study revealed some or no difference in mean performance for the four student/school groups, further analysis to discover items that exhibit DIF for either the paper/pencil or online groups should be addressed. Although the items were field tested and selected for the operational tests based on a lack of DIF between demographical groups, a follow-up look at DIF analysis would provide additional evidence of comparability. Is there an item difference in student performance by mode?

As mentioned in the discussion, the addition of variables could increase the percent of variance explained in the models. The inclusion of student variables that measure pervious mathematics performance such as NeSA mathematics scores from prior
years or norm referenced tests could further reduce the variability within schools. A revised model with mathematics performance variables might fit the data better. Likewise the addition of student level variables such as student motivation or gifted identification could result in better model fit and a reduction of variability. Studies incorporating school characteristics such as ratio of full-time teachers to number of students or the number mathematics course available in the school could explain variability between schools (Liu & Koirala, 2010). What other level 1 and level 2 variables could impact mathematics performance within and between schools?

The investigation of which factors determine the assessment mode for SPED students could explain test mode differences for these students. SPED students are provided certain accommodations for NeSA assessments. Some of the SPED accommodations include:

- content presentation such as audio presentation of directions, content, and test items to student;
- student response such as student responds orally to test items and test administrator records student responses; and
- timing/scheduling/setting such as test administrator provides multiple and frequent breaks during testing time.

Do accommodations supporting SPED students dictate their assessment mode? Are certain accommodations such as reading the questions only available for paper/pencil testing? The incorporation of accommodations in the model could better define this subgroup and show a possible impact on mathematics performance not related to assessment mode.
Likewise a similar study looking at accommodations utilized by ELL students could reveal potential factors on mathematics achievement affecting this subgroup. Some of the ELL accommodations include:

- Direct Linguistic Support with Test Directions such as test administrator reads directions aloud in English or native language,
- Direct Linguistic Support with Content and Test Items such as test administrator provides translated audio recording of content and test items in English or native language to be used in conjunction with paper/pencil test.
- Indirect Linguistic Support such as test administrator provides a flexible testing schedule.

Do accommodations supporting ELL students determine their assessment mode? The study could include a measure of the student’s level of English fluency as a variable in the model. Does their fluency level determine the assessment mode for ELL students? Is there a relationship between the student’s level of fluency and their mathematics performance? The addition of ELL accommodations could better explain the mathematics performance of ELL students.

**Conclusions and Implications**

The purpose of this study was to establish comparability between paper/pencil and computer based administration of the NeSA mathematics assessment and to determine the differential impact of the administrative mode on subgroups. Both student level and school level variables were considered in fitting the model to the data. This study helps mathematics educators, administrators, and policy makers determine whether
testing mathematics performance in different administrative modes produce equivalent scores or not.

After controlling for student and school level variables, the mathematics performance of white male students in schools with state average FRL rate and percent of white students not qualified for FRL or ELL support or SPED services who took online assessments in schools selecting online testing was not different from those who took paper/pencil assessments in schools selecting paper/pencil testing. The analysis yielded no significant difference and estimates equivalent scores regardless of the mode. Based on the evidence mathematics educators, administrators, and policy makers can be confident that any difference in mathematics performance is not based on the testing modes, but rather other school/student variables. Comparison of mathematics performance from the two test/assessment modes can be made after controlling for student and school level variables. Contrary to this finding, differences in mathematics performance in some grades were found for white male students not qualified for FRL or ELL support or SPED services who took online assessments in schools selecting paper/pencil testing or who took paper/pencil assessments in schools selecting online testing. These results suggest the need for further study of students taking assessments different from their school’s test selection.

Nebraska is poised to expand the online testing to more students just like many states who conduct their annual accountability assessments on computer. The state policy makers have considered several factors in addition to the comparability of online and paper/pencil testing in considering the transition to online testing. Other potential advantages for online testing over paper/pencil testing include (Pliskin, 2011, p. 5):
• improved efficiency by eliminating the handling, scanning and mailing of paper/pencil material,
• increased security with multilevel password protection and the reduction of paper booklets and printed answer keys, and
• enhanced timeliness of assessment reports and feedback.

Tim Davey (2011) identifies several conveniences for computerized testing to include automatic entry of scores into databases, flexible schedule of testing, and student preference for testing on computers. ETS suggests that online testing “creates opportunities to support accommodations for students with special needs” (Pliskin, 2011, p. 5). These advantages along with the comparable performance on both tests modes need to be considered by Nebraska Policy makers in moving forward with online assessments for Nebraska schools.
References


Appendix A

Tables Showing Mathematics Performance of Grades 4, 8, and 11
Table A1

Statistics for Grade 4 and Subgroups—Number, Mean, and Standard Deviation

<table>
<thead>
<tr>
<th>Test Mode</th>
<th>Student Paper/ School Paper</th>
<th>Student Online/ School Online</th>
<th>Student Paper/ School Online</th>
<th>Student Online/ School Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num</td>
<td>Mean</td>
<td>S.D.</td>
<td>Num</td>
<td>Mean</td>
</tr>
<tr>
<td>All</td>
<td>7961</td>
<td>41.27</td>
<td>9.66</td>
<td>12993</td>
</tr>
<tr>
<td>Male</td>
<td>4100</td>
<td>41.15</td>
<td>9.36</td>
<td>6526</td>
</tr>
<tr>
<td>Female</td>
<td>3861</td>
<td>41.41</td>
<td>9.36</td>
<td>6467</td>
</tr>
<tr>
<td>Sped</td>
<td>1378</td>
<td>35.65</td>
<td>10.37</td>
<td>2143</td>
</tr>
<tr>
<td>Not Sped</td>
<td>6583</td>
<td>42.45</td>
<td>9.07</td>
<td>10850</td>
</tr>
<tr>
<td>FRL</td>
<td>3970</td>
<td>38.22</td>
<td>9.81</td>
<td>5643</td>
</tr>
<tr>
<td>Not FRL</td>
<td>3991</td>
<td>44.32</td>
<td>8.40</td>
<td>7350</td>
</tr>
<tr>
<td>ELL</td>
<td>1126</td>
<td>36.94</td>
<td>9.29</td>
<td>1205</td>
</tr>
<tr>
<td>Not ELL</td>
<td>6835</td>
<td>41.99</td>
<td>9.53</td>
<td>11788</td>
</tr>
</tbody>
</table>
Table A2

*Statistics for Grade 8 and Subgroups—Number, Mean, and Standard Deviation*

<table>
<thead>
<tr>
<th>Test Mode</th>
<th>Group</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>8223</td>
<td>42.71</td>
<td>12.08</td>
<td>12014</td>
<td>43.79</td>
<td>11.83</td>
<td>277</td>
<td>28.50</td>
<td>10.77</td>
<td>29</td>
<td>18.93</td>
<td>7.19</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>4262</td>
<td>42.40</td>
<td>12.39</td>
<td>6063</td>
<td>43.57</td>
<td>12.19</td>
<td>181</td>
<td>27.88</td>
<td>10.58</td>
<td>22</td>
<td>18.45</td>
<td>6.50</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>3961</td>
<td>3.06</td>
<td>11.73</td>
<td>5951</td>
<td>44.01</td>
<td>11.44</td>
<td>96</td>
<td>29.66</td>
<td>11.09</td>
<td>7</td>
<td>20.49</td>
<td>9.47</td>
</tr>
<tr>
<td></td>
<td>Sped</td>
<td>1122</td>
<td>31.64</td>
<td>11.78</td>
<td>1479</td>
<td>33.08</td>
<td>12.27</td>
<td>238</td>
<td>27.34</td>
<td>9.96</td>
<td>14</td>
<td>16.36</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td>Not Sped</td>
<td>7101</td>
<td>44.46</td>
<td>11.16</td>
<td>10535</td>
<td>45.29</td>
<td>10.96</td>
<td>39</td>
<td>35.59</td>
<td>12.79</td>
<td>15</td>
<td>21.33</td>
<td>9.12</td>
</tr>
<tr>
<td></td>
<td>FRL</td>
<td>3661</td>
<td>37.71</td>
<td>12.11</td>
<td>4895</td>
<td>38.68</td>
<td>12.15</td>
<td>182</td>
<td>27.73</td>
<td>10.55</td>
<td>26</td>
<td>17.73</td>
<td>5.94</td>
</tr>
<tr>
<td></td>
<td>Not FRL</td>
<td>4562</td>
<td>46.72</td>
<td>10.30</td>
<td>7119</td>
<td>47.30</td>
<td>10.08</td>
<td>95</td>
<td>29.98</td>
<td>11.03</td>
<td>3</td>
<td>29.33</td>
<td>10.79</td>
</tr>
<tr>
<td></td>
<td>ELL</td>
<td>858</td>
<td>34.63</td>
<td>11.68</td>
<td>1036</td>
<td>35.64</td>
<td>12.14</td>
<td>20</td>
<td>30.15</td>
<td>10.82</td>
<td>3</td>
<td>19.33</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>Not ELL</td>
<td>7365</td>
<td>43.66</td>
<td>11.70</td>
<td>10978</td>
<td>44.56</td>
<td>11.45</td>
<td>257</td>
<td>28.37</td>
<td>10.77</td>
<td>26</td>
<td>18.88</td>
<td>7.55</td>
</tr>
</tbody>
</table>
Table A3

*Statistics for Grade 11 and Subgroups—Number, Mean, and Standard Deviation*

<table>
<thead>
<tr>
<th>Test Mode</th>
<th>Group</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
<th>Num</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>9404</td>
<td>36.98</td>
<td>13.68</td>
<td>11050</td>
<td>40.02</td>
<td>12.61</td>
<td>282</td>
<td>24.86</td>
<td>9.90</td>
<td>78</td>
<td>19.33</td>
<td>7.27</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>4779</td>
<td>36.96</td>
<td>14.06</td>
<td>5616</td>
<td>40.10</td>
<td>12.83</td>
<td>171</td>
<td>25.81</td>
<td>10.67</td>
<td>48</td>
<td>19.81</td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>4625</td>
<td>37.00</td>
<td>13.27</td>
<td>5434</td>
<td>39.94</td>
<td>12.38</td>
<td>111</td>
<td>23.29</td>
<td>8.37</td>
<td>30</td>
<td>18.57</td>
<td>9.12</td>
</tr>
<tr>
<td></td>
<td>Sped</td>
<td>1103</td>
<td>24.69</td>
<td>10.39</td>
<td>1093</td>
<td>27.52</td>
<td>10.68</td>
<td>100</td>
<td>24.10</td>
<td>10.43</td>
<td>29</td>
<td>35.00</td>
<td>5.90</td>
</tr>
<tr>
<td></td>
<td>FRL</td>
<td>3662</td>
<td>30.91</td>
<td>12.36</td>
<td>3461</td>
<td>34.03</td>
<td>12.18</td>
<td>144</td>
<td>23.32</td>
<td>6.45</td>
<td>53</td>
<td>18.62</td>
<td>6.45</td>
</tr>
<tr>
<td></td>
<td>Not FRL</td>
<td>5742</td>
<td>40.86</td>
<td>12.89</td>
<td>7589</td>
<td>42.75</td>
<td>11.75</td>
<td>138</td>
<td>26.46</td>
<td>11.17</td>
<td>25</td>
<td>20.84</td>
<td>6.84</td>
</tr>
<tr>
<td></td>
<td>ELL</td>
<td>936</td>
<td>29.21</td>
<td>11.59</td>
<td>497</td>
<td>30.74</td>
<td>10.76</td>
<td>49</td>
<td>21.72</td>
<td>7.32</td>
<td>4</td>
<td>17.00</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>Not ELL</td>
<td>8468</td>
<td>37.84</td>
<td>13.55</td>
<td>10553</td>
<td>40.46</td>
<td>12.48</td>
<td>233</td>
<td>25.52</td>
<td>10.25</td>
<td>74</td>
<td>19.46</td>
<td>7.32</td>
</tr>
</tbody>
</table>
Appendix B

Tables Showing Deviance Comparisons and Pseudo R2
<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value Deviance Diff</th>
<th>DF Diff</th>
<th>Exact $p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: E-only</td>
<td>159065.9</td>
<td>159069.9</td>
<td>159085.9</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b: Add School Rand Int</td>
<td>154098.7</td>
<td>154104.7</td>
<td>154117.6</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 1a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1a</td>
<td>4967.2</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2a: Add Fixed Student Online_Paper</td>
<td>153736.4</td>
<td>153744.4</td>
<td>153761.6</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1b</td>
<td>362.3</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2b: Add Fixed School Online_Paper</td>
<td>153544.1</td>
<td>153554.1</td>
<td>153575.6</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2a</td>
<td>192.3</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2c: Add Random Student Online_Paper</td>
<td>153480.0</td>
<td>153494.0</td>
<td>153524.1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2b</td>
<td>64.1</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>2d: Add Cross-Level Online_Paper</td>
<td>153471.4</td>
<td>153487.4</td>
<td>153521.4</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2c</td>
<td>8.6</td>
<td>1</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1b</td>
<td>627.3</td>
<td>5</td>
<td>0.0000</td>
</tr>
<tr>
<td>3a: Add Fixed Student FRlunch</td>
<td>152387.3</td>
<td>152405.3</td>
<td>152444.0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2d</td>
<td>1084.18</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>3b: Add Fixed School FRlunch</td>
<td>152329.5</td>
<td>152349.5</td>
<td>152392.5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 3a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3a</td>
<td>57.8</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>3c: Add Random Student FRlunch</td>
<td>152296.6</td>
<td>152322.6</td>
<td>152378.5</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3b</td>
<td>32.9</td>
<td>3</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table B1 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value</th>
<th>Deviance Diff</th>
<th>DF Diff</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3d: Add Cross-Level FRlunch (reject) Comparison of model 3c</td>
<td>152294.6</td>
<td>152322.6</td>
<td>152382.8</td>
<td>14</td>
<td>3c</td>
<td>2.0</td>
<td>1</td>
<td>0.1573</td>
<td></td>
</tr>
<tr>
<td>4a: Add Fixed Student Gender Comparison of model 3b</td>
<td>152291.3</td>
<td>152319.3</td>
<td>152379.5</td>
<td>14</td>
<td>3c</td>
<td>5.3</td>
<td>1</td>
<td>0.0213</td>
<td></td>
</tr>
<tr>
<td>4b: Add Fixed School Gender (reject) Comparison of model 4a</td>
<td>152291.0</td>
<td>152321.0</td>
<td>152385.5</td>
<td>15</td>
<td>4a</td>
<td>0.3</td>
<td>1</td>
<td>0.5839</td>
<td></td>
</tr>
<tr>
<td>5a: Add Fixed Student ELL Comparison of model 4a</td>
<td>152111.7</td>
<td>152143.5</td>
<td>152212.3</td>
<td>15</td>
<td>4a</td>
<td>179.3</td>
<td>1</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>5b: Add Fixed School ELL Comparison of model 5a (reject)</td>
<td>152107.6</td>
<td>152141.6</td>
<td>152214.7</td>
<td>16</td>
<td>5a</td>
<td>4.1</td>
<td>1</td>
<td>0.0429</td>
<td></td>
</tr>
<tr>
<td>6a: Add Fixed Student SPED Comparison of model 5a</td>
<td>150592.4</td>
<td>150624.4</td>
<td>150693.1</td>
<td>16</td>
<td>5a</td>
<td>1519.3</td>
<td>1</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>6b: Add Fixed School SPED (reject) Comparison of model 6a</td>
<td>150591.3</td>
<td>150625.3</td>
<td>150698.3</td>
<td>18</td>
<td>6a</td>
<td>1.1</td>
<td>2</td>
<td>0.5769</td>
<td></td>
</tr>
<tr>
<td>7a: Add Fixed Student Ethnicity Comparison of model 6a</td>
<td>150194.8</td>
<td>150238.8</td>
<td>150333.4</td>
<td>22</td>
<td>6a</td>
<td>397.6</td>
<td>6</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>7b: Add Fixed School Ethnicity – White Comparison of model 7a</td>
<td>150189.5</td>
<td>150235.5</td>
<td>150334.4</td>
<td>23</td>
<td>7a</td>
<td>5.3</td>
<td>1</td>
<td>0.0213</td>
<td></td>
</tr>
<tr>
<td>7c: Add Fixed School Ethnicity – Hispanic Comparison of model 7a</td>
<td>150176.6</td>
<td>150224.6</td>
<td>150327.8</td>
<td>24</td>
<td>7b</td>
<td>12.9</td>
<td>1</td>
<td>0.0003</td>
<td></td>
</tr>
</tbody>
</table>

Table B1 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value Deviance Diff</th>
<th>DF Diff</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>8a: Add interaction Sped with Student Online</td>
<td>150159.7</td>
<td>150209.7</td>
<td>150317.2</td>
<td>25</td>
<td>7c</td>
<td>16.9</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 7c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8b: Add interaction Gender with Student Online</td>
<td>150152.4</td>
<td>150204.4</td>
<td>150316.2</td>
<td>26</td>
<td>8a</td>
<td>7.3</td>
<td>1</td>
<td>0.0069</td>
</tr>
<tr>
<td>Comparison of model 8a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8c: Add interaction FRL with Student Online</td>
<td>150143.0</td>
<td>150197.0</td>
<td>150313.1</td>
<td>27</td>
<td>8b</td>
<td>9.4</td>
<td>1</td>
<td>0.0022</td>
</tr>
<tr>
<td>Comparison of model 8b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8d: Add interaction ELL with Student Online</td>
<td>150134.1</td>
<td>150190.1</td>
<td>150310.5</td>
<td>28</td>
<td>8c</td>
<td>8.9</td>
<td>1</td>
<td>0.0029</td>
</tr>
<tr>
<td>Comparison of model 8c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table B2

**Pseudo R2 for Grade 4**

<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FRLunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FRLunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: E-only</td>
<td>92.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b: Add Random School Intercept</td>
<td>69.26</td>
<td>22.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a: Add Fixed Student Online_Paper</td>
<td>67.65</td>
<td>30.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.32</td>
</tr>
<tr>
<td>Comparison of model 1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2b: Add Fixed School Online_Paper</td>
<td>67.48</td>
<td>22.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c: Add Random Student Online_Paper</td>
<td>67.13</td>
<td>24.93</td>
<td>20.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2d: Add Cross-Level Online_Paper</td>
<td>67.01</td>
<td>25.18</td>
<td>22.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>Comparison of model 2c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a: Add Fixed Student FRLunch</td>
<td>63.9</td>
<td>21.56</td>
<td>19.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.23</td>
</tr>
<tr>
<td>3b: Add Fixed School FRLunch</td>
<td>63.97</td>
<td>19.97</td>
<td>16.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.03</td>
</tr>
<tr>
<td>Comparison of model 3a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c: Add Random Student FRLunch</td>
<td>63.70</td>
<td>18.59</td>
<td>16.22</td>
<td>1.57</td>
<td></td>
<td></td>
<td>0.42</td>
<td>6.91</td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B2 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FR Lunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FR Lunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>3d: Add Cross-Level FR Lunch (reject) Comparison of model 3c</td>
<td>63.71</td>
<td>18.42</td>
<td>16.29</td>
<td>1.53</td>
<td>-0.02</td>
<td>0.91</td>
<td>-0.43</td>
<td>2.55</td>
</tr>
<tr>
<td>4a: Add Fixed Student Gender Comparison of model 3b</td>
<td>63.69</td>
<td>18.58</td>
<td>16.23</td>
<td>1.59</td>
<td>0.44</td>
<td>0.05</td>
<td>-0.06</td>
<td>-1.27</td>
</tr>
<tr>
<td>4b: Add Fixed School Gender (reject) Comparison of model 4a</td>
<td>63.69</td>
<td>18.42</td>
<td>16.18</td>
<td>1.57</td>
<td>0.00</td>
<td>0.86</td>
<td>0.31</td>
<td>1.26</td>
</tr>
<tr>
<td>5a: Add Fixed Student ELL Comparison of model 4a</td>
<td>63.16</td>
<td>18.52</td>
<td>16.89</td>
<td>1.37</td>
<td>0.83</td>
<td>0.32</td>
<td>-4.07</td>
<td>13.84</td>
</tr>
<tr>
<td>5b: Add Fixed Student ELL (reject) Comparison of model 5a</td>
<td>63.16</td>
<td>18.35</td>
<td>16.98</td>
<td>1.37</td>
<td>0.00</td>
<td>0.92</td>
<td>-0.53</td>
<td>0.00</td>
</tr>
<tr>
<td>6a: Add Fixed Student SPED Comparison of model 5a</td>
<td>58.84</td>
<td>16.32</td>
<td>12.89</td>
<td>1.28</td>
<td>6.84</td>
<td>11.88</td>
<td>23.70</td>
<td>6.57</td>
</tr>
<tr>
<td>6b: Add Fixed School SPED (reject) Comparison of model 6a</td>
<td>58.84</td>
<td>16.33</td>
<td>12.85</td>
<td>1.28</td>
<td>0.00</td>
<td>-0.06</td>
<td>0.29</td>
<td>0.00</td>
</tr>
<tr>
<td>7a: Add Fixed Student Ethnicity Comparison of model 6a</td>
<td>57.93</td>
<td>15.01</td>
<td>12.74</td>
<td>0.70</td>
<td>1.55</td>
<td>8.03</td>
<td>1.14</td>
<td>45.31</td>
</tr>
<tr>
<td>7b: Add Fixed School Ethnicity—White (final model) Comparison of model 7a</td>
<td>57.93</td>
<td>14.84</td>
<td>12.66</td>
<td>0.69</td>
<td>0.00</td>
<td>1.13</td>
<td>0.63</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Table B2 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FRLunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FRLunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>7c: Add Fixed School Ethnicity—Hispanic (reject)</td>
<td>57.92</td>
<td>15.18</td>
<td>12.56</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.02</td>
<td>-2.29</td>
<td>0.79</td>
<td>2.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.36</td>
<td>34.71</td>
<td>37.11</td>
<td>56.05</td>
</tr>
<tr>
<td>Overall Comparison of model 7b with 1b, 1b, 2c, 3c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8a: Add interaction Sped with Student Online</td>
<td>57.88</td>
<td>15.03</td>
<td>12.30</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.07</td>
<td>0.99</td>
<td>2.07</td>
<td>-4.48</td>
</tr>
<tr>
<td>8b: Add Interaction Gender with Student Online</td>
<td>57.86</td>
<td>15.06</td>
<td>12.29</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
<td>-0.20</td>
<td>0.08</td>
<td>-1.43</td>
</tr>
<tr>
<td>8c: Add Interaction FRL with Student Online</td>
<td>57.84</td>
<td>15.15</td>
<td>12.07</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.09</td>
<td>-0.60</td>
<td>-1.79</td>
<td>4.23</td>
</tr>
<tr>
<td>8d: Add Interaction ELL with Student Online</td>
<td>57.81</td>
<td>15.20</td>
<td>12.22</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
<td>-0.33</td>
<td>-1.16</td>
<td>-4.41</td>
</tr>
</tbody>
</table>
### Table B3

*Deviance Comparison for Grade 8 Models*

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value Deviance Diff</th>
<th>DF Diff</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: E-only</td>
<td>160066.7</td>
<td>160664.7</td>
<td>160680.6</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b: Add School Rand Int</td>
<td>156069.9</td>
<td>156075.9</td>
<td>156087.0</td>
<td>3</td>
<td>1a</td>
<td>3996.8</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 1a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a: Add Fixed Student Online_Paper</td>
<td>155835.4</td>
<td>155843.4</td>
<td>155858.2</td>
<td>4</td>
<td>1b</td>
<td>234.5</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2b: Add Fixed School Online_Paper</td>
<td>155716.2</td>
<td>155726.2</td>
<td>155744.8</td>
<td>5</td>
<td>2a</td>
<td>119.2</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 2a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c: Add Random Student Online_Paper</td>
<td>155632.1</td>
<td>155646.1</td>
<td>155670.0</td>
<td>7</td>
<td>2b</td>
<td>84.1</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2d: Add Cross-Level Online_Paper</td>
<td>155526.0</td>
<td>15542.0</td>
<td>155571.6</td>
<td>8</td>
<td>2c</td>
<td>106.1</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 2c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a: Add Fixed Student FRlunch</td>
<td>154092.6</td>
<td>154110.6</td>
<td>154143.9</td>
<td>9</td>
<td>1b</td>
<td>543.9</td>
<td>5</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b: Add Fixed School FRlunch</td>
<td>154053.5</td>
<td>154073.5</td>
<td>154110.6</td>
<td>10</td>
<td>3a</td>
<td>39.1</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 3a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c: Add Random Student FRlunch</td>
<td>153968.7</td>
<td>153994.7</td>
<td>154042.9</td>
<td>13</td>
<td>3b</td>
<td>84.8</td>
<td>3</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table B3 continues*
<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value Deviance Diff</th>
<th>DF Diff</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3d: Add Cross-Level FRlunch</td>
<td>153958.2</td>
<td>153986.2</td>
<td>154038.1</td>
<td>14</td>
<td>3c</td>
<td>10.5</td>
<td>1</td>
<td>0.0012</td>
</tr>
<tr>
<td>Comparison of model 3c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a: Add Fixed Student Gender</td>
<td>153951.7</td>
<td>153981.7</td>
<td>154037.3</td>
<td>14</td>
<td>3b</td>
<td>101.8</td>
<td>4</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a: Add Fixed Student ELL</td>
<td>153669.5</td>
<td>153701.5</td>
<td>153760.9</td>
<td>16</td>
<td>4a</td>
<td>282.2</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 4a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6a: Add Fixed Student SPED</td>
<td>150828.9</td>
<td>15062.9</td>
<td>150925.9</td>
<td>17</td>
<td>5a</td>
<td>2840.6</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 5a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6b: Add Fixed School SPED (reject)</td>
<td>150818.9</td>
<td>150854.9</td>
<td>150921.6</td>
<td>18</td>
<td>6a</td>
<td>10.0</td>
<td>1</td>
<td>0.0016</td>
</tr>
<tr>
<td>Comparison of model 6a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a: Add Fixed Student Ethnicity</td>
<td>150393.6</td>
<td>150441.6</td>
<td>150530.5</td>
<td>24</td>
<td>6a</td>
<td>425.3</td>
<td>6</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 6a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7b: Add Fixed School Ethnicity – White</td>
<td>150375.0</td>
<td>150425.0</td>
<td>150517.7</td>
<td>25</td>
<td>7a</td>
<td>18.6</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 7a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7c: Add Fixed School Ethnicity – White, Black (reject)</td>
<td>150374.1</td>
<td>150426.1</td>
<td>150522.4</td>
<td>26</td>
<td>7b</td>
<td>0.9</td>
<td>1</td>
<td>0.3428</td>
</tr>
<tr>
<td>Comparison of model 7a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7d: Add Fixed School Ethnicity—White, Hispanic</td>
<td>150374.4</td>
<td>150426.4</td>
<td>150522.7</td>
<td>26</td>
<td>7a</td>
<td>0.6</td>
<td>1</td>
<td>0.4386</td>
</tr>
<tr>
<td>Comparison of model 7a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B3 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value Deviance Diff</th>
<th>DF Diff</th>
<th>Exact $p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>8a: Add only Interactions for SPED, ELL, Black</td>
<td>150356.3</td>
<td>150412.3</td>
<td>150516.1</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 7b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7b</td>
<td>18.7</td>
<td>3</td>
<td>0.0003</td>
</tr>
<tr>
<td>8a: Interaction for SPED with Online Testing</td>
<td>150369.8</td>
<td>150421.8</td>
<td>150518.2</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 7c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7c</td>
<td>4.3</td>
<td>1</td>
<td>0.0381</td>
</tr>
<tr>
<td>8b: Interaction for ELL with Online Testing</td>
<td>150364.6</td>
<td>150418.6</td>
<td>150518.7</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 8a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8a</td>
<td>5.2</td>
<td>1</td>
<td>0.0226</td>
</tr>
<tr>
<td>8c: Interaction for Black with Online Testing</td>
<td>150356.3</td>
<td>150412.3</td>
<td>150516.1</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 8b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8b</td>
<td>8.3</td>
<td>1</td>
<td>0.0040</td>
</tr>
</tbody>
</table>
Table B4

*Pseudo R2 for Grade 8*

<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FRLunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FRLunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: E-only</td>
<td>145.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b: Add Random School Intercept</td>
<td>113.10</td>
<td>23.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a: Add Fixed Student Online_Paper</td>
<td>111.39</td>
<td>31.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.51</td>
<td>-34.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2b: Add Fixed School Online_Paper</td>
<td>111.15</td>
<td>23.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.22</td>
<td>25.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c: Add Random Student Online_Paper</td>
<td>110.21</td>
<td>17.73</td>
<td>22.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2d: Add Cross-Level Online_Paper</td>
<td>110.02</td>
<td>22.32</td>
<td>11.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.17</td>
<td>-25.93</td>
<td>47.22</td>
<td></td>
</tr>
<tr>
<td>3a: Add Fixed Student FRLunch</td>
<td>102.82</td>
<td>17.78</td>
<td>12.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7.49</td>
<td>23.34</td>
<td>43.03</td>
<td></td>
</tr>
<tr>
<td>3b: Add Fixed School FRLunch</td>
<td>102.83</td>
<td>14.84</td>
<td>13.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 3a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.01</td>
<td>16.51</td>
<td>-4.33</td>
<td></td>
</tr>
<tr>
<td>3c: Add Random Student FRLunch</td>
<td>102.01</td>
<td>13.26</td>
<td>14.01</td>
<td>4.51</td>
<td>0.80</td>
<td>10.68</td>
<td>-4.15</td>
<td></td>
</tr>
</tbody>
</table>

Table B4 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FRLunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FRLunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>3d: Add Cross-Level FRLunch</td>
<td>102.03</td>
<td>13.22</td>
<td>14.33</td>
<td>3.83</td>
<td>-0.02</td>
<td>0.27</td>
<td>-2.34</td>
<td>15.11</td>
</tr>
<tr>
<td>Comparison of model 3c</td>
<td>102.00</td>
<td>13.19</td>
<td>14.31</td>
<td>3.80</td>
<td>0.81</td>
<td>11.17</td>
<td>-6.41</td>
<td>15.71</td>
</tr>
<tr>
<td>4a: Add Fixed Student Gender</td>
<td>100.58</td>
<td>13.48</td>
<td>15.17</td>
<td>3.7</td>
<td>1.39</td>
<td>-2.21</td>
<td>-5.99</td>
<td>1.34</td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td>87.58</td>
<td>13.59</td>
<td>18.68</td>
<td>2.61</td>
<td>12.92</td>
<td>-0.84</td>
<td>-23.18</td>
<td>30.52</td>
</tr>
<tr>
<td>5a: Add Fixed Student ELL</td>
<td>87.57</td>
<td>13.88</td>
<td>18.70</td>
<td>2.54</td>
<td>0.01</td>
<td>-2.13</td>
<td>-0.10</td>
<td>2.51</td>
</tr>
<tr>
<td>Comparison of model 4a</td>
<td>85.93</td>
<td>13.18</td>
<td>18.84</td>
<td>1.75</td>
<td>1.88</td>
<td>3.05</td>
<td>-0.84</td>
<td>32.91</td>
</tr>
<tr>
<td>6a: Add Fixed Student SPED</td>
<td>85.93</td>
<td>12.58</td>
<td>17.80</td>
<td>1.72</td>
<td>0.01</td>
<td>4.52</td>
<td>5.52</td>
<td>1.60</td>
</tr>
<tr>
<td>Comparison of model 5a</td>
<td>85.9</td>
<td>12.6y7</td>
<td>18.04</td>
<td>1.71</td>
<td>24.02</td>
<td>45.83</td>
<td>21.37</td>
<td>61.83</td>
</tr>
<tr>
<td>6b: Add Fixed School SPED</td>
<td>85.93</td>
<td>13.88</td>
<td>18.70</td>
<td>2.54</td>
<td>0.01</td>
<td>4.52</td>
<td>5.52</td>
<td>1.60</td>
</tr>
<tr>
<td>Comparison of model 6a</td>
<td>85.93</td>
<td>12.58</td>
<td>17.80</td>
<td>1.72</td>
<td>24.02</td>
<td>45.83</td>
<td>21.37</td>
<td>61.83</td>
</tr>
<tr>
<td>7a: Add Fixed Student Ethnicity</td>
<td>85.93</td>
<td>13.18</td>
<td>18.84</td>
<td>1.75</td>
<td>1.88</td>
<td>3.05</td>
<td>-0.84</td>
<td>32.91</td>
</tr>
<tr>
<td>Comparison of model 6a</td>
<td>85.93</td>
<td>12.58</td>
<td>17.80</td>
<td>1.72</td>
<td>24.02</td>
<td>45.83</td>
<td>21.37</td>
<td>61.83</td>
</tr>
<tr>
<td>7b: Add Fixed School Ethnicity—White (final model)</td>
<td>85.93</td>
<td>12.58</td>
<td>17.80</td>
<td>1.72</td>
<td>0.01</td>
<td>4.52</td>
<td>5.52</td>
<td>1.60</td>
</tr>
<tr>
<td>Comparison of model 7a</td>
<td>85.93</td>
<td>12.58</td>
<td>17.80</td>
<td>1.72</td>
<td>24.02</td>
<td>45.83</td>
<td>21.37</td>
<td>61.83</td>
</tr>
<tr>
<td>Overall Comparison with model 1b, 1b, 2a, and 3c</td>
<td>85.93</td>
<td>12.58</td>
<td>17.80</td>
<td>1.72</td>
<td>0.01</td>
<td>4.52</td>
<td>5.52</td>
<td>1.60</td>
</tr>
<tr>
<td>8a: Add interaction Sped with Student Online</td>
<td>85.9</td>
<td>12.6y7</td>
<td>18.04</td>
<td>1.71</td>
<td>0.03</td>
<td>-0.70</td>
<td>-1.34</td>
<td>0.58</td>
</tr>
<tr>
<td>Comparison of model 7b</td>
<td>85.93</td>
<td>12.58</td>
<td>17.80</td>
<td>1.72</td>
<td>24.02</td>
<td>45.83</td>
<td>21.37</td>
<td>61.83</td>
</tr>
</tbody>
</table>

Table B4 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FRLunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FRLunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>8b: Add Interaction ELL with Student Online</td>
<td>85.88</td>
<td>12.7</td>
<td>18.17</td>
<td>1.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 8a</td>
<td></td>
<td>0.02</td>
<td>-0.24</td>
<td>-0.72</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8c: Add Interaction Black with Student Online</td>
<td>85.86</td>
<td>12.39</td>
<td>17.41</td>
<td>1.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 8b</td>
<td></td>
<td>0.02</td>
<td>2.44</td>
<td>4.18</td>
<td>2.34</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table B5

*Deviance Comparison for Grade 11 Models*

<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value Deviance Diff</th>
<th>DF Diff</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: E-only</td>
<td>166755.1</td>
<td>166759.1</td>
<td>166775.0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b: Add School Rand Int</td>
<td>163220.3</td>
<td>163226.3</td>
<td>163237.2</td>
<td>3</td>
<td>1a</td>
<td>3534.8</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2a: Add Fixed Student Online_Paper</td>
<td>163087.8</td>
<td>163095.8</td>
<td>163110.3</td>
<td>4</td>
<td>1b</td>
<td>132.5</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2b: Add Fixed School Online_Paper</td>
<td>163047.0</td>
<td>163057.0</td>
<td>163075.1</td>
<td>5</td>
<td>2a</td>
<td>40.8</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>2c: Add Random Student Online_Paper</td>
<td>162927.1</td>
<td>162941.1</td>
<td>162966.4</td>
<td>7</td>
<td>2b</td>
<td>119.9</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>2d: Add Cross-Level Online_Paper</td>
<td>162807.9</td>
<td>162823.9</td>
<td>162852.9</td>
<td>8</td>
<td>2c</td>
<td>119.2</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>3a: Add Fixed Student FRLunch</td>
<td>161497.3</td>
<td>161515.3</td>
<td>161547.9</td>
<td>9</td>
<td>2b</td>
<td>1549.7</td>
<td>4</td>
<td>0.0000</td>
</tr>
<tr>
<td>3b: Add Fixed School FRLunch</td>
<td>161436.8</td>
<td>161456.8</td>
<td>161492.9</td>
<td>10</td>
<td>3a</td>
<td>60.5</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>3c: Add Random Student FRLunch</td>
<td>161397.8</td>
<td>161423.8</td>
<td>161470.8</td>
<td>13</td>
<td>3b</td>
<td>39.0</td>
<td>3</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table B5 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Model Deviance</th>
<th>AIC</th>
<th>BIC</th>
<th>Model DF</th>
<th>Baseline Model</th>
<th>Abs Value Deviance Diff</th>
<th>DF Diff</th>
<th>Exact p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3d: Add Cross-Level FRlunch (reject)</td>
<td>161396.9</td>
<td>161424.9</td>
<td>161475.5</td>
<td>14</td>
<td>3c</td>
<td>0.9</td>
<td>1</td>
<td>0.3428</td>
</tr>
<tr>
<td>Comparison of model 3c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a: Add Fixed Student Gender</td>
<td>161397.5</td>
<td>161425.5</td>
<td>161476.2</td>
<td>14</td>
<td>3b</td>
<td>39.3</td>
<td>4</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a: Add Fixed Student ELL</td>
<td>161255.9</td>
<td>161285.9</td>
<td>161340.1</td>
<td>15</td>
<td>4a</td>
<td>141.6</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 4a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5b: Add Fixed School ELL</td>
<td>161248.9</td>
<td>161280.9</td>
<td>161338.8</td>
<td>16</td>
<td>5a</td>
<td>7.0</td>
<td>1</td>
<td>0.0082</td>
</tr>
<tr>
<td>Comparison of model 5a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6a: Add Fixed Student SPED</td>
<td>158941.0</td>
<td>158975.0</td>
<td>159036.5</td>
<td>17</td>
<td>5b</td>
<td>2307.9</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 5a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a: Add Fixed Student Ethnicity (No ELL school))</td>
<td>158372.6</td>
<td>158413.9</td>
<td>158497.1</td>
<td>22</td>
<td>6a</td>
<td>568.4</td>
<td>5</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 6a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7b: Add Fixed School Ethnicity – White (No ELL school)</td>
<td>158341.0</td>
<td>158387.0</td>
<td>1547.2</td>
<td>23</td>
<td>7a</td>
<td>31.6</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>Comparison of model 7a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8a: Add only Interactions for ELL, only white school level</td>
<td>158335.2</td>
<td>158383.2</td>
<td>158470.0</td>
<td>24</td>
<td>7b</td>
<td>5.8</td>
<td>1</td>
<td>0.0160</td>
</tr>
<tr>
<td>Comparison of model 7b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8b: Add Interactions with all level one factors, only white school level</td>
<td>158333.9</td>
<td>150211.8</td>
<td>150349.4</td>
<td>27</td>
<td>8a</td>
<td>1.3</td>
<td>3</td>
<td>0.7291</td>
</tr>
</tbody>
</table>
Table B6

*Pseudo R2 for Grade 11*

<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FRLunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FRLunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a: E-only</td>
<td>176.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1b: Add Rand School Intercept</td>
<td>145.09</td>
<td>24.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a: Add Fixed Student Online_Paper</td>
<td>144.05</td>
<td>26.00</td>
<td></td>
<td></td>
<td>0.72</td>
<td>-7.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 1b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2b: Add Fixed School Online_Paper</td>
<td>143.92</td>
<td>23.48</td>
<td></td>
<td></td>
<td>0.09</td>
<td>9.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2c: Add Random Student Online_Paper</td>
<td>142.35</td>
<td>25.33</td>
<td>30.97</td>
<td></td>
<td>0.08</td>
<td>-3.16</td>
<td>68.07</td>
<td></td>
</tr>
<tr>
<td>2d: Add Cross-Level Online_Paper</td>
<td>142.24</td>
<td>26.13</td>
<td>9.89</td>
<td></td>
<td>0.08</td>
<td>-3.16</td>
<td>68.07</td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a: Add Fixed Student FRLunch</td>
<td>133.94</td>
<td>18.35</td>
<td>4.37</td>
<td></td>
<td>6.93</td>
<td>21.85</td>
<td>85.89</td>
<td></td>
</tr>
<tr>
<td>Comparison of model 2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3b: Add Fixed School FRLunch</td>
<td>134.02</td>
<td>10.44</td>
<td>1.17</td>
<td></td>
<td>-0.06</td>
<td>43.11</td>
<td>73.23</td>
<td></td>
</tr>
<tr>
<td>Comparison of model 3a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3c: Add Random Student FRLunch</td>
<td>133.30</td>
<td>9.73</td>
<td>0.61</td>
<td>3.88</td>
<td>0.54</td>
<td>6.80</td>
<td>47.86</td>
<td></td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B6 continues
<table>
<thead>
<tr>
<th>Model</th>
<th>Residual Variance</th>
<th>Random Subject Intercept Variance</th>
<th>Random Online Slope Variance</th>
<th>Random FRLunch Slope Variance</th>
<th>% Residual Variance Reduced</th>
<th>% Random Subject Intercept Reduced</th>
<th>% Random Online Slope Reduced</th>
<th>% Random FRLunch Slope Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>4a: Add Fixed Student Gender</td>
<td>133.29</td>
<td>9.74</td>
<td>0.61</td>
<td>3.88</td>
<td>0.54</td>
<td>6.70</td>
<td>47.86</td>
<td>0.00</td>
</tr>
<tr>
<td>Comparison of model 3b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a: Add Fixed Student ELL</td>
<td>132.41</td>
<td>9.97</td>
<td>1.89</td>
<td>3.59</td>
<td>0.66</td>
<td>-2.36</td>
<td>-209.84</td>
<td>7.47</td>
</tr>
<tr>
<td>Comparison of model 4a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5b: Add Fixed School ELL</td>
<td>132.42</td>
<td>8.99</td>
<td>1.23</td>
<td>3.62</td>
<td>-0.01</td>
<td>9.83</td>
<td>34.92</td>
<td>-0.84</td>
</tr>
<tr>
<td>Comparison of model 5a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6a: Add Fixed Student SPED</td>
<td>118.42</td>
<td>10.09</td>
<td>2.59</td>
<td>2.96</td>
<td>10.57</td>
<td>-12.24</td>
<td>-110.57</td>
<td>18.23</td>
</tr>
<tr>
<td>Comparison of model 5a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7a: Add Fixed Student Ethnicity</td>
<td>115.43</td>
<td>9.48</td>
<td>3.90</td>
<td>1.86</td>
<td>2.52</td>
<td>6.05</td>
<td>-50.58</td>
<td>37.16</td>
</tr>
<tr>
<td>Comparison of model 6a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7b: Add Fixed School Ethnicity – White (final model)</td>
<td>115.43</td>
<td>6.37</td>
<td>2.92</td>
<td>1.88</td>
<td>0.00</td>
<td>32.81</td>
<td>25.13</td>
<td>-1.08</td>
</tr>
<tr>
<td>8a: Add Sign. Interaction with Student Online and ELL (reject)</td>
<td>115.40</td>
<td>6.41</td>
<td>3.11</td>
<td>1.87</td>
<td>0.03</td>
<td>-0.63</td>
<td>-6.45</td>
<td>0.34</td>
</tr>
</tbody>
</table>
Appendix C

Final Programs for Grade 4, 8, and 11 Models
Program C.1: SAS Program for Grade 4 Final Model

Title 'Random Intercept_fixed effect of student and school online_paper choice plus interaction';
Title2 'Add fixed effect of student FRL and school FRL plus random student FRL';
Title3 'Add fixed effect of student GEN';
Title4 'Add fixed effect of student ELL';
Title5 'Add fixed effect of student SPED';
Title6 'Add fixed effect of student ETHNICITY with Significant School Percent white';

*data file math performance for 4th grade, input data school id and student id from data file;
proc mixed data = John.comp4b NOCLPRINT NOITPRINT COVTEST NAMELEN = 100
   METHOD = ML;
class agencyid NDE_STUDENT_ID ;

*model dependent variable raw math scores for NeSA;
*Independent admin mode variables - Fixed school online/paper, school online/paper, and
interaction of student online/paper with school online/paper;
model RAW_MATH_SCORE = stud_online_paper SCHOOL_ONLINE_PAPER
   stud_online_paper*SCHOOL_ONLINE_PAPER

*Student level independent variables FRL, gender, ELL, SPED, and Ethnicities;
food_program_code2 GENDER_CODE
ellcode2 SPECIAL_EDUCATION_CODE
HISPdc AMINdc ASIAdc BLACdc NHPIdc TWOdc

*School level independent variables Fixed School Ethnicity – white,Fixed school percent FRL;
PCT_WH70 PCT_FRL47

*Random variables intercept, student online paper, FRL;
solution ddfm = satterthwaite;
random intercept stud_online_paper food_program_code2/type = UN subject = agencyid;

*Estimates for each admin mode - P/P, P/O, O/O, O/P.
Estimate "School Paper / Student Paper effect" stud_online_paper 0 SCHOOL_ONLINE_PAPER 0;
Estimate "School Paper / Student Online effect" stud_online_paper 1 SCHOOL_ONLINE_PAPER 0;
Estimate "School Online / Student Paper effect" stud_online_paper 0 SCHOOL_ONLINE_PAPER 1;
Estimate "School Online / Student Paper effect 2" stud_online_paper 0 SCHOOL_ONLINE_PAPER 1;
run;
Program C.2: SAS Program for Grade 8 Final Model

Title 'Random Intercept, fixed effect of student and school online_paper choice plus interaction';
Title2 'Add fixed effect of student FRL and school FRL plus random student FRL';
Title3 'Add fixed effect of student GEN';
Title4 'Add fixed effect of student ELL';
Title5 'Add fixed effect of student SPED and school SPED';
Title6 'Add fixed effect of student ETHNICITY and School Ethnicity White';

*data file math performance for 8th grade, input data school id and student id from data file;
proc mixed data = John.comp8b NOCLPRINT NOITPRINT COVTEST NAMELEN = 100
METHOD = ML;
class agencyid NDE_STUDENT_ID;

*model dependent variable raw math scores for NeSA;
*Independent admin mode variables - Fixed school online/paper, school online/paper, and interaction of student online/paper with school online/paper;
model RAW_MATH_SCORE = stud_online_paper SCHOOL_ONLINE_PAPER
stud_online_paper*SCHOOL_ONLINE_PAPER

*Student level independent variables FRL, gender, ELL, SPED, and Ethnicities;
food_program_code2 GENDER_CODE ellcode2
SPECIAL_EDUCATION_CODE HISPdc AMINdc ASIAdc BLACdc NJPIdc TWOdc

*School level independent variables Fixed School Ethnicity – white,Fixed school percent FRL and fixed school percent SPED;
PCT_WH72 PCT_FRL43 PCT_SPED14

*Random variables intercept, student online paper, FRL;
/solution ddfm = satterthwaite;
random intercept stud_online_paper food_program_code2/ type = UN subject = agencyid;

*Estimates for each admin mode - P/P, P/O, O/O, O/P.
Estimate "School Paper / Student Paper effect" stud_online_paper 0 SCHOOL_ONLINE_PAPER 0;
Estimate "School Paper / Student Online effect" stud_online_paper 1 SCHOOL_ONLINE_PAPER 0;
Estimate "School Online / Student Paper effect" stud_online_paper 1 SCHOOL_ONLINE_PAPER 1;
Estimate "School Online / Student Paper effect 2" stud_online_paper 0 SCHOOL_ONLINE_PAPER 1;

run;
Program C.3: SAS Program for Grade 11 Final Model

title 'Random Intercept_fixed effect of student and school online_paper choice plus interaction';
Title2 'Add fixed effect of student FRL and school FRL plus random student FRL';
Title3 'Add fixed effect of student GEN ';
Title4 'Add fixed effect of student ELL';
Title5 'Add fixed effect of student SPED';
Title6 'Add fixed effect of student ETHNICITY and school ethnicity percents - white';

*data file math performance for 11th grade, input data school id and student id from data file;
proc mixed data = John.comp11b NOCLPRINT NOITPRINT COVTEST NAMELEN = 100
METHOD = ML;
Input class agencyid NDE_STUDENT_ID ;

*model dependent variable raw math scores for NeSA;
*Independent admin mode variables - Fixed school online/paper, school online/paper, and
interaction of student online/paper with school online/paper;
model RAW_MATH_SCORE = stud_online_paper SCHOOL_ONLINE_PAPER
stud_online_paper*SCHOOL_ONLINE_PAPER

*Student level independent variables FRL, gender, ELL, SPED, and Ethnicities;
food_program_code2 GENDER_CODE
ellcode2 SPECIAL_EDUCATION_CODE
HISPdc AMINdc ASIAde BLACdc NHPIdc TWOdc

*School level independent variables Fixed School Ethnicity – white and Fixed school percent
FRL;
PCT_WH75 PCT_FRL35

*Random variables intercept, student online paper, FRL;
/solution ddfm = satterthwaite;
random intercept stud_online_paper food_program_code2/ type = UN subject = agencyid;

*Estimates for each admin mode - P/P, P/O, O/O, O/P.
Estimate "School Paper / Student Paper effect" stud_online_paper 0 SCHOOL_ONLINE_PAPER 0;
Estimate "School Paper / Student Online effect" stud_online_paper 1
SCHOOL_ONLINE_PAPER 0;
Estimate "School Online / Student Paper effect" stud_online_paper 1
SCHOOL_ONLINE_PAPER 1;
Estimate "School Online / Student Paper effect 2" stud_online_paper 0
SCHOOL_ONLINE_PAPER 1;

run;
Appendix D

IRB Approval Letter
August 22, 2011

John Moon  
Department of Educational Psychology  

Delwyn Harnisch  
Teaching, Learning and Teacher Education  
125A HECO, UNL, 68588-0800  

IRB Number: 20110811942 EX  
Project ID: 11942  
Project Title: Comparability of Mode for Mathematics Assessment - Online and Paper/Pencil  

Dear John:  

This letter is to officially notify you of the approval of your project by the Institutional Review Board (IRB) for the Protection of Human Subjects. It is the Board’s opinion that you have provided adequate safeguards for the rights and welfare of the participants in this study based on the information provided. Your proposal is in compliance with this institution’s Federal Wide Assurance 00002258 and the DHHS Regulations for the Protection of Human Subjects (45 CFR 46) and has been classified as Exempt Category 4.  

You are authorized to implement this study as of the Date of Final Approval: 08/22/2011.  

We wish to remind you that the principal investigator is responsible for reporting to this Board any of the following events within 48 hours of the event:  
* Any serious event (including on-site and off-site adverse events, injuries, side effects, deaths, or other problems) which in the opinion of the local investigator was unanticipated, involved risk to subjects or others, and was possibly related to the research procedures;  
* Any serious accidental or unintentional change to the IRB-approved protocol that involves risk or has the potential to recur;  
* Any publication in the literature, safety monitoring report, interim result or other finding that indicates an unexpected change to the risk/benefit ratio of the research;  
* Any breach in confidentiality or compromise in data privacy related to the subject or others; or
* Any complaint of a subject that indicates an unanticipated risk or that cannot be resolved by the research staff.

This project should be conducted in full accordance with all applicable sections of the IRB Guidelines and you should notify the IRB immediately of any proposed changes that may affect the exempt status of your research project. You should report any unanticipated problems involving risks to the participants or others to the Board.

If you have any questions, please contact the IRB office at 472-6965.

Sincerely,

Becky R. Freeman, CIP
for the IRB