ELEMENTARY SCHOOL BENCHMARK AND END-OF-COURSE ASSESSMENT ANALYSIS

Prepared for Durham Public Schools

December 2016

In the following report, Hanover Research adopts correlation tests and descriptive statistics to analyze the extent to which benchmark assessments given throughout the year are aligned with the end-of-course (EOC) exams for elementary school students.

HANOVER RESEARCH

www.hanoverresearch.com

TABLE OF CONTENTS

Executive Summary and Key Findings 3
Introduction 3
Key Findings 3
Section I: Data and Methodology 5
Data 5
Outcome Variables 5
Segmentation Variables 6
Methodology 7
Pearson Correlations 8
Achievement Level Comparisons 8
Section II: Topline Comparisons 10
Correlations 10
Achievement Level Comparisons 11
Descriptive Analysis 11
Section III: Correlations by Student Subgroups 16
Appendix: Results by Elementary School Campus 18

© 2016 Hanover Research

EXECUTIVE SUMMARY AND KEY FINDINGS

INTRODUCTION

In this report, Hanover Research (Hanover) evaluates the extent to which student performance on elementary school benchmark assessments is predictive of performance on standardized end-of-course (EOC) exams. We compare benchmark assessment outcomes to EOC outcomes for Reading, Mathematics (Math), and Science. This report builds upon Hanover's previous high school and middle school analyses for Durham Public Schools (Durham).

This report is organized as follows:

- **Section I: Data and Methodology** describes the data and methodology used for this analysis.
- **Section II: Topline Comparisons** analyzes the correlation between the benchmark assessments and the EOC exams for each of the three subjects and the correspondence of proficiency levels across the two assessments for a given subject.
- **Section III: Correlations by Student Subgroups** presents a correlational analysis segmented by student demographic and academic subgroups, as well as by grade levels.
- **Appendix: Results by Elementary School Campus** provides the segmentation variables and correlation analyses for each elementary school campus individually.

KEY FINDINGS

- **Strong and positive correlations exist between the benchmark assessments and the EOC exams for all three subjects, and all correlations are statistically significant at the 99 percent confidence level.** Specifically, the correlation for Science tests is the strongest with a coefficient of 0.82, while the correlation for Reading tests is the weakest with a coefficient of 0.79. However, these correlations are similar for all three subjects. Interpreting the relationships using the units of each assessment, we find that:
  - For the Math test, a 1-percentage-point increase in percent correct on the benchmark assessment is associated with a 0.40-point increase in the corresponding EOC exam's score.
  - For the Reading test, a 1-percentage-point increase in percent correct on the benchmark assessment is associated with a 0.46-point increase in the corresponding EOC exam's score.
  - For the Science test, a 1-percentage point increase in percent correct on the benchmark assessment is associated with a 0.50-point increase in the corresponding EOC exam's score.
- **Slightly more than half (52.4 percent) of all students are rated at the same achievement level by the two assessments.** For a given subject, the percentage of

© 2016 Hanover Research

students being rated at the same level ranges from 54.7 percent (Reading) to 46.7 percent (Science). Further, 88.8 percent of students are rated at the same achievement level or a "neighbor" level (i.e., one level above or below) by the two assessments.

- **Overall, more students are rated at a higher level on the EOC exam compared to the benchmark assessment,** which is similar to the findings of the high school and middle school analyses**.** Specifically, 26.4 percent of students are rated at a higher level on the EOC exam than on the benchmark assessment, whereas only 21.1 percent of students are rated at a lower level on the EOC exam than on the benchmark assessment. In terms of individual subject areas, the largest difference we observe is in Science, where 40.2 percent of students are rated at a higher level on the EOC exam than on the benchmark assessment, while 13.1 percent of students are rated at a lower level.
- **When examining students in different demographic and academic subgroups, we observe positive and significant correlations between the benchmark assessments and EOC exams.** Though the strength of the correlations varies across subgroups, all comparisons have a coefficient of 0.6 or higher. For example, Asian and Multiracial students exhibit the highest correlations (up to 0.81), while Academically Gifted students show the smallest correlations.

© 2016 Hanover Research

SECTION I: DATA AND METHODOLOGY

In this section, Hanover Research describes the data we use for this study and the methodologies we adopt.

DATA

Durham provided Hanover with data on student outcomes for Reading, Math, and Science benchmark assessments and EOC exams for the 2015-16 academic year. This analysis specifically focuses on students from Grade 3 to Grade 5 (elementary school). All students in these grades take the Math and Reading tests, while only students in Grade 5 take the Science tests. We extract the demographic information for each student from the EOC dataset.

For the benchmark assessment data, there were several duplicates at the student level, where we found incidences of two students with the same student ID. In one instance there were two students with the same student ID from the same school. In this case, however, one student had a different student ID in three of the datasets. As such, we assumed the more often used student ID was the correct version and adjusted the mistake in the lone dataset. We then combined all data into a single dataset in which each observation is a particular student and subject. Combining the datasets from the benchmark assessments and the EOC exams gives us 16,920 observations containing academic and demographic variables for 7,528 elementary school students.

OUTCOME VARIABLES

Both the benchmark assessments and EOC exams include multiple measures for each subject. Figure 1.1 describes the measures available for both assessments. As shown in the table, we use percent correct and achievement level as measurements for the benchmark assessments, and score and achievement level as measurements for the EOC exams. These are the same set of measurements used in the previous analyses for Durham high schools and middle schools.

Figure 1.1: Outcome Variables Available for Analysis

ASSESSMENT	MEASUREMENT	USED OR NOT
Benchmark Assessments	**Percent of Questions Answered Correctly**	**Use**
	Suggested Numeric Score	Does not exhibit as much variation as *Percent of Questions Answers Correctly*, so we use the *Percent Correct* → Do not use
	Suggested Letter Mark	Does not have comparable measures in the EOC outcome → Do not use
	Achievement Level	**Ranging from one to five, this functions as a measurement for student proficient level → Use**
	Achievement Level - Extend	Does not have comparable measures in the EOC outcome → Do not use

© 2016 Hanover Research

Assessment	Measurement	Used or Not
EOC	Score	Use
	Achievement Level	Ranging from one to five, this functions as a measurement for student proficient level → Use

Figure 1.2 presents the summary statistics of the outcome variables used in this analysis.

Figure 1.2: Outcome Variables Summary

Outcome	Math		Reading		Science		All Subjects	
	Mean	N	Mean	N	Mean	N	Mean	N
Benchmark Percent Correct	54.24%	7,306	51.55%	7,324	55.10%	2,290	53.19%	16,920
Benchmark Achievement Level	2.58	7,306	2.57	7,324	2.74	2,290	2.60	16,920
EOC Score	447.54	6,967	442.18	7,032	251.45	2,163	418.96	16,162
EOC Achievement Level	2.75	6,967	2.56	7,032	3.15	2,163	2.72	16,162

Segmentation Variables

In Section III, Hanover Research segments the correlational analysis by student subgroups. These groups are based on both demographic and academic factors.

In terms of demographics, we segment by gender, race/ethnicity, and disability status. Note that when segmenting results by race/ethnicity, we focus on whether the student is Asian, African American, Hispanic, Multiracial, or White, since Pacific Islander has a sample size that is too small to yield valid insights (Figure 1.3). We also examine the correlations for students that have limited English proficiency (LEP). Finally, for academic giftedness, we follow the previous analyses and include both a measure of overall academic giftedness as well as whether a student is academically gifted specifically in Reading or Math.

In Hanover's previous analysis for high school students, we segmented the students into three groups: "on track," "ahead of track," and "behind track."[1] However, unlike the high school analysis where the number of students taking the assessments exhibits great variation across grade levels, the number of students taking a given test is fairly balanced within each grade level for elementary school. Thus, for this analysis, we do not segment the students by whether they are on track or not.

Figure 1.3 on the following page presents the summary statistics for the variables we use to segment students. Due to the large number of elementary schools in the dataset, the summary of segmentation variables by campus location is presented in Figure A.1 in the Appendix.

[1] "On track" indicates the student is in the most common grade for students taking the subject; "ahead of track" indicates the student is in a lower grade than the most common grade for that subject; "behind track" indicates the student is in a higher grade than the most common grade for that subject.

© 2016 Hanover Research

Figure 1.3: Segmentation Variables Summary

Segmentation Variable	Math		Reading		Science		All Subjects	
	Pct.	N	Pct.	N	Pct.	N	Pct.	N
Demographic Factors								
Male	50.33%	7,024	50.18%	7,052	50.02%	2,201	50.22%	16,277
Female	49.67%	7,024	49.82%	7,052	49.98%	2,201	49.78%	16,277
Asian	2.35%	7,024	2.24%	7,052	2.27%	2,201	2.29%	16,277
African American	43.25%	7,024	43.35%	7,052	43.03%	2,201	43.26%	16,277
Hispanic	32.39%	7,024	32.35%	7,052	31.67%	2,201	32.27%	16,277
Multi-racial	3.02%	7,024	3.01%	7,052	3.13%	2,201	3.03%	16,277
Pacific Islander	0.37%	7,024	0.38%	7,052	0.50%	2,201	0.39%	16,277
White	18.62%	7,024	18.68%	7,052	19.40%	2,201	18.75%	16,277
Disability	11.72%	7,024	11.63%	7,052	13.04%	2,201	11.86%	16,277
Academic Factors								
Limited English Proficiency	23.12%	7,024	22.77%	7,052	22.90%	2,201	22.94%	16,277
Academically Gifted	20.52%	7,024	20.58%	7,052	22.94%	2,201	20.87%	16,277
Academically Gifted - Math	16.64%	7,024	16.70%	7,052	19.08%	2,201	15.25%	16,277
Academically Gifted - Reading	14.89%	7,024	14.93%	7,052	17.45%	2,201	16.99%	16,277
Grade Level								
Grade 3	34.74%	7,306	34.39%	7,324	0.00%	2,201	29.89%	16,920
Grade 4	32.38%	7,306	32.11%	7,324	0.00%	2,201	27.88%	16,920
Grade 5	29.02%	7,306	29.78%	7,324	96.11%	2,201	38.43%	16,920

METHODOLOGY

In this subsection, Hanover describes the methods we use for this analysis. This study adopts a similar strategy as the previous correlation analyses examining Durham's high school and middle school outcomes. The goal of this analysis is to determine the strength of the correlation between each pair of the assessments in each subject for elementary school students. However, correlational relationships between the benchmark assessments and EOC exams could result from a number of conditions which indicate various relationships between the assessments and may also reflect outcomes of Durham's academic programming.

For example, lower correlation could indicate that the benchmark assessment is different in content to the EOC exam. Further, since content-specific academic programming occurs during the time between the benchmark assessment and EOC exam, low correlation could also indicate that Durham's programming changes student outcomes such that students who perform poorly on the benchmark assessment perform comparatively better on the EOC exam. If programming decisions, such as interventions, are initiated in response to the benchmark assessment, this may further reduce the expected correlation between the two assessments. In short, there are multiple reasons for why we may observe a stronger or weaker correlation between these assessments.

© 2016 Hanover Research

PEARSON CORRELATIONS

Using the percent of questions answered correctly on the benchmark assessment and the scores on the EOC exam for each subject, we compute Pearson correlation coefficients separately for each subject. This correlation coefficient ranges from -1 to 1, with -1 indicating a perfect negative correlation (i.e., when one variable rises, the other falls) and 1 indicating a perfect positive correlation (i.e., the two variables rise or fall together). A coefficient of zero indicates that there is no correlation between the variables. Statistical significance is based on the null hypothesis that the correlation between the two variables is zero, meaning that three "significance stars" which indicate a p-value of less than .01 are evidence that there is a less than 1 percent probability that the relationship between the two variables is zero. Because the scores on the EOC exams are not scaled consistently across the subjects, we do not compute an overall correlation coefficient for the three assessments pooled together.

Building on the presentation of the Pearson correlations, we further describe the relationship between the benchmark assessments and EOC exams based on the results of a simple linear regression, where percent correct from the benchmark assessment is used to predict EOC score. Whereas the Pearson correlation coefficients are standardized, ranging from -1 to 1, unstandardized regression coefficients allow us to express the relationship between assessments in each subject in terms of the units of each assessment (e.g., an "X"-percentage point increase in percent correct on the benchmark assessment is associated with a "Y"-point increase in EOC scaled score).

We also provide scatter plots and trend lines illustrating the relationship between assessments, where the slope of the trend line is equal to the unstandardized regression coefficient for a given subject. Lastly, in Section III, we compute Pearson correlations separately for each subject and each student subgroup, where applicable.

ACHIEVEMENT LEVEL COMPARISONS

We analyze achievement levels by comparing the numbers and percentages of students who fall within each achievement level evaluated by each assessment.

As shown in Figure 1.4 on the following page, we compare the numbers and percentages of students who fall within each achievement level for each assessment. This allows the reader to observe the extent to which students are categorized within the same achievement level for both assessments or in different achievement levels. The numbers in the dark teal cells represent students who fall into the same achievement level for both assessments. The light teal cells represent students who fall into a "neighbor" category (i.e., an EOC level one level higher or lower than the benchmark assessment). The light red cells indicate students whose EOC achievement level is at least two levels above or below their achievement level on the benchmark assessment.

© 2016 Hanover Research

Figure 1.4: Achievement Level Comparisons (Template)

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1						
	2						
	3						
	4						
	5						
Total							

SECTION II: TOPLINE COMPARISONS

In this section, Hanover presents the results of comparing the benchmark assessments with the EOC exams separately for each subject. In the achievement level comparisons below, we also compare the subjects pooled into a single analysis.

CORRELATIONS

The correlation coefficients in Figure 2.1 indicate the Pearson correlation between the percent of questions answered correctly in the benchmark assessment and the scores on the EOC exam. For all three subjects, there is a strong positive correlation that is statistically significant beyond the 99 percent confidence level. These correlations range from 0.785 for Reading to 0.818 for Science.

Figure 2.1: Correlations between Percent Correct on Benchmark and EOC Score

SUBJECT	COEFFICIENT	N
Math	0.806***	6,967
Reading	0.785***	7,032
Science	0.818***	2,163

Asterisks denote statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1.

To further illustrate the relationships between benchmark assessments and EOC exams, we describe these correlations in terms of simple linear regression coefficients,[2] where we predict the EOC score using the percent correct value from the benchmark assessment.

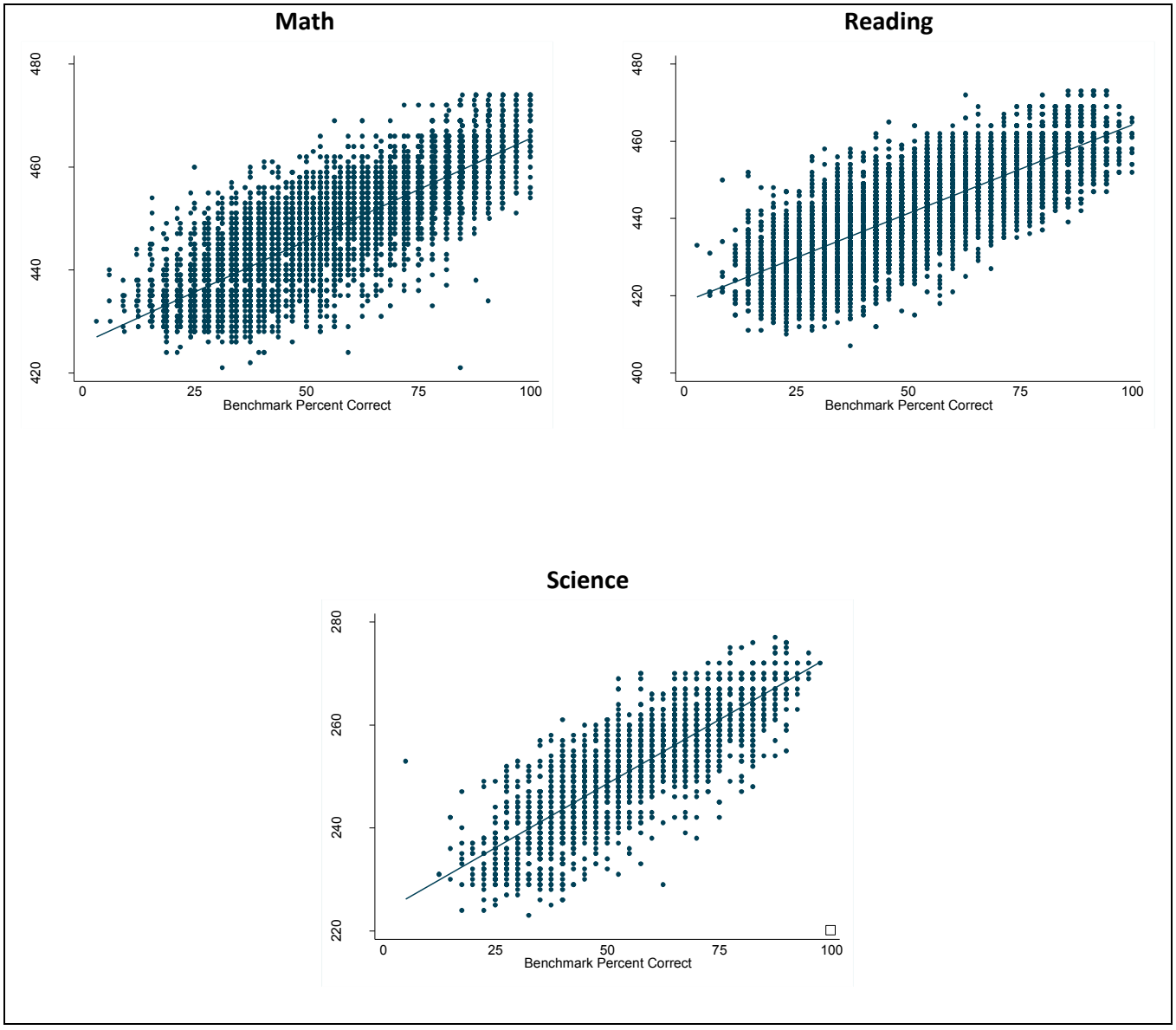
- We find that for the **Math** test, for every 1-percentage-point increase in percent correct on the benchmark assessment, we expect a 0.40-point increase in the corresponding EOC exam score.
- For **Reading**, a 1-percentage-point increase in percent correct on the benchmark assessment is associated with a 0.46-point increase in the corresponding EOC exam score.
- Lastly, every 1-percentage point increase in percent correct on the **Science** benchmark assessment is associated with a 0.50-point increase in the corresponding EOC exam score.

[2] Note that the Pearson correlation coefficients are equivalent to standardized regression coefficients of a simple linear regression involving percent correct on the benchmark assessment and the corresponding EOC assessment scaled scored. Here we discuss the unstandardized regression coefficients, as they allow us to estimate how a change in one unit of the predictor variable (percent correct on the benchmark assessment) relates to changes in units of the outcome variable (EOC scaled score).

© 2016 Hanover Research

We present the scatterplots of student performance on benchmark assessments and EOC exams for each subject in Figure 2.2.

Figure 2.2: Scatterplots of Benchmark Percent Correct and EOC Score



ACHIEVEMENT LEVEL COMPARISONS

DESCRIPTIVE ANALYSIS

Figure 2.3 presents a summary of the achievement analysis. Both assessments have achievement levels ranging from one to five, and according to Figure 2.3, when pooling all three subjects together, 52.4 percent of students are rated at the same level by both assessments. The highest percentage is seen in Reading (54.7 percent) and the lowest is in Science (46.7 percent). Further, 88.8 percent of students are rated at the same achievement

© 2016 Hanover Research

level or a "neighbor" level (i.e., one level above or below). In regards to the individual subject, Reading has the greatest percentage of students rated at the same or at a "neighbor" level by the two assessments (89.8 percent), followed by Science (88.1 percent) and Math (88.0 percent).

Similar to the findings for high school and middle school analyses, more students are rated at a higher achievement level on the EOC exam compared to the benchmark assessment. Specifically, 26.4 percent of students are rated at a higher level on the EOC exam than on the benchmark assessment, whereas only 21.1 percent of students are rated at a lower level on the EOC exam than on the benchmark assessment. The largest difference we observe is in Science, where 40.2 percent of students are rated at a higher level on the EOC exam than on the benchmark assessment, while 13.1 percent of students are rated at a lower level. In comparison, the difference for Math is around 10 percentage points. Interestingly, only 20.4 percent of students test at a higher level in Reading on the EOC exam compared to the benchmark assessment, while 24.9 percent test at a lower level.

Figure 2.3: Differences in Achievement Level from Benchmark to EOC, Summary

Difference	All Subjects	Math	Reading	Science
Same Level	52.43%	51.92%	54.69%	46.74%
Same Level or Neighbor Level	88.82%	88.03%	89.82%	88.12%
Same or Higher Level	78.87%	80.22%	75.06%	86.92%
Higher Level	26.44%	28.30%	20.36%	40.18%
Lower Level	21.13%	19.78%	24.94%	13.08%
Number of Observations	16,162	6,967	7,032	2,163

This figure represents a summary of the results presented in Figures 2.4-2.11.

The remaining figures in this section display detailed comparisons of achievement levels on the benchmark assessments and EOC exams for all subjects combined and each individual subject (summarized in Figure 2.3). Each subject area includes two figures: the first displays percentages, while the second displays counts. As described in Figure 1.4 in the methodology, numbers in the dark teal cells represent students who fall into the same achievement level for both assessments. The light teal cells represent students who fall into a "neighbor" category (i.e., an EOC level one level higher or lower than the benchmark assessment). The light red cells indicate students whose EOC achievement level is at least two levels above or below their achievement level on the benchmark assessment.

© 2016 Hanover Research

All Subjects

Figure 2.4: Achievement Level Comparisons, All Subjects[3]

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	19.60%	5.66%	1.05%	0.74%	0.01%	27.07%
	2	8.23%	10.45%	3.84%	5.04%	0.18%	27.73%
	3	0.88%	3.01%	2.19%	5.84%	0.43%	12.34%
	4	0.31%	2.30%	2.12%	12.94%	3.66%	21.31%
	5	0.02%	0.09%	0.15%	4.04%	7.25%	11.54%
Total		**29.03%**	**21.50%**	**9.34%**	**28.60%**	**11.53%**	**100.00%**

This figure includes all student-subject observations we analyze, meaning that students with data for assessments in multiple subjects are represented multiple times. N=16,162

Figure 2.5: Achievement Level Comparisons, All Subjects

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	3,168	915	170	120	2	4,375
	2	1,330	1,689	620	815	29	4,483
	3	142	486	354	944	69	1,995
	4	50	371	342	2,091	591	3,445
	5	3	14	24	653	1,172	1,866
Total		**4,693**	**3,475**	**1,510**	**4,623**	**1,863**	**16,162**

This figure includes all student-subject observations we analyze, meaning that students with data for assessments in multiple subjects are represented multiple times. N=16,162

Math

Figure 2.6: Achievement Level Comparison, Math

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	18.86%	7.18%	1.21%	0.89%	0.03%	28.16%
	2	7.68%	11.02%	3.62%	5.31%	0.23%	27.86%
	3	0.66%	3.01%	1.71%	5.53%	0.62%	11.52%
	4	0.34%	2.47%	1.58%	11.78%	3.72%	19.89%
	5	0.04%	0.10%	0.09%	3.80%	8.54%	12.57%
Total		**27.58%**	**23.78%**	**8.19%**	**27.31%**	**13.13%**	**100.00%**

N=6,967

[3] Figure 2.4 and Figure 2.5 describe achievement level comparisons for the benchmark assessments and EOC assessments for all subjects pooled. These comparisons allow Durham to observe the overall extent to which the benchmark assessments and EOCs assessments are comparable.

© 2016 Hanover Research

Figure 2.7: Achievement Level Comparison, Math

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	1,314	500	84	62	2	1,962
	2	535	768	252	370	16	1,941
	3	46	210	119	385	43	803
	4	24	172	110	821	259	1,386
	5	3	7	6	265	595	876
Total		**1,922**	**1,657**	**571**	**1,903**	**915**	**6,967**

N=6,967

READING

Figure 2.8: Achievement Level Comparison, Reading

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	22.52%	4.58%	0.95%	0.58%	0.00%	28.64%
	2	9.67%	10.48%	3.65%	4.01%	0.11%	27.93%
	3	1.17%	3.10%	2.29%	3.95%	0.11%	10.62%
	4	0.34%	2.57%	2.79%	12.92%	2.42%	21.04%
	5	0.00%	0.09%	0.26%	4.96%	6.47%	11.77%
Total		**33.70%**	**20.82%**	**9.94%**	**26.43%**	**9.11%**	**100.00%**

N=7,032

Figure 2.9: Achievement Level Comparison, Reading

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	1,584	322	67	41	0	2,014
	2	680	737	257	282	8	1,964
	3	82	218	161	278	8	747
	4	24	181	196	909	170	1,480
	5	0	6	18	349	455	828
Total		**2,370**	**1,464**	**699**	**1,859**	**641**	**7,032**

N=7,032

© 2016 Hanover Research

SCIENCE

Figure 2.10: Achievement Level Comparisons, Science

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	12.48%	4.30%	0.88%	0.79%	0.00%	18.45%
	2	5.32%	8.51%	5.13%	7.54%	0.23%	26.72%
	3	0.65%	2.68%	3.42%	12.99%	0.83%	20.57%
	4	0.09%	0.83%	1.66%	16.69%	7.49%	26.77%
	5	0.00%	0.05%	0.00%	1.80%	5.64%	7.49%
Total		**18.54%**	**16.37%**	**11.10%**	**39.81%**	**14.19%**	**100.00%**

N=2,163

Figure 2.11: Achievement Level Comparisons, Science

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	270	93	19	17	0	399
	2	115	184	111	163	5	578
	3	14	58	74	281	18	445
	4	2	18	36	361	162	579
	5	0	1	0	39	122	162
Total		**401**	**354**	**240**	**861**	**307**	**2,163**

N=2,163

© 2016 Hanover Research

SECTION III: CORRELATIONS BY STUDENT SUBGROUPS

In this section, Hanover describes correlations segmented by student subgroups based on demographic and academic factors.[4] In the Appendix, Figure A.2 presents these correlations by school campus location.

Figure 3.1 presents the correlation coefficients of different student subgroups. As shown in the table, all of the segmented results are positively correlated at the 99 confidence level. Approximately 83 percent of comparisons (35 out of 46 subgroup comparisons) exhibit a coefficient that is greater than 0.70, with the lowest correlation being 0.60.

Figure 3.1: Correlations by Student Subgroups

Segmentation Variable	Math		Reading		Science	
	Coefficient	N	Coefficient	N	Coefficient	N
Demographic Factors						
Female	0.770***	3,460	0.795***	3,499	0.739***	1,082
Male	0.780***	3,507	0.790***	3,533	0.785***	1,081
Asian	0.804***	156	0.758***	156	0.815***	47
African American	0.718***	3,031	0.739***	3,049	0.752***	934
Hispanic	0.736***	2,238	0.738***	2,272	0.703***	677
White	0.753***	1,305	0.767***	1,316	0.698***	425
Multi-racial	0.810***	211	0.794***	212	0.809***	69
Grade Levels						
Grade 3	0.777***	2,518	0.799***	2,512	-	-
Grade 4	0.811***	2,347	0.790***	2,345	-	-
Grade 5	0.795***	2,102	0.792***	2,175	0.763***	2,163
Academic Factors						
Disability	0.754***	819	0.784***	818	0.757***	272
Limited English Proficiency	0.743***	1,575	0.716***	1,594	0.676***	484
Academically Gifted	0.639***	1,440	0.657***	1,451	0.611***	504
Academically Gifted - Reading	0.659***	1,168	0.636***	1,178	0.633***	419
Academically Gifted - Math	0.600***	1,046	0.661***	1,053	0.607***	383
All Students						
All Students	0.775***	6,967	0.793***	7,032	0.763***	2,163

Asterisks denote statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1

Although all subgroup correlations are positive, the strength of the correlations vary, with Academically Gifted students demonstrating the smallest correlations and Asian and Multi-racial students exhibiting the strongest correlations. Correlations segmented by both grade and course range from 0.76 (Science – Grade 5) to 0.81 (Math – Grade 4). Correlations by ethnicity vary more; for example, correlations in Science range from 0.70 for White students

[4] Please see the methodology in Section I for descriptions of why specific subgroups are displayed.

© 2016 Hanover Research

to 0.82 for Asian students. Finally, segmentations by gender appear to vary by the smallest amount.

© 2016 Hanover Research

APPENDIX: RESULTS BY ELEMENTARY SCHOOL CAMPUS

Figure A.1: Summary of Enrolled School by Subject

School ID	Math (N=7,306)	Reading (N=7,324)	Science (N=2,290)	Total (N=16,920)
School ID 304	4.00%	3.89%	3.62%	3.90%
School ID 308	1.97%	1.99%	1.88%	1.97%
School ID 310	3.50%	3.40%	2.97%	3.39%
School ID 313	3.90%	3.88%	3.84%	3.88%
School ID 315	3.87%	3.81%	3.71%	3.82%
School ID 318	3.20%	3.22%	3.10%	3.20%
School ID 319	5.97%	5.91%	6.03%	5.95%
School ID 320	4.61%	4.57%	5.02%	4.65%
School ID 324	4.12%	4.04%	3.80%	4.04%
School ID 327	4.08%	4.04%	4.28%	4.09%
School ID 328	3.97%	3.85%	3.28%	3.82%
School ID 332	4.37%	4.21%	3.93%	4.24%
School ID 339	2.97%	2.94%	3.19%	2.98%
School ID 340	2.59%	2.58%	2.53%	2.58%
School ID 344	1.72%	1.69%	1.48%	1.68%
School ID 347	2.20%	2.20%	2.14%	2.19%
School ID 348	2.18%	2.12%	2.58%	2.20%
School ID 352	4.52%	4.49%	5.20%	4.60%
School ID 354	1.25%	1.24%	1.31%	1.25%
School ID 360	3.75%	3.73%	3.67%	3.73%
School ID 362	3.37%	3.33%	3.58%	3.38%
School ID 363	2.97%	2.96%	3.23%	3.00%
School ID 364	5.27%	5.31%	5.68%	5.34%
School ID 367	2.27%	2.25%	2.45%	2.29%
School ID 369	3.94%	3.93%	4.15%	3.97%
School ID 372	4.05%	3.97%	3.62%	3.96%
School ID 374	1.59%	1.60%	1.18%	1.54%
School ID 376	2.37%	3.55%	3.80%	3.07%
School ID 388	3.20%	3.09%	3.19%	3.15%
School ID 400	2.23%	2.20%	1.57%	2.13%

© 2016 Hanover Research

Figure A.2: Correlation between Benchmark and EOY Assessments by Subject and School

School ID	Math		Reading		Science	
	Coefficient	N	Coefficient	N	Coefficient	N
School ID 304	0.743***	269	0.756***	261	0.707***	79
School ID 308	0.742***	140	0.667***	142	0.712***	41
School ID 310	0.701***	227	0.669***	226	0.618***	61
School ID 313	0.785***	281	0.781***	281	0.838***	88
School ID 315	0.720***	271	0.736***	268	0.824***	83
School ID 318	0.772***	232	0.821***	234	0.705***	70
School ID 319	0.807***	411	0.829***	411	0.858***	128
School ID 320	0.748***	315	0.773***	316	0.722***	107
School ID 324	0.730***	287	0.736***	285	0.701***	82
School ID 327	0.783***	289	0.808***	289	0.750***	90
School ID 328	0.797***	276	0.715***	273	0.625***	73
School ID 332	0.812***	290	0.788***	291	0.788***	79
School ID 339	0.774***	204	0.706***	206	0.713***	68
School ID 340	0.668***	187	0.708***	187	0.659***	56
School ID 344	0.737***	122	0.773***	121	0.749***	30
School ID 347	0.766***	160	0.846***	160	0.824***	47
School ID 348	0.735***	157	0.700***	153	0.621***	57
School ID 352	0.707***	308	0.711***	310	0.785***	113
School ID 354	0.827***	91	0.878***	91	0.834***	29
School ID 360	0.698***	246	0.728***	246	0.556***	71
School ID 362	0.698***	234	0.750***	234	0.834***	79
School ID 363	0.804***	210	0.870***	209	0.795***	69
School ID 364	0.738***	383	0.736***	388	0.726***	130
School ID 367	0.594***	162	0.662***	161	0.655***	55
School ID 369	0.763***	287	0.800***	287	0.763***	95
School ID 372	0.756***	275	0.756***	273	0.670***	73
School ID 374	0.526***	111	0.764***	112	0.610***	26
School ID 376	0.675***	169	0.651***	253	0.738***	83
School ID 388	0.695***	228	0.777***	217	0.689***	71
School ID 400	0.705***	145	0.671***	147	0.682***	30

© 2016 Hanover Research

PROJECT EVALUATION FORM

Hanover Research is committed to providing a work product that meets or exceeds client expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.

http://www.hanoverresearch.com/evaluation/index.php

CAVEAT

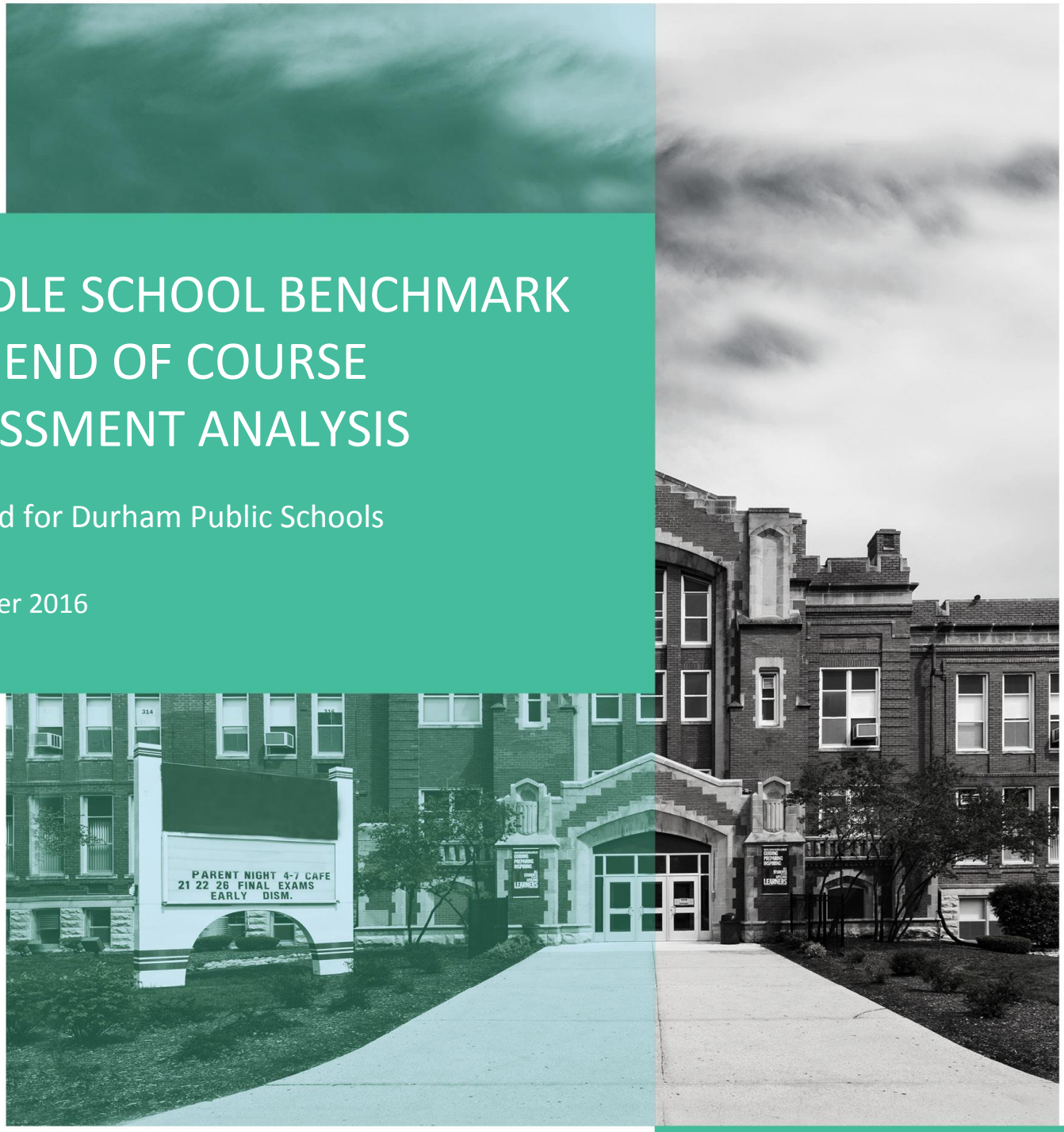
The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties that extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every client. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Clients requiring such services are advised to consult an appropriate professional.

4401 Wilson Boulevard, Suite 400
Arlington, VA 22203
P 202.559.0500 **F** 866.808.6585
www.hanoverresearch.com

MIDDLE SCHOOL BENCHMARK AND END OF COURSE ASSESSMENT ANALYSIS

Prepared for Durham Public Schools

November 2016



In the following report, Hanover Research uses correlation and descriptive analyses to evaluate the extent to which middle school benchmark assessment are predictive of standardized end-of-course (EOC) assessments.

HANOVER RESEARCH

www.hanoverresearch.com

TABLE OF CONTENTS

Executive Summary and Key Findings 3
Introduction 3
Key Findings 3
Section I: Data and Methodology 5
Data 5
Outcome Variables 5
Segmentation Variables 6
Methodology 8
Pearson Correlations 8
Achievement Level Comparisons 9
Section II: Topline Comparisons 10
Correlations 10
Achievement Level Comparisons 12
Descriptive Analysis 12
Section III: Correlations by Student Subgroups 16

© 2016 Hanover Research

EXECUTIVE SUMMARY AND KEY FINDINGS

INTRODUCTION

In this report, Hanover Research uses correlation and descriptive analyses to evaluate the extent to which middle school benchmark assessments are predictive of standardized end-of-course (EOC) assessments in Durham Public Schools (Durham). We compare benchmark assessment outcomes to EOC outcomes for Reading, Mathematics (Math), and Science. Similar to the high school analyses completed in May 2016, we observe strong correlations across subjects and student subgroups, with some variation in the strength of the correlations by subject/subgroup.

This report is organized in the following sections:

- **Section I: Data and Methodology** provides an overview of the data and methodology used in this analysis.
- **Section II: Topline Comparisons** presents overall correlations between benchmark assessments and EOC tests in each of the three subjects, as well as a descriptive analysis of the correspondence of proficiency levels across assessments within each subject.
- **Section III: Correlations by Student Subgroups** presents a correlational analysis segmented by student demographic and academic subgroups, as well as by grade level and school locations.

KEY FINDINGS

- **Strong and positive correlations exist between the benchmark assessments and the EOC exams for all three subjects, and all correlations are statistical significant beyond the 99 confidence level.** Specifically, the correlation for Science tests is the strongest with a coefficient of 0.84, while the correlation for Math tests is the weakest with a coefficient of 0.70. Interpreting the relationships using the units of each assessment, we find that:
  - In Math, a 1-percentage-point increase in percent correct on the benchmark assessment is associated with a 0.38-point increase on the corresponding EOC exam score.
  - In Reading, a 1-percentage-point increase in percent correct on the benchmark assessment is associated with a 0.51-point increase on the corresponding EOC exam score.
  - In Science, a 1-percentage point increase in percent correct on the benchmark assessment is associated with a 0.48-point increase on the corresponding EOC exam score.
- **In examining achievement levels, we find that approximately 56 percent of students are rated at the same level by the two assessments.** When examining the results by specific subjects, the percentage of students rated at the same achievement level ranges from 52.9 percent (Reading) to 59.4 percent (Math).

© 2016 Hanover Research

- Furthermore, 89.7 percent (Reading) to 92.6 percent (Math) of students achieve at the same level or at a "neighbor" level (i.e., one level above or one level below) on the EOC as the level they achieved on the benchmark assessment.
- **Overall, more students are rated at a higher achievement level on the EOC exam compared to the corresponding benchmark assessment,** which is similar to the findings of the high school analysis. Specifically, 26.6 percent of students are rated at a higher achievement level on the EOC exam than on the corresponding benchmark assessment, compared to only 17.5 percent of students are rated at a lower achievement level on the EOC exam than on the benchmark assessment.

- **When examining students by different subgroups, we observe positive and significant correlations between the benchmark assessment and EOC exam within each subject.** Though the strength of the correlation varies across subgroups, 94 percent of all comparisons have a coefficient of 0.5 or higher. In general, the correlation coefficients for Science are higher than the coefficients for Math or Reading tests, and exhibit narrower variation across subgroups.
  - Students with disabilities show the smallest correlations between the benchmark assessment and the EOC exam for Math (coefficient of 0.256) and Reading (coefficient of 0.376) tests, but exhibit a strong correlation in Science (coefficient of 0.754).

© 2016 Hanover Research

SECTION I: DATA AND METHODOLOGY

In this section, Hanover Research describes the data we use for this study and the methodologies we adopt.

DATA

Durham provided Hanover with data on student outcomes for Reading, Math, and Science benchmark assessments and EOC exams for the 2015-16 academic year. This analysis specifically focuses on students from Grade 6 to Grade 8 (middle school). For students in these three grades, all students take Math and Reading tests, while only students in Grade 8 take Science tests. We extract the demographic information for each student from the EOC dataset.

For Math benchmark assessment data, there are five duplicates at the student ID – school ID level, meaning that there are five incidences where two students share the same student ID and are from the same school but have different benchmark outcomes and different grade values. For these duplicates, we removed the observation where the student's grade level did not match the one in the EOC dataset. We then combine all data into a single dataset in which each observation is for a particular student and subject.

Combining the datasets gives us 16,850 observations for 7,216 students. Students with missing benchmark assessments or EOC exam scores are removed. In the final analytic dataset, we have **14,640 observations containing academic and demographic variables for 6,495 middle school students.**

OUTCOME VARIABLES

Both the benchmark assessments and EOC exams include multiple measures for each subject. Figure 1.1 on the following page describes the measures available for both assessments. As shown in the table, we use percent correct and achievement level as measurements for the benchmark assessments, and score and achievement level as measurements for the EOC exams. These are the same set of measurements used in the previous high school analysis .

© 2016 Hanover Research

Figure 1.1: Outcome Variables Available for Analysis

Assessment	Measurement	Used or Not
Benchmark Assessments	**Percent of Questions Answered Correctly**	**Use**
	Suggested Numeric Score	Does not exhibit as much variation as *Percent of Questions Answers Correctly*, so we use the *Percent Correct*[1] → Do not use
	Suggested Letter Mark	Does not have comparable measures in the EOC outcome → Do not use
	Achievement Level	**Ranging from one to five, this functions as a measurement for student proficiency level → Use**
	Achievement Level - Extend	Does not have comparable measures in the EOC outcome → Do not use
	"Domain of Knowledge" Outcomes	For the purpose of this analysis, we focus on the overall outcomes → Do not use
	Unit Achievement Level Outcomes	For the purpose of this analysis, we focus on the overall outcomes → Do not use
EOC	**Score**	**Use**
	Achievement Level	**Ranging from one to five, this functions as a measurement for student proficiency level → Use**

Figure 1.2 presents the summary statistics of the outcome variables used in this analysis.

Figure 1.2: Outcome Variables Summary

Outcome	Math		Reading		Science		All Subjects	
	Mean	N	Mean	N	Mean	N	Mean	N
Benchmark Percent Correct	46.1%	6,150	48.6%	6,280	49.1%	2,210	47.6%	14,640
Benchmark Achievement Level	2.1	6,150	2.4	6,280	2.9	2,210	2.4	14,640
EOC Score	446.5	6,150	451.7	6,280	247.8	2,210	2.5	14,640
EOC Achievement Level	2.3	6,150	2.5	6,280	2.9	2,210	418.7	14,640

SEGMENTATION VARIABLES

In Section III, Hanover Research segments the correlational analysis by student subgroups. These groups are based on both demographic and academic factors.

In terms of demographics, we segment by gender, race/ethnicity, and disability status. Note that when segmenting results by race/ethnicity, we focus on whether the student is African American, Hispanic, or White, since other race/ethnicity categories have too small of a sample sizes to yield valid insights.

For the academic giftedness, we follow the previous analysis' method to include an overall academic giftedness variable as well as whether a student is academically gifted in Reading

[1] The smaller variation is due to the suggested numeric score ranging from 50 to 100, whereas the percent correct variables range from 0 to 100. These two variables do not have a linear relationship; low values of the percent correct variable are clustered in the 50 to 70 range of the suggested score variable, while higher scores exhibit a highly linear relationship.

© 2016 Hanover Research

or Math. Specifically, we created an overall variable which considers a student as "academically gifted" as long as he/she is gifted in Reading (with value of "R" for the *aig* variable), Math (with a value of "M" for the *aig* variable), or both (with a value of "B" for the *aig* variable).

In the analysis for high school students, we segment the students to three groups: "on track," "ahead of track," and "behind the track."[2] However, unlike the high school analysis where the number of students taking each assessment exhibits great variation across grade levels, the number of students taking a given test is fairly balanced within each grade level for middle school. Thus, for this analysis, we do not segment the students by whether they are on track or not.

The analysis segments by grade level, to reveal any strong or weak correlations for specific grades. Lastly, note that when segmenting by school, we focus only on the four schools with the largest numbers of students taking these assessments.

Figure 1.3 presents the summary statistics for the variables we use to segment the students.

Figure 1.3: Segmentation Variables Summary

Segmentation Variable	Math		Reading		Science		All Subjects	
	Pct.	N	Pct.	N	Pct.	N	Pct.	N
Demographic Factors								
Female	50.7%	6,150	50.7%	6,280	51.3%	2,210	50.8%	14,640
Male	49.3%	6,150	49.3%	6,280	48.7%	2,210	49.2%	14,640
African American	48.7%	6,150	48.6%	6,280	51.0%	2,210	49.0%	14,640
Hispanic	27.4%	6,150	27.2%	6,280	26.2%	2,210	27.1%	14,640
White	18.0%	6,150	18.4%	6,280	17.6%	2,210	18.1%	14,640
Disability	12.1%	6,150	12.0%	6,280	11.0%	2,210	11.9%	14,640
Academic Factors								
Limited English Proficiency	8.4%	6,150	8.3%	6,280	9.8%	2,210	8.6%	14,640
Academically Gifted	25.7%	6,150	26.5%	6,280	27.4%	2,210	26.3%	14,640
Academically Gifted - Math	21.0%	6,150	21.7%	6,280	23.2%	2,210	21.6%	14,640
Academically Gifted - Reading	19.9%	6,150	20.7%	6,280	21.1%	2,210	20.4%	14,640
Grade Level								
Grade 6	31.4%	6.150	30.9%	6,280	-	2,210	26.5%	14,640
Grade 7	34.9%	6,150	34.2%	6,280	-	2,210	29.3%	14,640
Grade 8	33.6%	6,150	34.9%	6,280	100%	2,210	44.2%	14,640
Campus Location								
Rogers-Herr Middle School	9.9%	6,150	9.7%	6,280	9.3%	2,210	9.7%	14,640
Neal Middle School	11.0%	6,150	10.9%	6,280	11.3%	2,210	11.0%	14,640
Sherwood Githens Middle School	12.6%	6,150	12.3%	6,280	12.3%	2,210	12.4%	14,640
George L Carrington Middle School	14.9%	6,150	14.7%	6,280	14.4%	2,210	14.7%	14,640

Note: The percentages of the three grade levels do not add up to 100 percent in Math due to rounding.

[2] "on track": the student is in the most common grade for students taking the subject; "ahead of track": the student is in a lower grade than the most common grade for that subject; "behind track": the student is in a higher grade than the most common grade for that subject.

© 2016 Hanover Research

METHODOLOGY

In this subsection, Hanover describes the methods for making comparisons between the benchmark assessment and the EOC assessment.

The goal of this analysis is to determine the strength of the correlation between each pair of assessments in each subject. However, it should be noted that correlations between the benchmark assessment and EOC exam could result from a number of conditions which indicate various relationships between the assessments and may also reflect outcomes of Durham's academic programming. For example, lower levels of correlation could indicate that the benchmark assessment is different in content to the EOC exam. Further, since content-specific academic programming occurs during the time between the benchmark assessment and EOC exam, low levels of correlation could also indicate that Durham's programming improves student outcomes such that students who perform poorly on the benchmark assessment perform well on the EOC exam. If programming decisions, such as interventions, are initiated in response to the benchmark assessments, this may further reduce the expected correlation between the two assessments. In short, there are multiple reasons a stronger or weaker correlation between these assessments may be observed.

PEARSON CORRELATIONS

Using the percent of questions answered correctly on the benchmark assessment and the scores on the EOC exam for each subject, we compute Pearson correlation coefficients separately for each subject. This correlation coefficient ranges from -1 to 1, with -1 indicating perfect negative correlation (i.e., when one variable rises, the other falls) and 1 indicating perfect positive correlation (the two variables rise or fall together). A coefficient of zero indicates that there is no correlation between the variables. Statistical significance is based on the null hypothesis that the correlation between the two variables is zero, meaning that three significance "stars" indicate a p value of less than .01 and are evidence that there is a less than 1 percent probability that the relationship between the two variables is zero. Because the scores on the EOC exams are not scaled consistently across the subjects, we do not compute an overall correlation coefficient for the three assessments pooled together.

Building on the presentation of the Pearson correlations, we further describe the relationship between the benchmark and EOC assessments based on the results of a simple linear regression, where percent correct from the benchmark assessment is used to predict EOC exam score. Whereas the Pearson correlation coefficients are standardized, ranging from -1 to 1, unstandardized regression coefficients allow us to express the relationship between assessments in each subject in terms of the units of each assessment (e.g., an "X"-percentage point increase in percent correct on the benchmark assessment is associated with a "Y"-point increase in EOC scaled scored). We also provide scatter plots and trend lines illustrating the relationship between assessments, where the slope of the trend line is equal to the unstandardized regression coefficient for a given subject.

© 2016 Hanover Research

Lastly, in Section III, we compute Pearson correlations separately for each subject and each student subgroup, where applicable.

ACHIEVEMENT LEVEL COMPARISONS

We analyze achievement levels by comparing the numbers and percentages of students who fall within each achievement level for each assessment, as shown in Figure 1.4 below. This allows the reader to observe the extent to which students are categorized within the same achievement level for both assessments or in different achievement levels. The numbers in the dark teal cells represent students who fall into the same achievement level for both assessments, while students in the light teal cells represent students who fall into a higher achievement level on the EOC exam than on the benchmark assessment. The light red cells indicate students who fall into a lower achievement level on the EOC exam than on the benchmark assessment.

We also compute the number of students who stay at the same achievement level, stay at the same or a "neighbor" level (one level above or below), advance at least one level, and regress at least one level from the benchmark assessment to the EOC exam.

Figure 1.4: Achievement level Comparisons

SUBJECT		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1						
	2						
	3						
	4						
	5						
Total							

© 2016 Hanover Research

SECTION II: TOPLINE COMPARISONS

In this section, Hanover Research presents the results of the comparisons of the benchmark assessments with the EOC exams separately for each subject. In the achievement level comparisons below, we also compare the subjects pooled into a single analysis.

CORRELATIONS

The correlation coefficients in Figure 2.1 indicate the Pearson correlation between the percent of questions answered correctly in the benchmark assessment and the scores on the EOC exam. For all three subjects, there is a strong positive correlation that is statistically significant beyond the 99 percent level. These correlations range from 0.704 for Math to 0.841 for Science.

Figure 2.1: Correlations between Percent Correct on Benchmark and EOC Score

SUBJECT	COEF.	N
Math	0.704***	6,150
Reading	0.719***	6,280
Science	0.841***	2,210

Asterisks denote statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1.

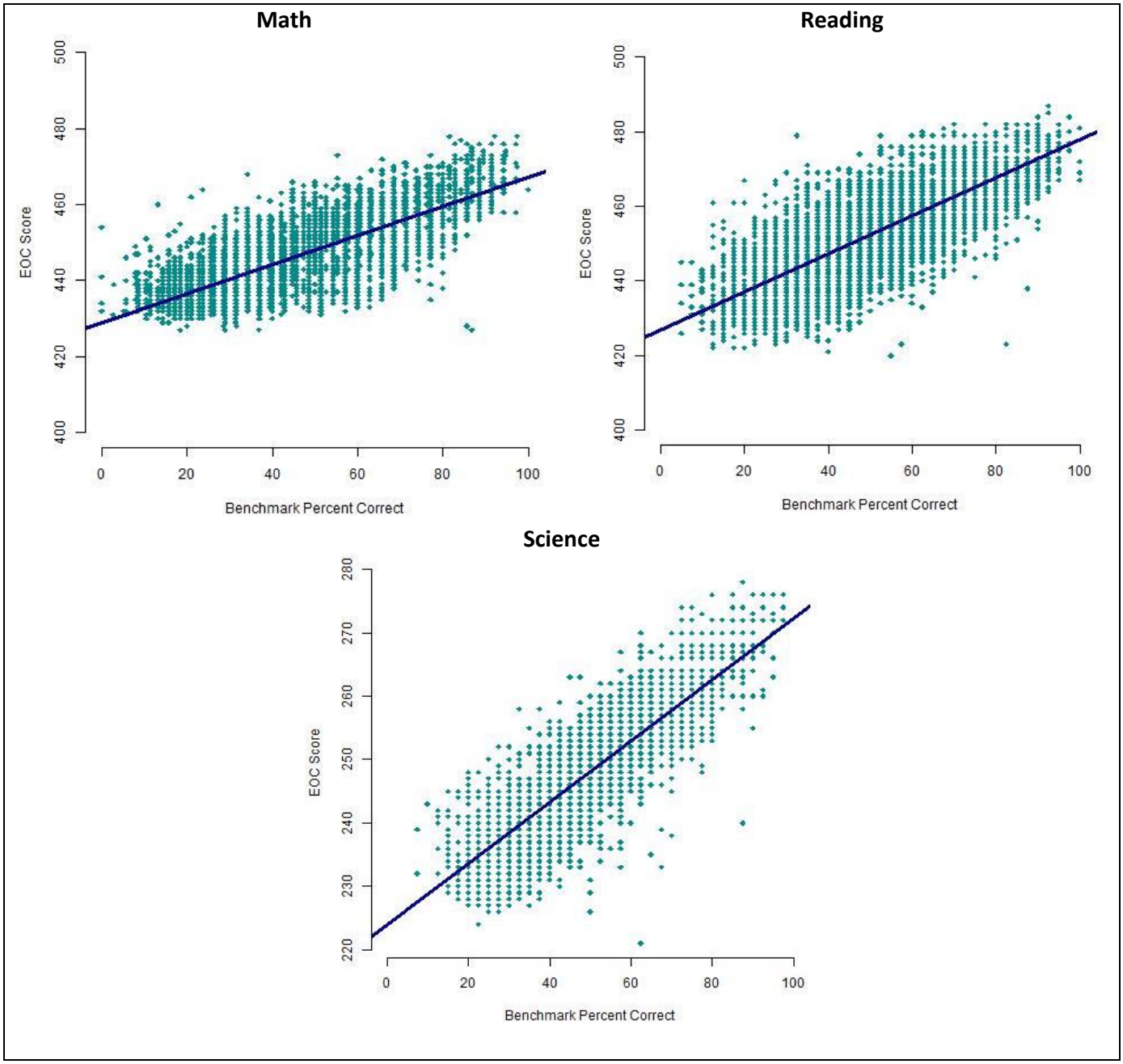
To further illustrate the relationships between benchmark and EOC assessments, we describe these correlations in terms of simple linear regression coefficients,[3] where we predict the EOC score using the percent correct value from the benchmark assessment. We find that for the Math test, for every 1-percentage-point increase in percent correct on the benchmark assessment, we expect a 0.38-point increase in the corresponding EOC exam score. For Reading, a 1-percentage-point increase in percent correct on the benchmark assessment is associated with a 0.51-point increase in the corresponding EOC exam score. Lastly, every 1-percentage point increase in percent correct on the Science benchmark assessment is associated with a 0.48-point increase in the corresponding EOC exam score.

We present the scatterplots of benchmark and EOC assessment performance for each subject in Figure 2.2 on the following page.

[3] Note that the Pearson correlation coefficients are equivalent to standardized regression coefficients of a simple linear regression involving percent correct on the benchmark assessment and the corresponding EOC assessment scaled scored. Here we discuss the unstandardized regression coefficients, as they allow us to estimate how a change in one unit of the predictor variable (percent correct on the benchmark assessment) relates to changes in units of the outcome variable (EOC scaled score).

© 2016 Hanover Research

Figure 2.2: Scatterplots of Benchmark Percent Correct and EOC Score



© 2016 Hanover Research

ACHIEVEMENT LEVEL COMPARISONS

DESCRIPTIVE ANALYSIS

Figure 2.3 presents a summary of the achievement analysis. Both assessments have achievement levels ranging from one to five, and according to Figure 2.3, when pooling all three subjects together, 55.9 percent of students are rated at the same level by both assessments. The highest percentage is seen in Math (59.3 percent) and the lowest is in Reading (52.9 percent). Further, 91 percent of students are rated at the same achievement level or at a "neighbor" level (i.e., one level above or below). In regards to individual subjects, Math has the greatest percentage of students rated at the same or at a "neighbor" level by the two assessments (59.3 percent), while Reading has the lowest percentage (52.9 percent).

Similar to the findings for high school analysis, more students are rated at a higher level on the EOC exam compared to the benchmark assessment. Specifically, 26.6 percent of students are rated at a higher level on the EOC exam than on the benchmark assessment, whereas only 17.5 percent of students are rated at a lower level on the EOC exam than on the benchmark assessment. The narrowest disparity we observe is in Science, where 23.9 percent of students are rated at a higher level on EOC exam than on benchmark assessment, while 21.3 percent of students are rated at a lower level. In comparison, the differences for Math and Reading are around 10 percentage points.

Figure 2.3: Differences in Achievement Level from Benchmark to EOC, Summary

DIFFERENCE	ALL SUBJECTS	MATH	READING	SCIENCE
Same Level	55.9%	59.4%	52.9%	54.8%
Same Level or Neighbor Level	91.0%	92.6%	89.7%	90.4%
Same or Higher Level	82.5%	84.6%	81.7%	78.7%
Higher Level	26.6%	25.3%	28.8%	23.9%
Lower Level	17.5%	15.4%	18.3%	21.3%
Number of Observations	14,640	6,150	6,280	2,210

This figure represents a summary of the results presented in Figures 2.4-2.11.

The remaining figures in this section display detailed comparisons of benchmark and EOC achievement levels for all subjects combined and for each individual subject (as summarized in Figure 2.3 above). Each subject area includes two figures: the first displays percentages while the second displays counts. As described in Figure 1.4 in the methodology, numbers in dark teal cells represent students classified within the same achievement level for both assessments, numbers in light teal cells represent students classified within a higher achievement level on the EOC exam than on the benchmark assessment, and numbers in light red cells indicate students classified within a lower achievement level on the EOC exam than on the benchmark assessment.

© 2016 Hanover Research

ALL SUBJECTS

Figure 2.4: Achievement Level Comparisons, All Subjects

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	27.4%	7.3%	1.0%	0.6%	0.0%	36.3%
	2	7.7%	10.2%	3.0%	3.8%	0.2%	24.8%
	3	1.1%	3.2%	2.2%	5.7%	0.5%	12.7%
	4	0.3%	1.4%	1.7%	10.8%	4.6%	18.9%
	5	0.0%	0.0%	0.1%	2.0%	5.3%	7.4%
Total		**36.6%**	**22.0%**	**7.8%**	**22.9%**	**10.6%**	**100.0%**

This figure includes all student-subject observations we analyze, meaning that students with data for assessments in multiple subjects are represented multiple times. N=14,640

Figure 2.5: Achievement Level Comparisons, All Subjects[4]

All Subjects		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	4,008	1,064	140	90	5	5,307
	2	1,128	1,488	435	552	25	3,628
	3	168	465	320	836	72	1,861
	4	51	202	246	1,586	673	2,760
	5	4	3	8	292	779	1,084
Total		**5,359**	**3,222**	**1,149**	**3,356**	**1,554**	**14,640**

This figure includes all student-subject observations we analyze, meaning that students with data for assessments in multiple subjects are represented multiple times. N=14,640

MATH

Figure 2.4: Achievement Level Comparison, Math

Math		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	36.3%	7.9%	0.6%	0.6%	0.0%	45.3%
	2	7.8%	9.1%	2.2%	3.3%	0.1%	22.5%
	3	1.0%	3.4%	1.6%	5.5%	0.7%	12.1%
	4	0.3%	0.9%	0.9%	7.8%	4.4%	14.4%
	5	0.0%	0.0%	0.0%	1.1%	4.6%	5.7%
Total		**45.3%**	**21.3%**	**5.3%**	**18.2%**	**10.0%**	**100.0%**

N=6,150

[4] Figure 2.4 and Figure 2.5 describe achievement level comparisons for the benchmark and EOC for all subjects pooled. These comparisons allow Durham to observe the overall extent to which the benchmark and EOCs are comparable.

© 2016 Hanover Research

Figure 2.5: Achievement Level Comparison, Math

Math		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	2,230	485	34	34	2	2,785
	2	480	557	136	201	9	1,383
	3	59	209	97	336	46	747
	4	16	56	56	482	271	881
	5	1	0	0	68	285	354
Total		**2,786**	**1,307**	**323**	**1,121**	**613**	**6,150**

N=6,150

READING

Figure 2.6: Achievement Level Comparison, Reading

Reading		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	22.5%	7.4%	1.1%	0.6%	0.0%	31.6%
	2	7.7%	13.1%	3.7%	4.7%	0.3%	29.5%
	3	0.9%	3.1%	2.4%	5.7%	0.4%	12.5%
	4	0.3%	1.9%	2.0%	10.7%	5.0%	19.9%
	5	0.0%	0.0%	0.1%	2.2%	4.1%	6.5%
Total		**31.4%**	**25.6%**	**9.3%**	**23.9%**	**9.8%**	**100.0%**

N=6,280

Figure 2.7: Achievement Level Comparison, Reading

Reading		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	1,412	467	67	38	3	1,987
	2	485	824	230	296	16	1,851
	3	59	195	151	358	23	786
	4	16	118	128	673	312	1,247
	5	2	3	8	137	259	409
Total		**1,974**	**1,607**	**584**	**1,502**	**613**	**6,280**

N=6,280

© 2016 Hanover Research

SCIENCE

Figure 2.8: Achievement Level Comparisons, Science

Science		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	16.6%	5.1%	1.8%	0.8%	0.0%	24.2%
	2	7.4%	4.8%	3.1%	2.5%	0.0%	17.8%
	3	2.3%	2.8%	3.3%	6.4%	0.1%	14.8%
	4	0.9%	1.3%	2.8%	19.5%	4.1%	28.5%
	5	0.0%	0.0%	0.0%	3.9%	10.6%	14.6%
Total		**27.1%**	**13.9%**	**11.0%**	**33.2%**	**14.8%**	**100.0%**

N=2,210

Figure 2.9: Achievement Level Comparisons, Science

Science		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1	366	112	39	18	0	535
	2	163	107	69	55	0	394
	3	50	61	72	142	3	328
	4	19	28	62	431	90	630
	5	1	0	0	87	235	323
Total		**599**	**308**	**242**	**733**	**328**	**2,210**

N=2,210

© 2016 Hanover Research

SECTION III: CORRELATIONS BY STUDENT SUBGROUPS

In this section, Hanover describes correlations segmented by student subgroups based on demographics and academic factors.[5]

Figure 3.1 presents the correlation coefficients of different student subgroups. As shown in the table, all of the segmented results are positively correlated at beyond the 99 confidence level. Approximately 94 percent of comparisons (49 out of 52 subgroup comparisons) exhibit a coefficient that is larger than 0.5.

Figure 3.1: Correlations by Student Subgroups

Segmentation Variable	Math		Reading		Science	
	Coef.	N	Coef.	N	Coef.	N
Demographic Factors						
Female	0.792***	3,118	0.784***	3,182	0.841***	1,133
Male	0.643***	3,032	0.662***	3,098	0.843***	1,077
African American	0.541***	2,994	0.613***	3,055	0.790***	1,128
Hispanic	0.752***	1,683	0.744***	1,708	0.800***	578
White	0.797***	1,107	0.740***	1,153	0.828***	389
Academic Factors						
Disability	0.256***	742	0.376***	756	0.754***	244
Limited English Proficiency	0.475***	519	0.552***	521	0.627***	216
Academically Gifted	0.682***	1,578	0.608***	1,662	0.789***	606
Academically Gifted - Reading	0.692***	1,222	0.577***	1,299	0.772***	466
Academically Gifted - Math	0.658***	1,289	0.615***	1,362	0.786***	512
Grade Levels						
Grade 6	0.874***	1,933	0.825***	1,939	-	-
Grade 7	0.631***	2,149	0.619***	2,147	-	-
Grade 8	0.801***	2,068	0.815***	2,194	0.841***	2,210
Campus Location						
Rogers-Herr Middle School	0.773***	606	0.770***	609	0.799***	206
Neal Middle School	0.733***	679	0.703***	686	0.764***	250
Sherwood Githens Middle School	0.792***	775	0.805***	772	0.857***	271
George L Carrington Middle School	0.786***	917	0.783***	923	0.829***	318
All Students						
All Students	0.704***	6,150	0.719***	6,280	0.841***	2,210

Asterisks denote statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1

Although all subgroup correlations are positive, the strength of the correlations vary. For example, students with disabilities show relatively low correlations for Math (coefficient of 0.256) and Reading (coefficient of 0.376) tests, but exhibit high correlation for Science tests (coefficient of 0.754). A similar scenario can also be seen for students with limited English proficiency: for these students, the correlation for Math tests is relatively low (coefficient of

[5] Please see the methodology in Section I for descriptions of why specific subgroups are displayed.

© 2016 Hanover Research

0.475) when compared to the correlations for Reading (coefficient of 0.552) and Science (coefficient of 0.627). In general, correlation coefficients of Science tests are higher than the coefficients of Math and Reading tests, and exhibit narrower variation across subgroups.

© 2016 Hanover Research

PROJECT EVALUATION FORM

Hanover Research is committed to providing a work product that meets or exceeds client expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.

http://www.hanoverresearch.com/evaluation/index.php

CAVEAT

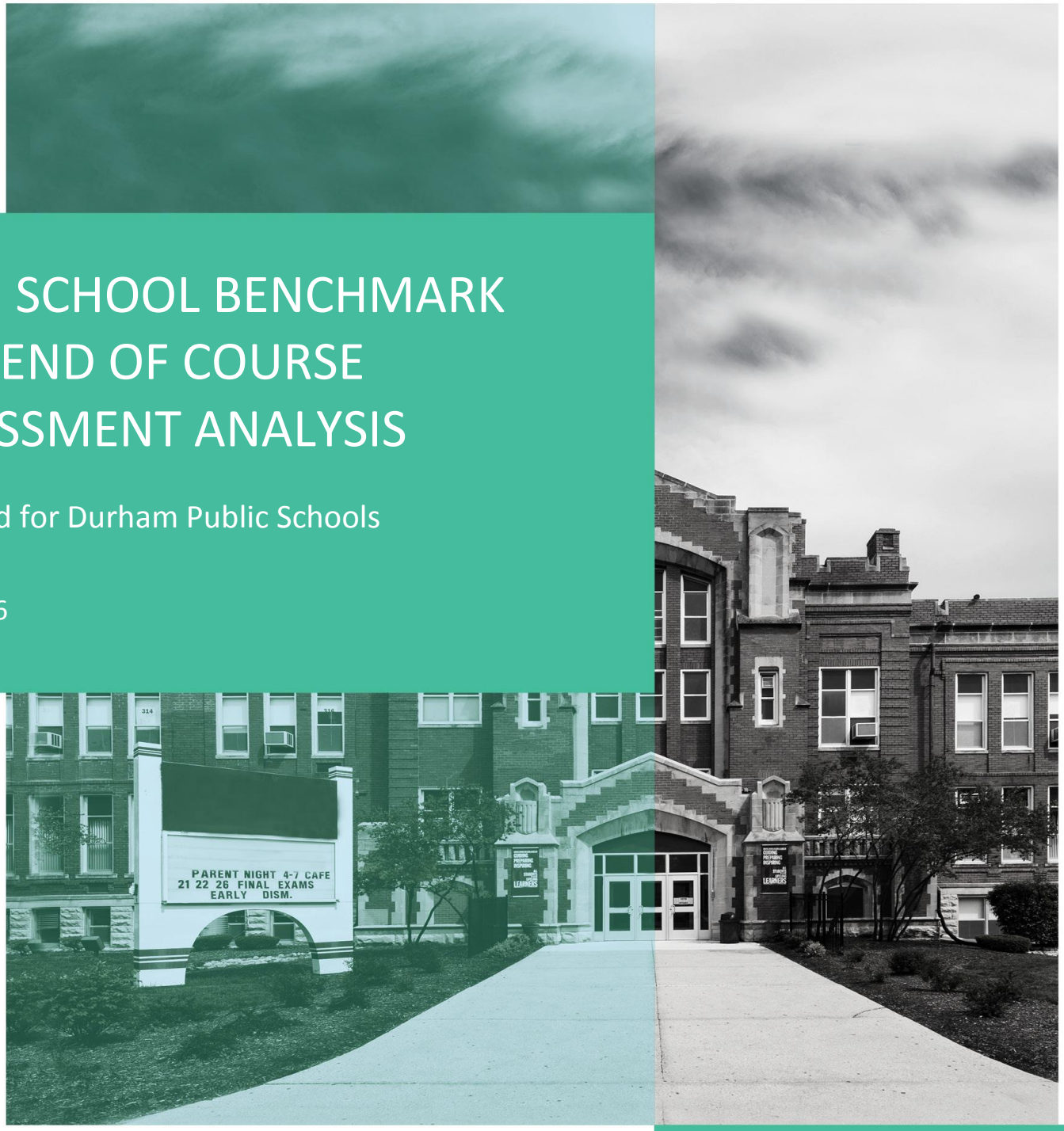
The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties that extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every client. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Clients requiring such services are advised to consult an appropriate professional.

4401 Wilson Boulevard, Suite 400
Arlington, VA 22203
P 202.559.0500 **F** 866.808.6585
www.hanoverresearch.com

HIGH SCHOOL BENCHMARK AND END OF COURSE ASSESSMENT ANALYSIS

Prepared for Durham Public Schools

May 2016



In the following report, Hanover Research uses correlation and descriptive analyses to evaluate the extent to which high school benchmark assessments are predictive of standardized end-of-course (EOC) assessments.

HANOVER RESEARCH

www.hanoverresearch.com

TABLE OF CONTENTS

Executive Summary and Key Findings 3
INTRODUCTION 3
KEY FINDINGS 3
Section I: Data and Methodology 5
DATA 5
Outcome Variables 5
Segmentation Variables 6
METHODOLOGY 7
Pearson Correlations 8
Achievement Level Comparisons 8
Section II: Topline Comparisons 10
CORRELATIONS 10
ACHIEVEMENT LEVEL COMPARISONS 12
All Subjects 13
Algebra I 13
Biology 14
English II 15
Section III: Correlations by Student Subgroups 16

© 2016 Hanover Research

EXECUTIVE SUMMARY AND KEY FINDINGS

INTRODUCTION

In this report, Hanover Research uses correlation and descriptive analysis to evaluate the extent to which high school benchmark assessments are predictive of standardized end-of-course (EOC) assessments at Durham Public Schools (Durham). We compare benchmark assessment outcomes to EOC outcomes for Algebra I, Biology, and English II. We find that there is strong correlation across subjects and student subgroups, with some variation in the strength of the correlations by subject/subgroup.

This report is organized as follows:

- **Section I: Data and Methodology** introduces the purpose of this report, reviews the data provided by Durham, and describes the methodologies used to analyze the data.
- **Section II: Topline Comparisons** presents overall correlations between benchmark assessments and EOC tests in each of the three subjects and a descriptive analysis of the correspondence of proficiency levels across assessments within each subject.
- **Section III: Correlations by Student Subgroups** presents a correlational analysis segmented by student demographic and academic subgroups, as well as by school.

KEY FINDINGS

- **In all three subjects, there is a strong positive correlation between benchmark assessments and EOC tests that is statistically significant beyond the 99 percent confidence level.** These correlations range from 0.653 for Biology to 0.752 for Algebra I when examining all students for whom data are available.[1] Expressing these relationships in terms of the units of each assessment, we find that:
  - Every 1-percentage-point increase in percent correct on the Algebra I benchmark assessment is associated with a 0.42-point increase in Algebra I EOC scaled score.[2]
  - Every 1-percentage point increase in percent correct on the Biology benchmark assessment is associated with a 0.37-point increase in Biology EOC scaled score.
  - Every 1-percentage point increase in percent correct on the English II benchmark assessment is associated with a 0.29-point increase in English II EOC scaled score.

[1] Note that the correlations are expressed in terms of Pearson correlation coefficients that range from -1 to 1, with -1 indicating perfect negative correlation (when one variable rises, the other falls) and 1 indicating perfect positive correlation (the two variables rise and fall together). A correlation coefficient of zero indicates that there is no correlation between the variables.

[2] When describing the relationships in terms of the units of each assessment (percent correct on the benchmark assessment and EOC scaled score), we use unstandardized regression coefficients, as they allow us to estimate how a change in one unit of the predictor variable (percent correct on the benchmark assessment) relates to changes in units of the outcome variable (EOC scaled score).

© 2016 Hanover Research

- **In terms of achievement levels (ranging from 1 to 5), the majority of students perform at the same or a similar level in a given subject on both the benchmark and EOC exams.**
  - Specifically, 45.2 percent to 53.2 percent of students who took both the benchmark and EOC assessment in a given subject were rated at the same achievement level on both assessments. Further, 80.6 percent to 86.2 percent of students achieved at the same level or at a "neighbor" level (i.e., one level above or one level below) on the EOC as they achieved on the benchmark assessment.
  - Compared to the benchmark assessment, more students performed better on the EOC assessment, with 25.7 percent to 39.8 percent achieving at a higher level on the EOC. By contrast, 10.3 percent to 23.7 percent of students achieved at a lower level on the EOC than the benchmark assessment, depending on the subject.
- **When segmenting the analyses by student subgroups (e.g., by gender, race/ethnicity, Limited English Proficiency, school, etc.), we still find that the benchmark assessment and EOC assessment within each subject are positively and significantly correlated.** The strength of the correlation varies by subgroup, though nearly all correlation coefficients were 0.5 or higher.
  - For example, we observed weaker correlations between benchmark and EOC assessments when focusing exclusively on students with a disability, with correlation coefficients ranging from 0.396 to 0.477. By contrast, when focusing on female students, we observed correlation coefficients of 0.627 to 0.724.

© 2016 Hanover Research

SECTION I: DATA AND METHODOLOGY

In this section, Hanover Research explains the data we analyze in this report and the methodologies we use to conduct the analysis.

DATA

Durham provided Hanover Research with student outcome data for the Algebra I, Biology, and English II benchmark and EOC assessments for Fall semester of the 2015-16 academic year. The benchmark dataset also includes demographic information for each student. In order to analyze the correlation between the assessments, Hanover combines all data into a single dataset in which each observation is for a particular student and subject.

We exclude one student record in which the student has a suggested numeric score for the benchmark assessment, but lacks a value for the percent correct variable for this assessment. We also exclude cases where a student is recorded as absent on the EOC assessment. In total, Hanover analyzed 1,773 student-subject observations.[3]

OUTCOME VARIABLES

Both the benchmark and EOC assessments include multiple measures for each subject. Outcomes for the benchmark assessment include a variable representing the percent of the questions the student answered correctly, achievement levels ranging from 1 to 5, unit achievement level outcomes, "domains of knowledge" outcomes, and suggested letter and numeric grades. For the purpose of this analysis, we focus on the overall outcomes and exclude unit achievement levels and domains of knowledge scores. Also, because the variable representing suggested numeric grades does not exhibit as much variation as the one representing the percent of questions answered correctly, we analyze the latter.[4] Like the benchmark assessment, the EOC assessment includes achievement levels on a scale of 1 to 5. The EOC assessment also includes a scaled score outcome. Figure 1.1 describes the outcome variables analyzed in this report.

[3] A student-subject level observation contains the information for that student, including both benchmark and EOC assessment outcomes for the particular subject, Algebra I, Biology, or English II.

[4] The smaller variation is due to the suggested numeric score ranging from 50 to 100, whereas the percent correct variable ranges from 0 to 100. These two variables do not have a linear relationship; low values of the percent correct variable are clustered in the 50 to 70 range of the suggested score variable, while higher scores exhibit a highly linear relationship.

© 2016 Hanover Research

Figure 1.1: Outcome Variables Summary

Outcome	Algebra I		Biology		English II		All Subjects	
	Mean	N	Mean	N	Mean	N	Mean	N
Benchmark Percent Correct	40.5%	427	40.8%	653	39.9%	693	40.4%	1,773
Benchmark Achievement Level	1.8	427	2.1	653	2.1	693	2.0	1,773
EOC Achievement Level	2.2	427	2.5	653	2.2	693	2.3	1,773
EOC Scaled Score	246.1	427	247.0	653	144.1	693	206.6	1,773

Segmentation Variables

In Section III, Hanover Research segments the correlational analysis by student subgroups. These groups are based on both demographic and academic factors.

With respect to demographics, we segment by gender, race/ethnicity, and disability status. Note that when segmenting results by race/ethnicity, we focus on whether the student is African American, Hispanic, or white. We do not segment by other student race/ethnicity categories because of the small sample sizes for these groups.

We also segment results by a number of academic factors, including Limited English Proficiency, academic giftedness, grade level ("on track," "ahead of track," "behind track" – described in greater detail below), and school.

Note that for academic giftedness, we include the overall variable as well as whether the student is academically gifted in math or reading. Because students may take Algebra I, Biology, and English in different grades, we also segment results based on whether the student is in the most common grade for students taking the subject ("on track"), is in a lower grade than the most common grade for that subject ("ahead of track"), or is in a higher grade than the most common grade for that subject ("behind track"). For example, the most common grade for Biology and English II is Grade 10. Therefore, students testing in these subjects in Grade 10 are considered "on track," while students testing in these subjects in Grade 9 are considered "ahead of track" and students testing in these subjects in Grades 11 or 12 are considered "behind track."[5] The most common grade for Algebra I is Grade 9, and because we only evaluate students in Grades 9 through 12, there are no students coded as "ahead of track" for this subject. Lastly, note that when segmenting by school site, we focus only on the four schools with the largest numbers of students taking these assessments.

[5] Note that these classifications of "on track," "ahead of track," and "behind track," are used exclusively for the purpose of forming subgroups of students by grade for this analysis (based on an examination of the grade level distributions existing in the assessment data) and are not classifications provided by Durham.

© 2016 Hanover Research

Figure 1.2: Segmentation Variables Summary

Segmentation Variable	Algebra I		Biology		English II		All Subjects	
	Mean	N	Mean	N	Mean	N	Mean	N
Demographic Factors								
Female	46.4%	427	50.5%	653	46.3%	693	47.9%	1,773
Male	53.6%	427	49.5%	653	53.7%	693	52.1%	1,773
African American	62.1%	427	54.7%	653	55.8%	693	56.9%	1,773
Hispanic	24.1%	427	26.6%	653	26.1%	693	25.8%	1,773
White	9.8%	427	14.1%	653	12.6%	693	12.5%	1,773
Disability	9.1%	427	6.4%	653	11.5%	693	9.1%	1,773
Academic Factors								
Limited English Proficiency	7.3%	427	5.1%	653	7.6%	693	6.6%	1,773
Academically Gifted	23.4%	427	18.5%	653	14.4%	693	18.1%	1,773
Academically Gifted - Math	18.7%	427	16.5%	653	12.7%	693	15.6%	1,773
Academically Gifted - Reading	14.8%	427	13.3%	653	10.1%	693	12.4%	1,773
On Track	62.8%	427	50.2%	653	88.6%	693	68.2%	1,773
Ahead of Track	-	-	11.0%	653	7.6%	693	9.3%	1,346[6]
Behind Track	37.2%	427	38.7%	653	3.8%	693	24.7%	1,773
Hillside High School	34.0%	427	18.7%	653	12.4%	693	19.9%	1,773
Northern High School	12.6%	427	19.0%	653	25.4%	693	20.0%	1,773
Riverside High School	26.7%	427	28.8%	653	23.7%	693	26.3%	1,773
Southern High School	25.5%	427	20.7%	653	30.4%	693	25.7%	1,773

METHODOLOGY

In this subsection, Hanover Research describes our methods for making comparisons between the benchmark assessment and the EOC assessment.

The goal of this analysis is to determine the strength of the correlation between each pair of assessments in each subject. However, it should be noted that correlations between the benchmark and EOC could result from a number of conditions which indicate various relationships between the assessments and may also reflect outcomes of Durham's academic programming. For example, lower levels of correlation could indicate that the benchmark is different in content to the EOC. However, since content-specific academic programming occurs during the time between administration of the benchmark and administration of the EOC, lower levels of correlation could also indicate that Durham's programming changes student outcomes such that students who perform poorly on the benchmark test perform well relative to other students on the EOC. If programming decisions, such as interventions, are initiated in response to benchmark results, this may further reduce the expected correlation between the two assessments. In short, there are multiple reasons a stronger or weaker correlation between these assessments may be observed.

[6] As noted above, the most common grade for Algebra I is Grade 9, and because we only evaluate students in Grades 9 through 12, there are no students coded "ahead of track" for this subject.

© 2016 Hanover Research

PEARSON CORRELATIONS

Using the percent of questions answered correctly on the benchmark and the scaled scores on the EOC for each subject, we compute Pearson correlation coefficients separately for each subject. This correlation coefficient ranges from -1 to 1, with -1 indicating perfect negative correlation (when one variable rises, the other falls) and 1 indicating perfect positive correlation (the two variables rise and fall together). A correlation coefficient of zero indicates that there is no correlation between the variables. Statistical significance is based on the null hypothesis that the correlation between the two variables is zero, meaning that three significance asterisks which indicate a p-value of less than 0.01 are evidence that there is a less than 1 percent probability that the relationship between the two variables is zero. Because the scaled scores on the EOC are not scaled consistently across the subjects, we do not compute an overall correlation coefficient for the three assessments pooled together.[7]

Building on the presentation of the Pearson correlations, we further describe the relationship between the benchmark and EOC assessments based on the results of a simple linear regression, where percent correct on the benchmark is used to predict EOC scaled score. Whereas the Pearson correlation coefficients are standardized, ranging from -1 to 1, unstandardized regression coefficients allow us to express the relationship between assessments in each subject in terms of the units of each assessment (e.g., an "*X*"-percentage point increase in percent correct on the benchmark assessment is associated with a "*Y*"-point increase in EOC scaled scored). We also provide scatter plots and trend lines illustrating the relationship between assessments, where the slope of the trend line is equal to the unstandardized regression coefficient for a given subject.

Lastly, in Section III, we compute Pearson correlations separately for each subject and each student subgroup, where applicable.

ACHIEVEMENT LEVEL COMPARISONS

We analyze achievement levels by comparing the numbers and percentages of students who fall within each achievement level for each assessment, as in Figure 1.3 below. This allows the reader to observe the extent to which students are categorized within the same achievement level for both assessments or in different achievement levels. The numbers in the dark teal cells represent students who fall into the same achievement level for both assessments. Those in the light teal cells represent students who fall into a higher achievement level on the EOC than on the benchmark. The light red cells indicate students who fall into a lower achievement level on the EOC than on the benchmark.

We also compute the number of students who stay at the same level, stay at the same or a "neighbor" level (one level above or below), advance at least one level, and regress at least one level.

[7] These different scales are observable by comparing the means of the scaled scores across the assessments in Figure 1.1, as well as the scatter plots displayed in Section II.

© 2016 Hanover Research

Figure 1.3: Achievement Level Comparisons

Subject		EOC Achievement Level					
		1	2	3	4	5	Total
Benchmark Achievement Level	1						
	2						
	3						
	4						
	5						
Total							

© 2016 Hanover Research

SECTION II: TOPLINE COMPARISONS

In this section, Hanover Research compares the results of the benchmark assessment with the EOC assessment separately for each subject. In the achievement level comparisons below, we also compare the subjects pooled into a single analysis.

CORRELATIONS

The correlation coefficients in Figure 2.1 indicate the Pearson correlation between the percent of questions answered correctly on the benchmark and the scaled scores on the EOC. In all cases there is a strong positive correlation that is statistically significant at beyond the 99 percent confidence level. These correlations range from 0.653 for Biology to 0.752 for Algebra I.

Figure 2.1: Correlations between Percent Correct on Benchmark and EOC Scaled Scores

SUBJECT	COEF.	N
Algebra I	0.752***	427
Biology	0.653***	653
English II	0.672***	693

Asterisks denote statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1.

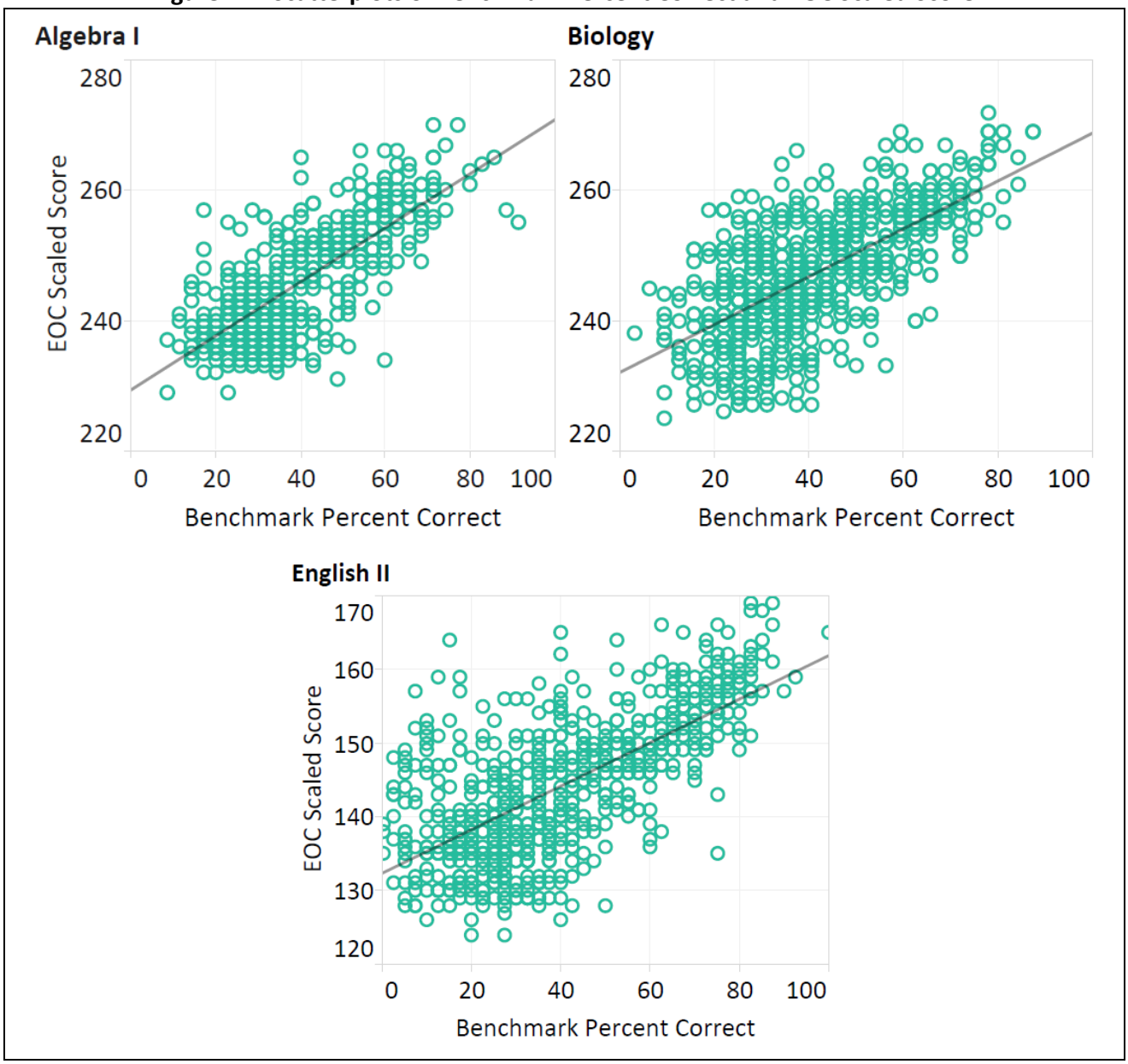
To further illustrate the relationship between benchmark and EOC assessments, we express these correlations in terms of simple linear regression coefficients,[8] where we use percent correct on the benchmark to predict EOC scaled scores. We find that for every 1-percentage-point increase in percent correct on the Algebra I benchmark assessment, we expect a 0.42-point increase in Algebra I EOC scaled score. For Biology, we find that every 1-percentage point increase in percent correct on the benchmark assessment is associated with a 0.37-point increase in the corresponding EOC assessment's scaled score. Lastly, for English II, every 1-percentage point increase in percent correct on the benchmark assessment is associated with a 0.29-point increase in the corresponding EOC assessment's scaled score.

Figure 2.2 on the following page presents scatterplots of benchmark and EOC assessment performance in each subject.

[8] Note that the Pearson correlation coefficients are equivalent to *standardized* regression coefficients of a simple linear regression involving percent correct on the benchmark assessment and the corresponding EOC assessment scaled scored. Here we discuss the *unstandardized* regression coefficients, as they allow us to estimate how a change in one unit of the predictor variable (percent correct on the benchmark assessment) relates to changes in units of the outcome variable (EOC scaled score).

© 2016 Hanover Research

Figure 2.2: Scatterplots of Benchmark Percent Correct and EOC Scaled Score



© 2016 Hanover Research

ACHIEVEMENT LEVEL COMPARISONS

In Figure 2.3, we provide a summary of our achievement level analysis. Each assessment has achievement levels ranging from 1 to 5, and Figure 2.3 shows that, overall, 49.2 percent of students are rated at the same achievement level by both the benchmark test and the EOC for the given subject. This percentage ranges from 45.2 percent for Biology to 53.2 percent for Algebra I. Further, 84.0 percent of students are rated at the same achievement level or a "neighbor" level (i.e., one level above or below). In terms of individual subjects, this percentage ranges from 80.6 percent in Biology to 86.2 percent in Algebra I.

In general, many more students are rated at a higher level on the EOC compared with the benchmark than those who are rated at a lower level on the EOC compared with the benchmark. Overall, 33.5 percent of students are rated at a higher level on the EOC than on the benchmark, whereas just 17.3 percent of students are rated at a lower level on the EOC than on the benchmark. We observe the greatest disparity in Algebra I, where 36.5 percent of students were rated at a higher level on the EOC than on the benchmark, compared to only 10.3 percent who were rated at a lower level. By contrast, the disparity is much narrower for English II, where 25.7 percent were rated at a higher level on the EOC than the benchmark, compared to 23.7 percent who were rated at a lower level.

Figure 2.3: Differences in Achievement Level from Benchmark to EOC, Summary

Difference	All Subjects	Algebra I	Biology	English II
Same Level	49.2%	53.2%	45.2%	50.6%
Same Level or Neighbor Level	84.0%	86.2%	80.6%	86.0%
Same or Higher Level	82.7%	89.7%	85.0%	76.3%
Higher Level	33.5%	36.5%	39.8%	25.7%
Lower Level	17.3%	10.3%	15.0%	23.7%
Number of Observations	**1,773**	**427**	**653**	**693**

Note: This figure represents a summary of the results presented in Figures 2.4-2.11.

The remaining figures in this section display detailed comparisons of benchmark and EOC achievement levels for all subjects combined and each individual subject (as summarized in Figure 2.3 above). Each subject area includes two figures: the first displays percentages while the second displays counts. As described in the methodology, numbers in dark teal cells represent students classified within the same achievement level for both assessments, numbers in light teal cells represent students classified within a higher achievement level on the EOC than on the benchmark, and numbers in light red cells indicate students classified within a lower achievement level on the EOC than on the benchmark.

© 2016 Hanover Research

ALL SUBJECTS

Figure 2.4: Achievement Level Comparisons, All Subjects[9]

ALL SUBJECTS		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	31.1%	10.5%	2.7%	3.5%	0.1%	47.9%
	2	6.7%	6.2%	3.3%	5.7%	0.4%	22.3%
	3	0.8%	3.5%	1.7%	5.1%	0.6%	11.7%
	4	0.5%	1.7%	1.8%	9.0%	1.5%	14.5%
	5	0.0%	0.0%	0.1%	2.3%	1.2%	3.6%
Total		**39.0%**	**21.9%**	**9.6%**	**25.7%**	**3.8%**	**100.0%**

Note: This figure includes all student-subject observations we analyze, meaning that students with data for assessments in multiple subjects are represented multiple times. N=1,773

Figure 2.5: Achievement Level Comparisons, All Subjects

ALL SUBJECTS		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	551	187	48	62	2	850
	2	118	110	59	101	7	395
	3	14	62	30	91	10	207
	4	8	30	32	160	27	257
	5	0	0	1	41	22	64
Total		**691**	**389**	**170**	**455**	**68**	**1,773**

Note: This figure includes all student-subject observations we analyze, meaning that students with data for assessments in multiple subjects are represented multiple times. N=1,773

ALGEBRA I

Figure 2.6: Achievement Level Comparisons, Algebra I

ALGEBRA I		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	39.3%	8.4%	1.6%	2.3%	0.0%	51.8%
	2	7.0%	5.6%	5.9%	8.0%	0.5%	26.9%
	3	0.5%	1.6%	1.4%	7.7%	0.7%	11.9%
	4	0.0%	0.2%	0.5%	6.3%	1.4%	8.4%
	5	0.0%	0.0%	0.0%	0.5%	0.5%	0.9%
Total		**46.8%**	**15.9%**	**9.4%**	**24.8%**	**3.0%**	**100.0%**

N=427

[9] Figure 2.4 and Figure 2.5 describe achievement level comparisons for the benchmark and EOC for all subjects pooled. These comparisons allow Durham to observe the overall extent to which the benchmark and EOCs are comparable.

© 2016 Hanover Research

Figure 2.7: Achievement Level Comparisons, Algebra I

ALGEBRA I		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	168	36	7	10	0	221
	2	30	24	25	34	2	115
	3	2	7	6	33	3	51
	4	0	1	2	27	6	36
	5	0	0	0	2	2	4
Total		**200**	**68**	**40**	**106**	**13**	**427**

N=427

BIOLOGY

Figure 2.8: Achievement Level Comparisons, Biology

BIOLOGY		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	24.2%	11.6%	3.5%	4.4%	0.3%	44.1%
	2	6.6%	8.3%	2.5%	6.6%	0.6%	24.5%
	3	1.2%	3.8%	1.2%	6.4%	1.1%	13.8%
	4	0.5%	1.2%	1.2%	9.5%	2.8%	15.2%
	5	0.0%	0.0%	0.0%	0.5%	2.0%	2.5%
Total		**32.5%**	**25.0%**	**8.4%**	**27.4%**	**6.7%**	**100.0%**

N=653

Figure 2.9: Achievement Level Comparisons, Biology

BIOLOGY		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	158	76	23	29	2	288
	2	43	54	16	43	4	160
	3	8	25	8	42	7	90
	4	3	8	8	62	18	99
	5	0	0	0	3	13	16
Total		**212**	**163**	**55**	**179**	**44**	**653**

N=653

© 2016 Hanover Research

ENGLISH II

Figure 2.10: Achievement Level Comparisons, English II

ENGLISH II		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	32.5%	10.8%	2.6%	3.3%	0.0%	49.2%
	2	6.5%	4.6%	2.6%	3.5%	0.1%	17.3%
	3	0.6%	4.3%	2.3%	2.3%	0.0%	9.5%
	4	0.7%	3.0%	3.2%	10.2%	0.4%	17.6%
	5	0.0%	0.0%	0.1%	5.2%	1.0%	6.3%
Total		**40.3%**	**22.8%**	**10.8%**	**24.5%**	**1.6%**	**100.0%**

N=693

Figure 2.11: Achievement Level Comparisons, English II

ENGLISH II		EOC ACHIEVEMENT LEVEL					
		1	2	3	4	5	TOTAL
Benchmark Achievement Level	1	225	75	18	23	0	341
	2	45	32	18	24	1	120
	3	4	30	16	16	0	66
	4	5	21	22	71	3	122
	5	0	0	1	36	7	44
Total		**279**	**158**	**75**	**170**	**11**	**693**

N=693

© 2016 Hanover Research

SECTION III: CORRELATIONS BY STUDENT SUBGROUPS

In this section, Hanover Research describes correlations segmented by student subgroups based on demographic and academic factors.[10]

Figure 3.1 shows that all of the segmented results are positively correlated at beyond the 95 percent confidence level (and most are significant at the 99 percent confidence level or higher). In terms of strength of the observed relationships, in nearly all cases (42 of 50 subgroup comparisons), the coefficient is above 0.5.

Figure 3.1: Correlations by Student Subgroups

Segmentation Variable	Algebra I		Biology		English II	
	Coef.	N	Coef.	N	Coef.	N
Demographic Factors						
Female	0.724***	198	0.627***	330	0.648***	321
Male	0.771***	229	0.676***	323	0.702***	372
African American	0.709***	265	0.653***	357	0.576***	387
Hispanic	0.758***	103	0.589***	174	0.641***	181
White	0.697***	42	0.666***	92	0.748***	87
Disability	0.403**	39	0.477**	42	0.396***	80
Academic Factors						
Limited English Proficiency	0.610***	31	0.556***	33	0.417**	53
Academically Gifted	0.588***	100	0.627***	121	0.590***	100
Academically Gifted - Math	0.591***	80	0.637***	108	0.584***	88
Academically Gifted - Reading	0.595***	63	0.572***	87	0.628***	70
On Track	0.741***	268	0.608***	328	0.660***	614
Ahead of Track[11]	-	0	0.702***	72	0.496***	53
Behind Track	0.555***	159	0.578***	253	0.611***	26
Hillside High School	0.735***	145	0.691***	122	0.600***	86
Northern High School	0.572***	54	0.343***	124	0.748***	176
Riverside High School	0.767***	114	0.723***	188	0.783***	164
Southern High School	0.627***	109	0.319***	135	0.432***	211
All Students						
All Students	0.752***	427	0.653***	653	0.672***	693

Asterisks denote statistical significance as follows: *** p<0.01, ** p<0.05, * p<0.1.

[10] Please see the methodology in Section I for descriptions of why specific subgroups are displayed.

[11] Note that "on track" represents students who took the benchmark/EOC assessments in a given subject in the most common grade for students testing in that subject (i.e., Grade 10 for Biology and English II and Grade 9 for Algebra I), while "ahead of track" represents students taking the assessment in an earlier grade and "behind track" represents students taking the assessment in a later grade. Further, as discussed in Section I, this report only analyzes students from Grade 9 to Grade 12. Since the greatest number of students taking the Algebra I assessments are in Grade 9, there are no students who are considered "ahead of track" for this subject.

© 2016 Hanover Research

While the correlations between the assessment outcomes are positive in all cases, they are stronger for some student subgroups than others. For example, correlations between the assessments are above average for all three subjects at Riverside High School (ranging from 0.723 for Biology to 0.783 for English II). This indicates that performance on the benchmark assessment by students at Riverside High School is a relatively strong predictor of performance on the EOC.

On the other hand, the correlations are weaker when disaggregating by some of the other groups. For example, students with disabilities exhibit relatively lower correlations for the three subjects, with coefficients ranging from 0.396 (English II) to 0.477 (Biology), though we also note that the sample size for this group is fairly small. Similarly, students with Limited English Proficiency (LEP) have lower correlations than those observed in some of the other subgroups, with correlation coefficients ranging from 0.417 (English II) to 0.610 (Algebra I), though again, relatively few LEP students were represented in the sample. Nevertheless, within each subgroup we observe a positive and moderate to strong correlation between benchmark and EOC performance in all three subjects.

© 2016 Hanover Research

PROJECT EVALUATION FORM

Hanover Research is committed to providing a work product that meets or exceeds client expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.

http://www.hanoverresearch.com/evaluation/index.php

CAVEAT

The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties that extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every client. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Clients requiring such services are advised to consult an appropriate professional.

4401 Wilson Boulevard, Suite 400
Arlington, VA 22203
P 202.559.0500 **F** 866.808.6585
www.hanoverresearch.com