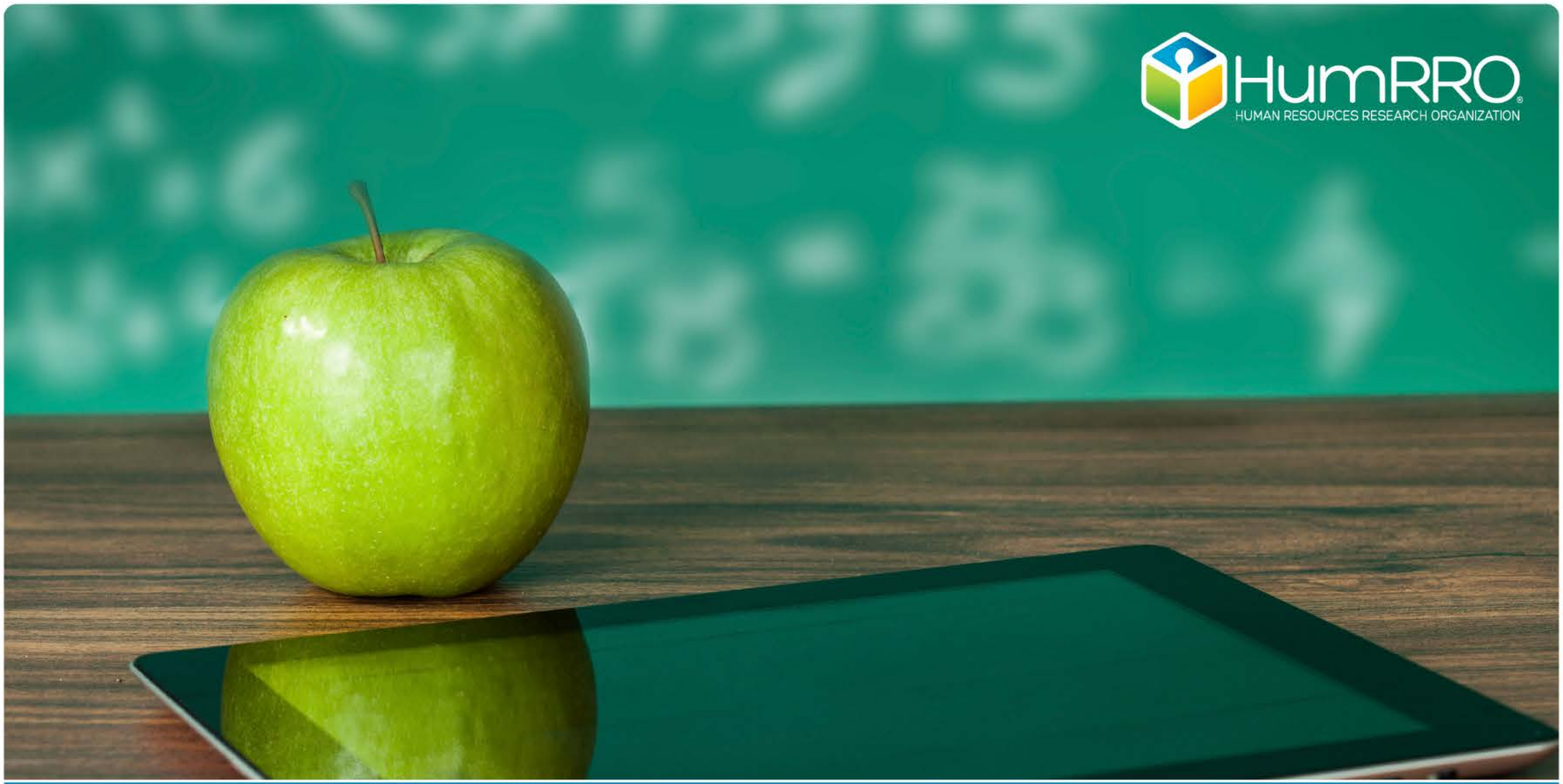

2018 No. 093

PARCC Grade Level Readiness Investigation

Draft Report

Prepared for: New Meridian Corporation
5800 Trailridge Dr.
Austin, TX 78731

Prepared under: New Meridian Corporation
5800 Trailridge Dr.
Austin, TX 78731

Authors: Justin D. Purl
Britany F. Crawford
Erin S. Banjanovic
Emily R. Dickinson
Lucy Chen
Arthur A. Thacker

Date: November 30, 2018

Executive Summary

Measuring college readiness is a major goal for the PARCC high school assessments (Steedle, Quesen, & Boyd, 2017). Grade-level readiness is an extension of this goal for assessments prior to high school. Specifically, scores at each grade level should indicate preparedness for the next grade level, and this link across grade levels would allow the PARCC elementary and middle school assessments to measure those “on track” to be college ready in high school. This study aimed to combine four years of student data to evaluate student proficiency trends over time. Furthermore, the study reports on the college readiness of high-school students given benchmarks on external tests (e.g., ACT, PSAT, SAT) compared to PARCC assessment proficiency in high school.

The evaluation of evidence for grade-level readiness within student performance data on the PARCC assessments involves three research questions:

1) To what extent is student performance related (linked) across grade levels from grade 3 through high school?

2) What is the nature of the trends in student performance across grade levels (3 through high school)?

3) Do high school tests have comparable college readiness benchmarks to those of external tests (e.g., College Board and ACT)?

To examine basic trends in student performance across years, the percentages of students proficient in a given year and the percent changing proficiency across years were evaluated. Less than 24% of students in a given grade level comparison have a prior year proficiency status that is incongruent with their current year proficiency status (i.e., proficient changed to not proficient and vice versa).

Hierarchical linear modeling (HLM) was used to estimate the trend in performance level across grade, where PARCC Level 4 or higher indicates proficiency. HLM accounts for the repeated measures design of the data (Bryk & Raudenbush, 1992), where each student is tested at multiple grade levels. A consistent (i.e., linear) upward trend in student performance level was found for ELA and a consistent downward trend for Math. In both subjects, the trend from grade 3 to high school is such that students are likely to attain a performance level similar to the performance level in the previous grade. In other words, proficient students (those with performance level 4 or 5) in the prior grade are likely to attain proficiency in the next grade level.

Using the work of Steedle, Quesen, and Boyd (2017), an expected external (e.g., SAT/PSAT and ACT) score cut was applied to the students in the PARCC population. The PARCC was a more rigorous benchmark for college readiness compared to external tests in all but two comparisons. Specifically, the estimated percent of college ready students according to the PARCC Level 4 cut was lower than the corresponding percent of college ready students expected from the external score cut. The exceptions were the ELA grade 11 expected cuts for the ACT English and Reading benchmarks, where the external cuts resulted in a lower percent of college ready students. Overall, students who are PARCC Level 4 or higher have a high likelihood of being academically prepared for college. Further investigation could compare PARCC scores directly to college course success to provide additional validity evidence for the link between PARCC scores and college readiness.

In summary, data from four consecutive years (i.e., 2015-2018) of the PARCC assessment, constituting nearly 4 million students in ELA and nearly 3 million students in Math, were used to evaluate the nature of the trend in student scores from grade 3 to high school. Findings imply that PARCC proficiency (i.e., attaining PARCC Level 4 or 5) can be used as an indicator of whether a student is “on track” to college readiness after high school.

PARCC Grade-Level Readiness Investigation

Table of Contents

List of Tables

List of Figures

PARCC Grade-Level Readiness Investigation

Introduction and Purpose

In the spring of 2018, the Human Resources Research Organization (HumRRO) was retained by the New Meridian Corporation to conduct a study to examine the grade-level readiness of students considered proficient on the Partnership for Assessment of Readiness for College and Careers (PARCC) assessments for Math and English Language Arts/Literacy (ELA) grades 3-8 and high school. Measuring college readiness is a major goal for the PARCC high school assessments (Steedle, Quesen, & Boyd, 2017). Grade-level readiness is an extension of this goal for assessments prior to high school. Specifically, scores at each grade level should indicate preparedness for the next grade level, and this link across grade levels would allow the PARCC elementary and middle school assessments to measure those "on track" to be college ready in high school. The need for a focus on college and career readiness as early as elementary school has been supported by researchers in career development and counseling (e.g., Auger, Blackhurst, & Wahl, 2005; Magnuson & Starr, 2000; Knight, 2015). Prior research on student achievement has estimated the growth trajectory of students from kindergarten to high school using national longitudinal data (i.e., Lee, 2012). Lee (2012) uses several assumptions, standardizations, and equating procedures to attempt linking scores from different samples of students across grade levels. The current study aimed to combine four years of PARCC student data across nine grade levels to evaluate student proficiency trends over time within a common sample. Furthermore, the study reports on the college readiness of high-school students given benchmarks on external tests (e.g., ACT, PSAT, SAT) compared to attaining performance level 4 or 5 on the PARCC high school assessments.

The evaluation of evidence for grade-level readiness within student performance data on the PARCC assessments involves three research questions:

1) To what extent is student performance related (linked) across grade levels from grade 3 through high school?
2) What is the nature of the trends in student performance across grade levels (3 through high school)?
3) Do high school tests have comparable college readiness benchmarks to those of external tests (e.g., ACT, PSAT, SAT)?

Research questions one and two both involve an evaluation of student performance trends from grade 3 through high school. Several trends may be observed on the PARCC assessments. Student proficiency in lower grade levels should be positively related to proficiency in later grade levels. Specifically, if a student is proficient in grade 3, they should be more likely than a student who did not reach proficiency in grade 3 to attain proficient status in grade 4. In general, this grade-level readiness hypothesis should emerge as a positive relationship between prior performance level/proficiency status and future level/status (i.e., research question one). Conversely, standards for each successive grade will be more demanding; thus, the likelihood of a student attaining proficiency may decrease as grade level increases. The increase in demand for students may also be more significant between certain grade-levels (e.g., elementary to middle school and middle to high school; West & Schwerdt, 2012). As such, key transition points may emerge resulting in student performance trends that are non-linear (i.e., research question two). Furthermore, trends across the grade span should be similar for students from diverse backgrounds. Several reasons may exist for groups of students having larger or smaller negative trends in proficiency between grade levels; however, similar student trends, on average, would support the idea that the assessment system is fair. Comparison of student trends by demographic characteristics is an exploratory evaluation meant to support research question two.

Evaluation of student score trends relies on an equivalent score design. The PARCC assessments strive to ensure year-to-year comparability of test scores through equating. This rigorous psychometric processes links grade-level and end-of-course scores across years and allows evaluation of trends in student scores. Specifically, a difference in scale score means between ELA grade 3 scores in 2017 and 2018 can be interpreted as a change in 3rd graders ELA performance on average. Without equating, it would be impossible to differentiate differences in means to changes in the difficulty of the test and changes in student performance. Comparable test scores that are the result of a strong equating design allow for information to be accumulated across years, yielding a voluminous data source for evaluating test properties and supporting education accountability policy implementation.

In contrast, vertical linking is a psychometric process used to make student scores across grade levels comparable. Specifically, a difference in means between ELA grade 3 and 4 can be interpreted as the average change in student performance between 3rd and 4th grade. As with year-to-year equating, the difficulty of the test can be eliminated as an alternative explanation. Although PARCC has not conducted a vertical linking study to date, efforts have been made to establish common interpretations of test performance across the grade levels. For example, the PARCC assessments were reviewed by state-nominated panels of experts during the summer of 2015. These panels matched threshold scores on each grade-level or high school assessment to the performance of a typical student such that students attaining a performance level threshold should be able to demonstrate command of grade-level standards. Additional studies provided further information on these threshold scores, which were the basis for the common performance levels adopted by state education chiefs from PARCC member states.[1] The five resulting performance levels are:

- Level 1: Did not yet meet expectations
- Level 2: Partially met expectations
- Level 3: Approached expectations
- Level 4: Met expectations
- Level 5: Exceeded expectations.

As a result, each grade level has five increasing performance levels based around thresholds representing typical students' grasp of the grade-level standards. Although this approach does not permit direct comparison of individual scores (e.g., a 728 in grade 3 should not be interpreted as equivalent to a 728 in grade 4); it does permit general comparisons of aggregate scores. For example, if 38% of students meet expectations according to 3rd grade standards (i.e., attained level 4 performance) and 42% of students meet expectations according to 4th grade standards, then we can conclude that the number of students meeting grade level standards has increased even if the average scale score is the same. Furthermore, performance levels across the grade level span should have similar grade appropriate interpretations. For this study, levels 4 and 5 are used to group students into a proficient category based on a threshold score of 750 for all subjects and grades.

[1] https://parcc-assessment.org/performance-levels/

The final research question involves the capstone to grade-level readiness: college readiness. At the high school level, the PARCC assessment proficiency threshold is designed to indicate that a student is college ready (Pearson, 2018). Thus, the grade 9-11 PARCC proficiency benchmarks should be comparable to the benchmarks established by external tests of college readiness (e.g., PSAT/SAT and ACT). Ideally, students who score at the benchmark for college readiness on PARCC assessments (i.e., achieve proficiency) will be just as qualified for college courses as those scoring at the benchmark set by the comparison assessments. Extreme differences between the percent college ready determined by the two cuts would indicate a misalignment between the PARCC performance levels and college readiness standards defined by external tests.

Altogether, support for these three research questions should demonstrate that student performance as measured by the PARCC assessments 1) is linked across grade levels, 2) has a stable trend (with or without major transitions), and 3) supports college readiness at the high school level.

Methods

Data from the 2015, 2016, 2017, and 2018 ELA and Math operational test administrations were used in this study. The procedures used to compile student data across years is described below. This is followed by a description of the sample and analyses used to evaluate the research questions.

Student Data Processing

Five states (or territories) provided data for the current study: The District of Columbia (DC), Illinois, Maryland, New Jersey, and New Mexico. HumRRO processed online and paper test administration data for ELA grades 3-11, Math grades 3-8, Algebra 1, Algebra 2, and Geometry 1. Student raw scores for each assessment were provided for four consecutive years (i.e., 2015-2018). Student records for a particular year were removed if they: (1) did not complete the test, (2) had an invalid score, (3) took a retest within the same year, (4) were identified as a duplicate record, or (5) took a non-English or accommodated form. The exclusion criteria are designed to limit the sample to students taking the most prevalent forms (unaccommodated, English forms) with the most typical circumstances (completing the test in full at one time). Raw scores were converted to scale scores using the raw score to scale score table corresponding to the year, grade, and form administered to the student and performance levels were assigned. Students were evaluated as proficient vs not proficient by collapsing the five proficiency levels. Students assigned proficiency levels 1-3 (did not yet meet expectations, partially met expectations, and approached expectations) were considered not proficient and students assigned levels 4 or 5 (met expectations and exceeded expectations) were considered proficient.[2] Records were then merged across years using a unique, encoded student identifier.

Description of Student Sample

The student sample size by test, grade, and year are summarized in Tables 1 and 2. In general, the number of students with scores was similar across the four-year span for each test and grade-level.

[2] A scale score cut of 750 is associated with proficiency level 4 across all grades, subjects, and years. Thus, students who scored below 750 were considered not proficient and students who score 750 or higher were considered proficient for this study.

Table 1. Number of Students Included in the ELA Sample by Grade and Year

Grade	Year			
	2015	2016	2017	2018
03	328,384	333,375	334,739	324,726
04	322,036	323,099	335,226	329,356
05	323,773	321,241	328,969	333,585
06	319,738	323,254	325,838	327,001
07	313,780	320,873	329,059	323,025
08	310,927	317,786	326,519	325,892
09	184,208	215,918	125,675	125,513
10	154,994	184,972	171,923	184,524
11	119,975	133,352	108,795	103,951
Total	2,377,815	2,473,870	2,386,743	2,377,573

Table 2. Number of Students Included in the Math Sample by Grade and Year

Grade	Year			
	2015	2016	2017	2018
03	276,443	266,696	280,558	269,871
04	269,945	261,209	284,377	274,985
05	273,881	239,216	257,046	257,929
06	254,173	251,685	264,568	263,717
07	243,819	243,144	261,068	254,447
08	204,480	203,997	208,015	205,848
Alg1	183,268	211,514	183,738	188,347
Alg2	128,277	120,275	111,360	111,072
Geo1	87,668	114,955	112,307	114,403
Total	1,921,954	1,912,691	1,963,037	1,940,619

Tables 3 and 4 provide further details about student demographics, including sex and race. Sample characteristics are comparable for both Math and ELA. The sample was evenly split between male and female students each year. The largest racial subgroup consisted of White students, followed by students who identify as Hispanic/Latino, Black or African American, Asian, with two or more races, American Indian or Alaskan Native, and then Native Hawaiian or Other Pacific Islander.

Table 3. Demographic Composition of ELA Student Sample

Demographic Variable/Subgroup	2015		2016		2017		2018	
	N	%	*N*	%	*N*	%	*N*	%
Sex								
Male	1,208,461	50.8%	1,253,397	50.7%	1,209,672	50.7%	1,203,500	50.6%
Female	1,168,238	49.1%	1,219,349	49.3%	1,175,687	49.3%	1,172,827	49.3%
Not Provided	1,116	0.1%	1,124	0.1%	1,384	0.1%	1,246	0.1%
Race								
American Indian/Alaska Native	29,474	1.2%	25,496	1.0%	26,164	1.1%	26,330	1.1%
Asian	154,390	6.5%	161,498	6.5%	159,867	6.7%	161,836	6.8%
Black or African American	444,042	18.7%	466,475	18.9%	451,861	18.9%	450,689	19.0%
Hispanic/Latino Ethnicity	613,623	25.8%	652,040	26.4%	650,384	27.2%	664,048	27.9%
Native Hawaiian or Other Pacific Islander	3,250	0.1%	3,503	0.1%	3,447	0.1%	3,633	0.2%
White	1,059,829	44.6%	1,084,932	43.9%	1,012,152	42.4%	985,938	41.5%
Two or More Races	50,049	2.1%	52,048	2.1%	61,607	2.6%	65,647	2.8%
Not Provided	23,158	1.0%	27,878	1.1%	21,261	0.9%	19,452	0.8%

Note. N = Number of students; % = percent of all students.

Table 4. Demographic Composition of Math Student Sample

Demographic Variable/Subgroup	2015		2016		2017		2018	
	N	%	*N*	%	*N*	%	*N*	%
Sex								
Male	969,701	50.5%	960,585	50.2%	983,420	50.1%	971,172	50.0%
Female	951,436	49.5%	951,281	49.7%	978,627	49.9%	968,579	49.9%
Not Provided	817	<0.1%	825	<0.1%	990	0.1%	868	<0.1%
Race								
American Indian/Alaska Native	22,668	1.2%	20,034	1.0%	21,405	1.1%	20,307	1.0%
Asian	125,370	6.5%	126,352	6.6%	138,453	7.1%	139,331	7.2%
Black or African American	344,738	17.9%	342,133	17.9%	357,358	18.2%	355,977	18.3%
Hispanic/Latino Ethnicity	460,337	24.0%	459,775	24.0%	484,568	24.7%	487,106	25.1%
Native Hawaiian or Other Pacific Islander	2,761	0.1%	2,821	0.1%	2,789	0.1%	2,943	0.2%
White	905,264	47.1%	898,258	47.0%	887,554	45.2%	863,892	44.5%
Two or More Races	47,578	2.5%	49,797	2.6%	55,345	2.8%	56,881	2.9%
Not Provided	13,238	0.7%	13,521	0.7%	15,565	0.8%	14,182	0.7%

Note. N = Number of students; % = percent of all students.

A critical feature of the current data set is that it includes longitudinal data for individual students. Tables 5 and 6 group data availability by the first year that the student appeared in the longitudinal dataset for ELA and Math, respectively. For ELA, over 27% of the students in the dataset had data available at four timepoints, while around 17% of the students had data available at only one timepoint. Compared to the ELA data, a smaller percentage of students (22%) have math data available at four timepoints. Approximately 21% of the students have data available at only one timepoint.

Table 5. Number of Students having Years with ELA Data Available

First Year Student is in Data	Years of Data Available	Students in Longitudinal Datasets (*N*=3,878,102)	
		N	%
2015	1	416,857	10.75%
	2	462,577	11.93%
	3	445,496	11.49%
	4	1,052,868	27.15%
2016	1	164,450	4.24%
	2	81,242	2.09%
	3	377,718	9.74%
2017	1	64,113	1.65%
	2	389,107	10.03%
2018	1	420,479	10.84%
Total	1	1,065,899	16.64%
	2	932,926	24.06%
	3	823,214	21.23%
	4	1,052,868	27.15%

Table 6. Number of Students having Years with Math Data Available

First Year Student is in Data	Years of Data Available	Students in Longitudinal Datasets (*N*=3,348,849)	
		N	%
2015	1	452,232	13.50%
	2	390,175	11.65%
	3	334,755	10.00%
	4	744,639	22.24%
2016	1	141,016	4.21%
	2	104,996	3.14%
	3	299,699	8.95%
2017	1	105,276	3.14%
	2	389,552	11.63%
2018	1	386,509	11.54%
Total	1	1,085,033	20.86%
	2	884,723	26.42%
	3	634,454	18.95%
	4	744,639	22.24%

A final assumption of the data is that scores for adjacent grade levels from consecutive years are correlated. These correlations represent the same cohort of students measured on similar content (convergent validity) or different content (divergent validity). Convergent validity correlations should be large in magnitude as is expected for a repeated measures design. Divergent validity estimates are expected to be lower than the convergent validity estimates but are still expected to be high as both ELA and Math measure student performance. Correlations between Math, Algebra, Geometry, and ELA scores for adjacent years are provided in Table 7 (2015 – 2016), Table 8 (2016 – 2017), and Table 9 (2017 – 2018). Correlations were computed for any instance where 50 or more students took a test in adjacent years. Correlations corresponding to a typical student path (i.e., where the next higher grade is taken one year later) are based on a sample size as low as 2,723 for Geometry following Algebra 2 between 2015 and 2016 and as high as 313,408 for ELA grade 5 following ELA grade 4 between 2017 and 2018. Correlations corresponding to an atypical path can also represent many students, such as the 52,958 students taking Algebra 1 following Math grade 7 between 2017 and 2018 (see Appendix A). These tables capture instances where students had to repeat a grade (e.g., scores for Math 3 were observed in 2015 and 2016), where students were accelerated (e.g., scores for Math 3 were observed in 2015 and scores for Math 5 were observed in 2016), and where students followed the typical academic trajectory (e.g., scores for Math 3 were observed in 2015 and scores for Math 4 were observed in 2016). These tables were produced separately for adjacent years of student data to examine the stability of the relationships across time and to gauge whether changes in instruction may have impacted the relationship among scores. Overall, the relationships appear to be relatively stable with no consistent differences across time. As would be expected, student scores for adjacent grade tests in the same subject (convergent validity estimates, bolded) exhibit the strongest relationships. Additionally, student scores for adjacent grade tests in the opposite subject (divergent validity estimates, underlined) exhibit strong relationships that are weaker than the convergent validity estimates. Support for this assumption is also preliminary support for research question one. Further analyses provide more stringent tests of the research questions.

Table 7. Correlations Between the 2015 and 2016 Scores

2015 Subject	Grade	2016 Subject																	
		Math						Algebra		Geometry	ELA								
		03	04	05	06	07	08	01	02	01	03	04	05	06	07	08	09	10	11
Math	03	.58	**.84**	.92	-	-	-	-	-	-	.41	.72	.87	-	-	-	-	-	-
	04	-	.72	**.85**	.91	-	-	-	-	-	-	.69	.73	.80	-	-	-	-	-
	05	-	-	.73	**.84**	.92	.78	.74	-	-	-	-	.67	.69	.89	-	-	-	-
	06	-	-	-	.67	**.85**	.83	.79	-	-	-	-	-	.67	.73	.71	-	-	-
	07	-	-	-	-	.59	**.81**	.79	-	.63	-	-	-	-	.75	.71	.77	-	-
	08	-	-	-	-	-	.61	**.69**	.80	.75	-	-	-	-	.51	.76	.64	.61	.48
Algebra	01	-	-	-	-	-	.80	.75	**.84**	**.78**	-	-	-	-	.57	.56	.64	.56	.42
	02	-	-	-	-	-	-	.61	.71	**.82**	-	-	-	-	-	-	.70	.55	.51
Geometry	01	-	-	-	-	-	-	.72	**.77**	.51	-	-	-	-	-	-	.59	.56	.47
ELA	03	.32	.70	.82	-	-	-	-	-	-	.53	**.80**	.86	-	-	-	-	-	-
	04	-	.58	.71	.86	-	-	-	-	-	-	.64	**.82**	.80	-	-	-	-	-
	05	-	-	.49	.73	.82	.58	.40	-	-	-	-	.64	**.81**	.88	-	-	-	-
	06	-	-	-	.45	.73	.63	.57	.46	.51	-	-	-	.58	**.82**	.71	-	-	-
	07	-	-	-	-	.52	.70	.63	.47	.47	-	-	-	-	.64	**.82**	.81	-	-
	08	-	-	-	-	-	.53	.59	.54	.54	-	-	-	-	-	.61	**.77**	.81	-
	09	-	-	-	-	-	-	.46	.60	.58	-	-	-	-	-	-	.59	**.76**	.71
	10	-	-	-	-	-	-	.37	.53	.41	-	-	-	-	-	-	.47	.57	**.69**
	11	-	-	-	-	-	-	.32	.42	.24	-	-	-	-	-	-	.16	.62	.68

Note. Correlations are only reported in instances where at least 50 students took both tests. Estimates of convergent validity are bolded and estimates of discriminant validity are underlined.

Table 8. Correlations Between the 2016 and 2017 Scores

2016 Subject	Grade	2017 Subject																	
		Math						Algebra		Geometry	ELA								
		03	04	05	06	07	08	01	02	01	03	04	05	06	07	08	09	10	11
Math	03	.57	**.84**	.87	-	-	-	-	-	-	.38	<u>.74</u>	.85	-	-	-	-	-	-
	04	-	.73	**.84**	.88	-	-	-	-	-	-	.63	<u>.73</u>	.72	-	-	-	-	-
	05	-	-	.84	**.84**	.90	-	.56	-	-	-	-	.70	<u>.71</u>	.74	-	-	-	-
	06	-	-	-	.76	**.84**	.83	.75	-	-	-	-	-	.70	<u>.74</u>	.75	-	-	-
	07	-	-	-	-	.63	**.81**	.78	-	.48	-	-	-	-	.81	<u>.72</u>	.85	-	-
	08	-	-	-	-	-	.74	**.70**	.71	.70	-	-	-	-	.59	.70	<u>.63</u>	.68	-
Algebra	01	-	-	-	-	-	.80	.75	**.86**	**.79**	-	-	-	-	.50	.61	.66	<u>.61</u>	.46
	02	-	-	-	-	-	-	.55	.69	**.85**	-	-	-	-	-	.58	.77	.62	<u>.57</u>
Geometry	01	-	-	-	-	-	-	.71	**.79**	.61	-	-	-	-	-	.57	.71	.61	.51
ELA	03	.34	<u>.72</u>	.83	-	-	-	-	-	-	.54	**.81**	.87	-	-	-	-	-	-
	04	-	.51	<u>.70</u>	.86	-	-	-	-	-	-	.67	**.82**	.83	-	-	-	-	-
	05	-	-	.50	<u>.74</u>	.82	-	.26	-	-	-	-	.60	**.82**	.81	-	-	-	-
	06	-	-	-	.48	<u>.71</u>	.77	.53	-	.34	-	-	-	.55	**.83**	.83	-	-	-
	07	-	-	-	-	.48	<u>.69</u>	.63	.44	.48	-	-	-	-	.63	**.83**	.83	-	-
	08	-	-	-	-	-	.49	<u>.58</u>	.53	.55	-	-	-	-	-	.64	**.82**	.81	-
	09	-	-	-	-	-	-	.42	<u>.63</u>	.58	-	-	-	-	-	-	.60	**.82**	.61
	10	-	-	-	-	-	-	.44	.51	<u>.45</u>	-	-	-	-	-	-	.46	.57	**.73**
	11	-	-	-	-	-	-	.31	.31	.31	-	-	-	-	-	-	.58	.44	.53

Note. Correlations are only reported in instances where at least 50 students took both tests. Estimates of convergent validity are bolded and estimates of discriminant validity are underlined.

Table 9. Correlations Between the 2017 and 2018 Scores

2017 Subject	Grade	2018 Subject: Math 03	Math 04	Math 05	Math 06	Math 07	Math 08	Algebra 01	Algebra 02	Geometry 01	ELA 03	ELA 04	ELA 05	ELA 06	ELA 07	ELA 08	ELA 09	ELA 10	ELA 11
Math	03	.60	**.84**	.89	-	-	-	-	-	-	.39	.73	.86	-	-	-	-	-	-
	04	-	.66	**.85**	.86	-	-	-	-	-	-	.51	.73	.77	-	-	-	-	-
	05	-	-	.66	**.84**	.90	-	.61	-	-	-	-	.50	.71	.78	-	-	-	-
	06	-	-	-	.68	**.85**	.82	.77	-	-	-	-	-	.66	.75	.86	-	-	-
	07	-	-	-	-	.63	**.80**	.83	-	.75	-	-	-	-	.76	.73	.72	-	-
	08	-	-	-	-	-	.59	**.74**	.80	.73	-	-	-	-	-	.74	.64	.71	-
Algebra	01	-	-	-	-	-	.80	.66	**.87**	**.82**	-	-	-	-	.27	.60	.68	.59	.46
	02	-	-	-	-	-	-	.50	.67	**.85**	-	-	-	-	-	-	.77	.64	.51
Geometry	01	-	-	-	-	-	-	.73	**.80**	.50	-	-	-	-	-	.36	.69	.62	.47
ELA	03	.37	.73	.81	-	-	-	-	-	-	.54	**.83**	.88	-	-	-	-	-	-
	04	-	.49	.72	.81	-	-	-	-	-	-	.64	**.83**	.78	-	-	-	-	-
	05	-	-	.37	.73	.80	-	.29	-	-	-	-	.63	**.82**	.82	-	-	-	-
	06	-	-	-	.47	.74	.76	.57	-	.40	-	-	-	.56	**.84**	.85	-	-	-
	07	-	-	-	-	.52	.69	.66	.45	.49	-	-	-	-	.61	**.85**	.78	-	-
	08	-	-	-	-	-	.43	.63	.61	.62	-	-	-	-	-	.59	**.85**	.83	-
	09	-	-	-	-	-	-	.49	.65	.59	-	-	-	-	-	-	.64	**.83**	.68
	10	-	-	-	-	-	-	.45	.53	.48	-	-	-	-	-	-	.40	.50	**.71**
	11	-	-	-	-	-	-	.27	.34	.34	-	-	-	-	-	-	.33	.52	.56

Note. Correlations are only reported in instances where at least 50 students took both tests. Estimates of convergent validity are bolded and estimates of discriminant validity are underlined.

The composition of the longitudinal dataset supports the evaluation of research questions across time and by demographic variables. The longitudinal data analysis is described next.

Grade-Level and College Readiness Analyses

To examine basic trends in student performance across years, the percentages of students attaining PARCC Level 4 or 5 (i.e., proficient) in a given year and the percent changing proficiency across years was evaluated. These percentages provide an estimate of overall trend in proficiency across levels and the relationship between grade levels. To evaluate the relationship between years, the percent of students with a consistent proficiency status in adjacent years (i.e., proficient to proficient and not proficient to not proficient) is evaluated as an indication of the likelihood that a given student's performance will remain the same. In this way, prior proficiency status is used to *predict* the proficiency status in the subsequent year. This percentage is then an indicator of model accuracy. This accuracy metric is related to classification accuracy (e.g., Rudner, 2005) in that both represent the likelihood that a student was classified correctly. Rudner's (2005) approach indexes the certainty that a student will fall above a cut given measurement error. The model accuracy approach indexes the certainty that a prediction will be correct.

The focal unit of analysis is student proficiency (i.e., Level 4 or 5 vs not Level 4 or 5) across grade levels. The year in which a student took an assessment (i.e., 2015, 2016, 2017, or 2018) is collapsed to optimize available data for the longitudinal analysis. Furthermore, performance levels are used in the analysis to optimize the variance in the dependent (focal) variable while maintaining comparability across grade levels. Hierarchical linear modeling (HLM) was used to estimate the trend in performance level across grade, where PARCC Level 4 or higher indicates proficiency. HLM accounts for the repeated measures design of the data (Bryk & Raudenbush, 1992), where each student is tested at multiple grade levels. Modeled using HLM, total variance is decomposed into between-student variance and within-student variance. Between-student factors, like achievement, will be more influential if the grade levels have consistent, strong relationships. Alternatively, within-student factors will be more influential if grade level specific factors (e.g., the introduction of new content in a particular grade level or the transition to middle or high school) are strongly related. Specifically, the Intraclass Correlation Coefficient (ICC) provides an estimate of the variance in performance level accounted for by between-student factors compared to within-student (grade level) factors. The larger the ICC, the more student performance across grade levels can be considered related in a systematic way (i.e., due primarily to student growth). Both latent growth modeling and HLM are typical for nested longitudinal designs; however, researchers recommend HLM when data for the time span (i.e., grade 3 through 11) are incomplete for most individuals (e.g., Ghisletta, Renaud, Jacot, & Courvoisier, 2015). Thus, HLM was conducted to account for most students in the sample having four or fewer grade levels available. Furthermore, the overall trend for students was also estimated in a latent growth modeling framework and resulted in the same conclusion as the HLM approach. As such, only HLM results are presented.

Through HLM, a linear trend was tested across grades 3 through 11 in ELA. For Math, two linear trends were modeled representing two possible trajectories through the high school assessments. Specifically, grade 3 through 8 were always followed by Algebra 1 but the order of Algebra 2 and Geometry were switched to create two possible course-taking patterns. The beta weight associated with the linear trend analysis can be interpreted as the expected change between each grade transition and assumes that the change between grades is constant. A beta weight of less than one suggests that no change in performance level is expected for the average student. Next, the change in performance level between each pair of grade levels was estimated as separate steps in a common model. This analysis allows the direct comparison of changes in performance level between pairs of grade levels to determine if any transitions are larger or in a different direction than the others. The beta weight for each step in the model can be interpreted as the expected change in performance level between adjacent grade levels. The analysis of transitions was particularly important for the grade 5 to 6 and grade 8 to 9 (or Algebra 1 in Math) performance level change estimates, as these pairs correspond to the transition from elementary to middle and middle to high school. Because of the focus on trend analysis, only students with three or four years of data (i.e., sufficient data to estimate variance within a student) were used in the models.

Finally, college readiness was evaluated through comparison between external test scores and PARCC assessment scores. College readiness is defined as achieving a benchmark linked to a 0.75 probability of earning at least a C (i.e., the minimum grade required to receive credit) in an entry-level college course for a given subject (Steedle, Quesen, & Boyd, 2017). Benchmarks are created by using scores on a high school standardized test to predict college freshman performance and determining the score at which 75% of students are expected to pass subject specific courses. The College Board and ACT developed expected scores for their respective assessments and tracked the percentage of students in a given year who are college ready. In 2017, as part of a broader study examining the validity of the PARCC performance levels, Pearson estimated expected scores on the PARCC assessments that would correspond to the college readiness benchmarks for ELA on multiple external tests: PSAT 10 evidence-based reading and writing (EBRW), PSAT/NMSQT EBRW, SAT EBRW, ACT English, and ACT Reading. Similarly, expected scores were calculated for benchmarks from the PSAT10 Math, PSAT/NMSQT, SAT Math, and ACT Math for PARCC Algebra 1, Algebra 2, and Geometry assessments. The PARCC proficiency threshold and the estimated threshold scores from the Pearson study were used to determine if students were ready for postsecondary pursuits based on their high school PARCC ELA and Math scale scores. College readiness rates according the PARCC assessments were ultimately compared to national college readiness rates reported by the College Board and ACT.

Results

Research questions one and two are interdependent and evaluated together as grade-level readiness research questions. The final research question is evaluated as support for college readiness. Results are presented separately by for grade-level and college readiness.

Grade Level Readiness

The percent of proficient students (meeting or exceeding grade-level expectations) for each year are presented in Figure 1 for ELA and Figure 2 for Math. The nature of the trend in proficiency (research question two) appears to be generally stable for ELA and generally decreasing for Math. Specifically, ELA scores seem to fluctuate around 40% proficiency from grade 3 to 11; Math scores seem to decrease from around 40% to around 30% between grades 3 and 8 with another drop between Algebra 1 and subsequent high school assessments (i.e., Algebra 2 and Geometry).

The percent proficient varies between years by less than 6% for ELA grades 3 through 8 and grade 11, with more students generally being identified as proficient each academic year. Grades 9 (range = 12.58%) and 10 (range = 10.96%) have slightly larger upward trends in proficiency across academic years. For Math, the percent proficient varies between years by more than 6% for all grades except grades 6 and 8, again, with more students generally identified as proficient each subsequent year. In descending order, Algebra 1 (range = 11.52%), grade 5 (range = 10.21%), Geometry (range = 9.29%), grade 4 (range = 8.71%), grade 3 (range = 7.86%), Algebra 2 (range = 6.74%), and grade 7 (range = 6.30%) all have upward trends between 2015 and 2018. Several reasons could produce an upward trend over four years. For example, increased emphasis on rigorous standards may yield higher levels of student achievement. Specifically, systemic changes in the way Math is taught has led to Algebra concepts being introduced earlier than previous years (National Council of Teachers of Mathematics, 2014).

Student performance trends for the average student are likely to follow the percent proficient trend; however, the population of students considered in each grade are different. To more closely evaluate student proficiency trends, students must be matched across grade levels.

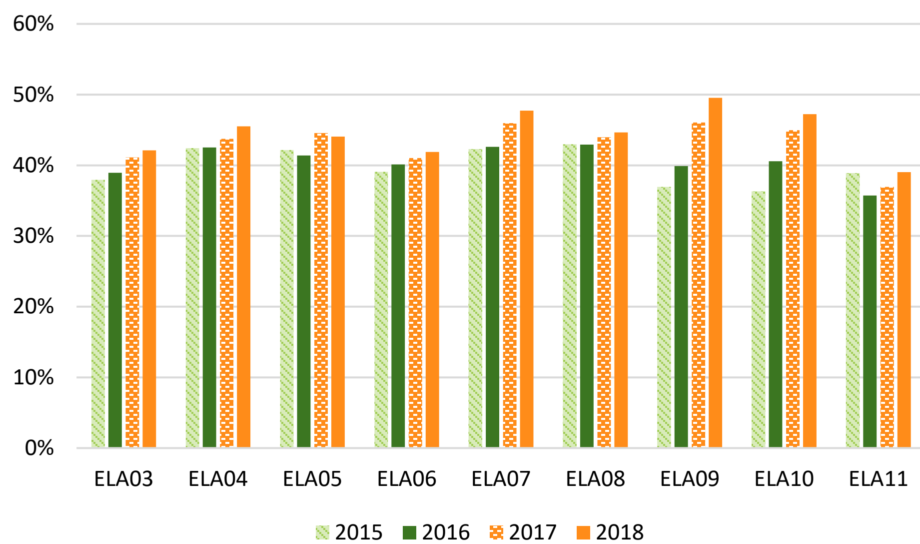

Figure 1. Student proficiency in ELA by grade and year.

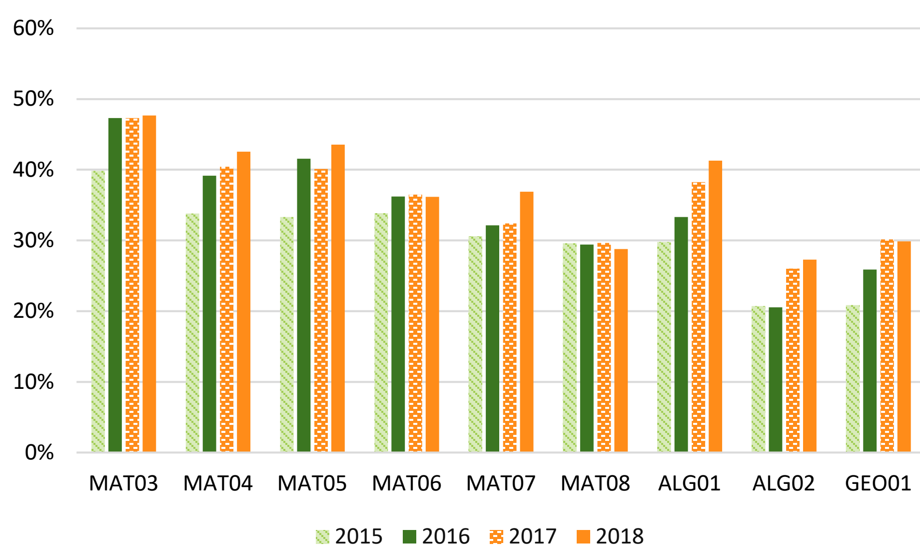

Figure 2. Student proficiency in Math by grade and year.

Trends in the overall percent proficient provide insights into population level changes between grade levels and years. Changes in students' proficiency status (i.e., proficient vs. not) between years gives deeper insight into the research question one: are students' who are proficient in the previous grade level more likely to be proficient in the current grade level and vice versa? Specifically, the model accuracy of the prior year proficiency status in predicting current year proficiency is used to evaluate this question. Model accuracy is defined as the number of students who are proficient at both grade levels plus the number of students who are not proficient at both grade levels quantity divided by the total number of students taking the focal grade levels during the focal years. Higher accuracy means that more students are correctly predicted by proficiency status in the prior year. Figures 3-8 examine model accuracy through the proportion of students who have a change in proficiency status. Patterned light green and solid orange indicate no change in status (adjacent proficient or not proficient performance, respectively); while solid dark green and patterned orange indicate a change in status. Numerically, ELA grade 10 to 11 have the lowest accuracy for all pairs of years with between 20-24% of students changing proficiency across years; whereas, the highest model accuracy for ELA is in grade 7 to 8 for all pairs of years with between 15-17% of students changing proficiency. For Math, accuracy is lowest for Algebra 1 to 2 with between 19-20% of student changing proficiency and highest for Geometry to Algebra 2 with between 11-13% of student changing proficiency for all pairs of years. Overall, less than 24% of students in a given grade level comparison have a prior year proficiency status that is incongruent with their current year proficiency status (see Appendix B).

This evaluation supports a positive relationship between student performance at adjacent grade levels for each pair of adjacent grades. To provide an overall measure of the relationship among student performance and to model average student trends, HLM analysis is conducted next.

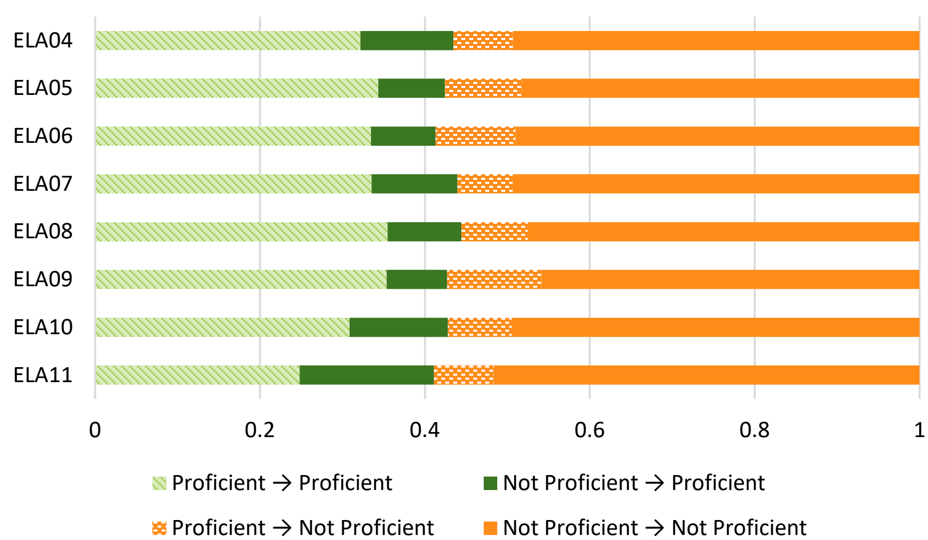

Figure 3. Change in student proficiency in ELA between 2015 and 2016.

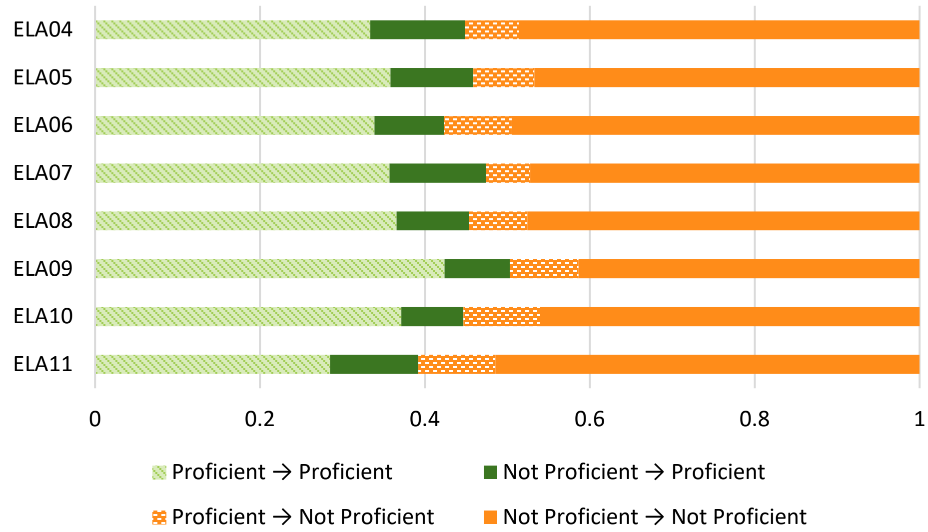

Figure 4. Change in student proficiency in ELA between 2016 and 2017.

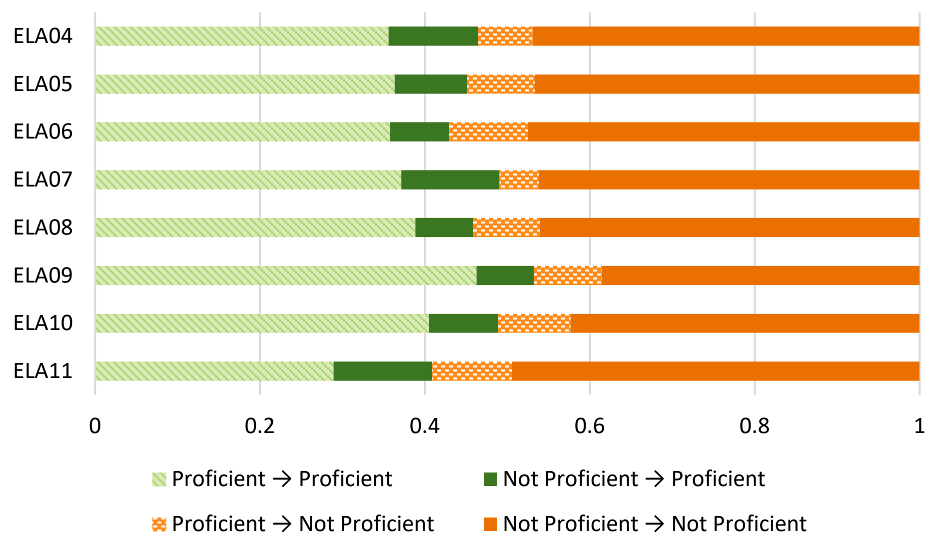

Figure 5. Change in student proficiency in ELA between 2017 and 2018.

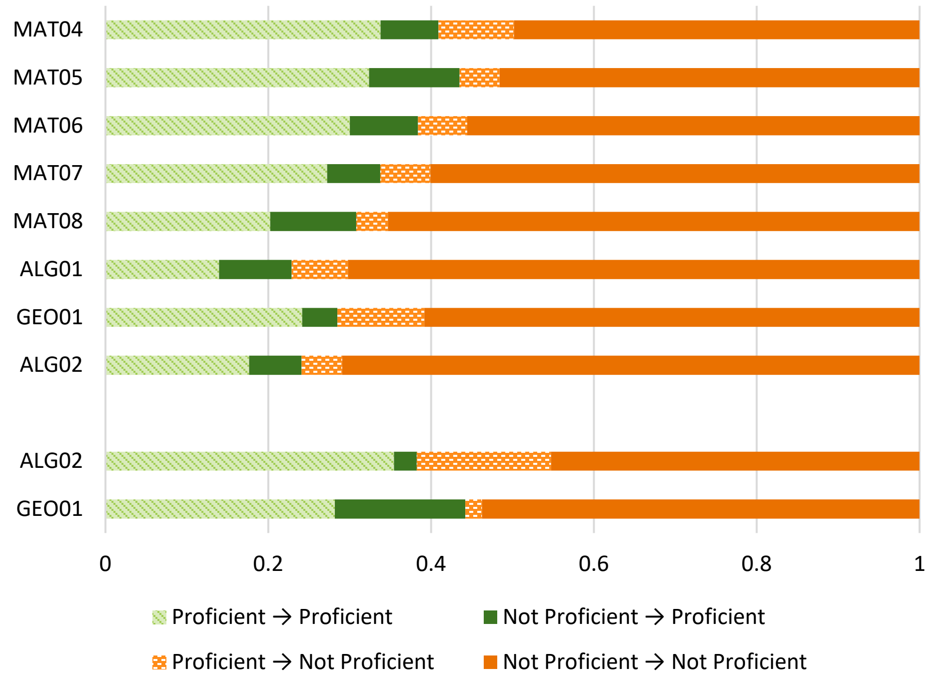

Figure 6. Change in student proficiency in Math between 2015 and 2016.

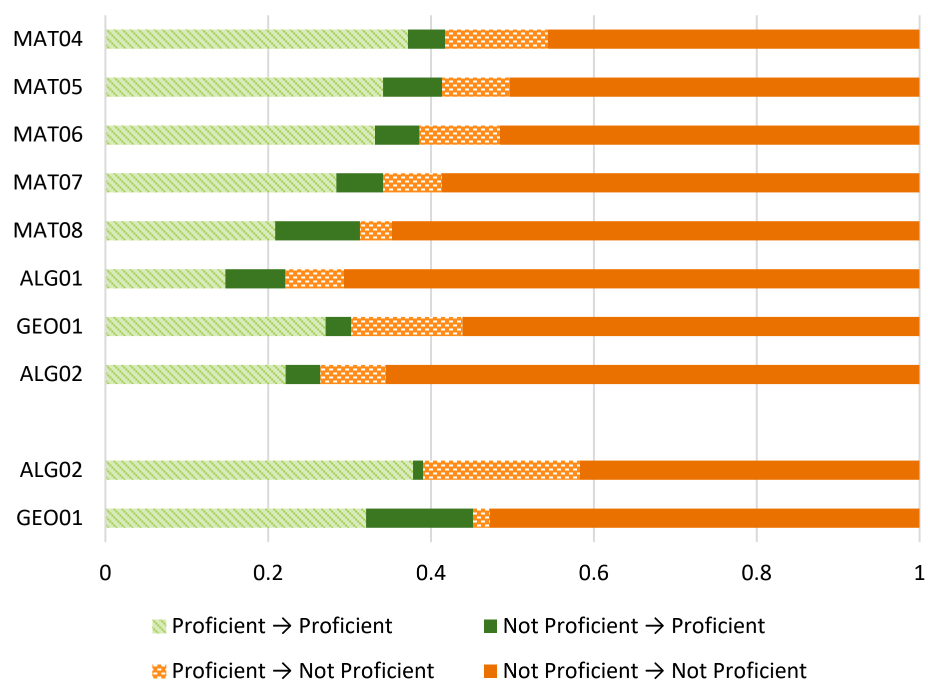

Figure 7. Change in student proficiency in Math between 2016 and 2017.

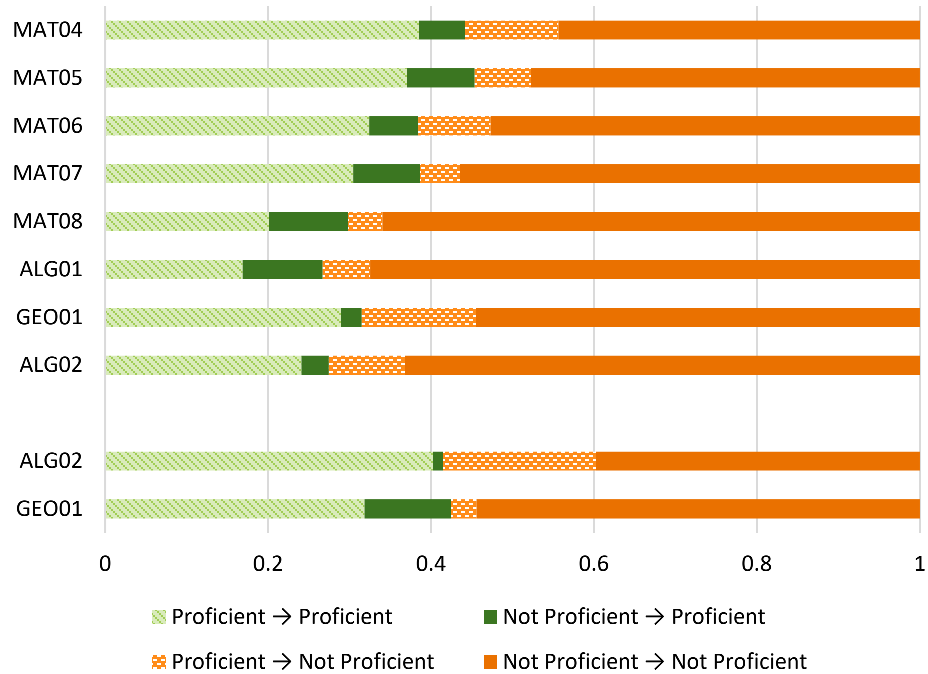

Figure 8. Change in student proficiency in Math between 2017 and 2018.

HLM was conducted for all students with three or more grade levels available in the data. This resulted in a sample size of 1,876,082 students (with over 6.6 million observations) for ELA and 1,379,093 students (with over 4.8 million observations) for Math. All model estimates are significant, which is likely due to the large sample size (Sullivan & Feinn, 2012). First, a null model — intercept only model — was fitted to estimate the ICC for ELA performance level and Math performance level across grade levels. The ICC estimates were 0.75 for both subjects indicating that 75% of the variance can be attributed to differences between students (e.g., student growth and achievement). This is robust evidence that performance levels are related across grade levels. In other words, student performance level at one grade level is highly related to performance level at the next grade level.

Next, two models were fit to the data representing two possible trends. The first model represents the case where student achievement builds slowly and systematically over grade levels. This case assumes that each grade level increase will result in a similar change to student performance. For a given student, this would be exemplified by a child who just meets expectations in grade 3 and slowly improves to exceeding expectations in high school, where the students score might be steadily increasing by several points each grade. On the average, students in the sample may be just below meets expectations in grade 3 and decrease to just above approaching expectations in high school. This average trend represents the prediction for the average student. For this model, student performance levels from grades 3 to high school are the outcome and the predictor is a linear trend modeled using a “dummy code” (i.e., 0-1-2-3-4-5-6-7-8). In contrast, students may experience sporadic fluctuations in achievement. For example, a student may be motivated in grade 4 and distracted by social pursuits in grade 5 then engaged in learning again in grade 6. The second model allows these grade level differences to drive the model prediction such that any grade level transitions that are systematically large will appear in the model results. Specifically, the elementary to middle and middle to high school transitions may be systematic resulting in sharp changes to the model predictions at those points. For this model, student performance level is predicted using eight trends (dummy codes) representing the transition for each pair of grade levels. In these dummy codes, the point at which the trend changes from zero to one is the transition that the trend is estimating:

Grade 3 to 4:	0-1-1-1-1-1-1-1-1
Grade 4 to 5:	0-0-1-1-1-1-1-1-1
Grade 5 to 6:	0-0-0-1-1-1-1-1-1
Grade 6 to 7:	0-0-0-0-1-1-1-1-1
Grade 7 to 8:	0-0-0-0-0-1-1-1-1
Grade 8 to 9:	0-0-0-0-0-0-1-1-1
Grade 9 to 10:	0-0-0-0-0-0-0-1-1
Grade 10 to 11:	0-0-0-0-0-0-0-0-1

These two models were used to evaluate whether changes between grade levels are consistent or sporadic. Consistent changes across grade levels (i.e., a linear trend) would suggest that grade level effects (e.g., content changes between grades) are small and systematic. Sporadic changes (i.e., a step function) would suggest that there are key transition points for students progressing through the subject matter. Linear trends were estimated and compared to models allowing each grade level increase to result in a performance level change (i.e., a step model). Results of the linear and step models are presented in Figure 9 for ELA and Figures 10 and 11 for Math. As the graphics demonstrate, the linear and step model trends overlap. For ELA, grades 4 and 7 appear to have larger increases (beta = 0.24 and 0.10) than other grade levels followed by subsequent drops (beta = -0.03 and -0.05). However, these changes are highly similar to the overall linear trend from grade 3 to 11, beta = 0.04 (see Appendix C). The increases are less than a full performance level suggesting the practical significance is small (i.e., the same performance level is expected for the typical student). Specifically, the average student is expected to increase less than a full performance level (approximately a third of a performance level) as they progress from ELA grade 3 to 11 suggesting that the average student would not change performance level.

For Math, grade 4 has a minor drop below the linear trend (beta = -0.17) and Algebra 2 has a major drop for both course-taking patterns (beta = -0.45 and -0.59); Algebra 1 has a small increase over the linear trend (beta = 0.15). Except for Algebra 2, the step pattern is consistent with the linear trend, beta = -0.04 and -0.03 (see Appendix C). In general, students are expected to drop less than a full performance level (approximately one third to one fifth of a performance level) as they progress from grade 3 to Algebra 2 or Geometry, which would result in no change in performance level for the average student. The results from the step model for Algebra 2 suggest that moving into Algebra 2 is a major transition point for students progressing from both Algebra 1 and Geometry.

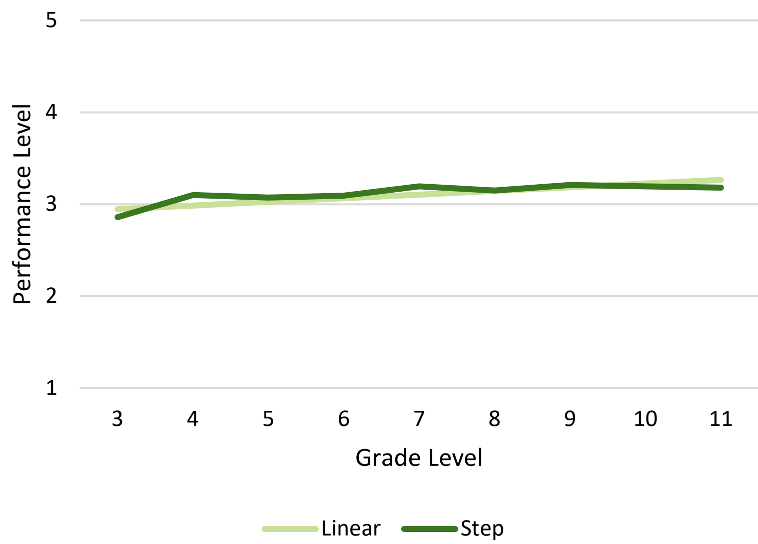

Figure 9. Comparison of predicted ELA performance level for linear and step models.

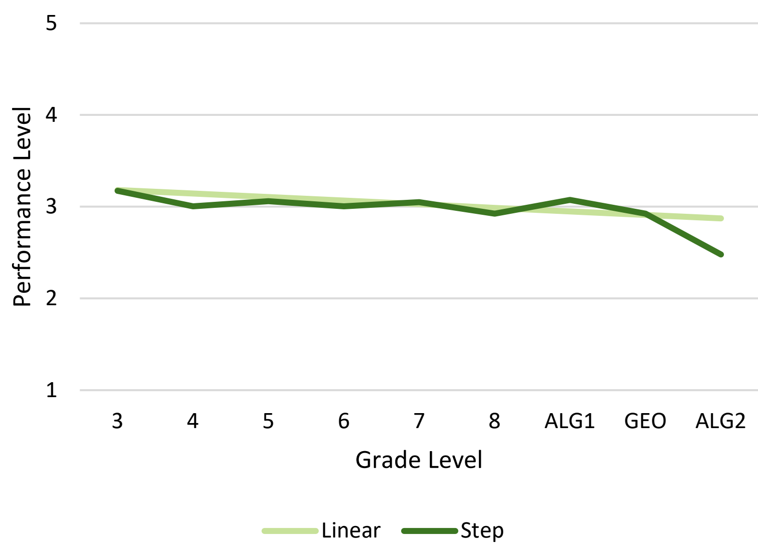

Figure 10. Comparison of predicted Math performance level for linear and step models with Algebra 2 as the terminal grade.

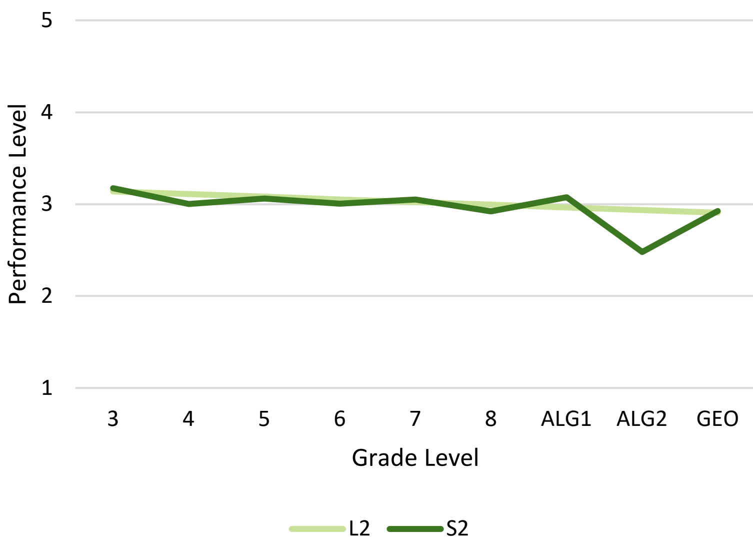

Figure 11. Comparison of predicted Math performance level for linear and step models with Geometry as the terminal grade.

Given the similarities between the linear and step models, performance level changes can be considered consistent for the average student. A further evaluation of research question two involves estimating the trend for specific groups of students. The consistent (linear) trend for the average student is positive 0.4 for ELA and negative 0.4 for Math where Algebra 2 is the terminal assessment.[3] Similar estimates can be derived for any group of students, and the trends can be plotted for comparison. In particular, an indicator of grade level and a dichotomous indicator representing group membership can be used with an interaction term to estimate the difference in performance level trends for each group in a common model. A comparison of the trends by student group are presented in Figures 12-16 for ELA and 17-21 for Math for sex, race, English language learner (ELL) status, socioeconomic status (SES), and disability status. The typical comparison group (i.e., male, White, non-ELL, students not categorized as low SES, and students not categorized as having a disability) is listed first and always colored light green for ease of comparison. As an example, the female group outperforms the male group in ELA by a small margin in grade 3 (nearly a fifth of a performance level) and the margin grows to a larger margin in grade 11 (over half of a performance level). Differences between the comparison group and another group are consistently less than a full performance level, except for American Indian/Alaska Native students compared to White students in ELA grades 3-6 and ELL student compared to non-ELL students in ELA grades 3-5. Also note that both the comparison group and the other student group are not proficient on average. The one exception is the Asian subgroups in ELA grades 9-11, which have an average performance level greater than four.

[3] Analyses were also conducted for Geometry as the terminal assessment. All conclusions were the same, and the results were similar enough to warrant the elimination of the second set of figures.

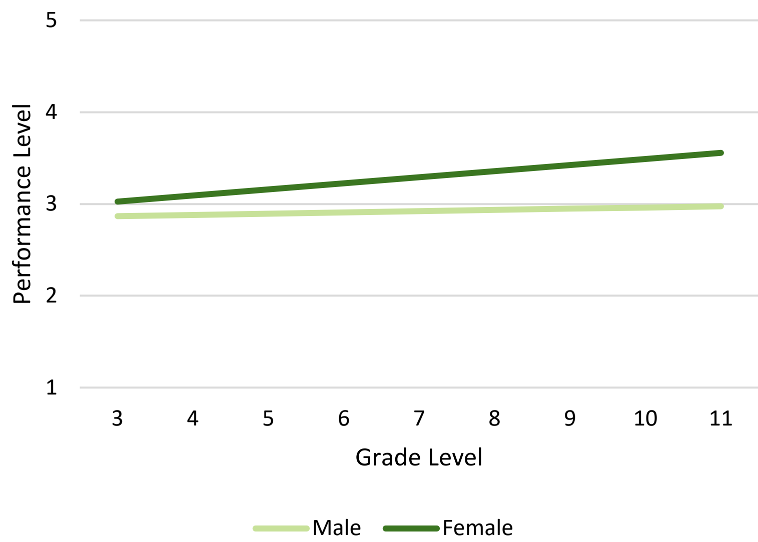

Figure 12. Comparison of linear predicted ELA performance level trends for males and females.

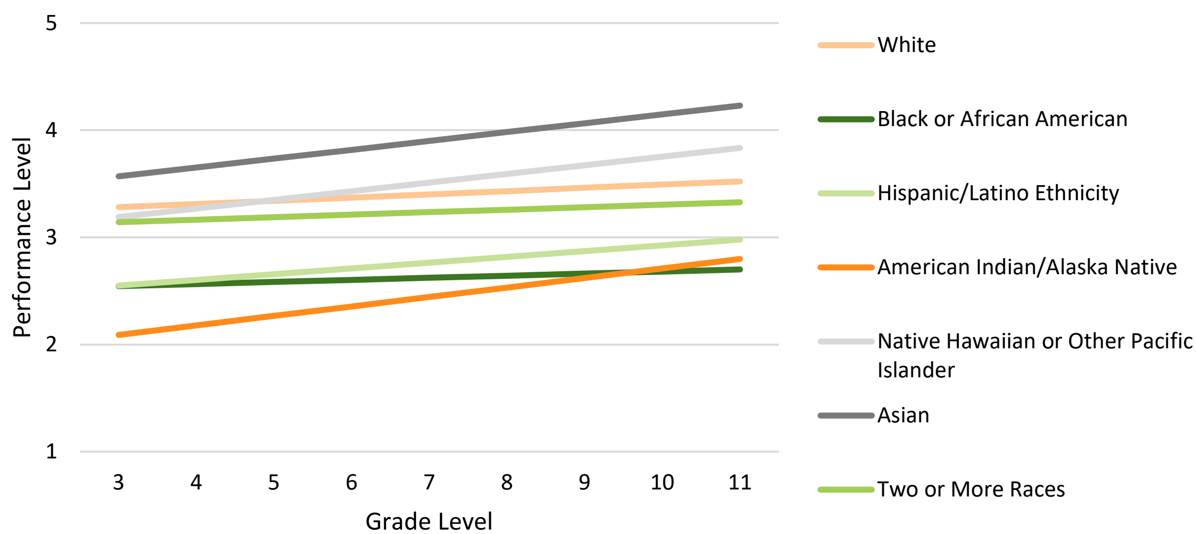

Figure 13. Comparison of linear predicted ELA performance level trends for race categories.

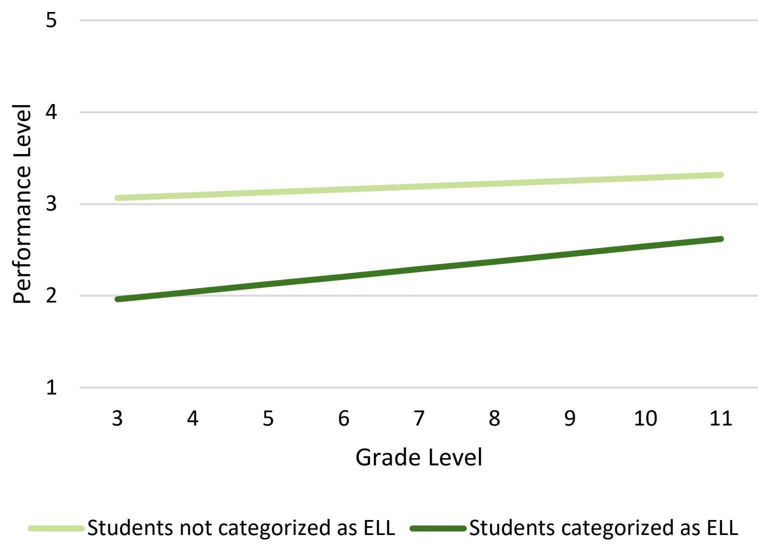

Figure 14. Comparison of linear predicted ELA performance level trends for English language learners (ELL) and non-ELL.

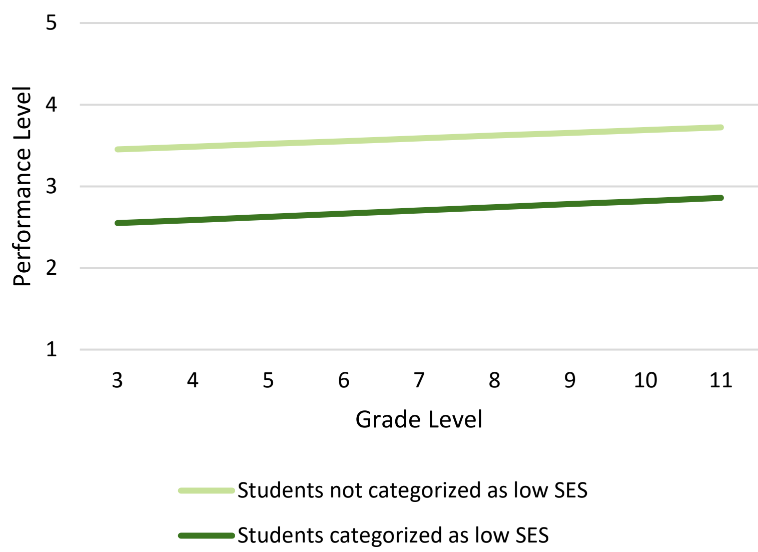

Figure 15. Comparison of linear predicted ELA performance level trends for students categorized as low socio-economic status (SES) and those not categorized as low SES.

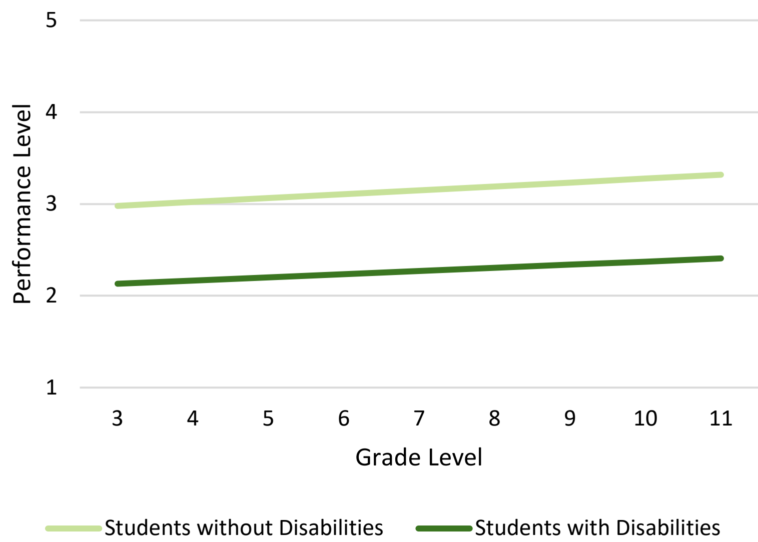

Figure 16. Comparison of linear predicted ELA performance level trends for students categorized as having a disability and those not categorized as having a disability.

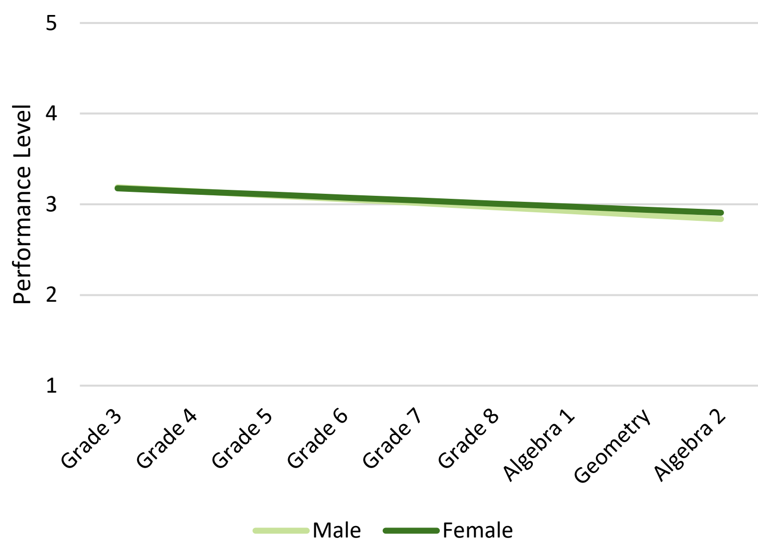

Figure 17. Comparison of linear predicted Math performance level trends for males and females.

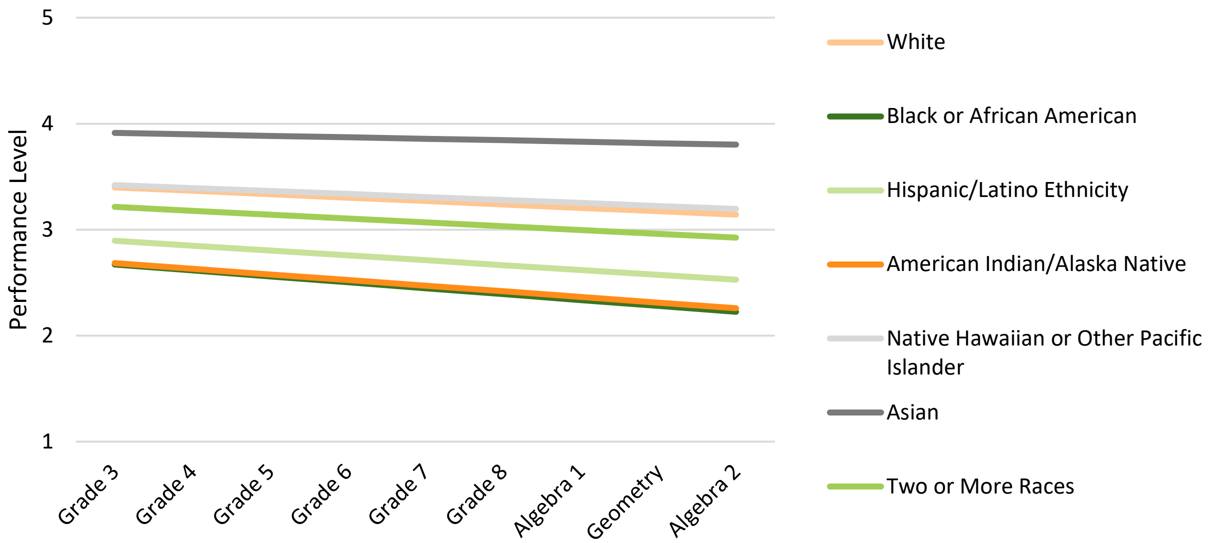

Figure 18. Comparison of linear predicted Math performance level trends for race categories.

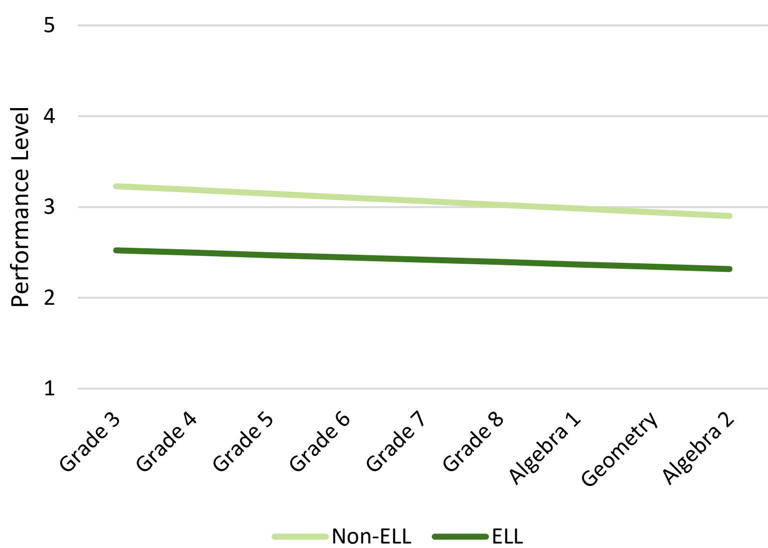

Figure 19. Comparison of linear predicted Math performance level trends for English language learners (ELL) and non-ELL.

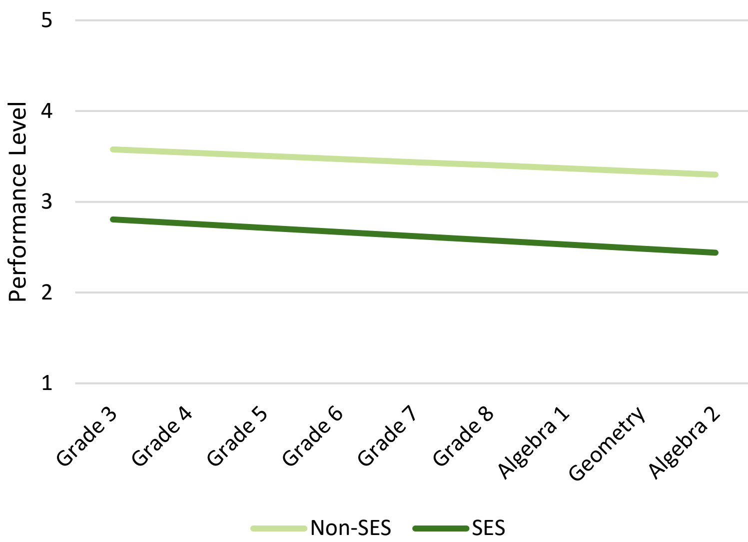

Figure 20. Comparison of linear predicted Math performance level trends for students categorized as low socio-economic status (SES) and those not categorized as low SES.

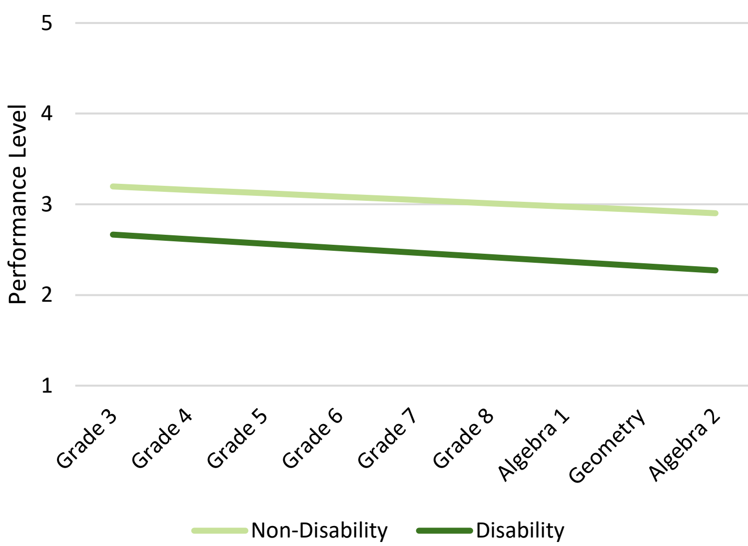

Figure 21. Comparison of linear predicted Math performance level trends for students categorized as having a disability and those not categorized as having a disability.

College Readiness

The grade-level readiness research questions were posed to determine if PARCC assessments are providing information on whether students are "on track" to college readiness in pre-high school assessments. The capstone to grade-level readiness is college readiness in high school. Specifically, students may be on track to proficiency (i.e., performance levels 4 and 5) in high school, but high school tests may not indicate college readiness. Previous studies analyzed students' PARCC scores to determine the PARCC score cut associated with likely college readiness according to external tests (e.g., PSAT and ACT; Steedle, Quesen, & Boyd, 2017). The current study used these estimates to compare the percent of students deemed college-ready or proficient according to the PARCC performance levels (with levels 4 and 5 being considered ready) to those deemed ready by the external tests according to the associated PARCC score. For example, the PARCC grade 9 ELA score of 747.9 was estimated to be equivalent to the PSAT 10 EBRW college-ready benchmark of 430. Students whose PARCC ELA 09 scale scores were at or above 747.9 were considered college ready according to the PSAT 10 EBRW criterion, and students whose scores fell below 747.9 were not considered college ready. For each external test, the percentage of students scoring at or above the estimated score associated with a specific external benchmark (i.e., college ready based on the external benchmark) was compared to the percentage of students scoring at or above Level 4 on the corresponding PARCC assessment (i.e., college ready based on the PARCC benchmark). Comparisons between the PARCC ELA and Math assessments and their corresponding external tests are provided in Figure 22 and 23, respectively. Except for ACT English and Reading, the PARCC proficiency cut is higher than the external benchmark equivalent, which results in a smaller percentage of students being considered college ready according to the PARCC proficiency cut.

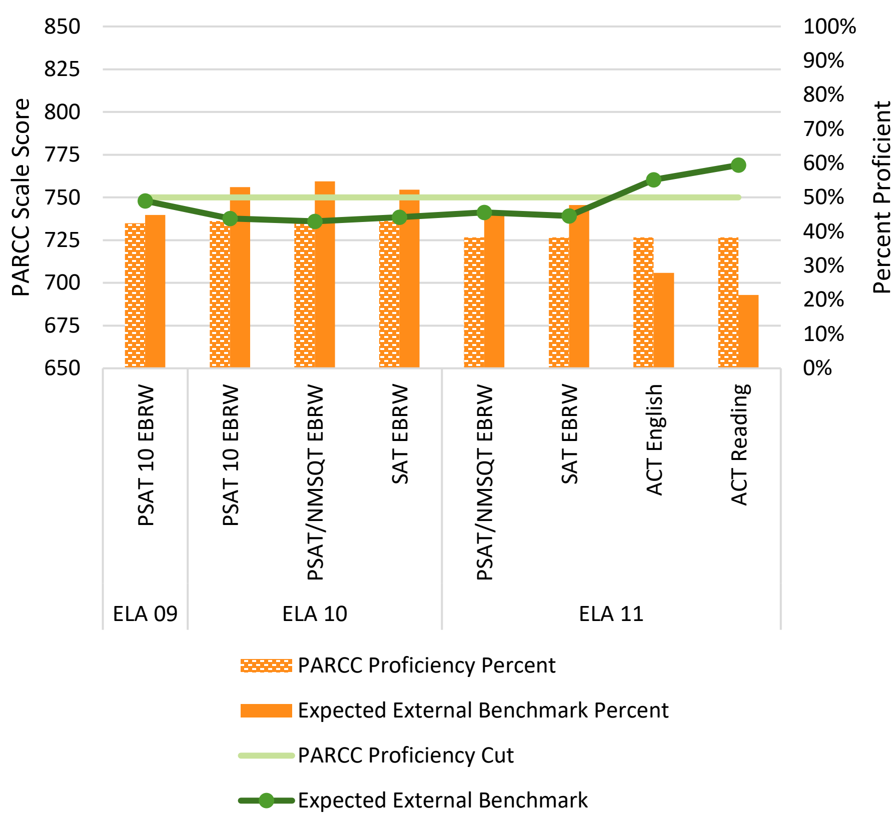

Figure 22. Percent college ready according to PARCC and external benchmarks for ELA.

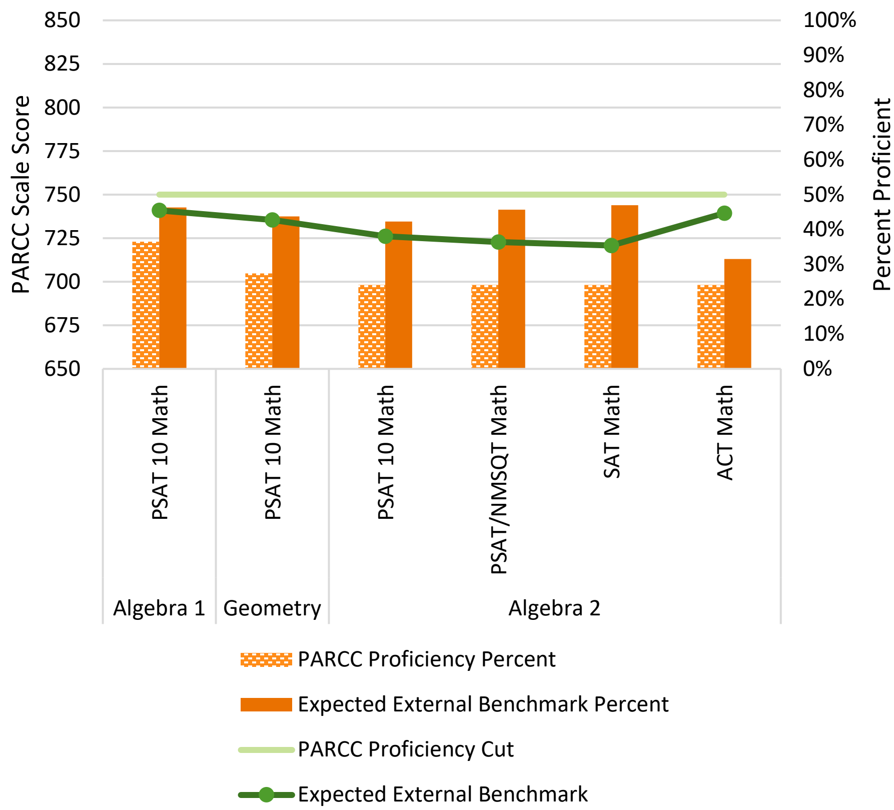

Figure 23. Percent college ready according to PARCC and external benchmarks for Math.

Summary and Conclusions

The current study examined PARCC assessment proficiency trends across grade levels for ELA and Math and estimated the relationship between grade levels to draw inferences about grade-level readiness. Additionally, college readiness was evaluated in a comparison of PARCC proficiency to external benchmarks (e.g., ACT, PSAT, SAT). Analysis focused on direct comparison of percentages and the modeling of grade level readiness through HLM. PARCC proficiency (i.e., attaining PARCC Level 4 or 5 vs. not) appears to indicate that a student is "on track" for the next grade level from grades 3 through high school, and the high school proficiency cut (i.e., 750) appears to be sufficiently rigorous to expect that proficient students will be college ready.

Overall, student performance on the PARCC assessments has a slight upward trend from grades 3 to 11 for ELA and has a slight downward trend from grades 3 to Geometry and Algebra 2 for Math that is not likely to result in a change to performance level for the typical student. Between grade levels, less than 24% of students are expected to change proficiency status between grade levels across ELA and Math. At the minimum, as few as 11% are expected to change, when transitioning from Geometry to Algebra 2. As such, the likelihood that a proficient student will remain proficient in the next grade level is high because student achievement compared to standards tends to be consistent across years. Performance levels across all grades are highly related, where student growth and achievement are the driving forces for both ELA and Math (i.e., ICC = 0.75). Altogether, the pattern of results supports the grade-level readiness research questions for the PARCC elementary, middle, and high school assessments. Additionally, the findings for grade-level readiness apply equally to all student groups (i.e., sex, race, ELL, SES, and disability groups).

These conclusions are limited because vertical linking is not conducted for the PARCC assessment scores. Changes in the overall difficulty of the tests between adjacent grade levels could have influenced the results. Specifically, the slight upward trend for ELA and the slight downward trend in Math performance level from grade 3 to high school could be due to changes in the difficulty of the tests. Additionally, the sporadic differences, though small, from the linear trend in proficiency level change from grade 3 to high school are more likely due to changes in curriculum content or test difficulty than drastic changes in student performance. Algebra 2 seems to require more attention than other grades given the relatively dramatic decrease in student performance. The current study cannot eliminate these alternative explanations; however, the general finding of consistent changes from grade 3 to high school implies that the impact of alternative explanations would be small at worst.

Using the work of Steedle, Quesen, and Boyd (2017), an expected external score cut was applied to the students in the PARCC population. PARCC proficiency was a more rigorous benchmark for college readiness compared to external tests, except for the ELA grade 11 expected cuts compared to the ACT English and Reading benchmarks. As such, students who are PARCC proficient have a high likelihood of being academically prepared for college. Further investigation could compare PARCC scores directly to college course success to establish true PARCC college readiness benchmarks.

In summary, data from four consecutive years (i.e., 2015-2018) of PARCC assessment constituting nearly 4 million students in ELA and nearly 3 million students in Math was used to establish the nature of the trend in student scores from grade 3 to high school. Findings imply that PARCC proficiency (i.e., attaining performance level 4 or 5) at each grade level can be used as an indicator of whether a student is “on track” to college readiness after high school.

References

Auger, R. W., Blackhurst, A. E., & Wahl, K. H. (2005). The development of elementary-aged children's career aspirations and expectations. *Professional School Counseling, 8*(4) 322-329.

Bryk, A., & Raudenbush, W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Ghisletta, P., Renaud, O., Jacot, N., & Courvoisier, D. (2015). Linear mixed-effects and latent curve models for longitudinal life course analysis. In C. Burton-Jeangros, S. Cullati, A. Sacker, et al. (Eds.), *A life course perspective on heath trajectories and transitions.* Cham, Switzerland: Springer. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK385371/ doi: 10.1007/978-3-319-20484-0_8

Knight, J. L. (2015). Preparing elementary school counselors to promote career development: Recommendations for school counselor education programs. *Journal of Career Development, 42*(2), 75-85. doi:10.1177/0894845314533745

Lee, J. (2012). College for all: Gaps between desirable and actual P–12 math achievement trajectories for college readiness. *Educational Researcher, 41*(2), 43-55.

Magnuson, C. & Starr, M. (2000). How early is too early to begin life career planning? The importance of the elementary school years. *Journal of Career Development, 27*(2).

National Council of Teachers of Mathematics. (2014). *What does research tell us about fostering algebraic thinking in arithmetic?* Reston, VA: The National Council of Teachers of Mathematics. Retrieved from: https://www.nctm.org/uploadedFiles/Research_and_Advocacy/research_brief_and_clips/NCTM%20Research%20Clip%20-%20Algebraic%20Thinking%20in%20School%20Arithmetic.pdf

Pearson. (2018). *PARCC: Final Technical Report for 2017 Administration.* Retrieved from: https://www.isbe.net/Documents/PARCC-2017-Technical-Report.pdf

Rudner, L. M. (2005). Expected classification accuracy. *Practical Assessment Research & Evaluation, 10*(13), 1-4.

Steedle, J., Quesen, S., & Boyd, A. (2017). Longitudinal study of external validity of the PARCC performance levels: Phase 1 report. Retrieved from: https://parcc-assessment.org/wp-content/uploads/2018/07/PARCC-PLS-Longitudinal-Validation-Study-Report-to-PARCC-v3.0.pdf

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279-282.

West, M., & Schwerdt, G. (2012). The middle school plunge: Achievement tumbles when young students change schools. *Education Next*, *12*(2), 62-68.

Appendix A

Table A1. Sample Size for Correlations Between the 2015 and 2016 Scores

2015 Subject	Grade	2016 Subject: Math 03	Math 04	Math 05	Math 06	Math 07	Math 08	Algebra 01	Algebra 02	Geometry 01
Math	03	1,925	**230,322**	82	2	-	-	-	-	-
	04	-	411	**210,472**	87	5	-	5	-	-
	05	-	1	345	**216,571**	270	87	129	-	1
	06	-	-	-	637	**213,152**	2,561	7,885	8	18
	07	-	-	-	2	758	**173,145**	41,371	31	531
	08	-	-	-	1	1	715	**101,153**	1,937	3,751
Algebra	01	-	-	-	3	22	1,740	8,980	**8,578**	**75,343**
	02	-	-	-	-	-	8	129	2,838	**2,723**
Geometry	01	-	-	-	-	-	5	594	**56,183**	1,616
ELA	03	1,997	243,787	111	3	-	-	-	-	-
	04	-	476	223,520	151	5	-	6	-	-
	05	-	1	344	232,588	300	166	130	-	4
	06	-	-	-	713	223,473	2,740	8,184	55	76
	07	-	-	-	2	823	183,165	43,337	347	5,952
	08	-	-	-	1	3	771	114,502	7,298	24,279
	09	-	-	-	1	-	35	7,435	21,885	51,602
	10	-	-	-	-	-	11	1,513	46,483	7,948
	11	-	-	-	-	-	-	186	4,133	514

(continued)

Table A1. Sample Size for Correlations Between the 2015 and 2016 Scores (continued)

2015 Subject	Grade	2016 Subject								
		ELA								
		03	04	05	06	07	08	09	10	11
Math	03	2,087	<u>252,748</u>	95	3	-	-	-	-	-
	04	1	488	<u>250,730</u>	97	3	-	-	-	-
	05	1	2	456	<u>252,239</u>	83	7	-	-	-
	06	-	-	-	767	<u>235,158</u>	199	6	-	-
	07	-	-	-	2	1,091	<u>225,686</u>	238	3	-
	08	-	-	-	1	97	904	<u>113,437</u>	752	69
Algebra	01	-	-	-	6	125	7,233	26,664	<u>74,043</u>	5,850
	02	-	-	-	-	1	35	383	8,779	<u>24,250</u>
Geometry	01	-	-	-	-	4	42	2,897	21,219	39,793
ELA	03	2,327	**300,243**	110	3	2	-	-	1	-
	04	1	599	**298,572**	114	4	-	-	-	-
	05	1	1	481	**297,274**	106	11	-	-	-
	06	-	-	-	1,562	**294,405**	311	16	-	1
	07	-	-	-	2	981	**290,281**	414	5	-
	08	-	-	-	1	37	704	**176,450**	578	13
	09	-	-	-	1	-	-	2,174	**95,374**	881
	10	-	-	-	-	-	-	133	1,651	**82,770**
	11	-	-	-	-	-	-	77	178	1,400

Note. Correlations are only reported in instances where at least 50 students took both tests. Estimates of convergent validity are bolded and estimates of discriminant validity are underlined.

Table A2. Sample Size for Correlations Between the 2016 and 2017 Scores

2016 Subject	Grade	2017 Subject								
		Math						Algebra		Geometry
		03	04	05	06	07	08	01	02	01
Math	03	1,717	**237,128**	59	1	-	-	1	-	-
	04	4	399	**214,664**	60	3	-	4	-	-
	05	-	-	340	**215,938**	168	38	168	-	2
	06	-	-	1	650	**221,115**	170	8,072	4	35
	07	-	-	-	-	719	**178,921**	42,412	19	371
	08	-	-	-	-	18	1,462	**74,621**	220	1,221
Algebra	01	-	-	-	-	18	1,963	7,795	**10,579**	**87,350**
	02	-	-	-	-	-	16	159	2,480	**2,770**
Geometry	01	-	-	-	-	-	6	816	**68,303**	1,862
ELA	03	1,813	266,311	69	3	-	-	1	-	-
	04	4	418	240,954	77	3	-	6	-	-
	05	-	-	321	245,585	203	37	193	-	8
	06	-	-	-	659	241,390	243	10,055	41	124
	07	-	-	1	1	751	191,897	51,636	566	7,821
	08	-	-	-	-	2	492	86,982	8,462	26,938
	09	-	-	-	-	-	42	5,948	24,292	56,091
	10	-	-	-	-	-	19	1,803	54,469	7,426
	11	-	-	-	-	-	2	264	4,708	801

(continued)

Table A2. Sample Size for Correlations Between the 2016 and 2017 Scores (continued)

2016 Subject	Grade	2017 Subject								
		ELA								
		03	04	05	06	07	08	09	10	11
Math	03	1,814	<u>249,882</u>	89	-	-	-	-	-	-
	04	2	445	<u>247,337</u>	69	2	-	-	-	-
	05	-	-	384	<u>224,826</u>	69	6	-	-	-
	06	-	-	1	706	<u>237,640</u>	138	1	-	-
	07	-	-	-	-	970	<u>229,038</u>	150	5	1
	08	-	-	-	-	170	2,976	<u>56,142</u>	686	26
Algebra	01	-	-	-	6	132	8,189	29,701	<u>81,450</u>	6,261
	02	-	-	-	-	-	53	321	6,520	<u>19,856</u>
Geometry	01	-	-	-	-	4	76	3,255	22,111	46,797
ELA	03	2,076	**311,866**	108	2	-	-	-	-	-
	04	2	537	**305,078**	91	4	-	-	-	-
	05	-	-	434	**300,965**	110	5	-	-	-
	06	-	-	-	1,406	**303,607**	326	2	-	-
	07	-	-	1	-	900	**301,919**	201	6	1
	08	-	-	-	-	3	572	**103,788**	690	6
	09	-	-	-	-	-	-	1,853	**100,273**	674
	10	-	-	1	-	-	-	229	1,953	**93,903**
	11	-	-	-	-	-	1	80	286	2,385

Note. Correlations are only reported in instances where at least 50 students took both tests. Estimates of convergent validity are bolded and estimates of discriminant validity are underlined.

Table A3. Sample Size for Correlations Between the 2017 and 2018 Scores

2017 Subject	Grade	2018 Subject								
		Math						Algebra		Geometry
		03	04	05	06	07	08	01	02	01
Math	03	1,644	**247,991**	68	2	2	-	-	-	-
	04	1	410	**233,364**	64	4	-	3	-	-
	05	-	-	287	**232,332**	184	6	263	-	3
	06	1	1	-	569	**230,283**	137	10,856	-	19
	07	-	-	-	2	645	**185,618**	52,958	4	241
	08	-	-	-	1	2	427	**77,644**	1,806	1,363
Algebra	01	-	-	-	-	6	1,667	13,697	**11,483**	**94,713**
	02	-	-	-	-	-	7	201	2,059	**3,015**
Geometry	01	-	-	-	-	-	8	792	**72,955**	1,799
ELA	03	1,724	260,059	78	5	3	-	-	-	-
	04	2	447	244,414	75	5	-	3	-	1
	05	-	-	305	247,648	213	6	264	-	9
	06	1	1	-	603	239,130	186	11,036	43	157
	07	-	-	-	1	669	193,330	54,245	626	8,025
	08	-	-	-	1	1	421	88,840	10,203	29,520
	09	-	-	-	-	-	12	5,795	28,921	57,903
	10	-	-	-	-	-	18	2,099	53,337	7,665
	11	-	-	-	-	-	-	262	3,923	681

(continued)

Table A3. Sample Size for Correlations Between the 2017 and 2018 Scores (continued)

2017 Subject	Grade	2018 Subject								
		ELA								
		03	04	05	06	07	08	09	10	11
Math	03	1,770	260,378	93	1	2	-	-	-	-
	04	3	457	266,725	82	4	-	-	-	-
	05	-	1	333	239,914	58	4	-	-	-
	06	1	1	1	615	247,641	137	2	-	-
	07	-	-	-	2	780	244,313	98	3	-
	08	-	-	-	1	32	551	58,737	462	28
Algebra	01	-	-	1	4	182	10,401	31,862	90,350	5,954
	02	-	-	-	-	-	39	329	8,620	19,887
Geometry	01	-	-	-	-	8	119	3,292	28,604	47,713
ELA	03	2,047	**309,870**	102	8	2	-	-	-	-
	04	3	608	**313,408**	103	5	-	-	-	-
	05	-	-	442	**305,724**	94	6	-	-	-
	06	1	1	-	1,118	**303,118**	253	1	-	-
	07	-	-	-	1	800	**307,158**	124	3	-
	08	-	-	-	1	1	551	**108,474**	859	11
	09	-	-	-	-	-	-	1,777	**109,304**	903
	10	-	-	-	-	-	-	161	6,520	**91,305**
	11	-	-	-	-	-	-	73	301	1,148

Note. Correlations are only reported in instances where at least 50 students took both tests. Estimates of convergent validity are bolded and estimates of discriminant validity are underlined.

Appendix B

Table B1. Percent of ELA Students Proficient/Not Proficient in 2016 Predicted Using Proficient/Not Proficient in 2015

Grade	Proficient (2016)		Not Proficient (2016)		Accuracy
	Proficient	Not Proficient	Proficient	Not Proficient	
04	0.32	0.11	0.07	0.49	0.81
05	0.34	0.08	0.09	0.48	0.83
06	0.33	0.08	0.10	0.49	0.82
07	0.34	0.10	0.07	0.49	0.83
08	0.35	0.09	0.08	0.48	0.83
09	0.35	0.07	0.11	0.46	0.81
10	0.31	0.12	0.08	0.49	0.80
11	0.25	0.16	0.07	0.52	0.76

Table B2. Percent of ELA Students Proficient/Not Proficient in 2017 Predicted Using Proficient/Not Proficient in 2016

Grade	Proficient (2017)		Not Proficient (2017)		Accuracy
	Proficient	Not Proficient	Proficient	Not Proficient	
04	0.33	0.12	0.07	0.49	0.82
05	0.36	0.10	0.07	0.47	0.83
06	0.34	0.08	0.08	0.49	0.83
07	0.36	0.12	0.05	0.47	0.83
08	0.37	0.09	0.07	0.48	0.84
09	0.42	0.08	0.08	0.41	0.84
10	0.37	0.08	0.09	0.46	0.83
11	0.29	0.11	0.09	0.51	0.80

Table B3. Percent of ELA Students Proficient/Not Proficient in 2018 Predicted Using Proficient/Not Proficient in 2017

Grade	Proficient (2018)		Not Proficient (2018)		Accuracy
	Proficient	Not Proficient	Proficient	Not Proficient	
04	0.36	0.11	0.07	0.47	0.83
05	0.36	0.09	0.08	0.47	0.83
06	0.36	0.07	0.10	0.48	0.83
07	0.37	0.12	0.05	0.46	0.83
08	0.39	0.07	0.08	0.46	0.85
09	0.46	0.07	0.08	0.39	0.85
10	0.40	0.08	0.09	0.42	0.83
11	0.29	0.12	0.10	0.49	0.78

Table B4. Percent of Math Students Proficient/Not Proficient in 2016 Predicted Using Proficient/Not Proficient in 2015

Grade	Proficient (2016)		Not Proficient (2016)		Accuracy
	Proficient	Not Proficient	Proficient	Not Proficient	
04	0.34	0.07	0.09	0.50	0.84
05	0.32	0.11	0.05	0.52	0.84
06	0.30	0.08	0.06	0.56	0.86
07	0.27	0.07	0.06	0.60	0.87
08	0.20	0.11	0.04	0.65	0.86
ALG01	0.14	0.09	0.07	0.70	0.84
GEO	0.24	0.04	0.11	0.61	0.85
ALG02	0.18	0.06	0.05	0.71	0.89
ALG02	0.35	0.03	0.16	0.45	0.81
GEO	0.28	0.16	0.02	0.54	0.82

Table B5. Percent of Math Students Proficient/Not Proficient in 2017 Predicted Using Proficient/Not Proficient in 2016

Grade	Proficient (2017)		Not Proficient (2017)		Accuracy
	Proficient	Not Proficient	Proficient	Not Proficient	
04	0.37	0.05	0.13	0.46	0.83
05	0.34	0.07	0.08	0.50	0.84
06	0.33	0.06	0.10	0.52	0.85
07	0.28	0.06	0.07	0.59	0.87
08	0.21	0.10	0.04	0.65	0.86
ALG01	0.15	0.07	0.07	0.71	0.85
GEO	0.27	0.03	0.14	0.56	0.83
ALG02	0.22	0.04	0.08	0.66	0.88
ALG02	0.38	0.01	0.19	0.42	0.80
GEO	0.32	0.13	0.02	0.53	0.85

Table B6. Percent of Math Students Proficient/Not Proficient in 2018 Predicted Using Proficient/Not Proficient in 2017

Grade	Proficient (2018)		Not Proficient (2018)		Accuracy
	Proficient	Not Proficient	Proficient	Not Proficient	
04	0.39	0.06	0.11	0.44	0.83
05	0.37	0.08	0.07	0.48	0.85
06	0.32	0.06	0.09	0.53	0.85
07	0.30	0.08	0.05	0.56	0.87
08	0.20	0.10	0.04	0.66	0.86
ALG01	0.17	0.10	0.06	0.67	0.84
GEO	0.29	0.03	0.14	0.55	0.83
ALG02	0.24	0.03	0.09	0.63	0.87
ALG02	0.40	0.01	0.19	0.40	0.80
GEO	0.32	0.11	0.03	0.54	0.86

Appendix C

Table C1. Fixed Effects Results for HLM predicting ELA Performance Level (Number of Observations = 6,626,480)

Model	Predictor	-2 log(*L*)	*β*	SE	*p*-value
Model 1		16,363,014			
	Intercept[a]		3.10440	0.000771	<.0001
Model 2		16,269,973			
	Intercept		2.94510	0.001071	<.0001
	Linear		0.03986	0.000208	<.0001
Model 3		16,271,687			
	Intercept		2.85840	0.001166	<.0001
	04		0.23990	0.001016	<.0001
	05		-0.02621	0.000842	<.0001
	06		0.02036	0.000805	<.0001
	07		0.10200	0.000822	<.0001
	08		-0.04776	0.000900	<.0001
	09		0.06176	0.001290	<.0001
	10		-0.01660	0.001459	<.0001
	11		-0.01304	0.001908	<.0001

Note. -2 log(L) = negative two times the log likelihood. SE = standard error. For all models, the intercept was the only random effect. Restricted Maximum Likelihood (REML) estimation was used for all models.

[a] The estimate in the intercept-only model reflects the average performance level for all grades; whereas, the intercept estimate in the subsequent models reflects the average performance level at grade 3.

Table C2. Fixed Effects Results for HLM predicting Math Performance Level (Number of Observations = 4,801,671)

Model	Predictor	-2 log(*L*)	*β*	SE	*p*-value
Model 1		11,051,487			
	Intercept[a]		3.01222	0.000838	<.0001
Model 2a (ALG Terminal)		11,000,633			
	Intercept		3.18170	0.001164	<.0001
	Linear		-0.03860	0.000210	<.0001
Model 2b (GEO Terminal)		11,026,589			
	Intercept		3.13810	0.001171	<.0001
	Linear		-0.02890	0.000203	<.0001
Model 3a (ALG Terminal)		10,889,658			
	Intercept[b]		3.17350	0.001268	<.0001
	04[b]		-0.17000	0.001107	<.0001
	05[b]		0.05738	0.000910	<.0001
	06[b]		-0.05720	0.000856	<.0001
	07[b]		0.04745	0.000870	<.0001
	08[b]		-0.12810	0.001032	<.0001
	ALG01[b]		0.15080	0.001250	<.0001
	GEO[b]		-0.14920	0.001456	<.0001
	ALG02[b]		-0.44520	0.001829	<.0001
Model 3b (GEO Terminal)		10,889,658			
	ALG02		-0.59440	0.001713	<.0001
	GEO		0.44520	0.001829	<.0001

Note. -2 log(L) = negative two times the log likelihood. SE = standard error. For all models, the intercept was the only random effect. Restricted Maximum Likelihood (REML) estimation was used for all models.

[a] The estimate in the intercept-only model reflects the average performance level for all grades; whereas, the intercept estimate in the subsequent models reflects the average performance level at grade 3. b These step model estimates are the same between both course taking patterns when estimated in separate models, and as such the results are only presented once.

Table C3. Fixed Effects Results for HLM predicting ELA Performance Level by Demographic Variable

Demographic Variable (Number of Observations)	Predictor	-2 log(*L*)	*β*	SE	*p*-value
Sex (6,623,919)		16,194,784			
	Intercept		2.86700	0.001502	<.0001
	Linear		0.01346	0.000290	<.0001
	Female		0.15870	0.002131	<.0001
	Linear*Female		0.05290	0.000412	<.0001
Race (6,590,669)		15,883,866			
	Intercept		3.28210	0.001551	<.0001
	Linear		0.02997	0.000310	<.0001
	American Indian/Alaska Native		-1.19280	0.010440	<.0001
	Asian		0.28760	0.004258	<.0001
	Black or African American		-0.73670	0.002844	<.0001
	Hispanic/Latino Ethnicity		-0.73160	0.002502	<.0001
	Native Hawaiian or Other Pacific Islander		-0.09219	0.026690	<.0001
	Two or More Races		-0.13990	0.006412	<.0001
	Linear*American Indian/Alaska Native		0.05862	0.001915	<.0001
	Linear*Asian		0.05253	0.000837	<.0001
	Linear*Black or African American		-0.01056	0.000573	<.0001
	Linear*Hispanic/Latino Ethnicity		0.02361	0.000496	<.0001
	Linear*Native Hawaiian or Other Pacific Islander		0.05063	0.005296	<.0001
	Linear*Two or More Races		-0.00682	0.001364	<.0001

(continued)

Table C3. Fixed Effects Results for HLM predicting ELA Performance Level by Demographic Variable (continued)

Demographic Variable (Number of Observations)	Predictor	-2 log(*L*)	*β*	SE	*p*-value
ELL (6,623,927)		16,124,959			
	Intercept		3.06550	0.001104	<.0001
	Linear		0.03146	0.000217	<.0001
	ELL		-1.10370	0.003406	<.0001
	Linear*ELL		0.05070	0.000738	<.0001
SES (6,625,892)		15,907,148			
	Intercept		3.45240	0.001508	<.0001
	Linear		0.03372	0.000301	<.0001
	SES		-0.90260	0.002028	<.0001
	Linear*SES		0.00490	0.000409	<.0001
Disability (6,615,560)		16,181,903			
	Intercept		2.97870	0.001088	<.0001
	Linear		0.04244	0.000213	<.0001
	Disability		-0.84760	0.005223	<.0001
	Linear*Disability		-0.00795	0.000896	<.0001

Note. -2 log(L) = negative two times the log likelihood. SE = standard error. For all models, the intercept was the only random effect. Restricted Maximum Likelihood (REML) estimation was used for all models.

Table C4. Fixed Effects Results for HLM predicting Math Performance Level by Demographic Variable

Demographic Variable (Number of Observations)	Predictor	-2 log(*L*)	*β*	SE	*p*-value
Sex (4,800,281)		10,996,599			
	Intercept		3.18740	0.001653	<.0001
	Linear		-0.04368	0.000298	<.0001
	Female		-0.01110	0.002328	<.0001
	Linear*Female		0.01009	0.000420	<.0001
Race (4,777,311)		10,690,317			
	Intercept		3.39860	0.001581	<.0001
	Linear		-0.03212	0.000294	<.0001
	American Indian/Alaska Native		-0.71430	0.012140	<.0001
	Asian		0.51440	0.004318	<.0001
	Black or African American		-0.72760	0.003175	<.0001
	Hispanic/Latino Ethnicity		-0.50310	0.002821	<.0001
	Native Hawaiian or Other Pacific Islander		0.02192	0.028850	<.0001
	Two or More Races		-0.18330	0.006630	<.0001
	Linear*American Indian/Alaska Native		-0.02106	0.002046	<.0001
	Linear*Asian		0.01827	0.000764	<.0001
	Linear*Black or African American		-0.02362	0.000591	<.0001
	Linear*Hispanic/Latino Ethnicity		-0.01384	0.000514	<.0001
	Linear*Native Hawaiian or Other Pacific Islander		0.00398	0.005178	<.0001
	Linear*Two or More Races		-0.00419	0.001306	<.0001

(continued)

Table C4. Fixed Effects Results for HLM predicting Math Performance Level by Demographic Variable (continued)

Demographic Variable (Number of Observations)	Predictor	-2 log(*L*)	*β*	SE	*p*-value
ELL (4,801,671)		10,966,063			
	Intercept		3.22870	0.001193	<.0001
	Linear		-0.04084	0.000215	<.0001
	ELL		-0.70570	0.004753	<.0001
	Linear*ELL		0.01511	0.000949	<.0001
SES (4,801,659)		10,736,693			
	Intercept		3.57750	0.001547	<.0001
	Linear		-0.03479	0.000284	<.0001
	SES		-0.77160	0.002203	<.0001
	Linear*SES		-0.01102	0.000410	<.0001
Disability (4,791,352)		10,958,709			
	Intercept		3.19710	0.001181	<.0001
	Linear		-0.03692	0.000214	<.0001
	Disability		-0.53100	0.006707	<.0001
	Linear*Disability		-0.01245	0.001076	<.0001

Note. -2 log(L) = negative two times the log likelihood. SE = standard error. For all models, the intercept was the only random effect. Restricted Maximum Likelihood (REML) estimation was used for all models.