The reliability of shorter assessments in New Jersey for group-level inferences

Lindsay Fox, Jacob Hartog, and Natalie Larkin

July 2021

For New Jersey Department of Education

-

This memo was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract ED-IES-17-C-0006 by Regional Educational Laboratory Mid-Atlantic administered by Mathematica. The content does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

The reliability of shorter assessments in New Jersey for group-level inferences

Lindsay Fox, Jacob Hartog, and Natalie Larkin

July 2021

Education policymakers must balance the reliability of assessments to measure academic knowledge and skills with the burdens that assessments place upon students, teachers, and schools. In 2019, New Jersey began using the New Jersey Student Learning Assessments (NJSLA), shorter assessments based on the Partnership for Assessment of Readiness for College and Careers (PARCC). Regional Educational Laboratory researchers examined the reliability of test results for the NJSLA by comparing results at the school, test, and subgroup levels from 2016 to 2019. The findings indicated a high degree of reliability across most measures the researchers examined; during the transition to the NJSLA, the reliability did not decrease for any test results—except the Algebra 2 test—reported by the New Jersey Department of Education. The instability of the Algebra 2 results was most likely not attributable to changes in the assessment but instead to changes in the student population that was required to test following a change in the state's testing requirements.

Why this study?

The Every Student Succeeds Act allows states the flexibility to adjust statewide testing to suit the needs of students, parents, teachers, principals, and other stakeholders in education. In spring 2019, the New Jersey Department of Education (NJDOE) transitioned to a shorter version of the Partnership for Assessment of Readiness for College and Careers (PARCC) test called the New Jersey Student Learning Assessments (NJSLA). New Meridian worked with NJDOE to develop the NJSLA. NJDOE officials want to learn how using a shorter student assessment affects the reliability of school-level performance measures, including performance subscores such as reading and writing for English language arts (ELA). New Jersey is one of multiple states, including Illinois (Thayer, 2019), that previously administered the PARCC but transitioned to a shorter test to reduce burden for stakeholders while continuing to address the state standards.

Key terms

PARCC: The Partnership Assessment of Readiness for College and Careers is a standardized assessment that was administered by the New Jersey Department of Education until spring 2019.

NJSLA: The New Jersey Student Learning Assessments is the assessment that replaced the PARCC in New Jersey in spring 2019.

Major claim: The main two subscores of the English language arts portion of the assessment into which the items are split: reading and writing. Students receive a scale score for each.

Subclaim: A subset of items that test a specific skill set within a subject. Students receive one of the following grades: met or exceeded expectations, approached expectations, or did not yet meet or partially met expectations. Subclaims remained consistent between the PARCC and NJSLA.

Toward the goal of reducing burden for statewide stakeholders, students' total testing time on the NJSLA was 25 to 33 percent shorter than their time on the PARCC (table 1). NJSLA reduces overall testing time through two approaches: decreasing the number of test units (portions of the test administered at one time) and decreasing the number of items included in a test unit. The NJSLA ELA and math assessments measure the same constructs and use the same item banks as the PARCC ELA and math assessments. As a result, NJDOE can report results using the same performance level and scales scores as prior years (NJDOE, 2019a). Reliability in estimates across schools and over time has important benefits for New Jersey policymakers. Reliable estimates allow NJDOE to identify schools where performance of student subgroups has improved or worsened over time. For New Jersey school

districts, reliable comparisons can help identify grade levels, curricular areas, or groups of students that need additional support or are currently thriving.

The NJSLA retained all three ELA units from the PARCC, but each student only receives two of the three units. Students randomly receive one of two test blueprints, both of which include a research simulation task, and one of two other possible task units. Students in grade 3 receive either a literary analysis task or a narrative writing task and short passage task, and students in grade 4 up to high school receive either a literary analysis task and short passage set or a narrative writing task and a long passage or a set of paired passages. These blueprints were vetted by NJDOE Office of Assessments to ensure that "scores are comparable" and tests are "similar in content and difficulty" (NJDOE, 2019b) across the randomly assigned versions. The intent of random test assignment is to ensure students and teachers prepare for all three test units.

On math tests for grades 3 to 5 and high school, the NJSLA includes one fewer unit than the PARCC—dropping from four to three. For grades 6 to 8, the number of units (four) stayed the same, but the testing time of each unit dropped 25 percent by reducing the number of items within each unit.

Table 1. The NJSLA decreased testing time by over 25 percent

Test	Description of change to test time/units	Maximum test time (minutes)		Percent of time reduced
		PARCC	NJSLA	
Math				
Grades 3 to 5	Decreased units from four to three; unit test time unchanged	240	180	25
Grades 6 to 8	Retained four units; decreased unit test time by 20 minutes each	240	180	25
High school	Decreased units from four to three	270	180	33
ELA				
Grade 3	Decreased units from three to two; unit test time unchanged	225	150	33
Grade 4 through high school	Decreased units from three to two; unit test time unchanged	270	180	33

ELA is English language arts. NJSLA is New Jersey Student Learning Assessments. PARCC is Partnership for Assessment of Readiness for College and Careers.
Source: Authors' analysis of data from the New Jersey Department of Education.

To help NJDOE understand how the reliability of school-level aggregates of its test results might have been affected by a shorter assessment, Regional Educational Laboratory researchers collaborated with NJDOE staff to explore the consistency of scores between the PARCC and the NJSLA across students, tests, and subgroups. Regional Educational Laboratory researchers worked with NJDOE staff to obtain school-level test performance data from the spring administration of the PARCC in 2016, 2017, and 2018 and the spring 2019 administration of the NJSLA.[1] There are many ways to assess the reliability of test outcomes. This analysis will only focus on school-, test-, and subgroup-level results, not the reliability of individual student results. For this study, researchers focused on measuring and analyzing the alternate-form reliability by calculating the change in the year-to-year correlation of test scores (reported at the school, test, and subgroup levels) before and after the test changed from the PARCC to the NJSLA.[2] Researchers also examined year-to-year movement across quintiles of school, test, and subgroup performance to evaluate whether the new, shorter test was associated with a decrease in the

[1] Year refers to the calendar year of the spring administration of the PARCC. For example, year 2016 refers to the spring of school year 2015/16.

[2] Alternate-form reliability is a measure of the consistency of scores across comparable forms of a test administered to the definition same group of individuals (Crocker & Algina, 1986). This analysis assumes the students within a single school and subgroup across adjacent years are similar enough that the alternate-form reliability can still be estimated across distinct groups of students taking the same subject test. As in the case of Algebra 2, in at least one test, the assumption of comparability of students across years within schools was violated, which was observable from the change in composition of students.

stability of school, test, and subgroup results. They examined these two measures for schoolwide aggregates at the test, subgroup, and outcome level to understand patterns in the variation of school-level scores. Outcomes included percentage proficient, mean scale scores, mean scale scores for reading and writing *major claims* (chief portions of tests), percentage meeting or exceeding expectations in ELA and math *subclaims*, and median student growth percentiles.

Research questions

Research approach. This study addressed the following research questions:

1. How much did school-level test scores change when schools switched from the PARCC to the NJSLA compared with the changes observed across years when the PARCC was in place?
2. How did changes in school-level test scores vary by test, student subgroup, group size, test subscore, and outcome measure over these years?

Box 1. Data sources, sample, and methods

Data sources. This study uses New Jersey Partnership for Assessment of Readiness for College and Careers (PARCC) performance data at the school, test, and subgroup levels from 2016 to 2018 and spring New Jersey Student Learning Assessments (NJSLA) performance data from 2019. The Office of Assessments provided assessment data for the study, including assessment results for both overall tests and test subclaims, which are measures of proficiency within a test. The Office of Performance Management provided median student growth percentiles at the school, test, and subgroup levels.

Sample. This study used performance data for all school, test, and subgroup combinations within New Jersey's public schools in accordance with the New Jersey Department of Education's (NJDOE) data suppression policy, under which results from groups of 10 or fewer students or that could disclose results for individual students are not reported. The sample for all analyses was limited to school and subgroup records that had complete, non-suppressed data for all four years of interest (2016 to 2019), which enabled the analysis to focus on the same school and subgroup combinations over time. The sample included subgroup-level results for 2,155 unique schools. The same students will appear in multiple observations because individual students have taken multiple tests in multiple years and are often in multiple subgroups.

When reporting results by subgroups, the researchers excluded subgroups with fewer than 20 schools so as not to make inferences on the basis of very small groups of schools, which would be unlikely to be stable regardless of test characteristics. Across all tests, this rule only excluded the two or more races subgroup, which accounted for 8.1 percent of test and subgroup combinations within the sample.

Methodology. This study used two measures of agreement to examine test reliability in the transition from the PARCC to the NJSLA:

Concordance correlation coefficient: The concordance correlation coefficient measures the level of agreement between two variables (Lin, 1989, 2000).[3] A coefficient of 1 indicates perfect agreement and a coefficient of 0 indicates no agreement (e.g., a school's rank in one year and its rank in the next year have no relationship). To probe for decreased inter-test reliability in the transition from the PARCC to NJSLA, the study team used the concordance correlations between two pairs of spring PARCC test administrations (2016 and 2017 and 2017 and 2018) as a benchmark of the degree of concordance that would otherwise be expected if the same test is used. The study team then compared this benchmark with the concordance correlation coefficient between years 2018 and 2019, when the test transitioned from the PARCC in spring 2018 to the NJSLA in spring 2019. Average differences in coefficients in the two sets of years before the test change were mainly but not entirely under 0.05 in absolute value; thus, patterns of differences of 0.1 or larger for particular outcomes, tests, or subgroups were

[3] The concordance correlation coefficient combines measures of precision and accuracy to determine how far the observed data deviate from the line of perfect concordance (that is, the line at 45 degrees on a square scatterplot). Lin's coefficient increases in value as a function of the nearness of the data's reduced major axis to the line of perfect concordance (the accuracy of the data) and of the tightness of the data about its reduced major axis (the precision of the data).

established by researchers and NJDOE to distinguish meaningful decreases in reliability. Correlations of 0.7 or higher were described as "high."

Movement between quintiles of performance: The study team used churn, defined here as the percentage of schools that move at least one performance quintile between a set of pairwise-years, to assess the change in school- and subgroup-level performance between the PARCC and NJSLA. Like the concordance correlation analysis, the degree of churn between years 2016 and 2017 and 2017 and 2018 served as a benchmark of expected movement. The study team then calculated the difference between this baseline and the churn between years 2018 and 2019 to examine whether more schools shifted performance level from year to year when the test transitioned from the PARCC to NJSLA. Patterns of differences above 5 percentage points were considered meaningful increases in churn. The team also conducted this analysis on the number of schools that shifted more than one performance level quintile per year.

The measure of churn does not incorporate the directionality of school performance quintile movement (that is, whether school performance improved or declined from year to year). To probe the direction of movement, the study team also examined transition matrices, which map schools' quintile of performance between two pairwise years.

To investigate whether year-to-year test reliability was varied across different levels and groups (research question 2), researchers examined concordance correlations and quintile movement by test, subgroup, type of performance measure (such as scale score and proficiency level), and subgroup size.

Findings

The research team compared the benchmark of expected concordance (average concordance correlation between 2016 and 2017 and 2017 and 2018) with the concordance correlation between 2018 and 2019 (box 1) and found that—with one exception—across tests, subgroups, and outcomes, there was **little change** in the year-to-year concordance correlation with the shift to a shorter test. Most year-to-year correlations in assessment outcomes at the test, subgroup, and school level remained high (more than 0.7 in most reported values apart from student growth percentiles). The discussion below focuses largely on percentage proficient, as a broadly reported outcome related to school-, test-, and subgroup- level performance. Outside of Algebra 2, there was no broad pattern of meaningful decreases in concordance correlation coefficients.

Finding 1. Concordance correlations in percentage proficient remained high from 2016 to 2019.

Averaging all tests and subgroups (excluding the total subgroup), the year-to-year concordance correlation for the percentage proficient was 0.87, and the decline in average concordance correlation was 0.007 (table 2). This was well below the 0.10 threshold for meaningful decreases in concordance correlation, and it was about the same magnitude as the change from 2016 and 2017 to 2017 and 2018, the years before the test change.

Table 2. Concordance correlations in percentage proficient remained stable over the years of the test change

Concordance correlation by pairwise years			Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
2016 and 2017	2017 and 2018	2018 and 2019		
0.876	0.882	0.872	-0.007	77,349

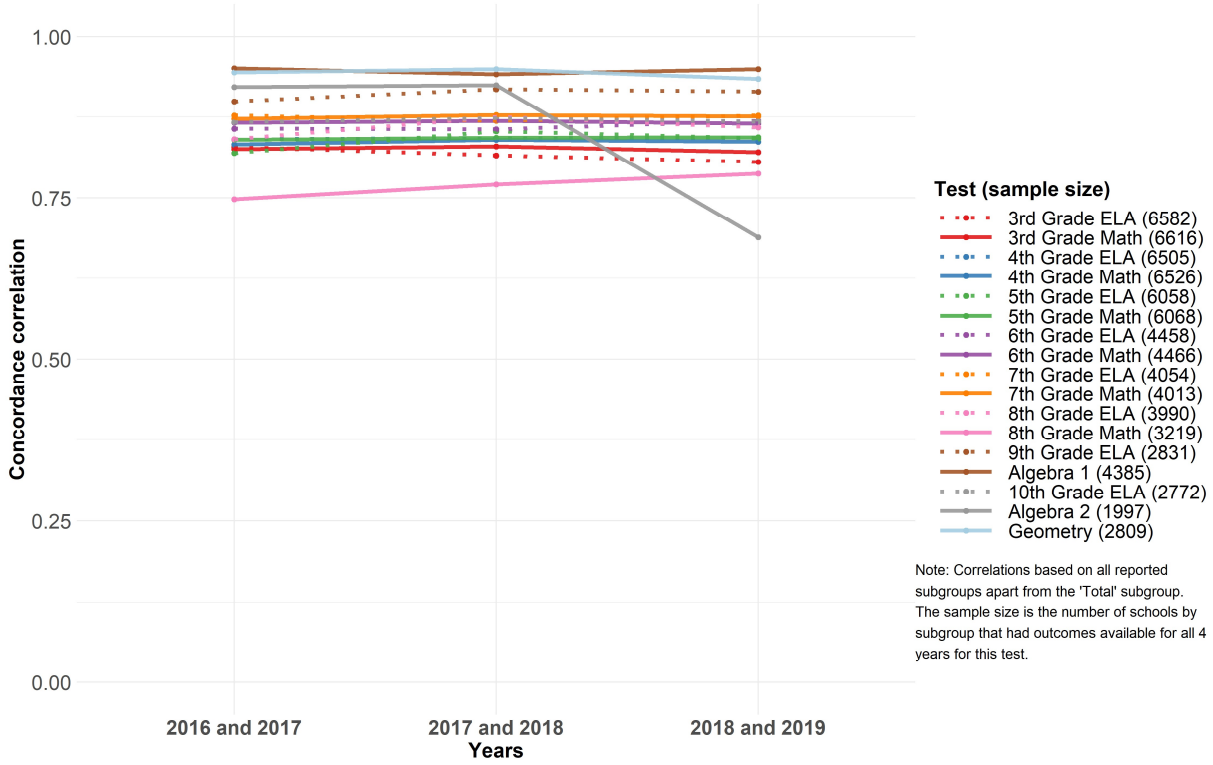
Note: Correlations are based on all reported subgroups apart from the total subgroup. The sample size is the number of school, test, and subgroup combinations that had outcomes available for all four years.
Source: Authors' analysis of data from the New Jersey Department of Education.

Finding 2. Concordance correlations in percentage proficient were stable for all tests except Algebra 2.

Except for Algebra 2, there were no large changes in the concordance correlation in percentage proficient over the years of the test change for any test (figure 1 and table 3). Apart from Algebra 2, the average changes in the

concordance correlation for percentage proficient across subgroups for each test ranged from a decline of 0.0165 (grade 3 ELA) to an increase of 0.0285 (grade 8 math). This indicates that reliability did not meaningfully decrease in these tests after New Jersey switched to the shorter test. There were, however, large decreases in the concordance correlation for Algebra 2 (by 0.2335), which were several times larger than any other average change.

Figure 1. Concordance correlations in percentage proficient were stable across tests from 2016 to 2019, except Algebra 2



ELA is English language arts.

Table 3. Across all tests but Algebra 2, the concordance correlation in percentage proficient was stable over the years of the test change

Test	Concordance correlation by pairwise years			Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
	2016 and 2017	2017 and 2018	2018 and 2019		
Grade 3 ELA	0.829	0.816	0.806	-0.016	6,582
Grade 4 ELA	0.841	0.842	0.843	0.002	6,505
Grade 5 ELA	0.820	0.853	0.842	0.005	6,058
Grade 6 ELA	0.857	0.856	0.869	0.012	4,458
Grade 7 ELA	0.877	0.869	0.877	0.004	4,054
Grade 8 ELA	0.841	0.875	0.859	0.001	3,990
Grade 9 ELA	0.899	0.918	0.914	0.005	2,831
Grade 10 ELA	0.867	0.873	0.868	-0.002	2,772
Grade 3 math	0.826	0.830	0.821	-0.007	6,616
Grade 4 math	0.833	0.840	0.837	0.000	6,526
Grade 5 math	0.840	0.843	0.844	0.003	6,068
Grade 6 math	0.866	0.869	0.865	-0.002	4,466
Grade 7 math	0.872	0.878	0.876	0.001	4,013
Grade 8 math	0.748	0.771	0.788	0.028	3,219
Algebra 1	0.950	0.941	0.949	0.003	4,385
Geometry	0.944	0.949	0.934	-0.012	2,809
Algebra 2	0.921	0.924	0.689	-0.234	1,997

ELA is English language arts.
Note: Correlations are based on all reported subgroups apart from the total subgroup. The sample size is the number of school and subgroup combinations that had outcomes available for all four years for each test.
Source: Authors' analysis of data from the New Jersey Department of Education.

Finding 2a. The concordance correlation in the percentage proficient for most subgroups decreased for Algebra 2 over the years of the test change.

For all subgroups except students with disabilities, the concordance correlation in percentage proficient in Algebra 2 decreased substantially over the years of the test change, with decreases of 0.225 to 0.3935 (figure 2; see table B18 in appendix B). The year of the test change coincided with a change in testing policy, which no longer required grade 11 students to take state exams. This is reflected in large changes in the composition of students taking Algebra 2 (table 4). The percentage of economically disadvantaged students taking the exam decreased by 12.5 percentage points, the percentage of students with disabilities decreased 7.8 percentage points, and the percentage of grade 11 students decreased by 53 percentage points. The change in high school testing policy is reflected in much smaller shifts in the composition of test takers for Algebra 1 and Geometry, with a 2 and 3 percentage point decrease in the share of economically disadvantaged students taking the test, respectively (see tables B31 and B32). This shift corresponds to a smaller decrease in correlation for Geometry between the baseline years and 2018 and 2019, and no decrease for Algebra 1.

Figure 2. Concordance correlations declined for almost all subgroups for Algebra 2 over the years of the test change

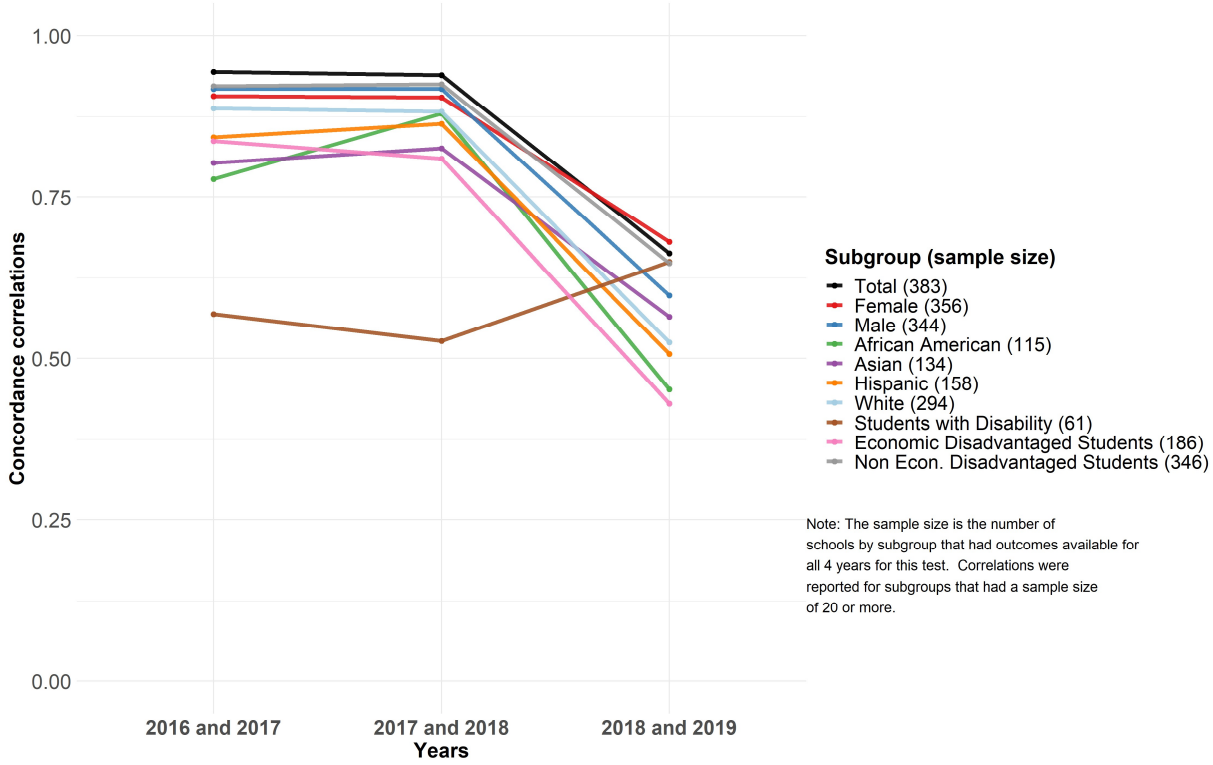


Table 4. The composition of Algebra 2 test takers shifted towards fewer economically disadvantaged students in 2019

Subgroup	Percentage of total students in 2016	Percentage of total students in 2017	Percentage of total students in 2018	Percentage of total students in 2019	Difference between 2019 and average of 2016–18 (percentage points)
Race and ethnicity					
White	47.3	45.6	46.3	50.7	4.3
African American	15.4	15.1	14.4	9.3	-5.7
Asian	11.1	11.3	11.5	20.4	9.1
Hispanic	24.8	26.0	26.2	17.4	-8.3
Gender					
Female	49.4	49.3	49.7	51.8	2.3
Male	50.6	50.7	50.3	48.2	-2.3
Subgroup					
English language learner	3.7	5.9	5.6	2.2	-2.8
Students with disabilities	11.6	11.7	11.7	4.0	-7.6
Economically disadvantaged	33.2	33.0	32.0	20.2	-12.5
Non-economically disadvantaged	66.8	67.0	68.0	79.8	12.5
Grade					
Grade 8	0.4	0.4	0.5	1.1	0.7
Grade 9	7.3	7.8	8.0	18.9	11.2
Grade 10	32.4	31.9	35.7	79.9	46.5

Subgroup	Percentage of total students in 2016	Percentage of total students in 2017	Percentage of total students in 2018	Percentage of total students in 2019	Difference between 2019 and average of 2016–18 (percentage points)
Grade 11	54.4	54.2	50.5	0.0	-53.0
Grade 12	5.4	5.6	5.2	0.0	-5.4
Sample size	74549	74931	78429	36845	

Source: Authors' analysis of data from the New Jersey Department of Education.

Finding 2b. The concordance correlation in median student growth percentiles subgroups did not decline over the year of the test change.

Year-to-year concordance correlations for median student growth percentiles were much lower than for other outcomes, ranging from 0.271 to 0.562—an expected result given that change measures tend to be less reliable than measures of proficiency levels and median student growth percentiles are known to be unreliable (Castellano & McCaffrey, 2019; Castellano, 2016). Over the years of the test change, the changes in concordance correlations averaged across all subgroups for each grade for which they were calculated (table 5) ranged from a decrease of 0.017 for grade 6 math to an increase in 0.062 for grade 8 ELA, which were below our 0.10 threshold for meaningful decreases. There was no pattern of meaningful decreases.

Table 5. Concordance correlations were lower for median student growth percentiles than for other test outcomes, but they did not decline substantially for any test over the years of the change

Test	Concordance correlation by pairwise years			Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size (school by subgroup combinations)
	2016 and 2017	2017 and 2018	2018 and 2019		
Grade 4 ELA	0.451	0.484	0.478	0.011	9,984
Grade 5 ELA	0.414	0.481	0.462	0.015	9,275
Grade 6 ELA	0.531	0.562	0.550	0.004	6,560
Grade 7 ELA	0.466	0.503	0.529	0.045	5,843
Grade 8 ELA	0.271	0.324	0.359	0.062	5,795
Grade 4 math	0.460	0.477	0.480	0.012	10,005
Grade 5 math	0.484	0.514	0.492	-0.007	9,288
Grade 6 math	0.523	0.513	0.501	-0.017	6,572
Grade 7 math	0.492	0.497	0.513	0.019	5,844

ELA is English language arts.
Note: Correlations are based on all reported subgroups apart from the total subgroup. The sample size is the number of schools by subgroup that had outcomes available for all four years for each test.
Source: Authors' analysis of data from the New Jersey Department of Education.

Finding 3. Concordance correlations by subgroup were about the same after the test change.

The changes in concordance correlations between subgroups over the years of the test change were small compared with the variation occurring over years in which schools used the same test (2016 and 2017 and 2017 and 2018; table 6). Changes in concordance correlation over the years of the test change varied from a decline of 0.025 to an increase of 0.013 (see table 6). The changes in concordance correlation for median student growth percentiles were also small relative to year-to-year variation (see tables B19 to B27 in appendix B).

Table 6. Finding 3: All tests (excluding Algebra 2) by subgroup; percentage proficient

Subgroup	Concordance correlation by pairwise years: 2016 and 2017	Concordance correlation by pairwise years: 2017 and 2018	Concordance correlation by pairwise years: 2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size (school by test combinations)
Total	0.890	0.898	0.900	0.006	13,886
Female	0.862	0.872	0.872	0.005	12,875
Male	0.858	0.865	0.860	-0.001	12,893
Economically disadvantaged students	0.747	0.759	0.763	0.010	8,552
Non-economically disadvantaged students	0.832	0.842	0.844	0.007	11,182
Students with disabilities	0.710	0.714	0.711	-0.001	5,904
African American	0.745	0.754	0.754	0.004	4,331
Asian	0.751	0.763	0.755	-0.002	2,714
Hispanic	0.747	0.767	0.770	0.013	7,142
Two or more races	0.763	0.755	0.734	-0.025	104
White	0.794	0.799	0.799	0.003	9,652

Note: The average concordance correlations for all reported subgroups across all tests (apart from Algebra 2) are shown in table 3. The sample size is the number of schools by test that had outcomes available for all four years for each subgroup. The total group contains school by test records of outcomes for all students, whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Finding 4. Concordance correlations were stable across schools of differing sizes.

To determine whether results from smaller schools showed reductions in stability across the years of the test change, we repeated the concordance correlation analysis by school size. We divided the results from the total subgroup into five quintiles based on the number of valid student test scores in 2019 that contributed to the total subgroup outcome. Then, we calculated concordance correlations separately for each quintile (table 7). The concordance correlations for percentage proficient did not decline substantially more for smaller reported groups (ranging from 10 to 44 students with valid scores) than for larger groups (ranging from 148 to 1178 students with valid scores). Changes in concordance correlation for all subgroups by quintiles of size are available in table B30 in appendix B.

Table 7. Schools with fewer valid scores in their total subgroup for tests had somewhat lower concordance correlations on average than larger groups, but there was no pattern of substantial declines over the year of the test change

Quintile of number of valid student scores	Range of number of valid student scores in quintile	Concordance correlation by pairwise years			Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
		2016 and 2017	2017 and 2018	2018 and 2019		
1	10–44	0.833	0.845	0.827	-0.012	2,939
2	45–65	0.900	0.910	0.899	-0.006	2,895
3	66–90	0.906	0.908	0.898	-0.009	2,815
4	91–147	0.916	0.919	0.916	-0.001	2,788
5	148–1178	0.934	0.941	0.935	-0.002	2,832

Note: Correlations are based on all reported total subgroups that reported the number of valid student scores for 2019 and had outcomes available for all four years.
Source: Authors' analysis of data from the New Jersey Department of Education

Finding 5. Across outcomes, the concordance correlation stayed about the same across the years of the test change, although some subclaims had small decreases.

Outcomes varied in average concordance correlation coefficients, but no outcomes showed patterns of dramatically decreasing coefficients above 0.1; no change amounted to greater than 0.03. Subclaims, such as Vocabulary (from ELA tests) or Modeling and Application (from math tests), as well as the ELA reading and writing major claims, had coefficients lower than percentage proficient or mean scale scores. However, coefficients for subclaims and major claims were also largely stable across years. Except in Algebra 2 (discussed above), the coefficients did not decrease by more than 0.03 for any outcome (table 8). The major claims for ELA (reading and writing) had no meaningful decreases in concordance correlation across the years of the test change, nor did any of the ELA or math subclaims.

Table 8. There was no pattern of substantial declines in concordance correlation for any test outcome, including major claims and subclaims across all grades and tests

	Concordance correlation by pairwise years			Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
	2016 and 2017	2017 and 2018	2018 and 2019		
Math outcomes					
Mean scale score	0.912	0.914	0.915	0.002	38,102
Percentage proficient	0.885	0.887	0.886	0.000	38,102
Met or exceeded expectations, Major Content subclaim	0.848	0.849	0.842	-0.007	38,102
Met or exceeded expectations, Additional & Supporting Content subclaim	0.837	0.823	0.804	-0.026	38,102
Met or exceeded expectations, Expressing Mathematical Reasoning subclaim	0.829	0.848	0.82	-0.019	38,102
Met or exceeded expectations, Modeling & Application subclaim	0.793	0.82	0.784	-0.022	38,102
ELA outcomes					
Mean scale score	0.888	0.894	0.893	0.002	37,250
Percentage proficient	0.852	0.861	0.857	0.000	37,250
Mean scale score, writing major claim	0.859	0.869	0.862	-0.002	37,250
Mean scale score, reading major claim	0.887	0.892	0.887	-0.002	37,250
Met or exceeded expectations, Knowledge & Use of Language Conventions subclaim	0.814	0.803	0.803	-0.005	37,250
Met or exceeded expectations, Literary Text subclaim	0.821	0.822	0.811	-0.010	37,250
Met or exceeded expectations, Informational Text subclaim	0.820	0.816	0.788	-0.030	37,250
Met or exceeded expectations, Vocabulary subclaim	0.807	0.800	0.774	-0.030	37,250
Met or exceeded expectations, Writing Expression subclaim	0.811	0.807	0.800	-0.009	37,250

ELA is English language arts.
Note: Correlations are based on all reported subgroups and tests apart from the total subgroup and Algebra 2. The sample size is the number of school, subgroup, and test combinations that had outcomes available for all four years within each subject.
Source: Authors' analysis of data from the New Jersey Department of Education.

Finding 6. Quintile movement was similar in 2018/19 to baseline years (2016/17 and 2017/18).

We found little difference when we measured the amount of churn before and after the test change. The difference in the percentage of schools that changed quintiles over the years of the test change was lower than 5 percentage points for all tests except Algebra 2 (table 9). For Algebra 2, there was a 7.9 percentage point increase in the percentage of schools that shifted one or more quintiles. The results from analyzing quintile movement are consistent with the results of the concordance correlations.

Table 9. School performance quintile change by test (percentage proficient)

Test	Percentage of schools changing quintiles of performance by year			Percentage point difference between average of 2016 and 2017 and 2017 and 2018 (baseline) and 2018 and 2019 (test change)
	2016 and 2017	2017 and 2018	2018 and 2019	
ELA				
Grade 3 ELA	59.0	59.1	58.5	-0.6
Grade 4 ELA	58.9	60.7	55.6	-4.2
Grade 5 ELA	59.4	54.3	57.7	0.9
Grade 6 ELA	57.4	54.5	54.8	-1.2
Grade 7 ELA	51.9	51.2	52.0	0.5
Grade 8 ELA	56.6	53.7	54.1	-1.0
Grade 9 ELA	52.4	47.9	47.7	-2.4
Grade 10 ELA	57.1	47.8	52.4	-0.1
Math				
Grade 3 math	58.7	56.6	58.1	0.4
Grade 4 math	59.4	57.0	59.4	1.1
Grade 5 math	56.4	57.1	56.4	-0.4
Grade 6 math	55.9	54.8	52.1	-3.2
Grade 7 math	55.0	45.8	55.1	4.8
Grade 8 math	58.8	58.6	57.5	-1.2
Algebra 1	43.7	42.1	39.6	-3.3
Geometry	44.2	43.8	45.8	1.7
Algebra 2	50.3	51.1	58.6	7.9

ELA is English language arts.
Source: Authors' analysis of data from the New Jersey Department of Education.

Implications

The findings from this study (summarized in Table 10, next page) show that New Jersey's school-level test results are remarkably stable from 2018 to 2019 compared with 2016 to 2018. Except Algebra 2, which was likely affected by changes in the composition of students taking the test, we do not see evidence that the change to a shorter test increased the variability or reduced the alternate-form reliability of scores at the school, test, and subgroup levels. This finding is true across tests, subgroups, group size, and outcomes. The primary implication of these findings is that New Jersey can feel confident that the aggregated school-level test results it reports from the NJSLA are similar in reliability to the test scores reported in the past from the PARCC and that it can use these aggregated results to make accountability-related decisions or other policy decisions without worry that the shorter assessment degraded their reliability.

Results for all tests showed stability over time except for Algebra 2, for which the concordance correlation dropped substantially for all subgroups. Because of changed testing requirements in New Jersey, economically disadvantaged students made up a smaller share of the student population taking the Algebra 2 exam in 2019 than in prior years and test scores increased. Even though the drop in stability might have mainly been attributable to these large compositional changes, this still implies that school-level aggregates of Algebra 2 test results from 2019 are not comparable to earlier results; thus, NJDOE should carefully consider how it uses school-level Algebra 2 test results—perhaps focusing on years after 2019 when the composition of test takers has stabilized.

Table 10. Summary of findings

Finding number	Finding
1	Concordance correlations remained high from 2016 to 2019.
2	Concordance correlations in percentage proficient were stable for all tests except Algebra 2.
2a	The concordance correlation in the percentage proficient for most subgroups decreased for Algebra 2 over the years of the test change.
2b	The concordance correlation in median student growth percentiles subgroups did not decline over the year of the test change.
3	Concordance correlations by subgroup were about the same after the test change.
4	Concordance correlations were stable across schools of differing sizes.
5	Across outcomes, the concordance correlation stayed about the same across the years of the test change, although some subclaims had small decreases.
6	Quintile movement was similar in 2018/19 to baseline years (2016/17 and 2017/18).

The variability of test results was also stable over time when broken down by the number of valid student scores reported in a school-, test-, and subgroup-level group. Even for the smallest quintile of reported group size, there was no associated drop in stability when the NJSLA was introduced. As expected, the concordance correlations for the smallest groups were lower in magnitude than for the larger groups, but the absence of a decline in this metric over time implies that the shorter assessment did not differentially affect results reported for smaller groups. In other words, NJDOE does not need to consider a larger cutoff for subgroup reporting to maintain the reliability of the results. The stability of median student growth percentiles was lower than for other test outcomes, but that higher degree of variability is a known property of growth measures compared with outcomes such as percentage proficient or average test scores.

Finally, the stability by student subgroup across all tests and outcomes (except Algebra 2) is reassuring. Similar to the findings by group size, this finding implies that the shorter assessment did not differentially affect results for particular subgroups. Therefore, NJDOE can feel confident that the reported results for subgroups are as reliable as they were in the past. The shift from the PARCC to the NJSLA may reduce burden on students, teachers, and other stakeholders, and it does not presently appear to have sacrificed the reliability in group-level average test results. Further investigation might clarify whether student-level reliability was affected by the change in tests.

Limitations

The primary limitation of this study is that it cannot disentangle changes that occurred because of the change in the assessment and changes because of other factors that could also affect year-to-year test variation, such as the change in students, teachers, or principals from year to year, changes in curriculum, or bad weather on testing day (Kane & Staiger, 2002). With data from three years before the test change and one year after the test change, the analyses could not separately identify the source of changes in variation, so researchers relied on the relative change in variation before and after the test change to make inferences. Though the study is not causal in nature, the patterns did not reveal any changes to be disentangled (except for Algebra 2, for which there is a plausible and likely explanation), and the patterns across the abundance of analyses conducted suggest the shortened assessment is as reliable as the longer one it replaced.

A secondary limitation of the analyses is that they relied on aggregate data rather than student-level data. Though more sophisticated analyses are possible with student-level data, employing the data NJDOE reports and uses for making policy decisions to conduct the analyses is a valid alternative given the constraints.

References

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores: Implications for test-based accountability systems. *Brookings Papers on Education Policy, 5*, 235–283.

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255–268.

Lin, L. I.-K. (2000). A note on the concordance correlation coefficient. *Biometrics*, *56*(1), 324–325, https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.0006-341X.2000.00324.x .

New Jersey Department of Education. (2018). *Statewide assessment outreach: Summary of findings, recommendations for next steps,* https://www.nj.gov/education/assessment/docs/outreach/AssessmentReportSummary.pdf.

New Jersey Department of Education. (2019a). *Spring 2019 blueprints for the New Jersey Student Learning Assessments (NJSLA).* https://www.nj.gov/education/broadcasts/2019/FEB/22/19683/Spring%202019%20Blueprints%20for%20the%20New%20Jersey%20Student%20Learning%20Assessments.pdf.

New Jersey Department of Education. (2019b). *New Jersey Student Learning Assessments-ELA spring 2019 companion guide.* https://www.nj.gov/education/assessment/resources/district/docs/FinalNJSLA-ELASpring2019CompanionGuide.pdf

Thayer, K. (2019, March 6). Illinois scraps controversial PARCC test in favor of shorter exam with new name. *Chicago Tribune*. https://www.chicagotribune.com/lifestyles/ct-life-iar-parcc-state-testing-20190227-story.html

REL 2021–106

July 2021

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0006 by Regional Educational Laboratory Mid-Atlantic administered by Mathematica. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Fox, L., Hartog, J., & Larkin N. (2021). *The reliability of shorter assessments in New Jersey for group-level inferences* (REL 2021–106). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

The reliability of shorter assessments in New Jersey for group-level inferences

Appendix A. About this study

Previous related studies of test reliability

Education stakeholders have long been interested in the reliability of student assessments, and this report fits into a broader context of studies using the correlation and volatility across assessments and years of administration to measure this reliability. In particular, research has examined changes in reliability with the shortening of an assessment; one study used individual level test re-test correlations to analyze differences between individuals who re-tested the GRE on the same version after a period of time and those who re-tested on a shortened version (Kingston & Turner, 1984). Another study examined the correlation between students' ACT and SAT 1 scores to measure the degree to which these scores were related to and could predict one another (Dorans, 1999). Like the study at hand, research has also assessed the reliability of assessments at the subgroup level, breaking down data along gender and race and ethnicity (Wilson, 1988).

While this study was not able to conduct analyses at the individual level for data availability reasons, previous research has also investigated reliability of the types of estimates used for accountability decisions and included in this study that pool results at the school-level. Kane and Staiger (2002) found that student test score data can be extremely volatile due to both small samples and other one-time factors, and pooling results at the school level across years and outcomes can make a measurement more consistent.

Technical details of the concordance correlation measure

The concordance correlation coefficient (CCC; Lin, 1989, 2000) is a measure of the agreement between continuous measures. The CCC does not assume the equality of variances between two measures (x, y). It is calculated using the standard deviation of each variable (σ), their mean values ($\bar{x}$, $\bar{y}$), and the Pearson correlation between the variables (r_{xy}). Lin's coefficient modifies the Pearson correlation coefficient by assessing the line of the best fit between the two variables and then computing the distance between that line fit and the 45-degree line through the origin representing perfect agreement. Lin's CCC is 1.0 when all the points lie exactly on the 45-degree line of the perfect agreement and decreases as the observations depart from this line and as the line of best fit departs from the 45-degree line.

$$CCC = \frac{2r_{xy}\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\bar{x} - \bar{y})^2}$$

Data sources

The Office of Assessments provided assessment data for the study, including assessment results for both overall tests and test subclaims, which are a subset of items that test a specific skill set within a subject. The Office of Performance Management provided median student growth percentiles at the school, test, and subgroup levels.

Table A1. Data sources

Source	Years	Tests	Reported subgroups	Outcomes
Office of Assessment	2016 to 2019	English language arts 3–10, Math 3–8, Algebra 1, Geometry, Algebra 2	African American, American Indian, Asian, economically disadvantaged students, female, Hispanic, male, missing race, non-economically disadvantaged students, Pacific Islander, students with disability, total, two or more races, White	Percentage proficient, mean scale score, mean writing scale score, mean reading scale score, percentage meeting expectations within subclaims: Informational Text, Narrative Text, Vocabulary and Written Expression (for English language arts), and Expressing Mathematical Reasoning, Modeling and Application, Major and Supporting Content (for math)
Office of Performance Management	2016 to 2019	English language arts 4–8, Math 4–7	American Indian or Alaska Native, Asian, Black or African American, economically disadvantaged students, English learners, female, Hispanic, male, non-economically disadvantaged students, non-English learners, students with disabilities, students without disabilities, total, two or more races, White	Median student growth percentile

References

Dorans, N. (1999). *Correspondences between ACT™ and SAT® I scores*. ETS Research Report Series, 1999. https://www.ets.org/Media/Research/pdf/RR-99-02-Dorans.pdf

Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores: Implications for test-based accountability systems. *Brookings Papers on Education Policy, 5*, 235–283.https://www.jstor.org/stable/20067246

Kingston, N., & Turner, N. (1984). *Analysis of score change patterns of examinees repeating the Graduate Record Examinations General Test1*. ETS Research Report Series, 1984. https://www.semanticscholar.org/paper/ANALYSIS-OF-SCORE-CHANGE-PATTERNS-OF-EXAMINEES-THE-Kingston-Turner/a69f8edfdc891f443c3d96a459cc3c0e56b6ad31

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255–268.

Lin, L. I.-K. (2000). A note on the concordance correlation coefficient. *Biometrics*, *56*(1), 324–325, https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.0006-341X.2000.00324.x

McBride, G. B. (2005). *A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient*. HAM2005-062: NIWA Client Report. https://www.medcalc.org/download/pdf/McBride2005.pdf.

Wilson, K. (1988). *A study of the long-term stability of GRE General Test scores*. ETS Research Report Series, 1988. https://onlinelibrary.wiley.com/doi/abs/10.1002/j.23308516.1988.tb00295.x

Appendix B. Supporting analysis

Table B1. Concordance correlations across years: Mean scale score by test

Test	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Grade 3 ELA	0.869	0.866	0.859	-0.008	6,582
Grade 4 ELA	0.878	0.877	0.884	0.007	6,505
Grade 5 ELA	0.861	0.891	0.886	0.010	6,058
Grade 6 ELA	0.889	0.890	0.900	0.011	4,458
Grade 7 ELA	0.904	0.897	0.909	0.008	4,054
Grade 8 ELA	0.880	0.899	0.890	0.001	3,990
Grade 9 ELA	0.923	0.937	0.934	0.004	2,831
Grade 10 ELA	0.895	0.895	0.897	0.002	2,772
Grade 3 math	0.868	0.868	0.864	-0.004	6,616
Grade 4 math	0.875	0.883	0.883	0.004	6,526
Grade 5 math	0.885	0.888	0.890	0.003	6,068
Grade 6 math	0.900	0.899	0.894	-0.005	4,466
Grade 7 math	0.903	0.909	0.907	0.001	4,013
Grade 8 math	0.817	0.837	0.837	0.010	3,219
Algebra 1	0.954	0.949	0.958	0.006	4,385
Geometry	0.948	0.957	0.941	-0.011	2,809
Algebra 2	0.932	0.939	0.759	-0.176	1,997

ELA is English language arts.
Note: Correlations are based on all reported subgroups apart from the total subgroup. The sample size is the number of schools by subgroup that had outcomes available for all four years for each test.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B2. Concordance correlations across years: Grade 3 ELA by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.856	0.854	0.840	-0.015	1,308
Female	0.825	0.812	0.785	-0.033	1,225
Male	0.802	0.801	0.790	-0.012	1,227
Economically disadvantaged students	0.662	0.617	0.635	-0.004	739
Non-economically disadvantaged students	0.691	0.678	0.659	-0.025	1,028
Students with disabilities	0.670	0.610	0.677	0.037	374
African American	0.664	0.606	0.637	0.002	315
Asian	0.514	0.625	0.573	0.003	206
Hispanic	0.674	0.659	0.668	0.001	575
White	0.671	0.654	0.618	-0.045	890

ELA is English language arts.

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.

Source: Authors' analysis of data from the New Jersey Department of Education.

Table B3. Concordance correlations across years: Grade 3 math by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.849	0.867	0.860	0.002	1,308
Female	0.800	0.809	0.804	0.000	1,226
Male	0.825	0.829	0.809	-0.018	1,228
Economically disadvantaged students	0.630	0.641	0.654	0.019	741
Non-economically disadvantaged students	0.728	0.738	0.724	-0.009	1033
Students with disabilities	0.667	0.638	0.649	-0.004	384
African American	0.624	0.717	0.654	-0.016	315
Asian	0.550	0.672	0.545	-0.066	210
Hispanic	0.640	0.664	0.648	-0.004	581
White	0.650	0.644	0.645	-0.002	895

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.

Source: Authors' analysis of data from the New Jersey Department of Education.

Table B4. Concordance correlations across years: Grade 4 ELA by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.869	0.875	0.875	0.003	1,274
Female	0.823	0.815	0.824	0.005	1,194
Male	0.821	0.820	0.814	-0.007	1,191
Economically disadvantaged students	0.662	0.680	0.696	0.025	738
Non-economically disadvantaged students	0.714	0.717	0.731	0.015	1,008
Students with disabilities	0.655	0.665	0.657	-0.003	432
African American	0.643	0.667	0.705	0.050	324
Asian	0.551	0.554	0.555	0.003	197
Hispanic	0.692	0.682	0.736	0.049	557
White	0.676	0.701	0.667	-0.021	857

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B5. Concordance correlations across years: Grade 4 math by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.863	0.868	0.875	0.010	1,275
Female	0.803	0.815	0.813	0.004	1,196
Male	0.824	0.826	0.824	-0.001	1,191
Economically disadvantaged students	0.625	0.650	0.641	0.004	741
Non-economically disadvantaged students	0.736	0.759	0.753	0.005	1,015
Students with disabilities	0.619	0.683	0.613	-0.038	428
African American	0.619	0.618	0.695	0.076	324
Asian	0.604	0.512	0.599	0.041	200
Hispanic	0.638	0.656	0.650	0.003	564
White	0.693	0.675	0.683	-0.001	860

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B6. Concordance correlations across years: Grade 5 ELA by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.828	0.879	0.867	0.014	1,177
Female	0.788	0.839	0.828	0.014	1,104
Male	0.795	0.827	0.808	-0.003	1,105
Economically disadvantaged students	0.632	0.677	0.662	0.007	678
Non-economically disadvantaged students	0.661	0.705	0.717	0.034	927
Students with disabilities	0.671	0.673	0.658	-0.014	414
African American	0.640	0.653	0.609	-0.038	306
Asian	0.638	0.564	0.581	-0.020	186
Hispanic	0.616	0.728	0.681	0.009	521
White	0.638	0.696	0.684	0.017	812

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B7. Concordance correlations across years: Grade 5 math by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.863	0.871	0.874	0.007	1,177
Female	0.806	0.823	0.815	0.000	1,105
Male	0.837	0.838	0.836	-0.001	1,103
Economically disadvantaged students	0.655	0.630	0.624	-0.019	681
Non-economically disadvantaged students	0.745	0.750	0.757	0.009	931
Students with disabilities	0.616	0.626	0.645	0.024	412
African American	0.654	0.642	0.627	-0.021	307
Asian	0.671	0.589	0.623	-0.007	187
Hispanic	0.635	0.652	0.645	0.002	524
White	0.673	0.675	0.690	0.016	813

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B8. Concordance correlations across years: Grade 6 ELA by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.861	0.856	0.872	0.013	795
Female	0.837	0.838	0.859	0.022	731
Male	0.842	0.828	0.842	0.007	728
Economically disadvantaged students	0.716	0.730	0.736	0.013	527
Non-economically disadvantaged students	0.718	0.711	0.765	0.051	613
Students with disabilities	0.636	0.653	0.651	0.006	400
African American	0.676	0.673	0.694	0.019	275
Asian	0.732	0.663	0.588	-0.11	167
Hispanic	0.715	0.714	0.729	0.015	459
White	0.690	0.694	0.745	0.053	541

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B9. Concordance correlations across years: Grade 6 math by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.871	0.881	0.876	0.000	796
Female	0.836	0.854	0.853	0.008	732
Male	0.862	0.861	0.854	-0.007	729
Economically disadvantaged students	0.723	0.718	0.691	-0.029	529
Non-economically disadvantaged students	0.754	0.783	0.777	0.009	618
Students with disabilities	0.601	0.562	0.620	0.038	396
African American	0.698	0.720	0.685	-0.024	274
Asian	0.700	0.670	0.618	-0.067	166
Hispanic	0.739	0.690	0.698	-0.016	464
White	0.711	0.704	0.696	-0.012	542

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B10. Concordance correlations across years: Grade 7 ELA by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.882	0.872	0.883	0.006	693
Female	0.858	0.853	0.855	0.000	647
Male	0.854	0.845	0.857	0.008	647
Economically disadvantaged students	0.744	0.737	0.756	0.016	479
Non-economically disadvantaged students	0.774	0.798	0.811	0.025	548
Students with disabilities	0.677	0.654	0.642	-0.023	374
African American	0.771	0.728	0.749	-0.001	276
Asian	0.691	0.728	0.695	-0.015	167
Hispanic	0.746	0.720	0.750	0.017	429
White	0.709	0.700	0.757	0.053	476

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B11. Concordance correlations across years: Grade 7 math by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.895	0.881	0.898	0.010	690
Female	0.842	0.853	0.856	0.009	644
Male	0.848	0.862	0.860	0.005	643
Economically disadvantaged students	0.722	0.726	0.730	0.006	476
Non-economically disadvantaged students	0.811	0.815	0.824	0.011	545
Students with disabilities	0.650	0.680	0.624	-0.041	377
African American	0.688	0.744	0.742	0.026	273
Asian	0.717	0.730	0.646	-0.078	149
Hispanic	0.713	0.723	0.664	-0.054	427
White	0.733	0.726	0.739	0.009	472

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B12. Concordance correlations across years: Grade 8 ELA by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.850	0.883	0.867	0.000	683
Female	0.810	0.844	0.834	0.007	643
Male	0.787	0.854	0.819	-0.002	642
Economically disadvantaged students	0.689	0.702	0.713	0.017	467
Non-economically disadvantaged students	0.723	0.788	0.781	0.025	550
Students with disabilities	0.673	0.711	0.657	-0.035	358
African American	0.654	0.668	0.577	-0.084	267
Asian	0.512	0.709	0.701	0.090	166
Hispanic	0.655	0.720	0.699	0.011	420
White	0.674	0.732	0.697	-0.006	474

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B13. Concordance correlations across years: Grade 8 math by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.774	0.790	0.815	0.033	598
Female	0.760	0.778	0.776	0.007	513
Male	0.685	0.738	0.779	0.068	535
Economically disadvantaged students	0.591	0.627	0.671	0.062	403
Non-economically disadvantaged students	0.730	0.746	0.777	0.039	452
Students with disabilities	0.523	0.571	0.587	0.040	319
African American	0.544	0.556	0.550	0.000	225
Asian	0.630	0.661	0.684	0.039	60
Hispanic	0.658	0.715	0.704	0.017	320
White	0.667	0.659	0.721	0.058	392

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B14. Concordance correlations across years: Grade 9 ELA by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.921	0.940	0.935	0.005	401
Female	0.895	0.924	0.914	0.005	394
Male	0.899	0.914	0.915	0.008	395
Economically disadvantaged students	0.783	0.783	0.775	-0.008	322
Non-economically disadvantaged students	0.874	0.910	0.892	0.000	372
Students with disabilities	0.652	0.772	0.731	0.019	319
African American	0.781	0.785	0.801	0.018	212
Asian	0.695	0.687	0.764	0.073	158
Hispanic	0.753	0.813	0.815	0.032	326
White	0.805	0.867	0.839	0.003	323

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B15. Concordance correlations across years: Grade 10 ELA by subgroup (percent proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.869	0.882	0.874	-0.001	399
Female	0.843	0.849	0.869	0.023	395
Male	0.848	0.860	0.841	-0.013	391
Economically disadvantaged students	0.776	0.825	0.767	-0.033	311
Non-economically disadvantaged students	0.834	0.821	0.820	-0.007	371
Students with disabilities	0.541	0.607	0.687	0.113	304
African American	0.761	0.756	0.758	0.000	213
Asian	0.694	0.674	0.634	-0.05	154
Hispanic	0.677	0.724	0.695	-0.006	305
White	0.765	0.755	0.741	-0.019	321

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B16. Concordance correlations across years: Algebra 1 by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.947	0.934	0.949	0.008	852
Female	0.946	0.932	0.941	0.002	725
Male	0.951	0.942	0.954	0.008	717
Economically disadvantaged students	0.912	0.904	0.900	-0.008	418
Non-economically disadvantaged students	0.941	0.930	0.942	0.006	747
Students with disabilities	0.734	0.760	0.797	0.050	331
African American	0.893	0.876	0.847	-0.038	238
Asian	0.883	0.902	0.900	0.007	188
Hispanic	0.901	0.871	0.885	-0.001	377
White	0.928	0.916	0.938	0.016	643

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B17. Concordance correlations across years: Geometry by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.967	0.974	0.963	-0.007	460
Female	0.935	0.943	0.932	-0.007	401
Male	0.951	0.954	0.933	-0.019	421
Economically disadvantaged students	0.800	0.804	0.746	-0.056	302
Non-economically disadvantaged students	0.959	0.966	0.955	-0.007	424
Students with disabilities	0.557	0.631	0.600	0.006	282
African American	0.827	0.712	0.705	-0.064	187
Asian	0.861	0.863	0.852	-0.010	153
Hispanic	0.779	0.787	0.748	-0.035	293
White	0.923	0.933	0.908	-0.02	341

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B18. Concordance correlations across years: Algebra 2 by subgroup (percentage proficient)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.944	0.939	0.663	-0.278	383
Female	0.907	0.905	0.681	-0.225	356
Male	0.918	0.918	0.598	-0.320	344
Economically disadvantaged students	0.837	0.810	0.430	-0.394	186
Non-economically disadvantaged students	0.922	0.925	0.648	-0.275	346
Students with disabilities	0.569	0.527	0.650	0.102	61
African American	0.778	0.880	0.452	-0.377	115
Asian	0.803	0.826	0.565	-0.250	134
Hispanic	0.843	0.864	0.507	-0.346	158
White	0.888	0.883	0.525	-0.360	294

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B19. Concordance correlations across years: Grade 4 ELA by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.489	0.536	0.530	0.018	1,270
Female	0.413	0.469	0.459	0.018	1,185
Male	0.436	0.461	0.445	-0.004	1,180
Economically disadvantaged students	0.375	0.387	0.356	-0.025	715
Non-economically disadvantaged students	0.408	0.466	0.473	0.036	985
English learners	0.227	0.199	0.366	0.153	42
Non-English learners	0.482	0.526	0.525	0.021	1,270
Students with disabilities	0.261	0.223	0.266	0.024	372
Students without disabilities	0.480	0.503	0.514	0.023	1,256
Black or African American	0.257	0.360	0.310	0.002	311
Hispanic	0.297	0.312	0.299	-0.006	541
White	0.413	0.423	0.447	0.029	855

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B20. Concordance correlations across years: Grade 4 math by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.492	0.513	0.519	0.017	1,270
Female	0.479	0.462	0.488	0.017	1,185
Male	0.407	0.468	0.454	0.017	1,180
Economically disadvantaged students	0.425	0.419	0.447	0.025	714
Non-economically disadvantaged students	0.418	0.480	0.429	-0.02	988
English learners	0.374	0.308	0.313	-0.028	58
Non-English learners	0.487	0.511	0.511	0.012	1,270
Students with disabilities	0.287	0.276	0.298	0.017	372
Students without disabilities	0.506	0.505	0.518	0.012	1,258
Black or African American	0.395	0.341	0.398	0.030	309
Hispanic	0.364	0.448	0.404	-0.002	544
White	0.417	0.423	0.423	0.003	855

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B21. Concordance correlations across years: Grade 5 ELA by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.444	0.516	0.483	0.003	1,176
Female	0.379	0.438	0.441	0.033	1,090
Male	0.389	0.484	0.419	-0.018	1,093
Economically disadvantaged students	0.367	0.374	0.435	0.064	665
Non-economically disadvantaged students	0.407	0.467	0.460	0.023	909
Non-English learners	0.438	0.512	0.482	0.007	1,175
Students with disabilities	0.248	0.309	0.282	0.004	370
Students without disabilities	0.421	0.506	0.492	0.028	1,161
Black or African American	0.231	0.315	0.312	0.039	299
Hispanic	0.365	0.372	0.399	0.031	506
White	0.433	0.448	0.431	-0.010	807

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B22. Concordance correlations across years: Grade 5 math by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.520	0.558	0.530	-0.009	1,176
Female	0.451	0.493	0.456	-0.016	1,088
Male	0.481	0.496	0.496	0.008	1,092
Economically disadvantaged students	0.457	0.456	0.426	-0.031	665
Non-economically disadvantaged students	0.503	0.536	0.508	-0.012	911
English learners	0.394	0.693	0.473	-0.071	34
Non-English learners	0.520	0.553	0.523	-0.013	1,174
Students with disabilities	0.340	0.415	0.378	0.001	367
Students without disabilities	0.516	0.561	0.532	-0.006	1,161
Black or African American	0.380	0.406	0.371	-0.022	299
Hispanic	0.445	0.415	0.425	-0.005	509
White	0.449	0.478	0.476	0.012	807

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B23. Concordance correlations across years: Grade 6 ELA by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.532	0.597	0.571	0.006	792
Female	0.489	0.531	0.532	0.022	723
Male	0.514	0.554	0.535	0.001	721
Economically disadvantaged students	0.492	0.534	0.512	-0.001	530
Non-economically disadvantaged students	0.516	0.540	0.589	0.061	602
Non-English learners	0.524	0.592	0.566	0.008	789
Students with disabilities	0.447	0.402	0.395	-0.029	362
Students without disabilities	0.520	0.565	0.556	0.014	775
Black or African American	0.416	0.399	0.336	-0.071	264
Hispanic	0.499	0.509	0.519	0.015	447
White	0.535	0.555	0.545	0.000	534

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B24. Concordance correlations across years: Grade 6 math by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.539	0.546	0.519	-0.023	791
Female	0.506	0.504	0.486	-0.019	723
Male	0.514	0.480	0.523	0.026	722
Economically disadvantaged students	0.496	0.480	0.442	-0.046	529
Non-economically disadvantaged students	0.505	0.521	0.543	0.030	603
English learners	0.179	0.421	0.203	-0.097	30
Non-English learners	0.535	0.542	0.513	-0.025	787
Students with disabilities	0.390	0.353	0.321	-0.050	362
Students without disabilities	0.544	0.530	0.509	-0.028	776
Black or African American	0.416	0.306	0.356	-0.005	263
Hispanic	0.452	0.465	0.469	0.010	449
White	0.503	0.493	0.502	0.004	534

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B25. Concordance correlations across years: Grade 7 ELA by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.481	0.515	0.545	0.047	689
Female	0.431	0.484	0.487	0.029	639
Male	0.483	0.484	0.542	0.059	640
Economically disadvantaged students	0.456	0.512	0.509	0.025	472
Non-economically disadvantaged students	0.419	0.499	0.526	0.067	532
Non-English learners	0.472	0.514	0.543	0.050	687
Students with disabilities	0.297	0.367	0.368	0.036	347
Students without disabilities	0.481	0.516	0.521	0.023	680
Black or African American	0.410	0.436	0.358	-0.065	265
Hispanic	0.442	0.382	0.540	0.128	411
White	0.427	0.505	0.540	0.074	467

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B26. Concordance correlations across years: Grade 7 math by subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.508	0.535	0.540	0.018	688
Female	0.498	0.460	0.500	0.021	637
Male	0.438	0.460	0.499	0.050	637
Economically disadvantaged students	0.440	0.376	0.435	0.027	473
Non-economically disadvantaged students	0.523	0.514	0.561	0.043	532
English learners	0.239	0.327	-0.028	-0.311	30
Non-English learners	0.510	0.531	0.544	0.024	685
Students with disabilities	0.381	0.441	0.338	-0.073	347
Students without disabilities	0.485	0.498	0.543	0.052	679
Black or African American	0.374	0.376	0.350	-0.025	261
Hispanic	0.403	0.432	0.429	0.012	411
White	0.512	0.521	0.528	0.012	464

Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B27. Concordance correlations across years: Grade 8 ELA subgroup (median student growth percentile)

Subgroup type	Concordance correlation by pairwise years: 2016 and 2017	2017 and 2018	2018 and 2019	Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
Total	0.268	0.342	0.364	0.059	681
Female	0.282	0.315	0.362	0.064	634
Male	0.260	0.321	0.335	0.045	638
Economically disadvantaged students	0.278	0.261	0.327	0.057	461
Non-economically disadvantaged students	0.245	0.311	0.361	0.083	540
Non-English learners	0.266	0.334	0.364	0.064	680
Students with disabilities	0.260	0.292	0.230	-0.046	339
Students without disabilities	0.267	0.338	0.355	0.052	673
Black or African American	0.254	0.211	0.277	0.045	259
Hispanic	0.248	0.248	0.287	0.039	399
White	0.221	0.331	0.403	0.127	471

ELA is English language arts.
Note: The sample size is the number of schools that had outcomes available for all four years for this test for each subgroup. Correlations were reported for subgroups that had a sample size of 20 or more. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B28. Concordance correlations across years: Aggregated across subject by outcome, including Algebra 2

	Concordance correlation by pairwise years			Difference between average of 2016 to 2017 and 2017 to 2018 (baseline) and 2018 to 2019 (test change)	Sample size
	2016 and 2017	2017 and 2018	2018 and 2019		
Math outcomes					
Mean scale score	0.916	0.919	0.901	-0.016	40,099
Percentage proficient	0.888	0.890	0.875	-0.014	40,099
Percentage proficient, Major Content	0.857	0.857	0.832	-0.025	40,099
Percentage proficient, Additional & Supporting Content	0.845	0.830	0.784	-0.053	40,099
Percentage proficient, Expressing Mathematical Reasoning 3	0.837	0.852	0.808	-0.036	40,099
Percentage proficient, Modeling & Application	0.805	0.829	0.783	-0.034	40,099
ELA outcomes					
Mean scale score	0.859	0.869	0.862	-0.002	37,250
Mean scale score, writing	0.887	0.892	0.887	-0.002	37,250
Mean scale score, reading	0.814	0.803	0.803	-0.005	37,250
Percentage proficient	0.888	0.894	0.893	0.002	37,250
Percentage proficient, Knowledge & Use of Language Conventions	0.852	0.861	0.857	0.000	37,250
Percentage proficient, Literary Text	0.821	0.822	0.811	-0.010	37,250
Percentage proficient, Informational Text	0.820	0.816	0.788	-0.030	37,250
Percentage proficient, Vocabulary	0.807	0.800	0.774	-0.030	37,250
Percentage proficient, Writing Expression	0.811	0.807	0.800	-0.009	37,250

ELA is English language arts.

Note: Correlations are based on all reported subgroups and tests apart from the total subgroup. The sample size is the number of schools by subgroup and by test that had outcomes available for all four years. The total group contains school by test records of outcomes for all students whereas all other groups contain school by test records of outcomes for students with a particular characteristic.

Source: Authors' analysis of data from the New Jersey Department of Education.

Table B29. Quintile change by test (percentage proficient)

Test	Average difference in percentage of schools changing one or more quintiles of performance between 2016 and 2017 and 2017 and 2018 (baseline) and 2018 and 2019 (test change) (percentage points)	Average difference in percentage of schools changing two or more quintiles of performance between 2016 and 2017 and 2017 and 2018 (baseline) and 2018 and 2019 (test change) (percentage points)
ELA		
Grade 3 ELA	-0.6	1.0
Grade 4 ELA	-4.2	-0.8
Grade 5 ELA	0.9	3.2
Grade 6 ELA	-1.2	-1.1
Grade 7 ELA	0.5	-0.8
Grade 8 ELA	-1.0	-0.5
Grade 9 ELA	-2.4	0.8
Grade 10 ELA	-0.1	-0.8
Math		
Grade 3 math	0.4	0.1
Grade 4 math	1.1	-0.3
Grade 5 math	-0.4	0.7
Grade 6 math	-3.2	-1.2
Grade 7 math	4.8	1.2
Grade 8 math	-1.2	-1.3
Algebra 1	-3.3	-1.0
Geometry	1.7	0.0
Algebra 2	7.9	5.7

ELA is English language arts.
Source: Authors' analysis of data from the New Jersey Department of Education.

Table B30. Concordance correlations across years: By size of reported subgroup (percentage proficient)

Quintile of number of valid student scores	Range of number of valid student scores in quintile	Concordance correlation by pairwise years			Difference between average of 2016 and 2017 and 2017 and 2018 (baseline) and 2018 and 2019 (test change)	Sample size
		2016 and 2017	2017 and 2018	2018 and 2019		
1	10–24	0.817	0.827	0.806	-0.016	18,213
2	25–36	0.864	0.869	0.862	-0.005	15,990
3	37–54	0.887	0.893	0.887	-0.003	16,762
4	55–92	0.910	0.912	0.907	-0.004	16,233
5	93–1,178	0.928	0.935	0.932	0.001	16,754

Note: Correlations are based on all reported subgroups that had valid score information for 2019.
Source: Authors' analysis of data from the New Jersey Department of Education

Table B31. Composition of students who took Algebra 1 over the years

Subgroup	Percentage of total students in 2016	Percentage of total students in 2017	Percentage of total students in 2018	Percentage of total students in 2019	Difference between 2019 and average of 2016 18 (percentage points)
Race/ethnicity					
White	46.9	45.7	44.9	44.2	-1.7
African American	15.5	15.3	15.3	14.6	-0.7
Asian	9.5	9.9	10.0	10.2	0.4
Pacific Islander	0.2	0.2	0.2	0.2	0.0
Hispanic	26.6	27.1	27.9	28.9	1.7
American Indian	0.1	0.1	0.1	0.1	0.0
Other	1.2	1.7	1.5	1.7	0.3
Gender					
Female	48.4	48.6	48.5	48.5	0.0
Male	51.6	51.4	51.5	51.5	0.0
Subgroup					
English learner	5.1	7.7	7.4	6.8	0.0
Current English learner	4.2	5.7	5.8	4.9	-0.3
Former English learner	0.9	1.9	1.7	1.8	0.4
Students with disabilities	14.8	14.9	15.1	15.0	0.0
Economically disadvantaged	36.6	35.3	35.9	33.9	-2.0
Non-economically disadvantaged	63.4	64.7	64.1	66.1	2.0
SE accommodation	14.1	13.4	13.9	13.9	0.1
Grade					
Grade 6	0.1	0.1	0.1	0.1	0.0
Grade 7	3.4	3.6	3.8	4.3	0.7
Grade 8	29.2	29.7	29.6	29.9	0.4
Grade 9	59.3	58.5	58.8	59.2	0.3
Grade 10	6.1	6.3	5.9	6.5	0.5
Grade 11	1.7	1.6	1.5	0.0	-1.6
Grade 12	0.3	0.3	0.3	0.0	-0.3

Source: Authors' analysis of data from the New Jersey Department of Education.

Table B32. Composition of students who took Geometry over the years

Subgroup	Percentage of total students in 2016	Percentage of total students in 2017	Percentage of total students in 2018	Percentage of total students in 2019	Difference between 2019 and average of 2016 18 (percentage points)
Race/ethnicity					
White	48.0	47.3	46.5	47.2	0.0
African American	15.4	14.9	15.0	13.6	-1.5
Asian	9.8	9.7	9.9	10.2	0.4
Pacific Islander	0.2	0.2	0.2	0.2	0.0
Hispanic	25.4	26.3	26.9	27.1	0.9
American Indian	0.1	0.1	0.1	0.1	0.0
Other	1.1	1.5	1.4	1.5	0.2
Gender					
Female	48.5	48.9	48.9	49.6	0.8
Male	51.5	51.1	51.1	50.4	-0.8
Subgroup					
English learner	4.5	6.4	6.3	5.5	-0.2
Current English learner	3.5	4.3	4.8	4.0	-0.2
Former English learner	1.0	2.0	1.4	1.5	0.0
Students with disabilities	14.7	14.7	15.0	13.2	-1.6
Economically disadvantaged	34.0	33.3	33.6	30.6	-3.0
Non-economically disadvantaged	66.0	66.7	66.4	69.4	3.0
SE accommodation	13.6	13.1	13.7	12.8	-0.7
Grade					
Grade 8	3.7	3.8	3.9	4.5	0.7
Grade 9	26.4	28.3	27.5	30.3	2.9
Grade 10	60.6	58.7	59.2	65.1	5.6
Grade 11	8.4	8.1	8.4	0.0	-8.3
Grade 12	0.8	1.0	1.0	0.0	-0.9

Source: Authors' analysis of data from the New Jersey Department of Education.