THE SCHOOL DISTRICT OF PHILADELPHIA

Relationship Between Star Assessments and PSSA Performance, 2022-23

Key Findings:

- During the 2022-23 school year, correlations between student performance on the Star Reading and the PSSA ELA assessments ranged from .74–.78, and correlations between student performance on the Star Math and the PSSA Math assessments ranged from .71–.80 across grades 3-8 and the Fall, Winter 1, and Spring testing windows. Overall, results suggested a strong positive relationship between Star Reading, Star Math, and the PSSA ELA and Math tests.
- Both Star Reading and Star Math were more accurate in classifying students who scored below the proficiency standard on the respective PSSA than they were in classifying students who scored proficient or higher: Across grades and testing windows, Star Reading correctly classified students who scored proficient or higher on the PSSA ELA between 60%–85% of the time, but correctly classified students who scored below the standard 86%–96% of the time. Star Math correctly classified students who scored proficient or higher on the PSSA Math between 55%–84% of the time, but correctly classified students who scored below the standard 91%–98% of the time.
- Among students who scored At/Above Benchmark (highest performance level) on Star Reading, between 73%–91% also scored proficient or higher on the PSSA ELA across grades and testing windows. For Star Math, between 68%–90% of students who scored At/Above Benchmark also scored proficient or higher on the PSSA Math.

Roland Reyes, MS
Senior Statistician

Office of Research and Evaluation

September 2024

Table of Contents

List of Tables

ACKNOWLEDGEMENTS

The author thanks Ebru Erdem, Michael Frisone, Andrew Munn, Soula Servello, Theodore Wills, and Joy Lesnick for helpful feedback and quality assurance support. The author also acknowledges the hard work of colleagues in the District Performance Office who maintain and process assessment data, and the hard work of colleagues across the Office of Evaluation, Research, and Accountability who support the administration of the Star Assessments and the PSSAs.

Introduction

The 2022-23 school year was the second year that the School District of Philadelphia (SDP or the District) used the suite of Star Assessments, published by Renaissance Learning Inc., as universal screeners for reading and math for students in grades K-12. This report focuses on student performance on two Star Assessments, Star Reading and Star Math, computer-adaptive tests (CATs) that are administered multiple times within the school year, and student performance on the end of year state standardized assessments, the Pennsylvania System of School Assessments or PSSAs.

The Star Reading and Star Math assessments serve multiple purposes in the District.[1] First, they are used to assess students' reading and math skills and to identify students who may need additional support in order to meet end-of-year state standards. Second, the Star CATs are used to track student growth over time. Third, they are used to monitor District-level progress toward meeting academic goals outlined in the Board of Education's Goals and Guardrails.[2] Specifically, because Star Reading and Star Math are administered multiple times within the school year, data from these tests are used to help understand how students are likely to perform on the end-of-year state assessment, the Pennsylvania System of School Assessments (PSSA) in English Language Arts (ELA) and Math. The purpose of this report is to examine the statistical relationship between Star Reading, Star Math, and the PSSAs.

About the Tests

Star Reading and Star Math are administered to all District students in grades 3-12 in testing windows that take place in the fall, winter, and spring. As CATs, the difficulty of items administered in a given testing session will depend on how well the student is performing on the test. Both tests comprise 34 multiple-choice questions that assess student performance across multiple domains aligned with the Common Core: [3,4,5]

- Star Reading Domains: Word Knowledge and Skills; Comprehension Strategies and Constructing Meaning; Analyzing Literary Text; Understanding Author's Craft; Analyzing Argument and Evaluating Text
- Star Math Domains: Numbers and Operations; Algebra; Geometry and Measurement; Data Analysis, Statistics, Probability

[1] See: https://www.philasd.org/era/assessment/star-information/
[2] For more information on the Board's Goals and Guardrails see: https://www.philasd.org/schoolboard/goals-and-guardrails/
[3] Renaissance Learning Inc. (2023). *Star Assessments for reading technical manual.* Renaissance Learning. https://renaissance.widen.net/view/pdf/xutuykoglg/SRRPTechnicalManual.pdf?t.download=true&u=zceria
[4] Renaissance Learning Inc. (2023). *Star Assessments for math technical manual.* Renaissance Learning. https://renaissance.widen.net/view/pdf/kxqxbuhbef/SMRPTechnicalManual.pdf?t.download=true&u=zceria
[5] For more details on the specific skills assessed by the Star CATs see: https://www.philasd.org/era/assessment/star-information/#1618402180282-71187e13-0e42

Results from the Star CATs are presented using multiple metrics that describe student achievement and growth.[6] The primary focus of this report are Star performance levels, which describe student performance using four categories: At/Above Benchmark, On Watch, Strategic Intervention, and Intensive Intervention. In the District, students who score At/Above Benchmark are considered on-track to meet end-of-year state standards on the PSSA.

The PSSA is a standards-based, criterion-referenced test administered near the end of the school year.[7] All Pennsylvania students in grades 3-8 are assessed in ELA and Math.[8] The PSSAs are designed to measure how well students acquired the knowledge and skills described in the Pennsylvania Anchor Content Standards as defined by the Eligible Content. Like Star, the PSSAs provide multiple metrics that describe student performance. The focus of this report are the PSSA performance levels: Advanced, Proficient, Basic, or Below Basic. Scores in the Advanced or Proficient range indicate that a student met grade level standards.[9]

Board Goals and Guardrails

In the 2020-21 school year, the Board of Education adopted the Goals and Guardrails, which outlines the Board's plan to improve student achievement.[10] Three of the five Goals in this framework focus on improving District-wide performance on the PSSA ELA and PSSA Math for all students in grades 3-8. To track progress toward these goals, metrics from Star Reading and Star Math are used to monitor student performance leading up to the PSSAs. These metrics, called *leading indicators*, include measures of student achievement and growth.[11] In this report, we focus on one of these leading indicators which uses the percentage of students who score At/Above Benchmark on Star as an estimate of the percentage of students likely to score Proficient or Advanced on the PSSA.

Research Questions

The purpose of this report is to examine the statistical relationship between student performance on Star Reading, Star Math, and the PSSAs. A prior SDP report demonstrated a statistical relationship using 2021-22 data, and the current report re-examines this relationship in the 2022-23 school year.[12]

---

[6] See: https://www.philasd.org/research/2022/06/09/star-tests-in-the-school-district-of-philadelphia-a-summary-of-metrics-that-describe-achievement-and-growth/

[7] Data Recognition Corporation. (2023). *2023 Pennsylvania System of School Assessment technical report: mathematics, English language arts, and science.* Data Recognition Corporation. https://www.education.pa.gov/Documents/K-12/Assessment%20and%20Accountability/PSSA/Technical%20Reports/2023%20PSSA%20Technical%20Report.pdf

[8] Students who are English Learners who attended school in the United States for less than 12 months by the end of the current year's PSSA testing window are not required to take the PSSA ELA. Students with significant cognitive disabilities might be eligible to take the Pennsylvania Alternate System of Assessment (PASA) instead of the PSSA. For more details see: https://www.education.pa.gov/K-12/Assessment%20and%20Accountability/PSSA/Pages/default.aspx

[9] For more information see: https://www.education.pa.gov/K-12/Assessment%20and%20Accountability/PSSA/Pages/DescriptorsCutScores.aspx

[10] See https://www.philasd.org/schoolboard/goals-and-guardrails/

[11] See https://www.philasd.org/era/goals-and-guardrails/

[12] See https://www.philasd.org/research/wp-content/uploads/sites/90/2023/07/Correlation-and-Classification-Accuracy-2021-22-Star-Computer-Adaptive-Tests-and-PSSAs-July-2023.pdf

To this end, we investigate the following research questions:

1. What was the correlation between student performance on the Star CATs and the PSSAs during the 2022-23 school year?
2. How accurately did performance on the Star CATs correctly classify students who scored Proficient or Advanced on the PSSAs during the 2022-23 school year?
3. What was the probability that students scored Proficient or Advanced on the PSSAs if they scored At/Above Benchmark on the Star CAT during the 2022-23 school year?

Method

Testing Windows

The Star tests were administered in four testing windows in the 2022-23 school year, but only the Fall, Winter 1, and the Spring testing windows were required (Table 1).[13] The PSSAs were administered in the spring of the 2022-23 school year. The following analyses focus on the three required Star testing windows.

Table 1. Testing windows in the 2022-23 school year

Testing Windows	Dates
Star Fall	9/6/2022 – 9/30/2022
Star Winter 1	1/4/2023 – 1/27/2023
Star Winter 2*	3/6/2023 – 3/24/2023
PSSA	4/24/2023 – 5/12/2023
Star Spring	5/1/2023 – 6/9/2023

*Optional

Note: The Fall, Winter 1, and Winter 2 testing windows were extended, and test results completed within those extensions were used in later analyses.

Source: 2022-23 School District of Philadelphia Assessment Calendar

Participants

All District students who took a PSSA test (ELA or Math) and the corresponding Star CAT (Reading or Math) in *any* required testing window were eligible for the analysis. For a given PSSA subject and grade level, three testing window samples were created by assigning students to the sample in which they had Star data (Fall, Winter 1, Spring). For instance, if student A took the PSSA ELA and had Star Reading data for the Fall, Winter 1, and Spring testing windows, then student A was included in all three testing window samples. If student B took the PSSA Math and had Star Math data for only the Fall and Winter 1 testing windows, then student B was only included in the Fall and Winter 1 testing window samples. Overall, 36 testing window samples were constructed (2 subjects X 3 testing windows X 6 grade levels). Note that if students took Star Reading or Star Math multiple times in a given testing window, then their latest and best score was used.

[13] For more information on District-wide participation on the Star and PSSA assessments in 2022-23 see: https://www.philasd.org/research/2024/07/08/a-summary-of-district-wide-assessments-administered-to-sdp-students-during-the-2022-23-school-year/

Students were excluded from the analyses for the following reasons. First, students who took the Pennsylvania Alternate System of Assessment (PASA) instead of the PSSA were excluded from the analysis in part because the PASA test does not report numeric test scores.[14] Second, scores from Star Spanish-Language tests were excluded because they are not on the same scale as scores from the Star English-Language tests.[15,16] This means that if a student took both a Spanish- and English-Language Star test in a given testing window, only the score from the English-Language test was used, even if the score from the Spanish-Language test was higher. If a student took only a Star Spanish-Language test for a given testing window, then they were excluded from that testing window sample.

Overall, at least 92% of students who took the PSSA ELA or PSSA Math test were included in any given Star testing window sample (Tables 2 and 3, respectively). Demographic characteristics and PSSA performance distributions for each testing window sample also did not deviate more than 2 percentage points from the overall District population of PSSA test-takers (see Appendix A for demographic details). The high percentage of students who took both the PSSAs and Star CATs was not surprising; participation for the individual tests was also high in 2022-23 (≥ 89% for the Star CATs across all three testing windows and ≥ 91% for the PSSAs).[17] Overall, the Winter 1 samples comprised the highest percentage of students who took the PSSAs (96%-97%), followed by Spring (95%-96%) then Fall (94%).

Table 2. Sample size for each Star Reading and PSSA ELA sample, 2022-23

Grade	Number who took PSSA ELA	Among students who took PSSA ELA, number who also completed Star Reading in the designated testing window					
		Fall		Winter 1		Spring	
		n	%	n	%	n	%
3	7,765	7,152	92%	7,515	97%	7,591	98%
4	7,836	7,519	96%	7,630	97%	7,670	98%
5	7,844	7,534	96%	7,661	98%	7,560	96%
6	7,352	6,943	94%	7,121	97%	6,979	95%
7	7,373	6,844	93%	7,068	96%	6,959	94%
8	7,403	6,883	93%	7,052	95%	6,775	92%
Total	45,573	42,875	94%	44,047	97%	43,534	96%

Note: % is the percentage of students who took the PSSA who also completed Star Reading in the given testing window. Only includes students with both a PSSA ELA score and a Star Reading score for the designated SDP testing window. Scores from a Star Spanish-Language test or from the PASA were excluded.

Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23; Qlik PSSA and Keystones app, accessed 11-6-23

[14] Students with significant cognitive disabilities might be eligible to take the Pennsylvania Alternate System of Assessment (PASA) instead of the PSSA. For more details on the PASA see: https://www.education.pa.gov/K-12/Special%20Education/Assessments/Pages/Pennsylvania-Alternate-System-of-Assessment-(PASA).aspx

[15] Renaissance Learning Inc. (2024). *Star Assessments for Spanish – Reading technical manual.* Renaissance Learning. https://renaissance.widen.net/view/pdf/isjf4ewjx8/SRSpTechnicalManual.pdf?t.download=true&u=zceria

[16] Renaissance Learning Inc. (2024). *Star Assessments for Spanish – Math technical manual.* Renaissance Learning. https://renaissance.widen.net/view/pdf/mllgnznrof/SMSpTechnicalManual.pdf?t.download=true&u=zceria

[17] Source: Qlik Academic Screeners app and Qlik PSSA and Keystone app, accessed 7/12/24

Table 3. Sample size for each Star Math and PSSA Math sample, 2022-23

Grade	Number who took PSSA Math	Among students who took PSSA Math, number who also completed Star Math in the designated testing window					
		Fall		Winter 1		Spring	
		n	%	n	%	n	%
3	7,946	7,485	94%	7,722	97%	7,759	98%
4	8,007	7,615	95%	7,738	97%	7,772	97%
5	7,965	7,543	95%	7,718	97%	7,628	96%
6	7,487	6,988	93%	7,204	96%	7,106	95%
7	7,497	6,887	92%	7,129	95%	7,026	94%
8	7,419	6,856	92%	6,962	94%	6,773	91%
Total	46,321	43,374	94%	44,473	96%	44,064	95%

Note: % is the percentage of students who took the PSSA who also completed Star Math in the given testing window. Only includes students with both a PSSA Math score and a Star Math score for the designated SDP testing window. Scores from a Star Spanish-Language test or from the PASA were excluded.
Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23; Qlik PSSA and Keystones app, accessed 11-6-23

Measures

Star Reading and Star Math

Star Unified scaled scores and Star performance levels were used for this analysis.[18,19] Star Unified scaled scores are on the Star Unified Scale, a vertical scale based on an item response theory (IRT) model that accounts for the difficulty of items. As a vertical scale, scores on the Unified Scale can be used to compare performance across grades and track growth over time. Star Unified scaled scores for Star Reading and Star Math range from 600-1400 across the entire vertical scale.

Star performance levels report student performance in one of four categories: At/Above Benchmark, On Watch, Strategic Intervention, and Intensive Intervention. Categories are based on students' national percentile rank (NPR), a score metric that indicates the student's standing on reading or math relative to a national sample of students who were in the same grade and who took the same Star test at around the same time. For the 2022-23 school year, the District's Star performance level cut scores were:

- Star Reading
  - At/Above Benchmark (≥ 40th NPR)
  - On Watch (25th to 39th NPR)
  - Strategic Intervention (24th to 10th NPR)
  - Intensive Intervention (< 10th NPR)

[18] Renaissance Learning Inc. (2023). *Star Assessments for reading technical manual.* Renaissance Learning. https://renaissance.widen.net/view/pdf/xutuykoglg/SRRPTechnicalManual.pdf?t.download=true&u=zceria
[19] Renaissance Learning Inc. (2023). *Star Assessments for math technical manual.* Renaissance Learning. https://renaissance.widen.net/view/pdf/kxqxbuhbef/SMRPTechnicalManual.pdf?t.download=true&u=zceria

- Star Math
  - At/Above Benchmark (≥ 70th NPR)
  - On Watch (25th to 69th NPR)
  - Strategic Intervention (24th to 10th NPR)
  - Intensive Intervention (< 10th NPR)

PSSA ELA and PSSA Math

PSSA scaled scores and PSSA performance levels were used for the analyses.[20] PSSA scaled scores are based on an IRT model that accounts for the difficulty of items, but scores are not on a vertical scale. This means that scores cannot be meaningfully compared across grade levels. All PSSA tests have a common minimum score of 600 but the max score depends on the grade, subject, and testing year. In 2023, the maximum score across grade levels for the PSSA Math was approximately 1525 and for PSSA ELA approximately 1596, on average.

PSSA performance levels report student performance in one of four categories: Advanced, Proficient, Basic, and Below Basic. Performance levels are defined by specific PSSA Scaled Scores, where a score greater than 1000 indicates a score of Proficient. Students who score Proficient or Advanced are considered to have met grade level standards.

Data Analysis

Two analyses were completed to investigate our research questions. The first analysis estimated Pearson correlations between Star Unified scaled scores and PSSA scaled scores for each subject, grade level, and testing window. The Pearson correlation coefficient ranges from -1 to +1 and expresses the linear relationship between two sets of scores.[21] Larger positive values, that is, values closer to +1, are desirable because they suggest stronger similarities in performance between the Star CATs and the PSSAs (e.g., students who performed well on one test were likely to perform well on the other test). We considered values of .70 or greater as suggesting a strong relationship between Star and PSSA performance.

The second analysis used classification accuracy metrics to evaluate (a) the accuracy with which the Star At/Above Benchmark performance level classified students who scored Proficient/Advanced or Basic/Below Basic on the PSSA and (b) the probability of scoring Proficient/Advanced on the PSSA given a score of At/Above Benchmark on Star. Four classification accuracy metrics were calculated: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).[22] All calculations were performed using collapsed versions of Star and PSSA performance levels: for Star,

[20] Data Recognition Corporation. (2023). *2023 Pennsylvania System of School Assessment technical report: mathematics, English language arts, and science.* Data Recognition Corporation. https://www.education.pa.gov/Documents/K-12/Assessment%20and%20Accountability/PSSA/Technical%20Reports/2023%20PSSA%20Technical%20Report.pdf

[21] Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory.* Waveland Press, Inc.

[22] Akobeng, A. K. (2007). Understanding diagnostic tests 1: Sensitivity, specificity, and predictive values. *Acta Paediatrica, 96,* 338-341. https://doi.org/10.1111/j.1651-2227.2006.00180.x. Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health, 5*(307), 1-7. https://doi.org/10.3389/fpubh.2017.00307

the resulting performance levels were At/Above Benchmark and Below Benchmark (includes On Watch, Strategic Intervention, Intensive Intervention), and for PSSA, the resulting performance levels were Proficient/Advanced and Basic/Below Basic.

- *Sensitivity* describes, among students who scored Proficient/Advanced on the PSSA, the proportion who scored At/Above Benchmark on Star.
- *Specificity* describes, among students who scored Basic/Below Basic on the PSSA, the proportion who scored Below Benchmark on Star.
- *Positive predictive value (PPV)* describes, among students who scored At/Above Benchmark on Star, the proportion who actually scored Proficient/Advanced on the PSSA.
- *Negative predictive value (NPV)* describes, among students who scored Below Benchmark on Star, the proportion who actually scored Basic/Below Basic on the PSSA.

Sensitivity and specificity were used to evaluate the accuracy with which Star classified students who scored Proficient/Advanced or Basic/Below Basic on the PSSA. High values of sensitivity mean that a large proportion of students who scored Proficient/Advanced on the PSSA were correctly classified by Star, and high values of specificity mean that a large proportion of students who scored Basic/Below Basic on the PSSA were correctly classified by Star. High values are desirable for both metrics because it reduces classification errors. But note that the metrics are inversely related, meaning that as one increases the other will decrease.

PPV and NPV (collectively, predictive values) were used to calculate the probability of scoring in one of the two collapsed PSSA performance levels given performance on Star. High values of PPV mean that there is a high probability of scoring Proficient/Advanced given a score of At/Above Benchmark, and high values of NPV means that there is a high probability of scoring Basic/Below Basic given a score of Below Benchmark. Note that PPV and NPV are partly dependent on sensitivity and specificity, and therefore should be interpreted together.

In summary, two analyses were completed to investigate our research questions. To investigate research question 1 ("What was the correlation between student performance on the Star CATs and the PSSAs during the 2022-23 school year?"), Pearson correlations were estimated. Research question 2 ("How accurately did performance on the Star CATs correctly classify students who scored Proficient or Advanced on the PSSAs during the 2022-23 school year?") was investigated through sensitivity and specificity metrics. To investigate research question 3 ("What was the probability that students scored Proficient or Advanced on the PSSAs if they scored At/Above Benchmark on the Star CAT during the 2022-23 school year?"), PPV and NPV were examined.

Results

Preliminary Analysis

Preliminary analyses suggested that students who were missing Star data in at least one testing window tended to have lower proficiency rates on the PSSAs when compared to students who took Star in all three windows. Because missing data can affect the results, additional analyses were

completed (not presented).[23] These analyses suggested that the results below were robust to missing data.

Star Reading and PSSA ELA

What was the correlation between Star Reading and the PSSA ELA?

The estimated correlation between Star Reading and the PSSA ELA was .77 across all grades and testing windows, on average (Table 4). Correlation coefficients ranged from .74 to .78 and tended to be stable across testing windows. Overall, results suggested a strong positive relationship between Star Reading and PSSA ELA performance.

Table 4. Correlation between Star Reading Unified Scale Scores and PSSA ELA scale scores in each Star testing window, 2022-23

Grade	Fall		Winter 1		Spring	
	n	Correlation	n	Correlation	n	Correlation
3	7,152	.74	7,515	.76	7,591	.75
4	7,519	.77	7,630	.77	7,670	.77
5	7,534	.77	7,661	.77	7,560	.78
6	6,943	.77	7,121	.77	6,979	.78
7	6,844	.77	7,068	.76	6,959	.77
8	6,883	.76	7,052	.77	6,775	.76

Note: Only includes students with a PSSA score and a Star score for the designated testing window. Star scores are the student's latest and best score. Scores from a Star Spanish-Language test or from the PASA were excluded. All correlations were statistically significant at $p < .001$.
Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23

How accurately did Star Reading classify students who scored Proficient or Advanced on the PSSA ELA?

Across all grades and testing windows, Star Reading tended to be more accurate when classifying students who scored Basic/Below Basic versus classifying students who scored Proficient/Advanced on the PSSA ELA (Table 5). Specifically, among students who scored Basic/Below Basic on the PSSA ELA, between 86%–96% also scored Below Benchmark, depending on grade and testing window (in other words, specificity ranged from 86%–96%). In comparison, between 60%–85% of students who

[23] Ancillary analyses involved using a statistical procedure called multiple imputation to address missingness. This process used 20 multiply imputed datasets that were generated using the R package MICE, with student demographic characteristics, PSSA performance, and Star performance as auxiliary variables. Correlation and classification accuracy analyses were repeated using the multiply imputed datasets and results differed less than .01 for correlation results and less than 2 percentage points for classification accuracy results, suggesting they were robust to missingness. For methodological details see: Collins, L. M., Schafer, J. L., & Kam, C. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6,* 330-351. https://doi.org/10.1037//1082-989X.6.4.330; Enders, C. K. (2022). *Applied missing data analysis* (2nd ed.). Guilford. For details on the MICE package see: van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45,* 1-67. https://www.jstatsoft.org/v45/i03/

scored Proficient/Advanced on the PSSA ELA were correctly classified by Star Reading across grades and testing windows (in other words, sensitivity ranged from 60%–85%). The relatively lower sensitivity rates mean that there was greater error when classifying students who scored Proficient/Advanced on the PSSA ELA.

Table 5. Classification accuracy metrics between Star Reading and PSSA ELA in each Star testing window, 2022-23

Grade	n	Star At/Above	PSSA Pro/Adv	Star – PSSA Difference	Correct Classifications	Sens	Spec	PPV	NPV
Fall									
3	7,152	29%	33%	-4	85%	71%	92%	81%	86%
4	7,519	29%	31%	-2	87%	76%	91%	79%	90%
5	7,534	27%	33%	-6	85%	69%	94%	84%	86%
6	6,943	27%	38%	-11	82%	62%	95%	88%	80%
7	6,844	27%	40%	-13	82%	61%	96%	91%	79%
8	6,883	26%	38%	-12	82%	60%	95%	89%	79%
Winter 1									
3	7,515	34%	32%	2	87%	81%	89%	78%	91%
4	7,630	34%	30%	4	87%	84%	88%	75%	93%
5	7,661	31%	33%	-2	86%	76%	91%	80%	89%
6	7,121	29%	38%	-9	84%	67%	94%	87%	83%
7	7,068	29%	39%	-10	83%	65%	95%	89%	81%
8	7,052	27%	38%	-11	82%	62%	95%	88%	80%
Spring									
3	7,591	35%	32%	3	87%	85%	88%	77%	93%
4	7,670	35%	30%	5	86%	85%	86%	73%	93%
5	7,560	32%	32%	0	86%	78%	90%	79%	90%
6	6,979	29%	37%	-8	84%	68%	94%	87%	83%
7	6,959	28%	40%	-12	83%	64%	96%	90%	80%
8	6,775	27%	38%	-11	83%	63%	95%	88%	81%

Note: Only includes students with both a PSSA ELA score and a Star Reading score for the designated SDP testing window. Star scores are the student's latest and best score. Scores from a Star Spanish-Language test or from the PASA were excluded. Star At/Above = Percentage scoring At or Above Benchmark on Star. PSSA P/A = Percentage scoring Proficient or Advanced on the PSSA. Star – PSSA Difference is the difference between the Star At/Above column and the PSSA P/A column. Correct Classifications = the percentage of students who a) scored At/Above Benchmark on Star and scored P/A on PSSA ELA and b) scored Below Benchmark on Star and scored Basic or Below Basic on the PSSA. Sens = Sensitivity. Spec = Specificity. PPV = Positive Predictive Value. NPV = Negative Predictive Value.

Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23

Sensitivity rates were consistently lower in grades 6-8 than in grades 3-5, meaning that there was a higher rate of error when classifying students in grades 6-8 who scored Proficient/Advanced on the PSSA ELA. Specifically, in grades 6-8, between 32%-40% of students who scored Proficient/Advanced on the PSSA ELA scored Below Benchmark on Star Reading; in grades 3-5, between 15%-31% of students who scored Proficient/Advanced on the PSSA ELA scored Below Benchmark on Star Reading. Most of these students scored in the On Watch category on Star rather than Strategic or Intensive Intervention.

Another way to interpret sensitivity and specificity metrics is to compare them against a threshold. While we are not aware of any universally agreed upon standard, the National Center for Intensive Intervention (NCII), which rates the technical adequacy of academic screening tools for the purpose of identifying students at risk for later academic difficulty, uses a minimum threshold of 70%, with values of 80% or greater being desirable (see also work by Forcht and Van Norman [2023] who analyzed the classification accuracy of the Star CATs against other state assessments).[24] For students in grades 6-8 at SDP, sensitivity rates never met the 70% threshold in any testing window, suggesting that lowering the cut score for At/Above Benchmark may improve sensitivity rates for these grade levels; however, note that changing the cut score will affect all other classification metrics (e.g., increased error when classifying students who scored Basic/Below Basic) and a thorough investigation of the consequences of changing cut scores should be considered before changes are made.[25]

Star Reading tended to be more accurate when classifying students who scored Proficient/Advanced in the Winter 1 and Spring testing windows than in the Fall. Sensitivity was lowest in the Fall (averaging 67% across grades) then increased and remained stable in Winter 1 (averaging 73%) and Spring (averaging 74%). Conversely, specificity was high throughout all testing windows, meaning that Star Reading was quite accurate when classifying students who scored Basic/Below Basic throughout the school year (averaging 94% across grades in the Fall, and averaging 92% in Winter 1 and Spring).

[24] National Center for Intensive Intervention (2020). *Academic screening tools chart rating rubric.* https://intensiveintervention.org/sites/default/files/NCII_AcademicScreening_RatingRubric_2020-06-30.pdf ; Forcht, E. R., & Van Norman, E. R. (2023). Comparison of screening methods for computer adaptive tests to predict reading and math performance. *Psychology in the Schools, 61,* 1590-1610. https://doi.org/10.1002/pits.23132

[25] Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45,* 117-135. https://doi.org/10.1016/j.jsp.2006.05.005; Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health, 5*(307), 1-7. https://doi.org/10.3389/fpubh.2017.00307

What was the probability that students scored Proficient or Advanced on the PSSA ELA if they scored At/Above Benchmark on Star Reading?

Results for PPV and NPV differed by testing window (Table 5). Average PPV across grade levels was highest in the Fall (85%) compared to Winter 1 (83%) and Spring (82%), but at the same time, average NPV was lowest in the Fall (83%) compared to Winter 1 (86%) and Spring (87%).

These results are related to changes in classification accuracy throughout the school year. In the Fall, fewer students scored At/Above Benchmark on Star, and the PPV shows that a higher percentage of those students also scored Proficient/Advanced on the PSSA ELA. But at the same time, there was a higher rate of error when classifying Proficient/Advanced students in the Fall (i.e., lower sensitivity), meaning that compared to the Winter 1 and Spring windows, a higher proportion of Proficient/Advanced students scored Below Benchmark. The consequence of this is reflected in the lower NPV, suggesting that in the Fall, there was a higher chance that students who scored Below Benchmark on Star still scored Proficient/Advanced on the PSSA ELA (17% chance in Fall versus 14% in Winter 1 and 13% in Spring).[26] As more students scored At/Above Benchmark in the Winter 1 and Spring testing windows, sensitivity rates also increased, meaning that the correct identification of Proficient/Advanced students increased; however, a somewhat higher percentage of Below/Basic students were also scoring At/Above Benchmark on Star, and these changes are reflected in changes in PPV and NPV. Thus, compared to the Fall, probability estimates in Winter 1 and Spring were associated with higher accuracy when classifying Proficient/Advanced students and somewhat lower accuracy when classifying Basic/Below Basic students.

Consistent with results for sensitivity and specificity, notable differences were observed by grade band. NPVs were lower for grades 6-8 than they were for grades 3-5, suggesting that a sizeable proportion of students in grades 6-8 who scored Below Benchmark on Star scored Proficient/Advanced on the PSSA ELA (most of these students had scored in the On Watch category on Star). For instance, in Winter 1, the NPV for 8th grade students was 80%, meaning that 20% of students who scored Below Benchmark still scored Proficient/Advanced on the PSSA. As noted above, further analyses are needed to examine how lowering the At/Above Benchmark cut score for grades 6-8 changes these predictive values.

[26] Calculated as (1 – NPV)

Star Math and PSSA Math

What was the correlation between Star Math and the PSSA Math?

The estimated correlation between Star Math and PSSA Math was .76 across all grades and windows, on average (Table 6). Correlation coefficients ranged from .71 to .80 and tended to be stable across windows. Overall, results suggested a strong positive relationship between Star Math and PSSA Math performance.

Table 6. Correlation between Star Math Unified Scale scores and PSSA Math scale scores in each Star testing window, 2022-23

Grade	Fall		Winter 1		Spring	
	n	Correlation	n	Correlation	n	Correlation
3	7,485	.80	7,722	.80	7,759	.79
4	7,615	.78	7,738	.80	7,772	.79
5	7,543	.77	7,718	.77	7,628	.76
6	6,988	.77	7,204	.76	7,106	.77
7	6,887	.75	7,129	.75	7,026	.74
8	6,856	.73	6,962	.71	6,773	.71

Note: Only includes students with a PSSA score and a Star score for the designated testing window. Star scores are the student's latest and best score within a given window. Scores from a Star Spanish-Language test or from the PASA were excluded. All results are statistically significant at $p < .001$.
Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23

How accurately did Star Math classify students who scored Proficient or Advanced on the PSSA Math?

Sensitivity and specificity results for Star Math followed similar trends to those for Star Reading, but recall that the cut score that defines At/Above Benchmark for Star Math was the 70th NPR while the cut score used for Star Reading was the 40th NPR. Like Star Reading, Star Math tended to be more accurate when classifying students who scored Basic/Below Basic versus classifying students who scored Proficient/Advanced, where specificity was consistently higher than sensitivity across all testing windows (Table 7). Specifically, among students who scored Basic/Below Basic on the PSSA Math, between 91%–98% also scored Below Benchmark on Star Math, depending on grade and testing window (in other words, specificity ranged from 91%–98%). In comparison, between 55%–84% of students who scored Proficient/Advanced on the PSSA Math were correctly classified (in other words, sensitivity ranged from 55%–84%, depending on grade and testing window). Thus, there was a higher rate of error when classifying students who scored Proficient/Advanced on the PSSA Math. Unlike Star Reading, however, clear differences in classification accuracy were not observed by grade band. This might be related to the different cut scores used to define At/Above Benchmark on Star Reading and Star Math.

Table 7. Classification accuracy metrics between Star Math and PSSA Math in each Star testing window, 2022-23

Grade	n	Star At/Above	PSSA Pro/Adv	Star – PSSA Difference	Correct Classifications	Sens	Spec	PPV	NPV
Fall									
3	7,485	17%	28%	-11	86%	56%	98%	90%	85%
4	7,615	17%	23%	-6	88%	60%	97%	84%	89%
5	7,543	14%	22%	-8	88%	56%	97%	85%	89%
6	6,988	12%	18%	-6	90%	55%	98%	83%	91%
7	6,887	14%	20%	-6	90%	61%	97%	85%	91%
8	6,856	13%	17%	-4	91%	61%	97%	81%	92%
Winter 1									
3	7,722	22%	27%	-5	89%	70%	96%	86%	90%
4	7,738	22%	23%	-1	89%	73%	93%	77%	92%
5	7,718	20%	21%	-1	90%	74%	95%	79%	93%
6	7,204	15%	17%	-2	90%	66%	95%	75%	93%
7	7,129	17%	20%	-3	91%	70%	96%	81%	93%
8	6,962	14%	16%	-2	91%	65%	96%	73%	93%
Spring									
3	7,759	24%	27%	-3	88%	72%	94%	82%	90%
4	7,772	25%	23%	2	89%	81%	91%	74%	94%
5	7,628	24%	21%	3	90%	84%	92%	73%	95%
6	7,106	21%	18%	3	90%	80%	92%	68%	96%
7	7,026	20%	20%	0	91%	77%	94%	76%	94%
8	6,773	17%	16%	1	91%	72%	94%	70%	95%

Note: Only includes students with both a PSSA Math score and a Star Math score for the designated SDP testing window. Star scores are the student's latest and best score. Scores from a Star Spanish-Language test or from the PASA were excluded. Star At/Above = Percentage scoring At or Above Benchmark on Star. PSSA P/A = Percentage scoring Proficient or Advanced on the PSSA. Star – PSSA Difference is the difference between the Star At/Above column and the PSSA P/A column. Correct Classifications = the percentage of students who a) scored At/Above Benchmark on Star and scored P/A on PSSA ELA and b) scored Below Benchmark on Star and scored Basic or Below Basic on the PSSA. Sens = Sensitivity. Spec = Specificity. PPV = Positive Predictive Value. NPV = Negative Predictive Value.

Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23

Like Star Reading, Star Math tended to become more accurate classifying students who scored Proficient/Advanced as the year progressed. Specifically, sensitivity averaged 58% across grades in the Fall, 70% in Winter 1, and 78% in the Spring. In contrast, Star Math accurately classified a high percentage of students who scored Basic/Below throughout the year, where specificity averaged 97% in the Fall, 95% in Winter 1, and 93% in the Spring.

What was the probability that students scored Proficient or Advanced on the PSSA Math if they scored At/Above Benchmark on Star Math?

PPV and NPV results for Star Math followed patterns similar to Star Reading. PPV and NPV differed by testing window, where average PPV across grades was highest in the Fall (85%) then decreased in Winter 1 (79%) and Spring (74%); at the same time, average NPV was lowest in the Fall (90%) and increased in Winter 1 (92%) and Spring (94%) (Table 7). Therefore, there was a higher chance that when students scored Below Benchmark in Star Math in the Fall, they would also score Proficient/Advanced on the PSSA Math (10% for Fall versus 8% in Winter 1 and 6% in Spring). Similar to Star Reading, probability estimates in the Winter 1 and Spring were associated with higher accuracy when classifying Proficient/Advanced students and somewhat lower accuracy when classifying Basic/Below Basic students when compared to the Fall.

Limitations

It is important to interpret the results presented here in light of their limitations. First, correlations between student performance on the Star CATs and the PSSA tests do not imply a causal relationship; that is, we cannot say that high performance on Star CATs will result in high performance on the PSSA tests. Rather, high correlations suggest an association between performance on both tests but it does not rule out that another factor (e.g., student motivation) drives that association. Second, we remind readers that our analyses were not able to include data from all District students and all testing windows due to missing Star data. Furthermore, students who were missing data in at least one Star testing window tended to have lower PSSA proficiency rates than students with Star scores in all three testing windows. However, testing window samples for all Star windows and tests were still representative of District PSSA test-takers in terms of gender, race/ethnicity, economic disadvantage status, English Learner status, special education status, and PSSA performance (see Appendix A), which supports the generalizability of results, despite missing data.

Summary

This analysis aimed to examine the statistical relationship between student performance on Star Reading, Star Math, and the PSSA ELA and Math tests, using data from the 2022-23 school year. Analyses focused on two key areas: the extent to which student performance on Star was similar to performance on the respective PSSA, and the accuracy with which the Star At/Above Benchmark performance level—defined as scores ≥ 40th NPR on Star Reading and ≥ 70th NPR on Star Math—classified student performance on the PSSAs.

Correlations between the Star CATs and the PSSAs suggested that student performance on the two tests was strongly related, where correlations between Star Reading and the PSSA ELA averaged .77 (range = .74–.78) and correlations between Star Math and PSSA Math averaged .76 (range = .71–.80)

across grades and testing windows. These results are similar to those from the 2021-22 school year, suggesting that the statistical relationship between the two tests was stable year over year.[27]

When evaluating the accuracy of the At/Above Benchmark performance level using sensitivity and specificity metrics, both Star Reading and Star Math were more accurate when classifying students who scored Basic/Below Basic than students who scored Proficient/Advanced on the respective PSSA. When classifying students who scored Proficient/Advanced, results showed that Star was least accurate in the Fall testing window, but accuracy improved throughout the school year; when classifying students who scored Basic/Below Basic, accuracy was consistently high (> 90%) across all testing windows. These results are consistent with results from the 2021-22 school year, suggesting that the cut scores defining the At/Above Benchmark performance level on Star Reading and Star Math do well to identify students who are at risk of scoring Basic/Below Basic but may also under-identify students who score Proficient/Advanced. It is important to keep in mind that the Star CATs are designed for multiple purposes, and while the focus of these analyses was on their role as leading indicators in the Board's Goals of Guardrails, results also suggested that Star Reading and Star Math do quite well for the purpose of identifying students who may not meet end-of-year standards.

PPV and NPV results revealed that PPV tended to be highest in the Fall and decreased throughout the school year, while NPV tended to be lowest in the Fall and increased throughout the year. These trends are related to changes in classification accuracy and are complex to interpret. Essentially, the predictive values partly reflect how well Star classified students in each testing window. In the Fall, At/Above rates tended to be at their lowest, and accompanying sensitivity rates suggested that this was partly because a higher proportion of Proficient/Advanced students were classified as Below Benchmark on Star in the Fall window. As the group of students who scored At/Above Benchmark increased in the Winter 1 and Spring testing windows, so did sensitivity rates, meaning that the correct identification of Proficient/Advanced students increased; at the same time, a higher proportion of Basic/Below Basic students were classified as At/Above Benchmark, reducing PPV rates while increasing NPV. Thus, compared to the Fall, probability estimates in the Winter 1 and Spring were associated with higher sensitivity and somewhat lower specificity.

Lastly, classification accuracy results showed that for Star Reading, sensitivity rates were consistently low for grades 6-8 compared to grades 3-5 (< 70%). While this suggests that lowering the At/Above Benchmark cut score for grades 6-8 may improve identification of students who scored Proficient/Advanced on the PSSA ELA, any modifications require a thorough consideration of the purposes of the Star assessment and the consequences of changing cut scores, among other factors.[28]

[27] For results using 2021-22 data see: https://www.philasd.org/research/2023/07/17/analysis-of-the-relationship-between-2021-22-star-assessment-and-pssa-assessment-performance/

[28] Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45,* 117-135. https://doi.org/10.1016/j.jsp.2006.05.005; Trevethan, R. (2017). Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health, 5*(307), 1-7. https://doi.org/10.3389/fpubh.2017.00307

Conclusion

Overall, results suggested that student performance on Star Reading and Star Math were strongly related to student performance on the respective PSSA test. Results were similar to those reported in 2021-22, suggesting that the relationship between the Star CATs and the PSSAs is stable year-over-year among SDP students. Results suggested that future analysis is needed to better understand the relationship between scoring Below Benchmark on Star Reading and Proficient or Advanced on the PSSA ELA. It may be of particular interest to focus on how scoring in the On Watch performance level (which indicates that students are only slightly below the benchmark) is related to PSSA performance.

Appendix A: Demographic Characteristics and Descriptive Statistics

As noted in the Participants section, each testing window sample was similar to the group of students who took the PSSA ELA (Table A1) or PSSA Math (Table A2).

Table A1. Demographic characteristics for all students who took the PSSA ELA and for each sample who took the PSSA ELA and Star Reading in the designated testing window, 2022-23

Demographic Characteristic	Demographic Category	PSSA (n = 45,573)	Star Testing Window		
			Fall (n = 42,875)	Winter 1 (n = 44,047)	Spring (n = 43,534)
Grade Level	3	17%	17%	17%	17%
	4	17%	17%	18%	17%
	5	17%	17%	18%	17%
	6	16%	16%	16%	16%
	7	16%	16%	16%	16%
	8	16%	16%	16%	16%
Gender	Female	49%	49%	49%	49%
	Male	51%	51%	51%	51%
Race/Ethnicity	Asian	10%	10%	10%	10%
	Black/African American	44%	44%	44%	44%
	Hispanic/Latino	26%	26%	26%	26%
	Multi-Racial/Other	4%	4%	4%	4%
	White	15%	15%	16%	15%
Economic Disadvantage	Economically Disadvantaged	79%	79%	78%	79%
	Not Economically Disadvantaged	21%	21%	22%	21%
English Learner	English Learner	17%	17%	17%	17%
	Not an English Learner	83%	83%	83%	83%
Special Education	Has an IEP	17%	17%	16%	16%
	Does not have an IEP	83%	83%	84%	84%

Note: The PSSA column includes all District students who completed the PSSA. Each Star Testing Window column only includes students with both a PSSA ELA score and a Star Reading score for the designated testing window. Scores from a Star Spanish-Language test or from the PASA were excluded. Students who are American Indian/Alaskan Native or Native Hawaiian/Pacific Islander each comprised < 1% of the sample and are included in the Multi-Racial/Other category. Non-Binary students comprised < 1% of the sample. IEP = Individualized Education Plan. Category *Has an IEP* does not include students with gifted IEPs. Categories may not sum to 100% due to rounding error.

Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23; Qlik PSSA and Keystones app, accessed 11-6-23; Qlik Report Library, Total Student Enrollment Yearly, downloaded 11-6-23

Table A2. Demographic characteristics for all students who took the PSSA Math and for each sample who took the PSSA Math and Star Math in the designated testing window, 2022-23

Demographic Characteristic	Demographic Category	PSSA (n = 46,321)	Star Testing Window		
			Fall (n = 43,374)	Winter 1 (n = 44,473)	Spring (n = 44,064)
Grade Level	3	17%	17%	17%	18%
	4	17%	18%	17%	18%
	5	17%	17%	17%	17%
	6	16%	16%	16%	16%
	7	16%	16%	16%	16%
	8	16%	16%	16%	15%
Gender	Female	49%	49%	49%	49%
	Male	51%	51%	51%	51%
Race/Ethnicity	Asian	10%	11%	10%	10%
	Black/African American	44%	44%	44%	44%
	Hispanic/Latino	26%	26%	26%	26%
	Multi-Racial/Other	4%	4%	4%	4%
	White	15%	16%	16%	16%
Economic Disadvantage	Economically Disadvantaged	78%	78%	78%	78%
	Not Economically Disadvantaged	22%	22%	22%	22%
English Learner	English Learner	18%	17%	18%	18%
	Not an English Learner	82%	83%	82%	82%
Special Education	Has an IEP	17%	16%	17%	16%
	Does not have an IEP	83%	84%	83%	84%

Note: The PSSA column includes all District students who completed the PSSA. Each Star Testing Window column only includes students with both a PSSA Math score and a Star Math score for the designated testing window. Scores from a Star Spanish-Language test or from the PASA were excluded. Students who are American Indian/Alaskan Native or Native Hawaiian/Pacific Islander each comprised < 1% of the sample and are included in the Multi-Racial/Other category. Non-Binary students comprised < 1% of the sample. IEP = Individualized Education Plan. Category *Has an IEP* does not include students with gifted IEPs. Categories may not sum to 100% due to rounding error.
Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23; Qlik PSSA and Keystones app, accessed 11-6-23; Qlik Report Library, Total Student Enrollment Yearly, downloaded 11-6-23

In terms of performance, differences in the distribution of PSSA ELA performance levels between all students District-wide and students in each PSSA ELA/Star Reading sample were smaller than 1 percentage point across grades and testing windows, on average, with the largest differences not exceeding more than 2 points (Tables A3 and A4). Overall, between 30% to 40% of students scored Proficient/Advanced on the PSSA ELA, and between 26% to 35% of students scored At/Above Benchmark on Star Reading across all testing window samples.

Table A3. Performance level distributions for students with both a PSSA ELA score and a Star Reading score in each Star testing window, 2022-23

		PSSA					Star			
Grade	**n**	**Pro + Adv**	**Adv**	**Pro**	**Basic**	**Below Basic**	**At/ Above**	**On Watch**	**Strategic**	**Intensive**
Fall										
3	7,152	33%	6%	27%	39%	28%	29%	11%	15%	44%
4	7,519	31%	10%	20%	34%	35%	29%	11%	15%	44%
5	7,534	33%	6%	27%	36%	31%	27%	13%	19%	42%
6	6,943	38%	11%	27%	46%	16%	27%	12%	18%	43%
7	6,844	40%	13%	27%	51%	9%	27%	12%	19%	43%
8	6,883	38%	11%	28%	37%	24%	26%	14%	19%	41%
Winter 1										
3	7,515	32%	6%	26%	38%	30%	34%	13%	15%	38%
4	7,630	30%	10%	20%	34%	36%	34%	11%	16%	38%
5	7,661	33%	6%	27%	36%	32%	31%	14%	18%	37%
6	7,121	38%	11%	27%	46%	16%	29%	13%	18%	39%
7	7,068	39%	12%	27%	52%	9%	29%	12%	19%	40%
8	7,052	38%	10%	28%	37%	25%	27%	14%	19%	40%
Spring										
3	7,591	32%	6%	26%	39%	30%	35%	11%	16%	39%
4	7,670	30%	10%	20%	34%	36%	35%	11%	15%	38%
5	7,560	32%	6%	27%	36%	32%	32%	12%	17%	39%
6	6,979	37%	11%	26%	47%	16%	29%	12%	18%	40%
7	6,959	40%	12%	27%	52%	9%	28%	12%	18%	42%
8	6,775	38%	10%	28%	37%	25%	27%	11%	18%	43%

Note: Only includes students with both a PSSA ELA score and a Star Reading score for the designated SDP testing window. Star scores are the student's latest and best score. Scores from a Star Spanish-Language test or from the PASA were excluded. For the PSSA performance levels, Adv = Advanced, Pro = Proficient, Basic = Basic, Below Basic = Below Basic. For the Star performance levels, At/Above = At or Above Benchmark, On Watch = On Watch, Strategic = Strategic Intervention, Intensive = Intensive Intervention.

Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23

Table A4. PSSA ELA performance level distributions District-wide, 2022-23

Grade	N	Advanced	Proficient	Basic	Below Basic
3	7,765	6%	25%	38%	30%
4	7,836	10%	20%	34%	36%
5	7,844	6%	26%	36%	32%
6	7,352	11%	26%	46%	17%
7	7,373	12%	26%	52%	10%
8	7,403	10%	27%	37%	26%

Note: Rows may not sum to 100% due to rounding.

Source: Qlik PSSA and Keystones app, accessed 11-6-23. PASA excluded

Similarly, differences in the distribution of PSSA Math performance levels between all students District-wide and students in each PSSA Math/Star Math sample were less than 1 percentage point across grades and testing windows, on average, with the largest differences not exceeding more than 2 points (Tables A5 and A6). Overall, between 16% to 28% of students scored Proficient/Advanced on the PSSA Math, and between 12% to 25% of students scored At/Above Benchmark on Star Math across all testing window samples.

Table A5. Performance level distributions for students with both a PSSA Math score and a Star Math score in each Star testing window, 2022-23

		PSSA					Star			
Grade	**n**	**Pro + Adv**	**Adv**	**Pro**	**Basic**	**Below Basic**	**At/ Above**	**On Watch**	**Strategic**	**Intensive**
Fall										
3	7,485	28%	10%	18%	22%	50%	17%	29%	16%	38%
4	7,615	23%	7%	16%	27%	50%	17%	26%	20%	38%
5	7,543	22%	7%	14%	26%	52%	14%	27%	18%	40%
6	6,988	18%	6%	11%	26%	56%	12%	29%	20%	39%
7	6,887	20%	8%	12%	20%	59%	14%	29%	21%	36%
8	6,856	17%	7%	10%	17%	67%	13%	33%	22%	32%
Winter 1										
3	7,722	27%	10%	17%	22%	51%	22%	31%	16%	31%
4	7,738	23%	7%	16%	27%	50%	22%	29%	19%	31%
5	7,718	21%	7%	14%	26%	52%	20%	29%	17%	34%
6	7,204	17%	6%	11%	26%	57%	15%	31%	19%	35%
7	7,129	20%	8%	11%	20%	60%	17%	30%	19%	33%
8	6,962	16%	6%	10%	16%	68%	14%	35%	21%	30%
Spring										
3	7,759	27%	10%	17%	22%	51%	24%	28%	17%	32%
4	7,772	23%	7%	16%	27%	50%	25%	28%	15%	32%
5	7,628	21%	7%	14%	26%	52%	24%	26%	16%	33%
6	7,106	18%	6%	11%	26%	56%	21%	28%	16%	35%
7	7,026	20%	8%	11%	20%	60%	20%	29%	17%	34%
8	6,773	16%	6%	10%	16%	67%	17%	33%	18%	33%

Note: Only includes students with both a PSSA Math score and a Star Math score for the designated testing window. Star scores are the student's latest and best score. Scores from a Star Spanish-Language test or from the PASA were excluded. For the PSSA performance levels, Adv = Advanced, Pro = Proficient, Basic = Basic, Below Basic = Below Basic. For the Star performance levels, At or Above = At or Above Benchmark, On Watch = On Watch, Strategic = Strategic Intervention, Intensive = Intensive Intervention.

Source: Qlik Report Library, Academic Screeners, downloaded 8-27-23; Qlik Report Library, PSSA & Keystone, downloaded 10-25-23

Table A6. PSSA Math performance level distributions District-wide, 2022-23

Grade	N	Advanced	Proficient	Basic	Below Basic
3	7,946	9%	17%	22%	52%
4	8,007	7%	16%	26%	51%
5	7,965	7%	14%	26%	53%
6	7,487	6%	11%	26%	57%
7	7,497	8%	11%	20%	61%
8	7,419	6%	9%	16%	68%

Note: Rows may not sum to 100% due to rounding.
Source: Qlik PSSA and Keystones app, accessed 11-6-23. PASA excluded