RENAISSANCE®

Understanding fluctuations in test scores

When interpreting the results of any test instrument, it is important to remember that the scores represent estimates of a student's ability level. Test scores are not absolute or exact measures of performance, nor is a single test score infallible in the information that it provides. When analyzing data for a student who did not score as expected, consider fidelity of test administration, the student's mindset, and statistical reasoning.

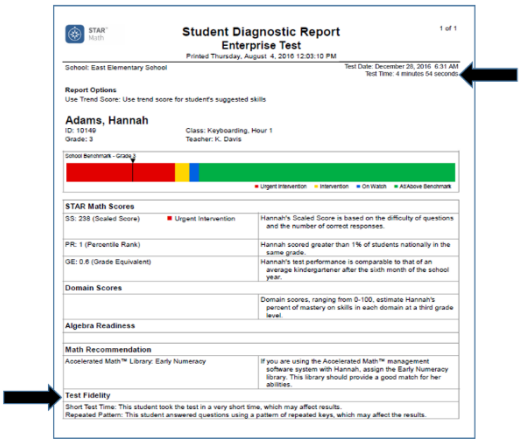
Fidelity of test administration and student mindset

Most questions surrounding an unexpected test score can be answered by looking at the fidelity of test administration and the student's mindset at the time of the test.

- Did testing conditions change from the previous testing event?
- How long did it take the student to complete the test?
- Were there distractions during testing?
- What was the student's frame of mind during testing?
- Was the student excited or distressed about an upcoming event?

To ensure reliable data, it is important that Renaissance Star™ assessments are administered with consistency and under conditions that set students up for success. In the software, you will find pretest instructions for each Star assessment that are designed to help you match the norming conditions. They include a checklist, script, and screenshots to help explain to students how to take the test. For additional considerations about test fidelity, view How to Administer the Test with Fidelity.

The Student Diagnostic Report (Renaissance Star Reading® and Renaissance Star Math®) displays a variety of test fidelity indicators. The amount of time a student took to complete the assessment is located near the top of the report. Students generally complete a Star Reading assessment in approximately 15 minutes, and a Star Math test in approximately 20 minutes. Test times may vary depending on the grade level. At the bottom of the Student Diagnostic report, a test fidelity alert may be present. This signals that the student took the test in a short amount of time and/or answered the questions using a pattern of repeated keys, which may affect the results.


Student Diagnostic Report
Enterprise Test
Printed Thursday, August 4, 2016 12:03:10 PM
1 of 1
School: East Elementary School
Test Date: December 20, 2016 6:31 AM
Test Time: 4 minutes 54 seconds
Report Options
Use Trend Score: Use trend score for student's suggested skills
Adams, Hannah
ID: 10149
Grade: 3
Class: Keyboarding, Hour 1
Teacher: K. Davis
School Benchmark - Grade 3
Urgent Intervention
Intervention
On Watch
At/Above Benchmark
STAR Math Scores
SS: 298 (Scaled Score)
Urgent Intervention
Hannah's Scaled Score is based on the difficulty of questions and the number of correct responses.
PR: 1 (Percentile Rank)
Hannah scored greater than 1% of students nationally in the same grade.
GE: 0.6 (Grade Equivalent)
Hannah's test performance is comparable to that of an average kindergartener after the sixth month of the school year.
Domain Scores
Domain scores, ranging from 0-100, estimate Hannah's percent of mastery on skills in each domain at a third grade level.
Algebra Readiness
Math Recommendation
Accelerated Math™ Library: Early Numeracy
If you are using the Accelerated Math™ management software system with Hannah, assign the Early Numeracy library. This library should provide a good match for her abilities.
Test Fidelity
Short Test Time: This student took the test in a very short time, which may affect results.
Repeated Pattern: This student answered questions using a pattern of repeated keys, which may affect the results.

Statistical reasoning

When fluctuations in test scores cannot be explained by test fidelity or student mindset, we then look to some statistical considerations.

Standard error of measurement: When interpreting the results of any test instrument, it is important to remember that the scores represent estimates of a student's true achievement level. Test scores are not absolute or exact measures of performance. The standard error of measurement (SEM) describes the extent to which scores would be expected to fluctuate. For example, in Star Reading, the average SEM is approximately 57, which means that that if a student were tested again right away, the score could fluctuate up to 57 points higher or lower than the original score about two-thirds of the time. If the change in a student's Star score is less than 1.5 times the SEM, there is little reason for concern. Chances are that students with outlying scores will show scores closer to the trend the next time the test is

given. Information regarding Standard Error of Measurement can be found in the Star Early Literacy Technical Manual, Star Reading Technical Manual, and Star Math Technical Manual.

Even though a 57 SEM might seem high, remember the Star scale is 1400 points. It is widely understood that computer-adaptive assessments have lower SEMs, than traditional, fixed-form assessments, particularly at the highest and lowest ends of the test scale, where measurement precision of conventional tests is weakest.

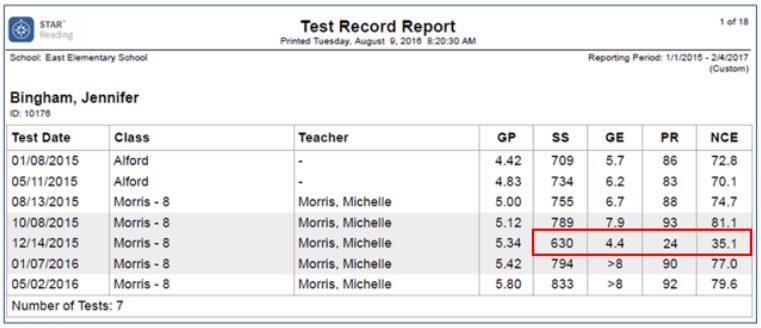
Average Conditional SEMs for the Enterprise scaled score: Star Early Literacy- 42, Star Reading- 57, and Star Math- 29.

Regression to the mean: The statistical phenomenon known as regression to the mean refers to the occurrence in which a score is extreme for the first test but closer to average for subsequent tests. For example, when students are tested twice, there is a tendency for students with the highest scores on the first test to score somewhat closer to average (and therefore lower) on the second test. The reverse happens on the low end of the scale—students who scored lowest on the first test tend to score somewhat higher on the second test. There is a tendency for all scores to regress to the mean, but this effect occurs more significantly for the highest and lowest scores. If score fluctuation occurs with a low or high scoring student, it may well be a regression to the mean issue. After the next testing event, check the student's score to see if data points are indeed 'regressing' to the mean.

Look for trends in data

Trend line: It is important to collect information and look for patterns, or *trends*, in the information instead of placing too much emphasis on a single score. Looking at the trend line and student performance over time will provide a more accurate picture of student performance and growth. A trend line marginalizes score fluctuations and allows you to see patterns of achievement. We would not expect the student to always perform their best or the same at each test administration. More data points help ensure a highly accurate trend line. Trend lines can appear on the following reports: Annual Progress, State Performance, and Student Progress Monitoring, as well as on the Reading and Math Dashboards. See Trend Line for more information.

Historical data: If the "View Student's Historical Star Assessments" capability has been granted by your Renaissance Place administrator, an educator can include historical data (student test data from previous school years) in reports. The Test Record report shows individual student results for every Star test taken during the time period you chose. It is the easiest way to get a detailed look at a student's test history.



STAR Reading

Test Record Report

Printed Tuesday, August 9, 2016 8:20:30 AM

1 of 18

School: East Elementary School

Reporting Period: 1/1/2015 - 2/4/2017 (Custom)

Bingham, Jennifer

ID: 10176

Test Date	Class	Teacher	GP	SS	GE	PR	NCE
01/08/2015	Alford	-	4.42	709	5.7	86	72.8
05/11/2015	Alford	-	4.83	734	6.2	83	70.1
08/13/2015	Morris - 8	Morris, Michelle	5.00	755	6.7	88	74.7
10/08/2015	Morris - 8	Morris, Michelle	5.12	789	7.9	93	81.1
12/14/2015	Morris - 8	Morris, Michelle	5.34	630	4.4	24	35.1
01/07/2016	Morris - 8	Morris, Michelle	5.42	794	>8	90	77.0
05/02/2016	Morris - 8	Morris, Michelle	5.80	833	>8	92	79.6

Number of Tests: 7

In this example, an educator is able to have a frame of reference for the decrease in Scaled Score (SS) on 12/14/15. One might conclude that there were extrinsic factors that accounted for the decrease in Scaled Score (SS) since it does not fit the student's score trend.

Conclusion

Educators know their students best. If there is a concern about large, individual score fluctuations, one possibility is to administer the test again, under closely monitored conditions. Progress is measured best when the same method is used every time the assessment is given. Additionally, look for trends across data sources that confirm or contradict test results that are in question.