

June 30, 2016

# Relating STAR Reading™ and STAR Math™ to the Florida Standards Assessments (FSA) Performance



## Quick reference guide to the STAR Assessments™



**STAR™**  
Reading

**STAR Reading™**—used for screening and progress-monitoring assessment—is a reliable, valid, and efficient computer-adaptive assessment of general reading achievement and comprehension for grades 1–12. STAR Reading provides nationally norm-referenced reading scores and criterion-referenced scores. A STAR Reading assessment can be completed without teacher assistance in about 10 minutes and repeated as often as weekly for progress monitoring.



**STAR™**  
Math

**STAR Math™**—used for screening, progress-monitoring, and diagnostic assessment—is a reliable, valid, and efficient computer-adaptive assessment of general math achievement for grades 1–12. STAR Math provides nationally norm-referenced math scores and criterion-referenced evaluations of skill levels. A STAR Math assessment can be completed without teacher assistance in less than 15 minutes and repeated as often as weekly for progress monitoring.

National Center on  
**INTENSIVE INTERVENTION**

at American Institutes for Research ■



**National Center on Response to Intervention**  
[www.rti4success.org](http://www.rti4success.org)



STAR Reading and STAR Math received the highest possible ratings for screening and progress monitoring by the **National Center on Response to Intervention**, are highly rated for progress monitoring by the **National Center on Intensive Intervention**, and meet all criteria for scientifically based progress-monitoring tools set by the **National Center on Student Progress Monitoring**.

All logos, designs, and brand names for Renaissance Learning's products and services, including but not limited to Renaissance Learning, Renaissance Place, STAR, STAR Assessments, STAR Math, STAR Math Enterprise, STAR Reading, and STAR Reading Enterprise are trademarks of Renaissance Learning, Inc., and its subsidiaries, registered, common law, or pending registration in the United States and other countries. All other product and company names should be considered the property of their respective companies and organizations.

©Copyright 2016 by Renaissance Learning, Inc. All rights reserved. Printed in the United States of America. This publication is protected by U.S. and international copyright laws. It is unlawful to duplicate or reproduce any copyrighted material without authorization from the copyright holder. For more information, contact:

RENAISSANCE LEARNING  
P.O. Box 8036  
Wisconsin Rapids, WI 54495-8036  
(800) 338-4204  
[www.renaissance.com](http://www.renaissance.com)

## Project purpose

Educators face many challenges; chief among them is making decisions regarding how to allocate limited resources to best serve diverse student needs. A good assessment system supports teachers by providing timely, relevant information that can help address key questions about which students are on track to meet important performance standards and which students may need additional help. Different educational assessments serve different purposes, but those that can identify students early in the school year as being at-risk to miss academic standards can be especially useful because they can help inform instructional decisions that can improve student performance and reduce gaps in achievement. Assessments that can do that while taking little time away from instruction are particularly valuable.

Indicating which students are on track to meet later expectations is one of the potential capabilities of a category of educational assessments called “interim” (Perie, Marian, Gong, & Wurtzel, 2007). They are one of three broad categories of assessment:

- Summative – typically annual tests that evaluate the extent to which students have met a set of standards. Most common are state-mandated tests such as the Florida Standards Assessments (FSA).
- Formative – short and frequent processes embedded in the instructional program that support learning by providing feedback on student performance and identifying specific things students know and can do as well as gaps in their knowledge.
- Interim – assessments that fall in between formative and summative in terms of their duration and frequency. Some interim tests can serve one or more purposes, including informing instruction, evaluating curriculum and student responsiveness to intervention, and forecasting likely performance on a high-stakes summative test later in the year.

This project focuses on the application of interim test results, notably their power to inform educators about which students are on track to succeed on the year-end summative state test and which students might need additional assistance to reach proficiency. Specifically, the purpose of this project is to explore statistical linkages between Renaissance Learning interim assessments<sup>1</sup> (STAR Reading and STAR Math) and the FSA. If these linkages are sufficiently strong, they may be useful for:

1. The early identification of students at risk of failing to make yearly progress goals in reading and math, which could help teachers decide to adjust instruction for selected students.
2. Forecasting percentages of students at each performance level on the state assessments sufficiently in advance to permit redirection of resources and serve as an early warning system for administrators at the building and district level.

## Assessments

### Florida Standards Assessments (FSA)

This report is concerned with the FSA English language arts and mathematics assessments in grades 3 through 8. The choice of these two subjects was made because they coincide with the content of the STAR interim assessments, STAR Reading and STAR Math.

FSA report scaled scores to describe a student’s location on the achievement continuum ranging from 240 to 403 for ELA and mathematics in grades 3-8. FSA results are used to classify students into five achievement

---

<sup>1</sup> For an overview of the STAR tests and how they work, please see the References section for a link to download *The research foundation for STAR Assessments* report. For additional information, full technical manuals are available for each STAR assessment by contacting Renaissance Learning at [research@renaissance.com](mailto:research@renaissance.com)

levels, ranging from 1 (lowest) to 5 (highest) each with descriptors that convey the knowledge and skills expected of students at the differing levels of performance. In general, Level 1 indicates the student demonstrates inadequate level of success; level 2 indicates the student demonstrates a below satisfactory level of success; level 3 indicates the student demonstrates a satisfactory level of success; level 4 indicates the student demonstrates an above satisfactory level of success; and level 5 indicates the student demonstrates mastery. The five achievement levels are defined by ranges of students' FSA scaled scores, displayed for ELA and math in Tables 1a and 1b, respectively.<sup>2</sup>

**Table 1a. FSA achievement level score ranges: ELA**

Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	240 - 284	285 - 299	300 - 314	315 - 329	330 - 360
4	251 - 296	297 - 310	311 - 324	325 - 339	340 - 372
5	257 - 303	304 - 320	321 - 335	336 - 351	352 - 385
6	259 - 308	309 - 325	326 - 338	339 - 355	356 - 391
7	267 - 317	318 - 332	333 - 345	346 - 359	360 - 397
8	274 - 321	322 - 336	337 - 351	352 - 365	366 - 403

**Table 1b. FSA achievement level score ranges: Mathematics**

Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	240 - 284	285 - 296	297 - 310	311 - 326	327 - 360
4	251 - 298	299 - 309	310 - 324	325 - 339	340 - 376
5	256 - 305	306 - 319	320 - 333	334 - 349	350 - 388
6	260 - 309	310 - 324	325 - 338	339 - 355	356 - 390
7	269 - 315	316 - 329	330 - 345	346 - 359	360 - 391
8	273 - 321	322 - 336	337 - 352	353 - 364	365 - 393

## STAR Reading™ and STAR Math™

STAR Assessments are nationally normed, computer adaptive measures of general achievement. STAR Reading and STAR Math are intended for use as interim assessments that can be administered at multiple points throughout the school year for purposes such as screening, placement, progress monitoring, and outcomes assessment. Renaissance Learning recommends that STAR tests be administered two to five times a year for most purposes, and more frequently when used in progress monitoring programs. Recent changes to the STAR test item banks and software make it possible to test as often as weekly, for short term progress monitoring in programs such as RTI (response to intervention).

<sup>2</sup> For more detailed descriptors associated with each achievement level, see: <http://www.fldoe.org/core/fileparse.php/5663/urlt/2015FSARangeSummary.pdf>

## Method

### Data collection

Analysis included the evaluation of correlations and statistical linkages between scores on FSA and STAR Reading and STAR Math. Such analyses require matched data, with student records that include both FSA and STAR test scores. Using a secure data-matching procedure compliant with the federal Family Educational Rights and Privacy Act (FERPA), staff from six large districts provided Renaissance Learning with FSA scores for students who had taken STAR Reading and/or STAR Math during the 2014–15 school year. Each record in the resulting matched data file included a student's FSA scores as well as scores on any STAR Reading or STAR Math tests taken during that same year.

Linkages between the STAR and FSA score scales were developed by applying equipercentile linking analysis (Kolen & Brennan, 2004) at each grade. The FSA score scale was linked to the STAR score scale yielding a table of equivalent FSA scores for each possible STAR score. This type of analysis requires students take both assessments at about the same time.

### Sample

The matched STAR-FSA data was divided into two samples. Linking was completed using a concurrent sample, which included STAR tests taken within 30 days before or after the mid-date of the FSA administration window specific to each district. The concurrent sample consisted of a total of over 26,000 students with matched FSA ELA and STAR Reading scores and over 9,000 students with matched FSA mathematics and STAR Math scores across grades 3 through 8. Of the concurrent sample, 10% of the students in each grade was reserved as part of a holdout sample which was used exclusively to evaluate the linking, and was not included in the sample used to compute it.

STAR tests taken prior to the +/-30 day FSA window were included in a predictive sample, which was used to evaluate the accuracy of using the linking results to predict FSA performance using STAR data from earlier in the school year. In the predictive sample, STAR scaled scores were projected to the mid-date of the FSA testing window using national growth norms (Renaissance Learning, 2016a, 2016b). National growth norms are based on grade and initial performance, and are updated annually using a three-year period of data which includes millions of students. They provide typical growth rates for students based on their starting STAR test score. For each STAR score in the predictive sample, the number of weeks between the STAR administration date and the FSA mid-date was calculated. The number of weeks between the two tests was multiplied by the student's expected weekly scaled score growth (based on national growth norms). The expected growth was then added to the observed scaled score to determine the projected STAR score at the in the middle of the FSA administration window.

Tables 2a through 2d contain sample sizes and descriptive statistics for each subject and sample.

**Table 2a. Descriptive statistics for STAR™ and FSA ELA test scores by grade (concurrent sample)**

Grade	Sample Size			FSA ELA		STAR Reading	
	Hold Out	Linking	Total	M	SD	M	SD
3	730	6,574	7,304	302.5	20.4	460.4	169.6
4	692	6,232	6,924	313.5	19.5	573.0	199.6
5	619	5,573	6,192	321.3	20.0	671.0	228.3
6	231	2,079	2,310	321.7	22.2	696.4	258.5
7	182	1,646	1,828	329.7	23.1	761.9	287.0
8	193	1,739	1,932	334.8	20.7	832.7	288.2

**Table 2b. Descriptive Statistics for STAR™ and FSA Mathematics Test Scores by Grade (concurrent sample)**

Grade	Sample Size			FSA Math		STAR Math	
	Hold Out	Linking	Total	M	SD	M	SD
3	150	1,358	1,508	303.5	20.1	620.4	86.8
4	194	1,750	1,944	317.4	21.5	701.1	94.3
5	263	2,374	2,637	321.8	22.3	746.4	97.9
6	143	1,291	1,434	322.0	22.8	751.4	108.0
7	121	1,090	1,211	331.5	21.8	781.5	100.5
8	93	843	936	333.6	22.1	759.1	106.6

**Table 2c. Descriptive statistics for STAR™ and FSA ELA test scores by grade (predictive sample)**

Grade	Sample Size	FSA ELA		STAR Reading	
		M	SD	M	SD
3	10,685	302.3	20.7	466.3	145.9
4	10,065	312.7	19.4	569.2	170.9
5	9,937	321.3	20.5	670.2	202.0
6	3,706	323.1	22.3	712.7	225.8
7	3,442	330.6	22.4	775.2	253.3
8	3,343	336.0	21.1	833.5	257.8

**Table 2d. Descriptive statistics for STAR™ and FSA mathematics test scores by grade (predictive sample)**

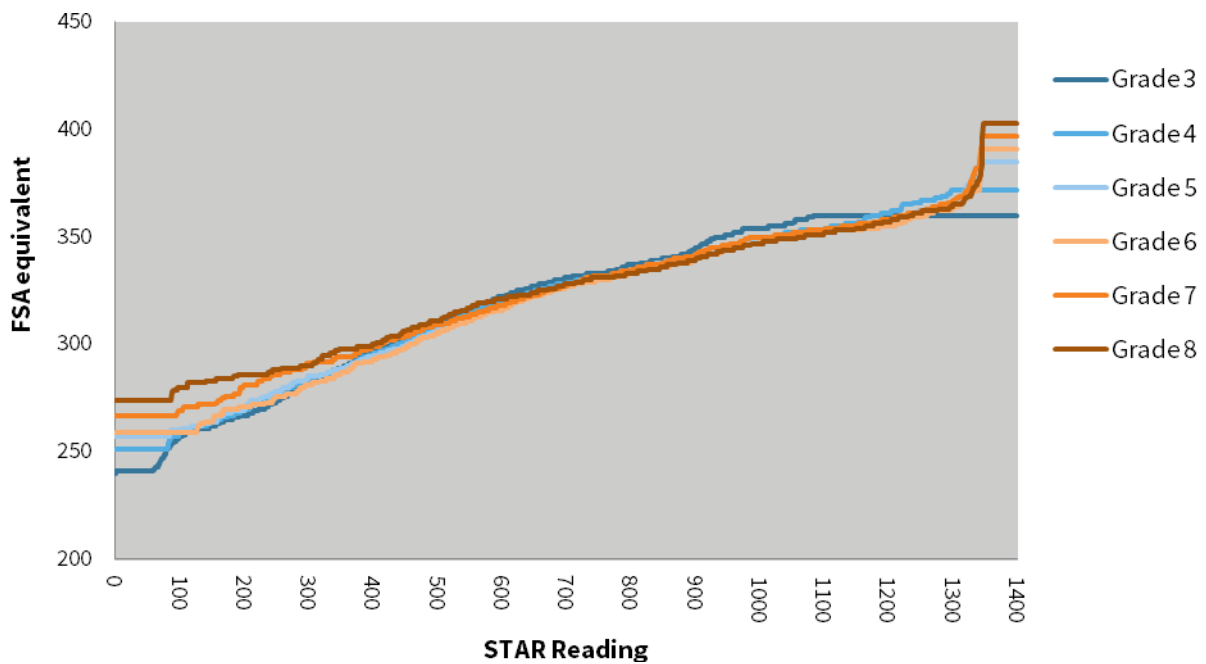
Grade	Sample Size	FSA Math		STAR Math	
		M	SD	M	SD
3	4,188	302.8	19.9	624.1	65.0
4	4,133	316.0	21.7	693.4	75.2
5	4,107	322.7	21.8	749.5	81.5
6	1,398	322.7	23.2	759.9	91.0
7	1,267	329.7	22.6	774.5	85.9
8	978	333.3	21.7	754.5	93.7

## Results

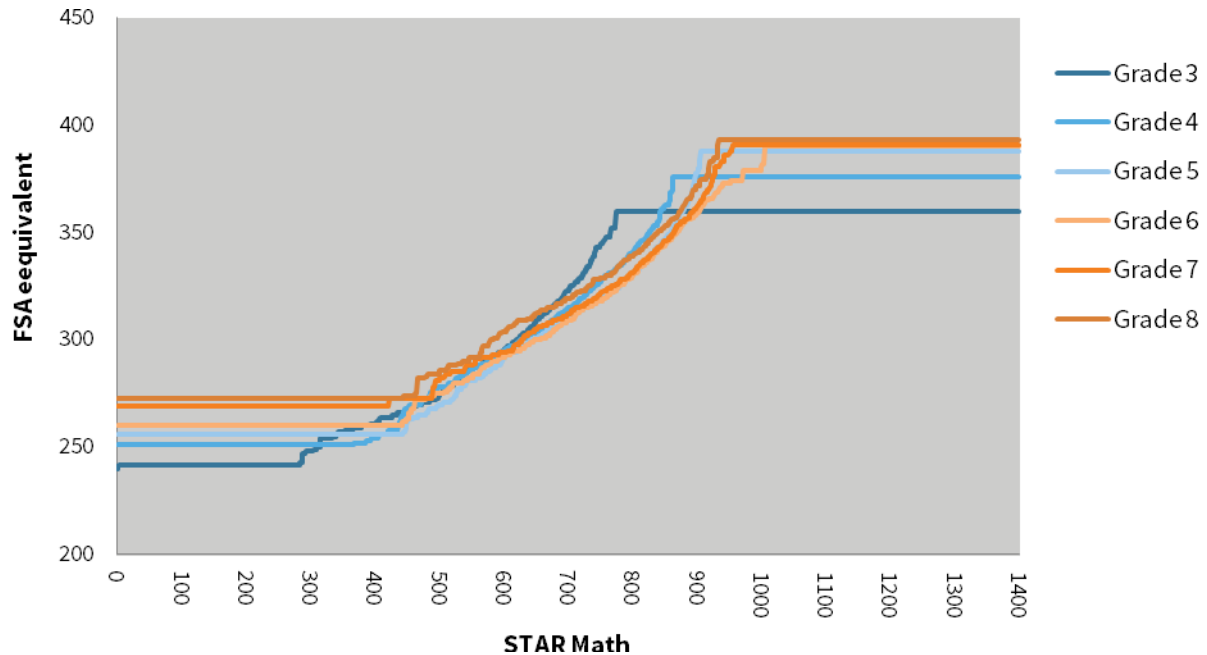
### Scale linkage

Equipercntile linking was used to develop linkages between STAR and FSA scales for reading and math. The result of the analysis was a set of tables yielding equivalent FSA scores for each possible STAR score. These results allow the user to look up the FSA ELA or mathematics test score that corresponds to every possible STAR Reading or STAR Math score (see Figures 1a through 1b).

**Figure 1a. Linkage of FSA ELA to the STAR Reading™ scale**



**Figure 1b. Linkage of FSA mathematics to the STAR Math™ scale**



### Correlations

Two sets of correlations were computed: one between the FSA scores and concurrent STAR scores, and another between FSA scores and the FSA score equivalents (obtained from the linking). Tables 3a and 3b display these correlations for reading and math respectively.

For reading, the correlations between the FSA ELA and STAR averaged .81 and ranged from .79 to .83. The correlations between FSA and FSA score equivalents were similar, averaging .82 and ranging from .80 to .83.

**Table 3a. Pearson correlations between STAR Reading™ and FSA ELA (concurrent sample)**

Grade	FSA ELA score correlation with:	
	Concurrent STAR Reading scale scores	FSA ELA score equivalents
3	0.82	0.83
4	0.80	0.81
5	0.79	0.80
6	0.82	0.82
7	0.83	0.83
8	0.81	0.81
Average	0.81	0.82



For math, correlations between the FSA Mathematics and STAR were similar to those for reading, averaging .79 and ranging from .71 to .84. The correlations between FSA and FSA score equivalents averaged .81 and ranged from .72 to .86.

**Table 3b. Pearson correlations between STAR Math™ and FSA mathematics (concurrent sample)**

Grade	FSA Math score correlation with:	
	Concurrent STAR Math scale scores	FSA Math score equivalents
3	0.78	0.79
4	0.79	0.81
5	0.82	0.83
6	0.84	0.86
7	0.82	0.85
8	0.71	0.72
Average	0.79	0.81

### STAR equivalents to FSA achievement level cut scores

The principal purpose for linking STAR and FSA ELA and mathematics scales was to identify the scores on STAR Reading and STAR Math that are approximately equivalent to the cut-off scores that separate achievement levels on the FSA tests. Tables 4a and 4b display those cut scores by grade for reading and math in grades 3 through 8, respectively.

**Table 4a. Equivalent STAR™ score achievement level ranges: Reading**

Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	< 319	319 - 427	428 - 539	540 - 681	>= 682
4	< 410	410 - 519	520 - 650	651 - 874	>= 875
5	< 477	477 - 627	628 - 810	811 - 1068	>= 1069
6	< 526	526 - 684	685 - 894	895 - 1210	>= 1211
7	< 592	592 - 769	770 - 941	942 - 1215	>= 1216
8	< 608	608 - 856	857 - 1100	1101 - 1314	>= 1315

**Table 4b. Equivalent STAR™ score achievement level ranges: Math**

Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	< 542	542 - 605	606 - 655	656 - 710	>= 711
4	< 621	621 - 676	677 - 740	741 - 794	>= 795
5	< 673	673 - 752	753 - 810	811 - 857	>= 858
6	< 706	706 - 777	778 - 830	831 - 883	>= 884
7	< 719	719 - 792	793 - 847	848 - 892	>= 893
8	< 720	720 - 793	794 - 849	850 - 893	>= 894

RMSEL and mean differences

Accuracy of the scale linkage was evaluated two ways. The same scores used to complete the linking were used to compute the root mean squared errors of linking (RMSEL). Additionally, the holdout sample (i.e., concurrent scores not used to complete the linking) were used to evaluate differences between observed FSA scores and FSA score equivalents. Tables 5a and 5b display these statistics by grade for reading and math respectively.

**Table 5a. Summary statistics from the ELA linkage (concurrent sample)**

Grade	Linking Sample RMSEL	Holdout Sample Difference Scores				
		N	Mean	SD	Min	Max
3	16.90	730	-0.44	12.07	-58.00	45.00
4	16.94	692	0.09	11.41	-38.00	56.00
5	17.46	619	0.22	11.74	-38.00	49.00
6	17.94	231	-2.28	13.27	-51.00	28.00
7	17.82	182	-0.91	14.25	-42.00	41.00
8	18.04	193	-0.45	12.35	-39.00	39.00

**Table 5b. Summary statistics from the math linkage (concurrent sample)**

Grade	Linking Sample RMSEL	Holdout Sample Difference Scores				
		N	Mean	SD	Min	Max
3	16.25	150	0.47	12.55	-25.00	48.00
4	17.23	194	0.98	12.47	-31.00	56.00
5	18.15	263	-1.52	11.80	-60.00	33.00
6	16.76	143	-0.76	11.56	-44.00	26.00
7	16.75	121	-0.02	14.06	-33.00	53.00
8	19.18	93	1.84	15.31	-45.00	36.00

## Classification accuracy

The predictive sample was used in analyses exploring the accuracy of using STAR tests taken earlier in the school year to predict FSA performance based on STAR cutscores identified in the linking analysis.

Two sets of correlations were calculated to summarize the predictive power of the STAR test scores: raw correlations between the projected STAR and observed FSA scale scores, and equated-score correlations between the FSA score equivalents obtained from the linking and the observed FSA scores. The predictive sample correlations were similar in magnitude to the correlations presented earlier for the concurrent sample, indicating that projected STAR scores are reliable estimates of FSA performance. Tables 6a and 6b display these correlations for reading and math, respectively.

For reading, the raw correlations averaged .84 (ranging from .83 to .85) and the correlations between FSA and FSA score equivalents averaged .85 (ranging from .83 to .86).

**Table 6a. Pearson correlations between projected STAR Reading™ scores and FSA ELA (predictive sample)**

Grade	FSA ELA score correlation with:	
	Projected STAR Reading scale scores	FSA ELA score equivalents
3	0.85	0.86
4	0.84	0.85
5	0.84	0.86
6	0.83	0.84
7	0.85	0.85
8	0.83	0.83
Average	0.84	0.85

For math, the raw correlations averaged .81 (ranging from .73 to .84) and the correlations between FSA and FSA score equivalents averaged .82 (ranging from .74 to .86).

**Table 6b. Pearson correlations between projected STAR Math™ scores and FSA mathematics (predictive sample)**

Grade	FSA Math score correlation with:	
	Projected STAR Math scale scores	FSA Math score equivalents
3	0.81	0.82
4	0.82	0.83
5	0.81	0.83
6	0.84	0.86
7	0.83	0.86
8	0.73	0.74
Average	0.81	0.82

Different evaluation data are presented for two-category (proficient vs. not proficient) versus five-category (performance level) projections. For the two-category projections, standard statistical classification diagnostics were calculated. Because these can be quite complex in the multi-category situation, a simpler approach was taken with five achievement levels.

**Two-category proficiency status projections.** Classification diagnostics were derived from counts of correct and incorrect classifications that could be made when using STAR scores to predict whether or not a student would be proficient on the FSA. The classification diagnostic formulas are outlined in Table 7a and the types of classifications are summarized in Table 7b.

**Table 7a. Descriptions of classification diagnostic accuracy measures**

Measure	Formula	Interpretation
Overall classification accuracy	$\frac{TP + TN}{N}$	Percentage of correct classifications
Sensitivity	$\frac{TP}{TP + FN}$	Percentage of proficient students identified as such using STAR
Specificity	$\frac{TN}{TN + FP}$	Percentage of not proficient students identified as such using STAR
Positive predictive value (PPV)	$\frac{TP}{TP + FP}$	Percentage of students STAR finds proficient who actually are proficient
Negative predictive value (NPV)	$\frac{TN}{FN + TN}$	Percentage of students STAR finds not proficient who actually are not proficient
Observed proficiency rate (OPR)	$\frac{TP + FN}{N}$	Percentage of students who achieve proficient
Projected proficiency rate (PPR)	$\frac{TP + FP}{N}$	Percentage of students STAR finds proficient
Proficiency status projection error	PPR - OPR	Difference between projected and observed proficiency rates

**Table 7b. Schema for a fourfold table of classification diagnostic data**

		FSA Result		Total
		Proficient	Not	
STAR Estimate	Proficient	True Positive (TP)	False Positive (FP)	Projected Proficient (TP + FP)
	Not	False Negative (FN)	True Negative (TN)	Projected Not (FN + TN)
Total		Observed Proficient (TP + FN)	Observed Not (FP + TN)	N = TP+FP+FN+TN

Classification accuracy diagnostics are presented in tables 8a and 8b for reading and math, respectively.

On average, students were correctly classified as either proficient or not (i.e., overall classification accuracy) 84% of the time for reading and 83% of the time for math. For reading, the forecasts were accurate between 82% and 85% of the time depending on the grade, while for math they were accurate between 77% and 85% of the time.

Sensitivity statistics (i.e., the percentage of proficient students correctly forecasted) averaged 84% for reading, with a range from 78% to 89%. For math, sensitivity statistics averaged 83%, with a range from 67% to 89%. Specificity statistics (i.e., the percentage of not proficient students correctly forecasted) were similar to sensitivity, averaging 83% for reading and 81% for math. Specificity is negatively related to observed proficiency rate, so grades with higher observed proficiency rates tend to have lower specificity.

Positive predictive values averaged 85% for reading and ranged from 84% to 86%. For math, positive predictive values averaged 85% and ranged from 80% to 87%. Therefore, when STAR scores forecasted students to be proficient, they actually were proficient 85% of the time for reading and 85% of the time for math.

For reading, negative predictive values were similar to positive predictive values, averaging 83% and ranging from 80% to 84%. For math, negative predictive values were also similar to positive predictive values, averaging 80% and ranging from 75% to 86%. The negative predictive value results indicated that when STAR scores forecasted that students were not proficient, they actually were not proficient 83% of the time for reading and 80% of the time for math.

For reading, differences between the observed and projected proficiency rates (i.e., proficiency status projection error) indicated that projected STAR Reading scores tended to accurately predict proficiency rates across grades. Positive values of proficiency status projection error indicate over-prediction and negative values indicate under-prediction. For reading, proficiency status projection errors averaged 0% and ranged from -4% to 3%. For math, proficiency status projection errors averaged -1% and ranged from -7% to 3%.

Finally, the area under the ROC curve (AUC) is a summary measure of diagnostic accuracy. The National Center on Response to Intervention has set an AUC of 0.85 or higher as indicating convincing evidence that an assessment can accurately predict another assessment result or outcome. In this study, both STAR Reading and STAR Math met or well exceeded that standard. For reading the average AUC was .92, with a range from .91 to .93, and for math the average AUC was .91, with a range from .85 to .93, indicating that STAR scores did a very good job of discriminating between which students scored proficient on the FSA and which did not.

**Table 8a. Classification diagnostics for reading**

Measure	Grade					
	3	4	5	6	7	8
Overall classification accuracy	85%	84%	84%	84%	85%	82%
Sensitivity	89%	88%	86%	83%	82%	78%
Specificity	79%	79%	83%	84%	88%	86%
Positive predictive value (PPV)	85%	84%	85%	84%	86%	85%
Negative predictive value (NPV)	84%	84%	84%	84%	84%	80%
Observed proficiency rate (OPR)	57%	55%	52%	49%	48%	50%
Projected proficiency rate (PPR)	60%	58%	53%	49%	46%	46%
Proficiency status projection error	3%	3%	1%	0%	-2%	-4%
Area Under the ROC Curve	0.93	0.92	0.93	0.93	0.93	0.91

**Table 8b. Classification diagnostics for math**

Measure	Grade					
	3	4	5	6	7	8
Overall classification accuracy	83%	84%	83%	85%	84%	77%
Sensitivity	89%	88%	84%	85%	82%	67%
Specificity	73%	78%	82%	84%	87%	85%
Positive predictive value (PPV)	85%	87%	86%	83%	86%	80%
Negative predictive value (NPV)	80%	79%	80%	86%	83%	75%
Observed proficiency rate (OPR)	63%	63%	57%	49%	51%	47%
Projected proficiency rate (PPR)	66%	64%	56%	50%	48%	39%
Proficiency status projection error	3%	1%	-1%	1%	-2%	-7%
Area Under the ROC Curve	0.91	0.91	0.92	0.93	0.93	0.85

**Five-category achievement level projections.** This section deals with the accuracy of forecasts of a more finely-grained performance classification of student performance according to the five FSA achievement levels. Compared to the two-category analyses, the five-level analyses were more straightforward. They focused on two components: 1) the percentage of agreement between classifications based on FSA test scores and those based on projected STAR scores and 2) comparisons of the projected and actual percentages of students who achieved each level on the FSA.

Table 9 lists two measures of accuracy of the achievement level classifications based on linked projected STAR scores. The first is the percent of perfect agreement between classifications based on FSA scores and those based on projected STAR scores. The second measure is the percent of agreement within +/- one achievement level.

Classification accuracy was similar for reading and math. For reading, perfect agreement rates averaged 56% and ranged from 53% to 58%. For math, they averaged 53% and ranged from 49% to 57%. The percent agreement within +/- one achievement level was above 90% for reading and math in all grades. These trends in classification errors suggest that STAR tests taken months before FSA can accurately predict a student’s general FSA performance (within one level above or below their actual achievement level) but are less precise when used to predict a student’s specific FSA achievement level.

**Table 9. Accuracy of projected STAR™ scores for predicting FSA achievement levels**

Grade	Reading		Math	
	Perfect Agreement	Agreement Within +/- One Level	Perfect Agreement	Agreement Within +/- One Level
3	57%	97%	51%	96%
4	56%	97%	53%	95%
5	58%	97%	53%	96%
6	55%	97%	57%	97%
7	56%	96%	55%	97%
8	53%	95%	49%	91%

Tables 10a and 10b list the percentages of students who were in each FSA achievement level versus those forecasted to be in each level for reading and math, respectively.

For both reading and math, the achievement level projection errors (i.e., differences between the percentage of actual and forecasted achievement) averaged 0%, but tended to slightly over-predict moderate performance (e.g. Levels 2 and 3) and slightly under-predict extreme performance (e.g., Levels 1 and 5). The math projection errors ranged from -7% to 7%, and the reading errors ranged from -6% to 7%.

**Table 10a. Actual versus forecasted percentages of students for each FSA ELA achievement level**

Grade	Level Achieved					Level Forecasted					Difference (Forecasted - Achieved)				
	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5
3	18%	25%	29%	20%	8%	14%	27%	31%	22%	7%	-5%	2%	3%	2%	-2%
4	19%	26%	28%	19%	8%	14%	28%	31%	21%	5%	-5%	2%	3%	2%	-2%
5	18%	29%	27%	19%	6%	14%	33%	30%	19%	4%	-4%	4%	3%	0%	-2%
6	26%	25%	24%	20%	6%	19%	31%	27%	19%	3%	-6%	7%	4%	0%	-3%
7	26%	26%	22%	17%	9%	24%	30%	20%	19%	6%	-2%	5%	-2%	3%	-3%
8	23%	27%	26%	16%	7%	20%	35%	28%	15%	3%	-4%	7%	2%	-1%	-4%

**Table 10b. Actual versus forecasted percentages of students for each FSA mathematics achievement level**

Grade	Level Achieved					Level Forecasted					Difference (Forecasted - Achieved)				
	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5	Level 1	Level 2	Level 3	Level 4	Level 5
3	16%	20%	29%	24%	11%	10%	24%	33%	28%	6%	-6%	3%	4%	5%	-5%
4	19%	18%	30%	21%	12%	14%	22%	36%	22%	5%	-4%	4%	7%	1%	-7%
5	19%	24%	29%	19%	10%	16%	29%	33%	19%	4%	-4%	5%	5%	0%	-6%
6	27%	24%	23%	19%	7%	23%	28%	29%	17%	4%	-5%	4%	6%	-1%	-4%
7	25%	24%	27%	15%	9%	23%	29%	28%	17%	3%	-3%	5%	1%	2%	-5%
8	27%	26%	28%	13%	6%	29%	32%	27%	10%	2%	2%	6%	-1%	-3%	-4%



## Conclusions and applications

The equipercentile linking method was used to link the STAR Reading and FSA ELA scaled scores and STAR Math and FSA mathematics score scales in grades 3 through 8. The result of each linkage analysis was an estimate of the approximately equivalent FSA score for that grade. Using the tables of linked scores, we identified STAR Reading and STAR Math scores that were linked to the cutscores for FSA achievement levels (reported in Tables 4a and 4b). Because the linking was done using a sample from just six districts, and may not be representative of the statewide student population, these cutscores should be considered approximations that can be updated with greater precision as more data become available in the future.

Correlations indicated a strong relationship between the STAR and FSA tests. On average, the correlation between FSA and concurrent STAR scores (i.e., STAR tests taken within +/- 30 days of the FSA mid-date) was .81 for reading and .79 for math. Similarly, the average correlation between FSA and predictive STAR scores (i.e., STAR tests taken earlier and projected to the FSA mid-date) was .84 for reading and .81 for math. When projecting STAR scores to estimate FSA performance, students were correctly classified as either proficient or not 83% of the time for reading and 84% for math.

The statistical linkages between STAR interim assessments and the FSA for ELA and mathematics provide a means of forecasting student achievement on the FSA based on STAR scores obtained earlier in the school year. Example STAR Reading and STAR Math reports that utilize the STAR-FSA linking are provided in the Appendix. They include individualized reports, which compare each student's STAR performance to the growth trajectory that typically would lead to proficiency on the FSA, as well as group-level performance reports that forecast of the number of students that are expected to score at each achievement level on the FSA. Both types of information can be used to help educators determine early and periodically which students are on track to reach proficiency and make decisions accordingly.

## References

Kolen, M. J. & Brennan, R. R. (2004). *Test equating scaling and linking: Methods and practices*. New York, NY: Springer Science+Business Media.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Aspen, CO: Aspen Institute.

Renaissance Learning. (2016a). *STAR Math: Technical manual*. Wisconsin Rapids, WI: Author. Available from Renaissance Learning by request to [research@renaissance.com](mailto:research@renaissance.com)

Renaissance Learning. (2016b). *STAR Reading: Technical manual*. Wisconsin Rapids, WI: Author. Available from Renaissance Learning by request to [research@renaissance.com](mailto:research@renaissance.com)

Renaissance Learning. (2014). *The research foundation for STAR Assessments: The science of STAR*. Wisconsin Rapids, WI: Author. Available online from <http://doc.renlearn.com/KMNet/R003957507GG2170.pdf>

## Independent technical reviews of STAR Reading and STAR Math

U.S. Department of Education: National Center on Intensive Intervention. (2016a). *Review of progress monitoring tools* [Review of STAR Math]. Washington, DC: Author. Retrieved from <http://www.intensiveintervention.org/chart/progress-monitoring>

U.S. Department of Education: National Center on Intensive Intervention. (2016b). *Review of progress monitoring tools* [Review of STAR Reading]. Washington, DC: Author. Retrieved from <http://www.intensiveintervention.org/chart/progress-monitoring>

U.S. Department of Education: National Center on Response to Intervention. (2010a). *Review of progress monitoring tools* [Review of STAR Math]. Washington, DC: Author. Retrieved from <https://web.archive.org/web/20120813035500/http://www.rti4success.org/pdf/progressMonitoringGOM.pdf>

U.S. Department of Education: National Center on Response to Intervention. (2010b). *Review of progress monitoring tools* [Review of STAR Reading]. Washington, DC: Author. Retrieved from <https://web.archive.org/web/20120813035500/http://www.rti4success.org/pdf/progressMonitoringGOM.pdf>

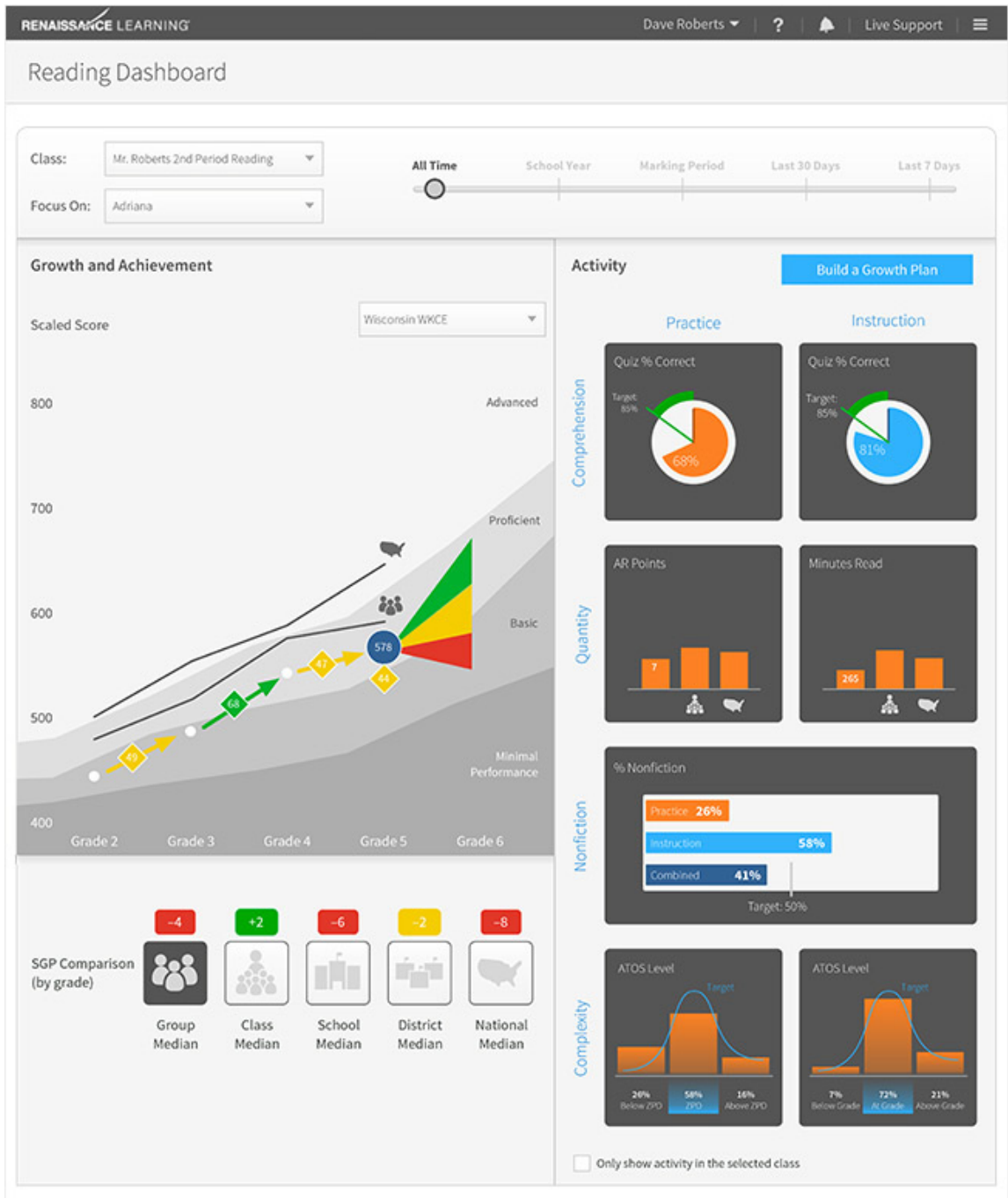
U.S. Department of Education: National Center on Response to Intervention. (2011a). *Review of screening tools* [Review of STAR Math]. Washington, DC: Author. Retrieved from <http://www.rti4success.org/resources/tools-charts/screening-tools-chart>

U.S. Department of Education: National Center on Response to Intervention. (2011b). *Review of screening tools* [Review of STAR Reading]. Washington, DC: Author. Retrieved from <http://www.rti4success.org/resources/tools-charts/screening-tools-chart>

# Appendix: Sample reports <sup>3</sup>

## Reading dashboard

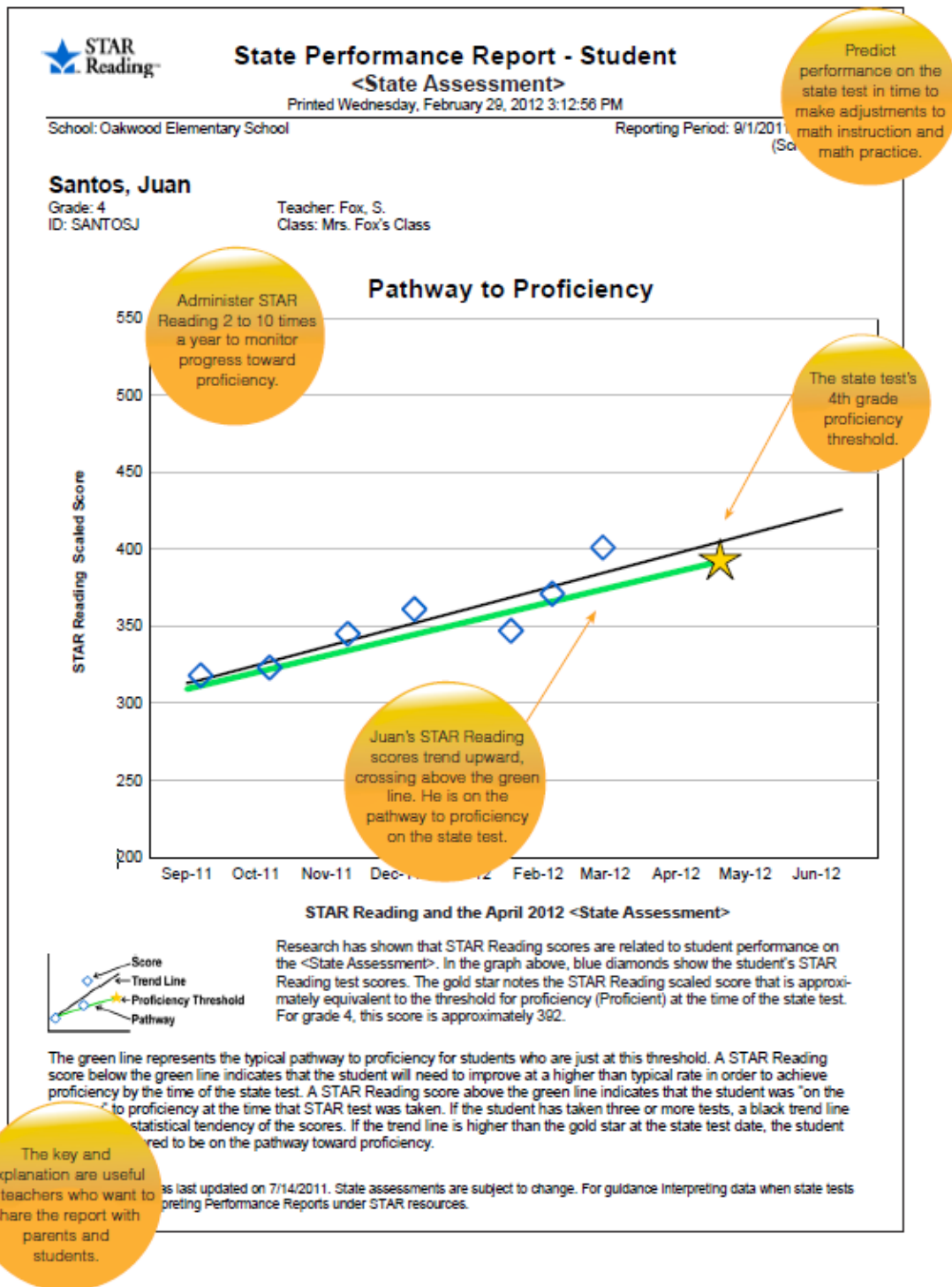
The Reading Dashboard combines key data from across Renaissance's products to provide a 360 degree view of student performance and progress toward goals. At-a-glance data paired with actionable insight helps teachers drive student growth to meet standards and ensure college and career readiness. Educators can review data at the student, group, or class level and various timeframes to determine what's working, what isn't, and most importantly, where to focus attention to drive growth.



<sup>3</sup> Reports are regularly reviewed and may vary from those shown as enhancements are made.

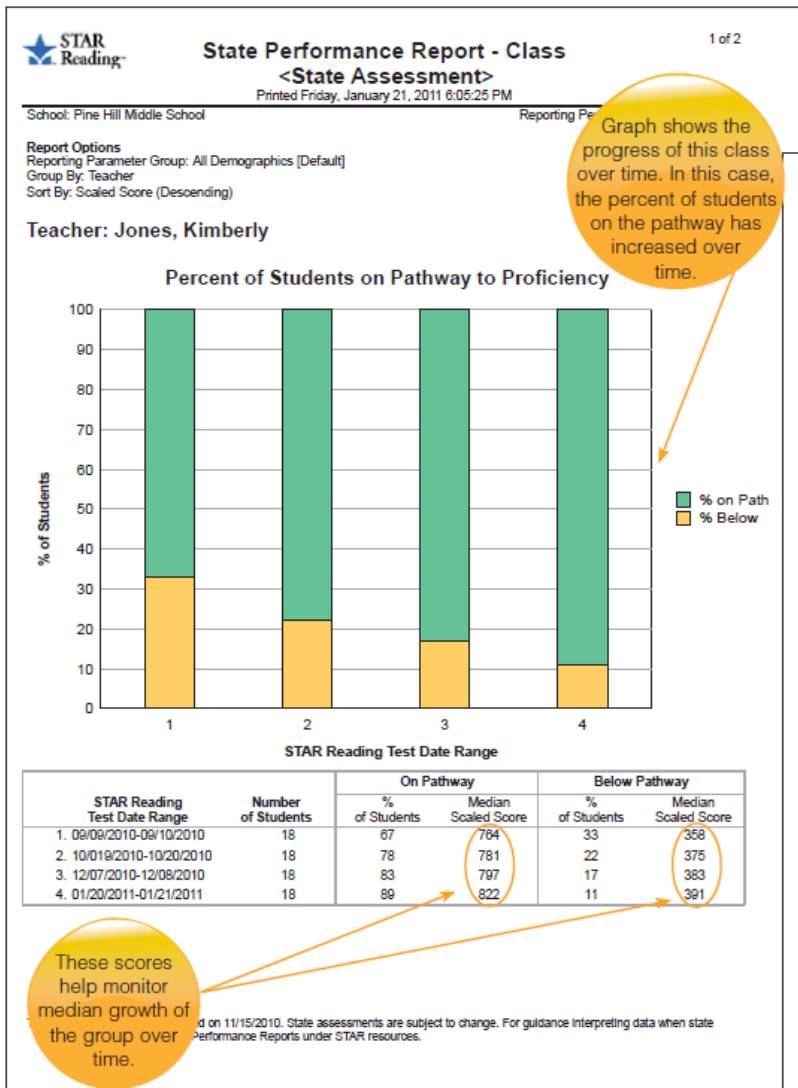
## Performance reports focusing on the pathway to proficiency

The report graphs the student's STAR Reading or STAR Math scores and trend line (projected growth) for easy comparison with the pathway to proficiency.



## Group performance reports

The Group Performance Report compares students' performance on the STAR Assessments to the pathway to proficiency for annual state tests and summarizes the results. It helps you see how groups of students (whole class, for example) are progressing toward proficiency. The report displays the most current data as well as historical data as bar charts so that you can see patterns in the percentages of students on the pathway to proficiency and below the pathway.



Predict performance on the state test in time to make adjustments to reading instruction and practice.

**STAR Reading** State Performance Report - Class <State Assessment>  
 Printed Friday, January 21, 2011 6:05:25 PM

School: Pine Hill Middle School Reporting Period: 9/01/2010-4/14/2011 (Outlook RP)

### On Pathway to Proficiency

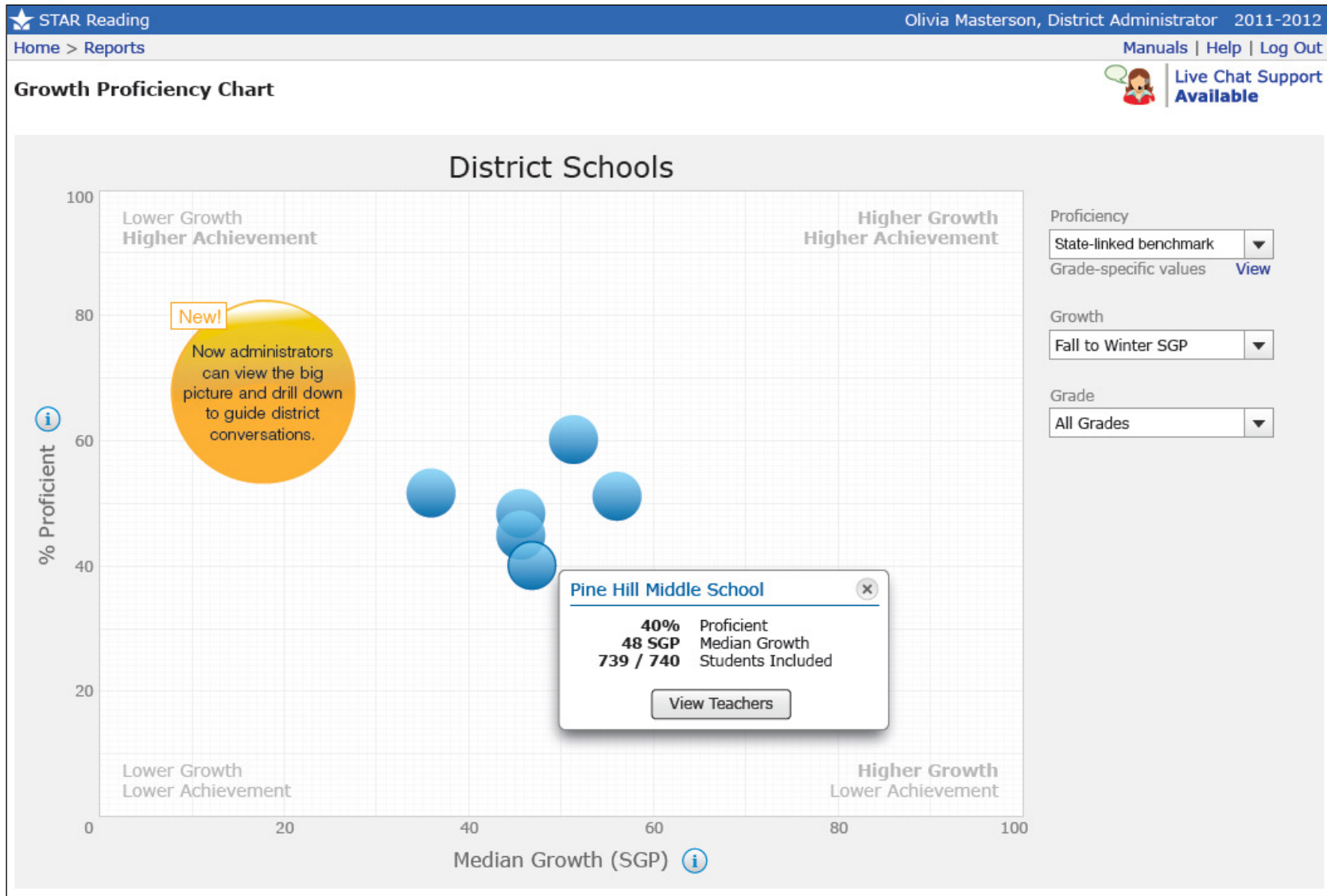
Student	Most Recent Test Date	STAR Reading Scaled Score
Rice, Heather	1/21/2011	1388
Curtis, Jason	1/21/2011	1269
Lao, Jose	1/21/2011	1200
Johnson, Tim	1/21/2011	955
Mackowski, Gregory	1/21/2011	927
Waldenmaler, Dean	1/21/2011	920
Reyes, Christina	1/21/2011	913
Frisch, Dena	1/21/2011	886
O'Rourke, Sean	1/21/2011	877
Statz, Madeline	1/21/2011	839
Major, Jasmine	1/21/2011	802
Atkinson, Rebecca	1/21/2011	790
Kahl, Robert	1/21/2011	603
Hanneman, David	1/21/2011	569
Bussey, Walter	1/21/2011	535
Farnes, Cathy	1/21/2011	501

### Below Pathway to Proficiency

Student	Most Recent Test Date	STAR Reading Scaled Score
Okada, Casey	1/21/2011	404
Locke, Kimberly	1/21/2011	377


## Growth proficiency chart

STAR Assessments' Student Growth Percentiles and expected state assessment performance are viewable by district, school, grade, or class. In addition to Student Growth Percentiles, other key growth indicators such as grade equivalency, percentile rank, and instructional reading level are also available to help educators identify best practices that are having a significant educational impact on student growth.



## High-level performance reports

Reports for administrators provide periodic, high level forecasts of student performance on the state tests. It includes a performance outlook for each achievement level and options for how to group and list information.



### STAR Reading™ Performance Report <State Assessment>

Printed Friday, January 14, 2011 08:19:05 AM

---

District: Renaissance District

#### Oakwood Elementary School

Grade	Student Performance Outlook <sup>2</sup> On the April 2011 <State Assessment>						STAR Reading Participation 09/01/2010-04/15/2011			
	Less Than Proficient		Proficient				Tested		Not Tested	
	Standard Not Met		Met Standard		Commended Performance		Total	%	Total	%
	Total	%	Total	%	Total	%	Total	%	Total	%
3	35	24	54	37	57	39	146	94	10	6
4	23	23	42	24	34	53	99	90	11	10
5	18	16	53	47	41	37	112	93	8	7
Summary	76	21	149	42	132	37	357	92	29	8

STAR uses advanced growth modeling to predict how each student will score on the state test.

Administrators can track district progress toward proficiency on the state test with this report.

#### Pine Hill Middle School

Grade	Student Performance Outlook <sup>2</sup> On the April 2011 <State Assessment>						STAR Reading Participation 09/01/2010-04/15/2011			
	Less Than Proficient		Proficient				Tested		Not Tested	
	Standard Not Met		Met Standard		Commended Performance		Total	%	Total	%
	Total	%	Total	%	Total	%	Total	%	Total	%
6	67	28	81	34	93	38	241	98	5	2
7	63	27	86	34	89	39	238	97	8	3
8	77	32	83	35	79	33	239	96	9	4
Summary	207	29	250	30	261	37	718	97	22	3

Monitor the percentage of students in each proficiency level on the state test.

Through extensive research, STAR Reading has been linked to individual state tests.

<sup>2</sup>The Student Performance Outlook is based on STAR Reading tests taken from the beginning of the school year until the start of the state testing period. STAR Reading tests taken after that period are not reflected in the report. Test information was last updated on 11/15/2010. State assessments are subject to change. For guidance interpreting data when state tests change, see Interpreting Performance Reports under STAR resources.

## Acknowledgments

The following experts have advised Renaissance Learning in the development of the STAR Assessments.



**Thomas P. Hogan, Ph.D.**, is a professor of psychology and a Distinguished University Fellow at the University of Scranton. He has more than 40 years of experience conducting reviews of mathematics curricular content, principally in connection with the preparation of a wide variety of educational tests, including the Stanford Diagnostic Mathematics Test, Stanford Modern Mathematics Test, and the Metropolitan Achievement Test. Hogan has published articles in the *Journal for Research in Mathematics Education* and *Mathematical Thinking and Learning*, and he has authored two textbooks and more than 100 scholarly publications in the areas of measurement and evaluation. He has also served as consultant to a wide variety of school systems, states, and other organizations on matters of educational assessment, program evaluation, and research design.



**James R. McBride, Ph.D.**, is vice president and chief psychometrician for Renaissance Learning. He was a leader of the pioneering work related to computerized adaptive testing (CAT) conducted by the Department of Defense. McBride has been instrumental in the practical application of item response theory (IRT) and since 1976 has conducted test development and personnel research for a variety of organizations. At Renaissance Learning, he has contributed to the psychometric research and development of STAR Math, STAR Reading, and STAR Early Literacy. McBride is co-editor of a leading book on the development of CAT and has authored numerous journal articles, professional papers, book chapters, and technical reports.



**Michael Milone, Ph.D.**, is a research psychologist and award-winning educational writer and consultant to publishers and school districts. He earned a Ph.D. in 1978 from The Ohio State University and has served in an adjunct capacity at Ohio State, the University of Arizona, Gallaudet University, and New Mexico State University. He has taught in regular and special education programs at all levels, holds a Master of Arts degree from Gallaudet University, and is fluent in American Sign Language. Milone served on the board of directors of the Association of Educational Publishers and was a member of the Literacy Assessment Committee and a past chair of the Technology and Literacy Committee of the International Reading Association. He has contributed to both [readingonline.org](http://readingonline.org) and *Technology & Learning* magazine on a regular basis. Over the past 30 years, he has been involved in a broad range of publishing projects, including the SRA reading series, assessments developed for Academic Therapy Publications, and software published by The Learning Company and LeapFrog. He has completed 34 marathons and 2 Ironman races.