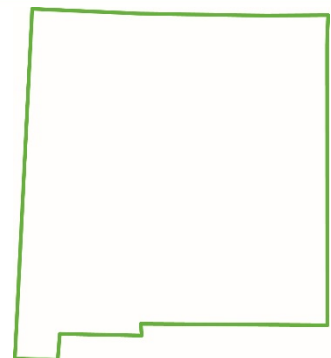


Relating Star Reading® and Star Math® to the New Mexico Partnership for Assessment of Readiness for College and Careers (PARCC) Assessments Performance



Contents

Introduction.....	3
Main Findings.....	3
Study.....	3
Results.....	6
Appendix A: About Star Reading and Star Math.....	8
Appendix B: PARCC achievement levels.....	8
References.....	9

Figures

Figure 1. Star Reading and Star Math scores highly correlate with PARCC Assessments.....	4
Figure 2. Projected scores from Star Reading and Star Math highly correlate with PARCC.....	6

Tables

Table 1. Star Reading and Star Math score equivalents for each PARCC achievement level range.....	5
Table 2. Proficiency forecasting using Star Reading and Star Math scores yields accurate results.....	7
Table B1. PARCC achievement level score ranges.....	8

Introduction

At Renaissance we know that as an educator, chief among your responsibilities is making decisions about how to allocate limited resources to best serve diverse student needs. A good assessment system supports your efforts, by providing timely, relevant information to help address key questions about which students are on track to meet important standards and who may need additional assistance.

Assessments that identify early any students at risk of missing academic standards are especially useful, as they inform instructional decisions to improve student performance and reduce gaps in achievement. Assessments that do this while taking little time away from instruction are particularly valuable. *Interim assessments*, one of three broad categories of educational assessment,¹ indicate which students are on track to meet later expectations (Perie, Marion, Gong, & Wurtzel, 2007).

This linking study applied results from two interim assessments, Renaissance Star Reading® and Renaissance Star Math®, to help you predict whether individual students are on track or need more assistance to succeed on the year-end summative Partnership for Assessment of Readiness for College and Careers (PARCC) assessments in English language arts/literacy in grade 3 through 10 and mathematics in grades 3 through 8.²

Assessments that identify early any students at risk of missing academic standards are especially useful.

Main Findings

Results from the linking analysis revealed that Star Reading and Star Math are accurate predictors of the PARCC assessments, meaning as an educator you can use Star scores to:

1. Identify early in the year students likely to miss reading and math yearly progress goals in time to make meaningful adjustments to instruction well before the year-end test.
2. Forecast the percent of students at each PARCC performance level to serve as an early warning system for building and district administrators and allow redirection of resources as needed.

Study

To determine if Star Reading and Star Math can predict student achievement on the end-of-year PARCC assessments in English language arts/literacy and mathematics, we began by linking the score scales for each assessment.

¹ **Formative assessments** are short and frequent processes, embedded in instruction, that support learning and provide specific feedback on what students know and can do versus where gaps in knowledge exist. **Summative assessments** evaluate whether students have met a set of standards, and serve most commonly as year-end state-mandated tests. **Interim assessments** represent the middle ground, in terms of duration and frequency and can serve purposes including informing instruction, evaluating curriculum and student responsiveness to intervention, and forecasting performance on high-stakes summative year-end tests.

² Technical manuals are available for Star Reading and Star Math by request to research@renaissance.com.

Data collection

Using a secure data-matching procedure compliant with the federal Family Educational Rights and Privacy Act (FERPA), staff from eight large districts in seven states (Arkansas, Colorado, Illinois, New Mexico, New Jersey, Ohio, and Rhode Island) provided Renaissance with state summative test scores for students who had taken Star Reading or Star Math during the 2014–2015 school year. Each record included a student's PARCC scores and was matched with all Star scores for that year.

Sample characteristics

Renaissance divided the PARCC data into two samples. The **concurrent** sample included students' scores for all Star tests taken within 30 days before or after the mid-date of the PARCC administration window. This sample numbered 32,029 students in grades 3–10 with matched PARCC and Star Reading scores and 26,770 students in those grades 3–8 with matched PARCC and Star Math scores. In each grade, we then set aside scores from a subset of these students—10%—as a holdout sample to use only to evaluate the scale linkage.

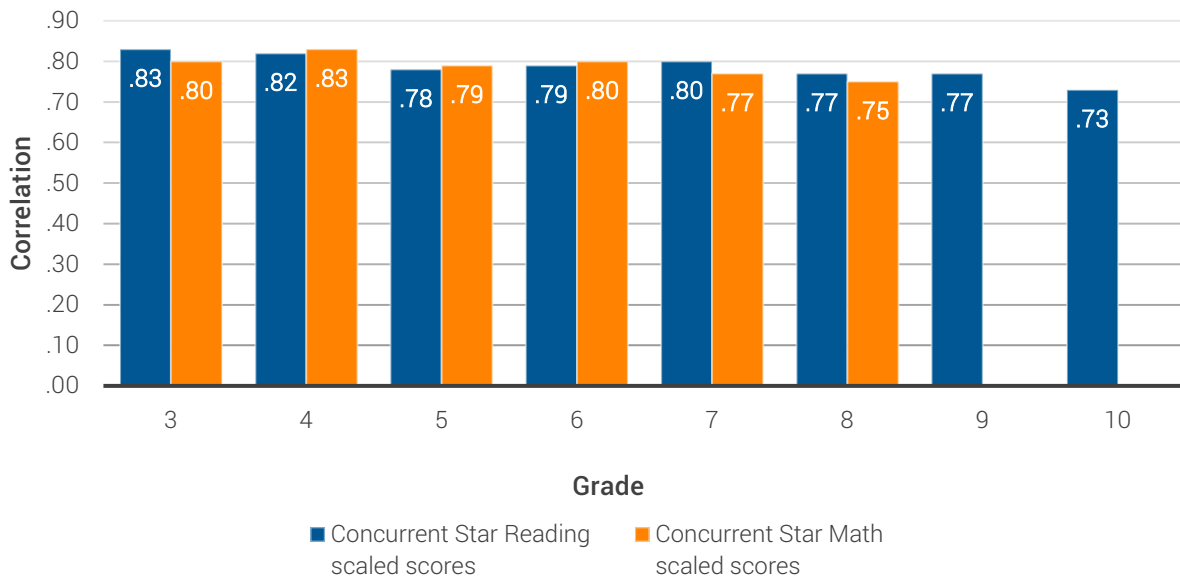
The linking analysis revealed that Star Reading and Star Math are accurate predictors of the PARCC Assessments.

The **predictive** sample, which included 27,415 students for reading and 23,260 students for math, included Star scores for tests taken more than 30 days before the mid-date in the PARCC testing window.

Correlations

Before linking Star tests with the PARCC Assessments, we ensured there was a strong relationship between the test scales. As seen in figure 1, the correlations were positive, averaging .79 and .79 between PARCC and Star Reading and Star Math, respectively.

Figure 1. Star Reading® and Star Math® scores highly correlate with PARCC Assessments



Scale linkage

Renaissance then linked the score scales for the Star Reading/Star Math and the PARCC Assessments in English language arts/literacy and mathematics by applying equipercentile linking analysis (Kolen & Brennan, 2004) in grades 3–10 in reading and grades 3–8 in math. The concurrent sample (sans the holdout sample) was used in the linking (scores from all Star tests taken within 30 days before or after the PARCC testing mid-date), and the result was a table of PARCC scores for each possible Star score.

The predictive sample was then used to evaluate if the linking results could accurately predict student performance on the PARCC Assessment with Star data from earlier in the school year. To do so, we took students' Star scores from tests taken more than 30 days prior to the PARCC testing mid-date and used national growth norms (Renaissance, 2016a, 2016b) to project what their Star scores would be at the mid-date. Then the scale linkage table was used to look up the projected Star scores (or the average of the projected scores for students with multiple Star scores in the predictive sample) to see how they translated to the PARCC scale.

PARCC cut scores and corresponding Star score equivalents

PARCC results are reported in scaled scores that describe each student's location on an achievement continuum ranging from approximately 650 to 850 and using five achievement levels: *Level 1, Level 2, Level 3, Level 4, and Level 5*.

A main purpose in linking Star Reading and Star Math to the PARCC Assessments was to identify Star scores approximately equivalent to the cut-off scores that separate the PARCC achievement levels. Table 1 displays these equivalent Star scores for grade 3–10 in reading and grades 3–8 in math. The corresponding PARCC cut scores can be found in the Appendix B.

Table 1. Star Reading® and Star Math® score equivalents for each PARCC achievement level range

Star Reading® cut-score equivalents					
Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	< 304	304–412	413–513	514–923	>= 924
4	< 335	335–441	442–570	571–1019	>= 1020
5	< 372	372–507	508–662	663–1178	>= 1179
6	< 417	417–560	561–776	777–1251	>= 1252
7	< 467	467–614	615–829	830–1292	>= 1293
8	< 527	527–678	679–908	909–1320	>= 1321
9	< 597	597–825	826–1035	1036–1324	>= 1325
10	< 685	685–900	901–1075	1076–1325	>= 1326
Star Math® cut-score equivalents					
Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	< 504	504–588	589–648	649–742	>= 743
4	< 562	562–661	662–735	736–850	>= 851
5	< 614	614–720	721–804	805–901	>= 902
6	< 650	650–756	757–826	827–925	>= 926
7	< 687	687–777	778–860	861–960	>= 961
8	< 739	739–815	816–875	876–1007	>= 1008

Results

Accuracy of scale linkage confirmed

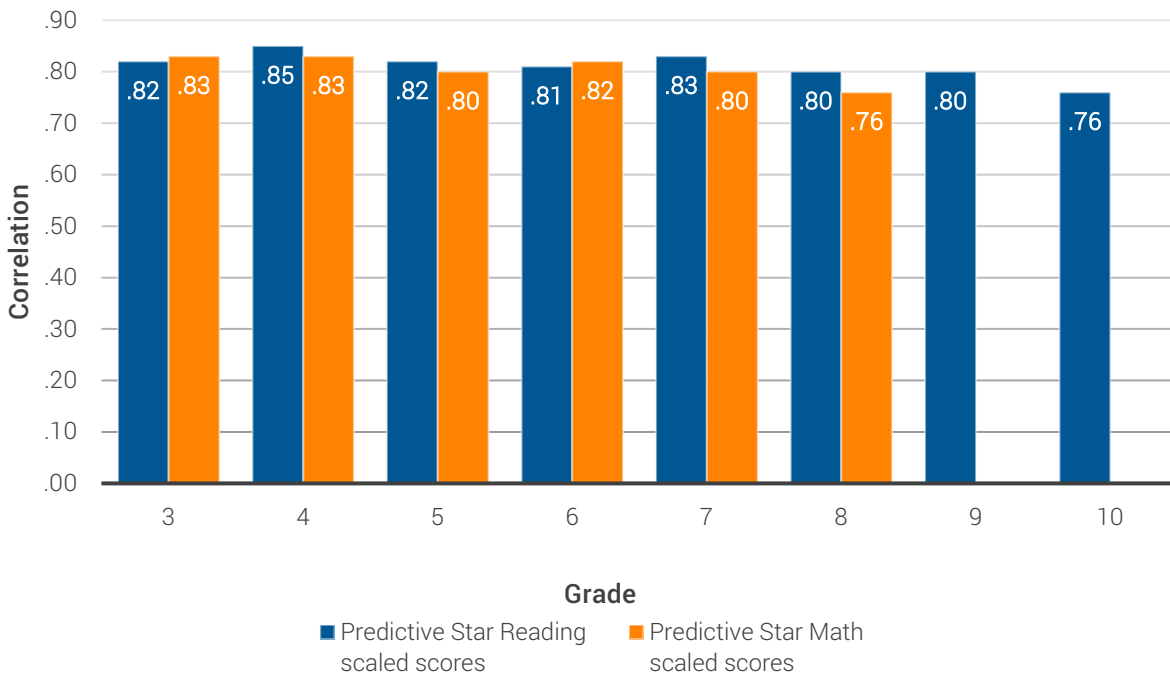
In evaluating the accuracy of the scale linkage, we used two methods to examine the differences between students' observed (actual) PARCC scores and our Star equivalents: (1) computing the RMSEL (the root mean squared errors of linking) using the scores from the linking study, and (2) applying the holdout sample, consisting of the subset of concurrent scores not used in the linking, to the linking results. Results showed that our linking computation performed as intended.

Predictive Star scores correlate highly with actual PARCC scores

To summarize the predictive power of Star Reading and Star Math, we calculated raw correlations between observed (actual) PARCC scores and projected Star scores. As figure 2 shows, the predictive correlation showed a strong relationship between the assessments (similar to the correlations from the concurrent sample, see figure 1, p. 4), indicating that earlier Star scores have a strong relationship with end-of-year PARCC scores. For reading, the correlations averaged .81 and for math, the associations were also high, averaging .81.

Star scores have a strong relationship with end-of-year PARCC scores.

Figure 2. Projected scores from Star Reading® and Star Math® highly correlate with PARCC scores



Star scores discriminate well between students who score proficient or not

Using the sample of actual PARCC scores, we were able to compare how our projected Star scores aligned with the observed PARCC scores. Table 2 displays classification diagnostics about whether students were correctly or incorrectly classified as proficient or not on the PARCC Assessments using projected Star scores. On average, students were correctly classified (i.e., overall classification accuracy) 86% of the time for reading and 91% of the time for math.

For Area Under the ROC Curve (AUC), a summary measure of diagnostic accuracy, Star Reading and Star Math averaged .91 and .94, respectively (also displayed in table 2). The AUCs far exceed the .85 standard set by the National Center on Response to Intervention to indicate convincing evidence that an assessment can accurately predict another assessment result or outcome.

Table 2. Proficiency forecasting using Star Reading® and Star Math® scores yields accurate results

Star Reading®								
Measure	Grade							
	3	4	5	6	7	8	9	10
Overall classification accuracy (percentage of correct classifications)	86%	87%	86%	86%	86%	83%	86%	85%
Area Under the ROC Curve	0.91	0.93	0.91	0.92	0.92	0.90	0.92	0.90
Star Math®								
Measure	Grade							
	3	4	5	6	7	8		
Overall classification accuracy (percentage of correct classifications)	89%	90%	92%	91%	91%	90%		
Area Under the ROC Curve	0.94	0.94	0.94	0.95	0.95	0.94		

Other diagnostic accuracy measures studied:

- ✓ **Sensitivity** represents the percentage of proficient students that were correctly forecasted, which for Star Reading averaged 70% and for Star Math averaged 62%.
- ✓ **Specificity** represents the percentage of not-proficient students that were correctly forecasted, which for Star Reading averaged 91% and for Star Math averaged 96%.
- ✓ **Positive predictive values** indicate that when Star scores forecasted students to be proficient, they actually were proficient 74% of the time for Star Reading and 79% of the time for Star Math.
- ✓ **Negative predictive values** indicate that when Star scores forecasted students to miss proficiency, they actually weren't proficient 89% of the time for reading and 92% of the time for math.
- ✓ **Proficiency status projection error**, the difference between actual and projected proficiency rates, indicates how well scores accurately predict proficiency within each grade. Star Reading average -1% and Star Math averaged -4% (negative scores indicate under-prediction while positive scores show over-prediction).

Appendix A: About Star Reading® and Star Math®

The computer-adaptive Star Reading and Star Math assessments serve multiple purposes including screening, progress monitoring, instructional planning, forecasting proficiency, standards mastery, and measuring growth. These highly reliable, valid, and efficient standards-based measures of student performance in reading and math provide valuable information regarding the acquisition of skills along a continuum of learning expectations. The assessments can be completed in about 20 minutes, and we recommend administering them two to five times a year for most purposes and more frequently—as often as weekly—when used in progress monitoring programs.



Star Reading and Star Math are highly rated for progress monitoring by the National Center on Intensive Intervention, and received high ratings for screening and progress monitoring by the National Center on Response to Intervention.

National Center on
INTENSIVE INTERVENTION

at American Institutes for Research ■



National Center on Response to Intervention
www.rti4success.org

Appendix B: PARCC Assessments achievement levels

Table B1. PARCC achievement level score ranges

PARCC achievement level score ranges: English language arts/literacy					
Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	650–699	700–724	725–749	750–809	810–850
4	650–699	700–724	725–749	750–789	790–850
5	650–699	700–724	725–749	750–798	799–850
6	650–699	700–724	725–749	750–789	790–850
7	650–699	700–724	725–749	750–784	785–850
8	650–699	700–724	725–749	750–793	794–850
9	650–699	700–724	725–749	750–790	791–850
10	650–699	700–724	725–749	750–793	794–850
PARCC achievement level score ranges: Mathematics					
Grade	Level 1	Level 2	Level 3	Level 4	Level 5
3	650–699	700–724	725–749	750–789	790–850
4	650–699	700–724	725–749	750–795	796–850
5	650–699	700–724	725–749	750–789	790–850
6	650–699	700–724	725–749	750–787	788–850
7	650–699	700–724	725–749	750–785	786–850
8	650–699	700–724	725–749	750–800	801–850

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

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>