April 22, 2016

Relating STAR Reading™ and STAR Math™ to the ACT® College Readiness Benchmarks

RENAISSANCE LEARNING™

Quick reference guide to the STAR Assessments™

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™—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.

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**.

RENAISSANCE LEARNING
P.O. Box 8036
Wisconsin Rapids, WI 54495-8036
(800) 338-4204
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.

- 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 tests of college readiness and which students might need additional assistance to reach benchmarks. Specifically, the purpose of this project is to explore statistical linkages between Renaissance Learning interim assessments[1] (STAR Reading and STAR Math) and the English, reading, and mathematics sections of the ACT®. If these linkages are sufficiently strong, they may be useful for:

1. The early identification of students at risk of failing to make college-readiness goals in reading and math, which could help teachers decide to adjust instruction for selected students.

2. Forecasting percentages of students at the benchmark performance levels on the ACT assessments sufficiently in advance to permit redirection of resources and serve as an early warning system for administrators at the building and district level.

---

[1] 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.

Assessments

The ACT®

This report is concerned with the ACT® sections of English, reading and mathematics. The choice of those subjects was made because they coincide with the content of the STAR interim assessments, STAR Reading and STAR Math.

The ACT reports scaled scores to describe a student's location on the achievement continuum ranging from 1 to 36. The ACT college readiness benchmarks for the sections of English, reading and mathematics are indicated in Table 1.

Table 1. ACT® test benchmarks

ACT® Subject-Area Test	The ACT® Test Benchmark
English	18
Reading	22
Mathematics	22

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).

Method

Data collection

Analysis plans included the evaluation of correlations and statistical linkages between scores on the ACT® subject areas of English, reading and mathematics, and STAR Reading and STAR Math. Such analyses require matched data, with student records that include both the ACT and STAR test scores. Using a secure data-matching procedure compliant with the federal Family Educational Rights and Privacy Act (FERPA), staff from 76 districts in 6 states (Alabama, Arkansas, Kentucky, North Carolina, Tennessee, and Wisconsin) provided Renaissance Learning with ACT test scores for students in grades 9-12 who had taken STAR Reading and/or STAR Math during school years ranging from 2007/08 to 2014/15. Each record in the resulting data file included a student's ACT test scores for the English, reading, and mathematics subject areas as well as scores on any STAR Reading or STAR Math tests taken during that same school year.

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

Sample

The matched STAR-ACT data was divided into two samples. Linking was completed using a concurrent sample, which included all STAR tests taken within 30 days before or after the date of the ACT test administration. The concurrent sample consisted of a total of 2,671 records with matched ACT English and STAR Reading scores, 2,670 records with matched ACT Reading and STAR Reading scores, and 1,495 records with matched ACT Mathematics and STAR Math scores. Of the concurrent sample, 10% of the records were 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 outside the +/-30 day ACT window were included in a predictive sample, which was used to evaluate the accuracy of using the linking results to predict ACT performance using STAR data from earlier in the school year.[2] In the predictive sample, STAR scaled scores were projected to the date of the ACT test administration using national growth norms (Renaissance Learning, 2016a, 2016b). National growth norms are based on grade and initial performance, and are updated annually using a five-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 ACT date was calculated. Then 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 time of the ACT. If a student had multiple STAR tests in the predictive sample, then all the projected scores were averaged. If a student had taken more than one ACT test, ACT scores from each administration were used in the analysis and projected STAR scores were averaged for each ACT administration.

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

Table 2a. Descriptive statistics for STAR™ and ACT® test scores (concurrent sample)

Linkage	Sample Size			ACT®		STAR™	
	Hold Out	Linking	Total	M	SD	M	SD
STAR Reading™ – ACT® English	267	2,404	2,671	16.88	5.89	944.19	333.41
STAR Reading™ – ACT® Reading	267	2,403	2,670	18.11	5.64	944.38	333.32
STAR Math™ – ACT® Mathematics	150	1,345	1,495	17.28	3.95	800.69	133.46

[2] A vast majority (>= 80%) of the predictive sample for English, reading, and mathematics consisted of STAR scores obtained prior to the date of the ACT test. STAR scores occurring after the date of the ACT test were included in the predictive sample and projected to the date of the ACT test by subtracting the expected growth from the student's scaled score.

Table 2b. Descriptive statistics for STAR™ and ACT® test scores (predictive sample)

Linkage	Sample Size	ACT®		STAR™	
		M	SD	M	SD
STAR Reading™ – ACT® English	14,246	17.50	6.03	981.18	295.17
STAR Reading™ – ACT® Reading	14,228	18.66	5.71	981.54	294.91
STAR Math™ – ACT® Mathematics	6,328	18.26	4.51	832.69	106.59

Correlations

Two sets of correlations were obtained from the sample: one between the ACT scores and concurrent STAR scores, and another between ACT scores and the ACT score equivalents (obtained from the linking). Table 3 displays these correlations for English, reading, and mathematics respectively.

For English, the correlation between the ACT English and STAR Reading was .71 and the correlation between ACT English and ACT English score equivalents was .75.

For Reading, the correlation between the ACT Reading and STAR Reading was .66 and the correlation between ACT Reading and ACT Reading score equivalents was .72.

For Mathematics, the correlation between the ACT Mathematics and STAR Math was .64 and the correlation between ACT Mathematics and ACT Mathematics score equivalents was .79.

Table 3. Pearson correlations between STAR™ scale scores and ACT® scale scores (concurrent sample)

Linkage	ACT® Score Correlation With:	
	Concurrent STAR™ Scale Scores	ACT® Score Equivalents
STAR Reading™ – ACT® English	.71	.75
STAR Reading™ – ACT® Reading	.66	.72
STAR Math™ – ACT® Mathematics	.64	.79

STAR™ equivalents to ACT® benchmark scores

A principal purpose of the linkage between STAR and ACT English, reading, and mathematics test scores was to identify the scores on STAR Reading and STAR Math that are approximately equivalent to the benchmark scores that on the ACT tests. Table 4 displays those scores for English, reading, and mathematics.

Because the linking was done using a sample from just 6 states, these benchmark scores should be considered approximations that can be updated with greater precision as more data become available in the future.

Table 4. Equivalent STAR™ score and ACT® benchmark scores

Subject Area	The ACT® test benchmark	STAR™
English	18	1068
Reading	22	1262
Mathematics	22	915

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 ACT scores and ACT score equivalents. Table 5 displays these statistics for English, reading, and mathematics respectively.

Table 5. Summary statistics from the ACT® English, reading, and mathematics linkage samples

Linkage	Linking Sample RMSEL	Holdout Sample Difference Scores				
		N	Mean	SD	Min	Max
STAR Reading™ – ACT® English	2.77	267	-0.04	4.34	-16	20
STAR Reading™ – ACT® Reading	2.82	267	0.30	4.31	-16	18
STAR Math™ – ACT® Mathematics	1.82	150	-0.03	2.76	-9	10

Classification accuracy

The predictive sample was used in analyses exploring the accuracy of using STAR tests to predict ACT performance based on STAR cutscores identified in the linking analysis.[3]

Two sets of correlations were calculated to summarize the predictive power of the STAR test scores: raw correlations between the projected STAR and observed ACT scale scores, and equated-score correlations between the ACT score equivalents obtained from the linking and the observed ACT scores. Table 6 displays these correlations for English, reading, and mathematics, respectively.

These correlations were similar in magnitude to the correlations presented earlier for the concurrent sample, indicating that projected STAR scores are reliable estimates of ACT performance.

For the STAR Reading – ACT English linkage, the raw correlation between STAR Reading projected scores and ACT English scores was .68 and the correlation between ACT English and ACT English score equivalents was .73.

For the STAR Reading – ACT Reading linkage, the raw correlation between STAR Reading projected scores and ACT Reading scores was .64 and the correlation between ACT Reading and ACT Reading score equivalents was .70.

For the STAR Math – ACT Mathematics linkage, the raw correlation between STAR Math projected scores and ACT Mathematics scores was .70 and the correlation between ACT English and ACT English score equivalents was .83.

Table 6. Pearson correlations between STAR™ projected scale scores and ACT® scale scores

Linkage	ACT® Score Correlation With:	
	Projected STAR™ Scale Scores	ACT® Score Equivalents
STAR Reading™ – ACT® English	.68	.73
STAR Reading™ – ACT® Reading	.64	.70
STAR Math™ – ACT® Mathematics	.70	.83

For the projection of students meeting ACT benchmarks, standard statistical classification diagnostics were calculated.

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 meet benchmarks on the ACT test. The classification diagnostic formulas are outlined in Table 7a and the types of classifications are summarized in Table 7b.

[3] Accuracy analyses were also conducted using a modified predictive sample in which only STAR scores occurring before the ACT test date were included. These accuracy results were highly similar to the results detailed in this section, which involve the predictive sample containing all STAR scores occurring more than 30 days before or after the ACT test date.

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 students meeting benchmark identified as such using STAR
Specificity	$\frac{TN}{TN + FP}$	Percentage of students not meeting benchmark identified as such using STAR
Positive predictive value (PPV)	$\frac{TP}{TP + FP}$	Percentage of students STAR projects will meet benchmark who actually meet benchmark
Negative predictive value (NPV)	$\frac{TN}{FN + TN}$	Percentage of students STAR projects will not meet benchmark who actually do not meet benchmark
Observed proficiency rate (OPR)	$\frac{TP + FN}{N}$	Percentage of students who meet benchmark
Projected proficiency rate (PPR)	$\frac{TP + FP}{N}$	Percentage of students STAR projects will meet benchmark
Proficiency status projection error	PPR - OPR	Difference between projected and observed proficiency rates

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

		ACT® Result		Total
		Met Benchmark	Did Not Meet Benchmark	
STAR™ Estimate	Will Meet Benchmark	*True Positive* (TP)	*False Positive* (FP)	*Projected Proficient* (TP + FP)
	Will Not Meet Benchmark	*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 Table 8 for English, reading, and mathematics, respectively.

On average, students were correctly classified as meeting the ACT benchmark or not (i.e., overall classification accuracy) 80% of the time for ACT English, 83% of the time for ACT Reading, and 89% of the time for mathematics.

Sensitivity statistics (i.e., the percentage of students who met ACT benchmarks students correctly forecasted) were 76% for ACT English, 62% for ACT Reading, and 67% for ACT Mathematics. Specificity statistics (i.e., the percentage of students who did not meet ACT benchmarks correctly forecasted) were higher than sensitivity, averaging 82% for ACT English, 90% for ACT Reading, and 96% for ACT Mathematics. Specificity is negatively related to observed proficiency rate, so lower observed proficiency rates tend to have higher specificity.

Positive predictive values were 79% for ACT English, 71% for ACT Reading, and 83% for ACT Mathematics. Therefore, when STAR scores forecasted students to be proficient, they actually were proficient between 71% of the time and 83% of the time depending upon the ACT subject area.

Negative predictive values were higher than positive predictive values, with values of 80% for ACT English, 86% for ACT Reading, and 91% for ACT Mathematics. The negative predictive value results indicated that when STAR scores forecasted that students would not meet ACT benchmarks, they actually did not meet the benchmarks 80% of the time for ACT English, 86% of the time for ACT Reading, and 91% of the time for ACT Mathematics.

Differences between the observed and projected proficiency rates (i.e., proficiency status projection error) indicated that STAR Reading and STAR Math scores tended to accurately predict the percent of students who met ACT benchmarks across the ACT subject areas. The proficiency status projection error was -1% for ACT English, -4% for ACT Reading, and -4% for ACT Mathematics. Negative values of proficiency status projection error indicate under-prediction.

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 exceeded that standard. The AUC was 0.87 for ACT English and 0.86 for ACT Reading, indicating that STAR Reading scores did a very good job of discriminating between which students met benchmarks for ACT English and reading and which did not. For ACT Mathematics the AUC was 0.93, indicating that STAR Math scores did a very good job of discriminating between which students met benchmarks on the ACT Mathematics section and which did not.

Table 8. Classification diagnostics for linkages

Measure	Linkage		
	STAR Reading™ – ACT® English	STAR Reading™ – ACT® Reading	STAR Math™ – ACT® Mathematics
Overall classification accuracy	80%	83%	89%
Sensitivity	76%	62%	67%
Specificity	82%	90%	96%
Positive predictive value (PPV)	79%	71%	83%
Negative predictive value (NPV)	80%	86%	91%
Observed proficiency rate (OPR)	46%	28%	23%
Projected proficiency rate (PPR)	45%	24%	19%
Proficiency status projection error	-1%	-4%	-4%
Area Under the ROC Curve	0.87	0.86	0.93

Conclusions and applications

The equipercentile linking method was used to link the STAR Reading score scales to the ACT subject-area tests of English and reading and the STAR Math score scales to the ACT Mathematics subject-area test. The result of each linkage analysis was an estimate of the approximately equivalent ACT score. Using the tables of linked scores, we identified STAR Reading and STAR Math scores that were linked to the benchmarks for the ACT subject areas of English, reading, and mathematics achievement levels (reported in Table 4). Because the linking was done using a sample from just 6 states, and are not representative of the nationwide student population, these equivalent benchmark scores 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 ACT subject-area tests. The correlation between ACT scores and concurrent STAR scores (i.e., STAR tests taken within +/- 30 days of the ACT test date) was .71, .66 and .64 for ACT English, ACT Reading and ACT Mathematics respectively. Similarly, the average correlation between ACT scores and predictive STAR scores (i.e., STAR tests projected to the ACT test date) was .68, .64 and .70 for ACT English, reading, and mathematics respectively. When projecting STAR scores to estimate ACT performance, students were correctly classified as meeting the ACT benchmark or not 80% of the time for ACT English, 83% of the time for ACT Reading, and 89% of the time for ACT Mathematics.

The statistical linkages between STAR interim assessments and the ACT subject-area tests in English, reading, and mathematics provide a means of forecasting student achievement on the ACT subject area tests based on STAR scores obtained earlier in the school year.

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

Renaissance Learning. (2016b). *STAR Reading: Technical manual*. Wisconsin Rapids, WI: Author. Available from Renaissance Learning by request to 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. Available online 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. Available online 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. Available online 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. Available online 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. Available online 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. Available online from http://www.rti4success.org/resources/tools-charts/screening-tools-chart

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 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.