RESEARCH REPORT

September 2022

Assessing the Predictive Validity of the IXL Real-Time Diagnostic Using Star Assessments as Criterion

Christina Schonberg, Ph.D

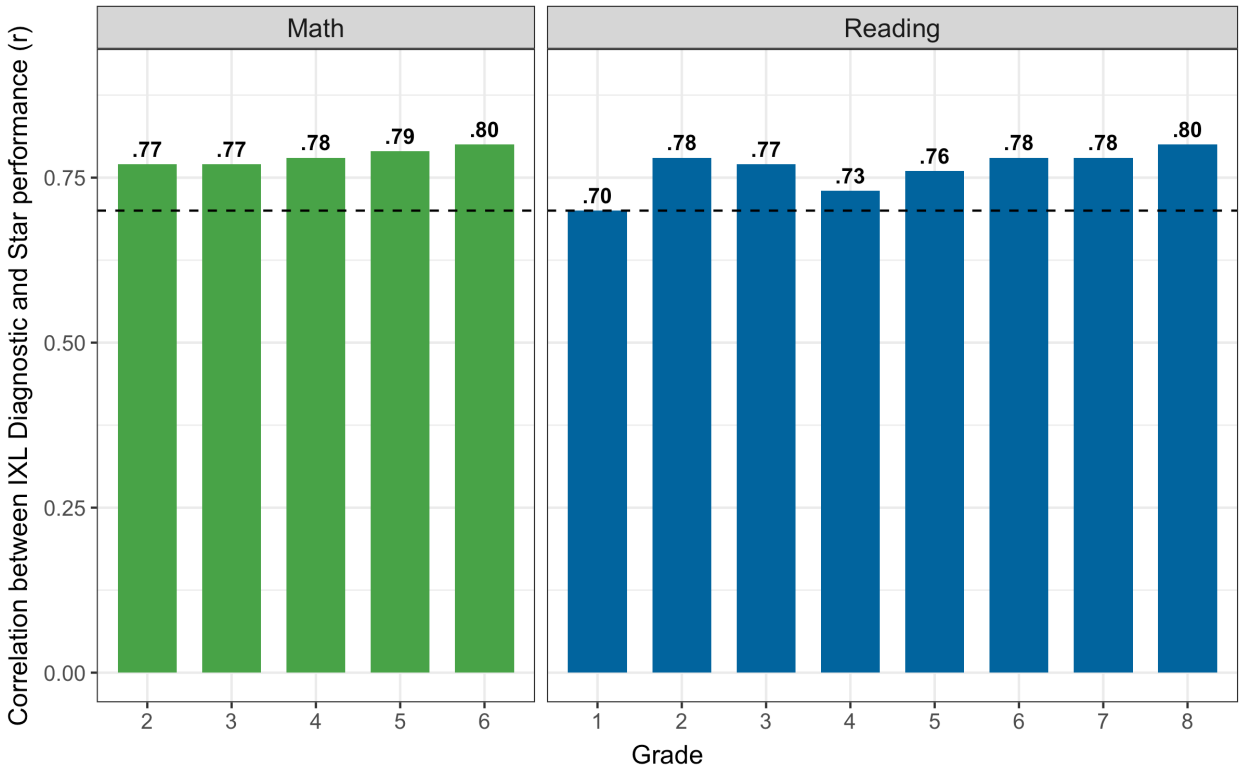
IXL LEARNING 777 Mariners Island Blvd., Suite 600, San Mateo, CA 94404
650-372-4040 | www.ixl.com

Executive Summary

IXL is an end-to-end teaching and learning solution that engages learners in grades Pre-K through 12 with a comprehensive curriculum and a first-of-its-kind assessment suite. A core component of IXL's assessment suite is the IXL Real-Time Diagnostic, an interim assessment designed to provide students and educators with in-the-moment information about students' subject area knowledge. The diagnostic was designed by a team of educators and mathematicians and uses Item Response Theory (Lord, 1980) to automatically adjust question difficulty based on students' response patterns. The diagnostic provides immediate insights about students' current knowledge levels and, using these results, creates personalized action plans that help students close gaps quickly.

IXL's Real-Time Diagnostic has been validated with other standardized assessments including the NWEA MAP Growth, ILEARN, FSA, and SOL assessments (An, 2021, 2022; IXL Learning, 2020b, 2020a; Schonberg, 2021). The goal of this study was to assess the validity of IXL's Real-Time Diagnostic in both math and reading using a new criterion measure: the Star Assessment. We analyzed math and reading data from students in grades 1-8 attending public schools in a large, suburban Oklahoma district. Within each subject, we found:

- Strong, positive correlations between IXL's Diagnostic and Star performance (all *rs* ≥ .70)
- A high degree of overlap in student proficiency classifications by the IXL Diagnostic and Star Assessment



These results provide further evidence that IXL's Real-Time Diagnostic is a valid measure of student grade-level proficiency.

Assessing the Predictive Validity of the IXL Real-Time Diagnostic Using Star Assessments as Criterion

Background

IXL is an end-to-end teaching and learning solution that is used by 1 in 4 students in the US and over 13 million students worldwide as of this writing. It includes a comprehensive curriculum for students in grades Pre-K through 12 as well as a first-of-its-kind assessment suite. A core component of IXL's assessment suite is the IXL Real-Time Diagnostic, an interim assessment that was developed by a collaborative team of educators and mathematicians and covers material aligned with state standards (see Bashkov et al., 2021).

IXL's Diagnostic is especially valuable because it provides immediate insights for students and educators about current knowledge levels in key strands of math and ELA. The diagnostic uses Item Response Theory (Lord, 1980) to analyze student response patterns. Based on students' responses, the diagnostic then provides personalized action plans that lay out clear next steps for students to reach their learning goals.

It is important that the information provided by IXL's Diagnostic maps onto real-world outcomes such as performance on other assessments—thus, collecting validity evidence for the diagnostic is essential. Prior validity research correlating the IXL Real-Time Diagnostic with other assessments has yielded favorable results. For example, studies using the NWEA MAP Growth assessment as criterion have found high predictive validity for IXL's Diagnostic (An, 2021; IXL Learning, 2020b). Similarly, the IXL Diagnostic was shown to have high predictive validity with the Indiana Learning Evaluation Assessment Readiness Network (ILEARN) assessment (IXL Learning, 2020a), the Virginia state Standards of Learning (SOL; Schonberg, 2021), and the Florida Standards Assessment (FSA; An, 2022). The goal of this study was to validate IXL's Diagnostic with a new external criterion measure: the Star Assessments, which are a widely-used set of formative assessments. Specifically, we examined the Star Math and Star Reading assessments as criterion measures.

RESEARCH QUESTIONS

This study aimed to answer the following research questions for math and reading, separately:

1. How strong are the correlations between scores on the IXL Real-Time Diagnostic assessment and the Star Assessment?
2. What is the degree of overlap in student grade-level proficiency classifications by IXL's Real-Time Diagnostic and the Star Assessment? That is, to what extent do students receive the same classifications (at or above grade level vs. below) by IXL's Diagnostic and the Star Assessment?

Study Design and Methodology

PARTICIPANTS

To be included in the math or reading analyses, students needed to have both a pinpointed IXL Real-Time Diagnostic score in math or reading and a Star Assessment score in the corresponding subject. Using these criteria, we included 7,524 students in the math analyses and 10,847 students in the reading analyses. Descriptive statistics for student demographic characteristics and performance on IXL's Diagnostic as well as the Star Assessment can be found in Appendix A.

DATA SOURCES

Star Assessment Data

To validate the IXL Real-Time Diagnostic with the Star Assessment, we obtained Spring 2022 student-level Star Assessment performance data from 30 schools in a large, suburban Oklahoma school district. In this district, students in grades 1 through 8 completed the Star Reading assessment, and students in grades 2 through 6 completed the Star Math assessment. The Star Assessments are widely used as formative assessments in districts nationwide and have been extensively validated. Both Star Reading and Star Math are computer adaptive tests (CATs), meaning that question difficulty changes based on students' patterns of correct and incorrect responses. The specific Star measure used as criterion in this study is percentile rank, which is a norm-referenced score obtained by comparing a student's scaled score to that of same-grade peers in a national sample. For more information about the Star Assessments, see https://www.renaissance.com/products/star-assessments/.

IXL Diagnostic Data

IXL Real-Time Diagnostic data were obtained from IXL's database. When a student completes a sufficient number of questions in a subject (math or ELA) in IXL's Diagnostic, they receive a pinpointed score that indicates their overall grade-level proficiency in that subject. For example, a score of 350 indicates that the student has acquired about 50% of third-grade material, whereas a score of 400 indicates that the student is ready to learn fourth-grade material. Within the ELA diagnostic, students receive an overall pinpointed ELA score as well as a pinpointed reading score. We obtained all available pinpointed diagnostic scores in math and reading for students in this Oklahoma district across the 2021-22 school year and used those closest to the end-of-year Star Assessment date in each subject.

ANALYTIC APPROACH

Research Question 1: Correlations Between IXL's Real-Time Diagnostic and Star Assessment Performance

We analyzed data from 30 elementary and middle schools in one Oklahoma school district where students completed IXL's Real-Time Diagnostic in math or reading and Star Assessments in math or reading. IXL's Diagnostic uses a continuous vertical scale across grades Pre-K through 12 ranging from 0 to 1300, indicating grade-level proficiency. In contrast, the Star Assessment measure we used—percentile rank—has values that range from 1 to 99.

Given this difference in scaling and range of scores, we used correlations (Pearson's *r*) to assess test-criterion relations within each grade separately. Correlation values can range from -1.00 to

+1.00 and indicate the strength of a linear relationship between two variables; values above .70 are considered to indicate a strong, positive relationship (Ratner, 2009). Thus, *r* coefficients above .70 between students' IXL Real-Time Diagnostic scores and their Star Assessment scores would provide additional predictive validity evidence for the IXL Real-Time Diagnostic as a measure of grade-level proficiency using the Star Assessments as criterion.

Research Question 2: Alignment of Student Proficiency Classifications Across Measures

In addition to correlations between IXL Real-Time Diagnostic and Star Assessment scores, we also examined alignment of student proficiency classifications across the two measures. Specifically, we conducted chi-square tests to investigate the degree to which students who performed at or above grade level on the IXL Real-Time Diagnostic also performed at or above benchmark (i.e., passing) on the Star Assessment (or, conversely, whether students' performance was below grade level on both measures). In addition, we conducted logistic regressions using IXL proficiency status (i.e., an IXL Real-Time Diagnostic score at or above grade level) to predict the likelihood of achieving proficiency (i.e., a classification of at or above benchmark) on the Star Assessment.

Results

MATH

We found strong, statistically significant correlations between IXL Real-Time Diagnostic math score and Star Math percentile rank in each grade (smallest $r = .77$, all $ps < .001$). See Figure 1 for scatterplots of these correlations by grade.

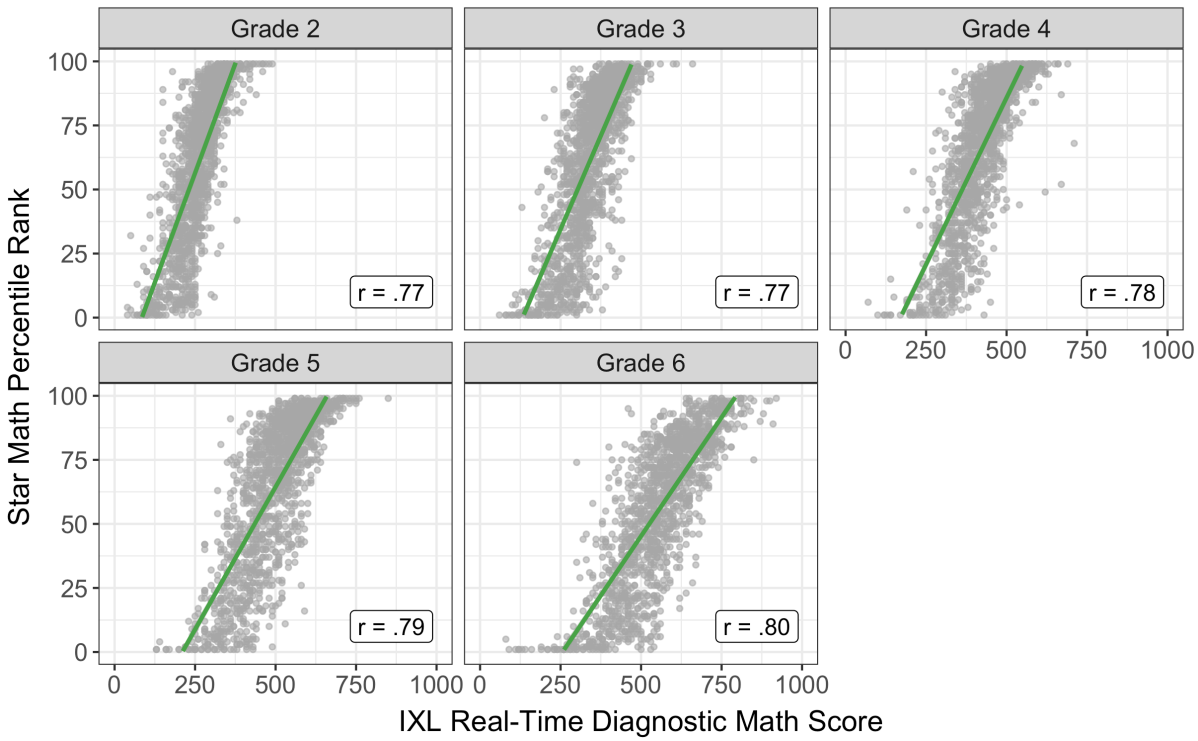


Figure 1. Correlations between IXL Real-Time Diagnostic score and Star percentile rank in math. All correlations are statistically significant at p < .001.

Next, we examined the alignment of student proficiency classifications across the two measures. We found that IXL's Real-Time Diagnostic and the Star Math assessment aligned in how they classified the majority (77%) of students, $\chi^2(1) = 1805.52$, $p < .001$. That is, 77% of students were classified by both measures as performing at or above grade level, or by both measures as performing below grade level. In line with this finding, a logistic regression model predicting Star Math proficiency status from IXL proficiency status showed that students who were classified as proficient by the IXL Real-Time Diagnostic were 11.46 times more likely to be proficient (i.e., at or above benchmark) on Star Math than students who were not proficient based on IXL's Diagnostic (see Table B1).

ELA

As in math, we found strong, statistically significant correlations between IXL Real-Time Diagnostic reading score and Star Reading percentile rank in each grade (smallest $r = .70$, all $ps < .001$). See Figure 2 for scatterplots of these correlations by grade.

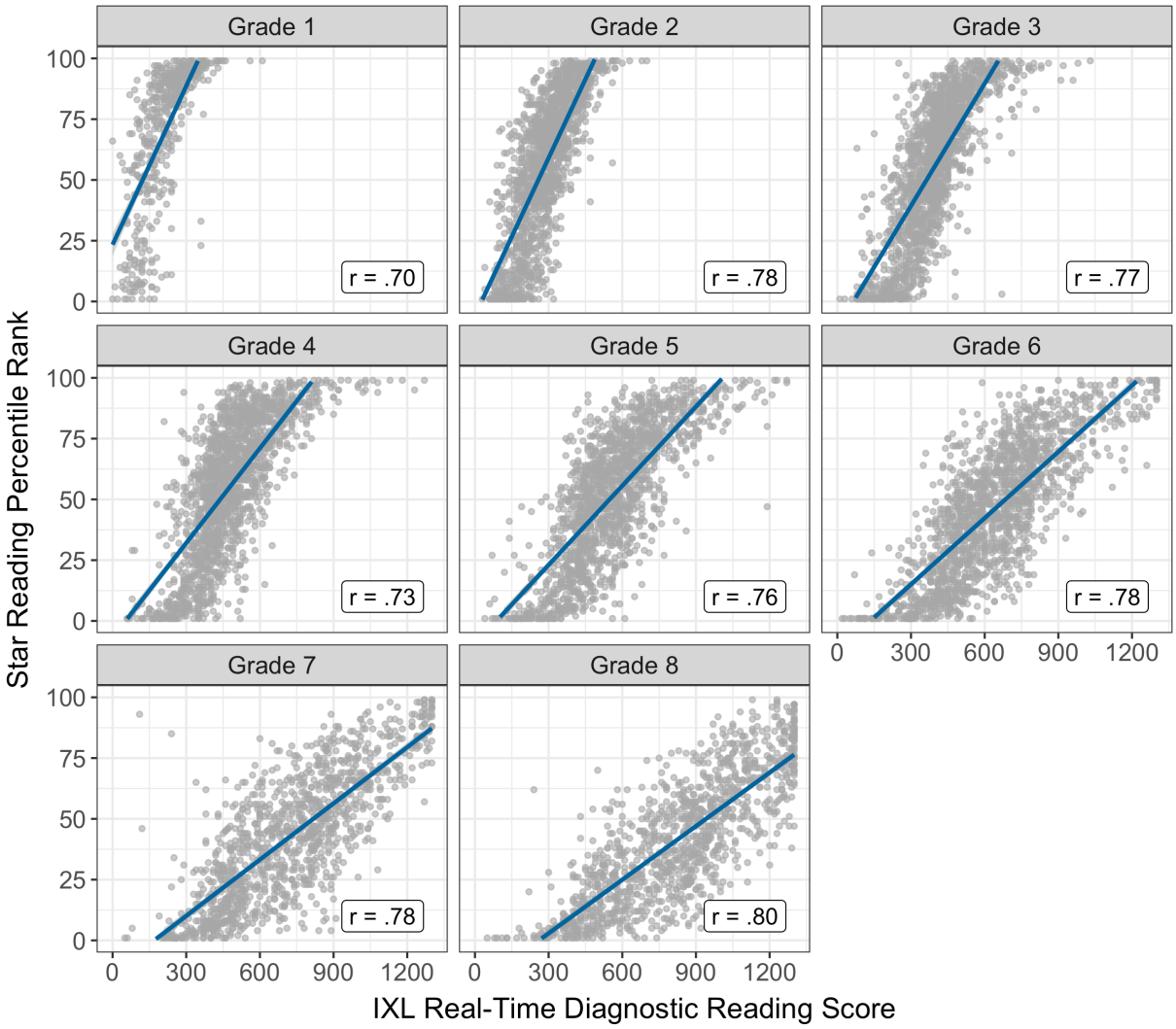


Figure 2. Correlations between IXL Real-Time Diagnostic score in Reading and Star Reading percentile rank. All correlations are statistically significant at p < .001.

Next, we examined the alignment of student proficiency classifications across measures. As in math, we found that IXL's Real-Time Diagnostic in reading and the Star Reading assessment aligned in how they classified the majority (79%) of students, $\chi^2(1) = 3283.65$, $p < .001$. That is, 79% of students were classified by both measures as performing at or above grade level, or by both measures as performing below grade level. Furthermore, a logistic regression model predicting Star Reading proficiency status from IXL Real-Time Diagnostic proficiency status found that students classified as proficient by IXL's Diagnostic were 13.22 times more likely to be proficient (i.e., at or above benchmark) on Star Reading, compared to students who were not classified as proficient by the IXL Diagnostic (see Table B2).

Conclusion

In this study, we investigated the predictive validity of the IXL Real-Time Diagnostic using a novel set of assessments as criterion: the Star Assessments in Math and Reading. We examined test-criterion relationships between the two measures in each subject and found strong correlations between the IXL Real-Time Diagnostic and the Star Assessment. In addition, we found a high degree of alignment in the two measures' classifications of students into proficiency levels (on or above vs. below). Coupled with prior examinations of content validity (Bashkov et al., 2021), construct validity (IXL Learning, 2020a), and predictive validity (An, 2022; IXL Learning, 2020b; Schonberg, 2021), these findings corroborate a strong program of validity for the IXL Real-Time Diagnostic.

References

An, X. (2021). *Assessing the predictive validity of the IXL Real-Time Diagnostic among marginalized students* (pp. 1-10). https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study-Marginalized-Students.pdf

An, X. (2022). *Examining the predictive validity of the IXL Real-Time Diagnostic using the Florida Standards Assessments as criterion* (pp. 1-8). https://www.ixl.com/materials/us/research/IXL_Real-Time_Diagnostic_Validation_Study_(FL_FSA).pdf

Bashkov, B. M., Mattison, K., & Hochstein, L. (2021). *IXL design principles: Core features grounded in learning science research* (pp. 1–18). https://www.ixl.com/research/IXL_Design_Principles.pdf

IXL Learning. (2020a). *Assessing the construct validity and the predictive validity of the IXL Real-Time Diagnostic* (pp. 1–15). https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study-ILEARN.pdf

IXL Learning. (2020b). *Validation study of the IXL Real-Time Diagnostic using MAP Growth Assessments* (pp. 1–11). https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study.pdf

Lord, F. (1980). *Applications of Item Response Theory to practical testing problems.* Lawrence Erlbaum Associates.

Ratner, B. (2009). The correlation coefficient: Its values range between +1/−1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing, 17*(2), 139-142.

Schonberg, C. (2021). *Demonstrating grade-level predictive validity of the IXL Real-Time Diagnostic using the Virginia SOL as criterion* (pp. 1–14). https://www.ixl.com/materials/us/research/IXL_Real-Time_Diagnostic_Grade-Level_Validation_Study_(VA_SOL).pdf

Appendix A: Descriptive Statistics of Student Demographics and Performance

Table A1. Student grade-level characteristics

	Math		Reading	
# of students (total)	7,524		10,846	
	N	%	N	%
Grade level				
Grade 1	--	--	494	4.55
Grade 2	1,449	19.26	1,567	14.44
Grade 3	1,586	21.08	1,699	15.66
Grade 4	1,542	20.49	1,620	14.94
Grade 5	1,461	19.42	1,558	14.36
Grade 6	1,486	19.75	1,580	14.57
Grade 7	--	--	1,170	10.79
Grade 8	--	--	1,159	10.68

Table A2. Means (standard deviations) for math assessments

	IXL Real-Time Diagnostic math score	Star Math percentile rank
Grade 2	255.64 (65.88)	58.34 (29.25)
Grade 3	340.21 (79.19)	60.84 (30.08)
Grade 4	418.42 (83.83)	64.57 (27.97)
Grade 5	494.71 (104.59)	63.23 (29.46)
Grade 6	554.68 (123.09)	55.38 (28.50)

Note. N = 7,524. IXL Real-Time Diagnostic scores and Star Math percentile ranks are from the end of the school year.

Table A3. Means (standard deviations) for reading assessments

	IXL Real-Time Diagnostic reading score	Star Reading percentile rank
Grade 1	212.11 (92.92)	69.50 (29.04)
Grade 2	287.33 (103.46)	56.30 (28.76)
Grade 3	387.93 (131.67)	54.11 (28.57)
Grade 4	476.88 (153.90)	55.04 (27.37)
Grade 5	555.03 (185.72)	50.73 (26.50)
Grade 6	635.85 (225.39)	45.58 (26.17)
Grade 7	722.35 (259.62)	42.66 (25.69)
Grade 8	825.70 (280.64)	41.53 (25.72)

Note. N = 10,846. IXL Real-Time Diagnostic scores and Star Reading percentile ranks are from the end of the school year.

Appendix B: Chi-Square and Logistic Regression Results

Table B1. 2x2 table, chi-square, and logistic regression for math

		Star Math		χ^2	Odds ratio
		Below benchmark	At or above benchmark		
IXL Real-Time Diagnostic Math	Below grade level	1,488 (53%)	1,313 (47%)	1805.52***	11.46***
	At or above grade level	425 (9%)	4,298 (91%)		

Note. N = 7,524.

*** $p < .001$.

Table B2. 2x2 table, chi-square, and logistic regression for reading

		Star Reading		χ^2	Odds ratio
		Below benchmark	At or above benchmark		
IXL Real-Time Diagnostic Reading	Below grade level	2,665 (74%)	916 (26%)	3283.65***	13.22***
	At or above grade level	1,311 (18%)	5,955 (82%)		

Note. N = 10,846.

*** $p < .001$.