RESEARCH REPORT

October 2020

Assessing the Construct Validity and the Predictive Validity of the IXL Real-Time Diagnostic

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

Executive Summary

IXL is a personalized learning platform designed to help students build academic skills. Embedded in IXL, the IXL Real-Time Diagnostic assesses students across concepts in the K-12 curriculum and delivers up-to-the-minute insights on their grade-level proficiency in math and English language arts (ELA). Using these insights, IXL creates personalized action plans that guide every learner to the exact skills that will help him or her grow.

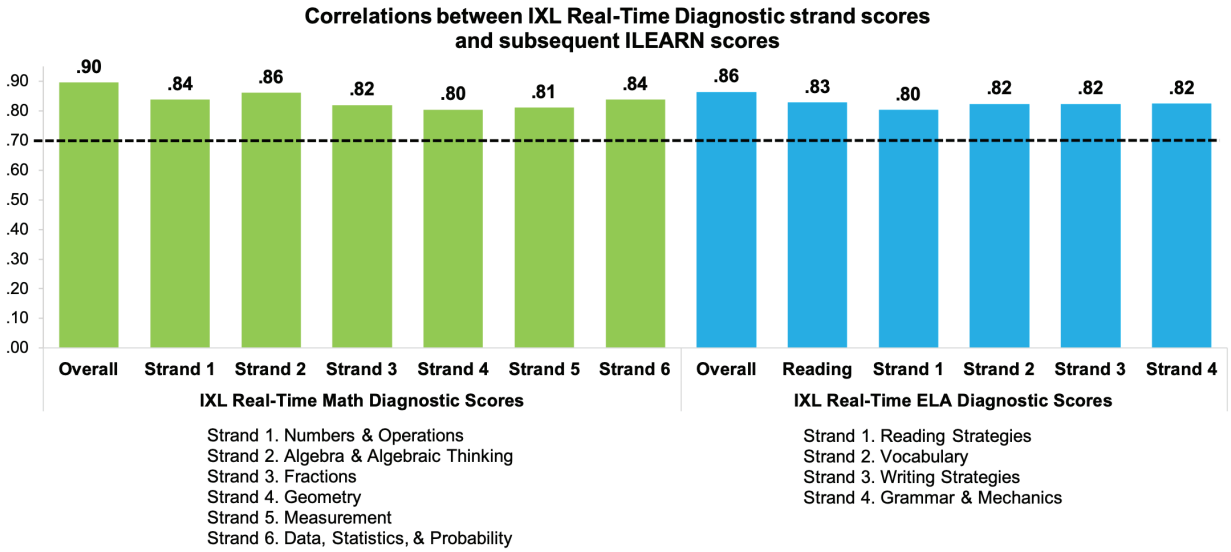
To assess a student's knowledge levels, the IXL Real-Time Diagnostic applies item response theory (IRT) models to estimate the numeric scores for a set of strands (i.e., broad categories of skills). For math, the strands include (a) Numbers & Operations; (b) Algebra & Algebraic Thinking; (c) Fractions; (d) Geometry; (e) Measurement; and (f) Data, Statistics, & Probability. For ELA, the strands include (a) Reading Strategies, (b) Vocabulary, (c) Writing Strategies, and (d) Grammar & Mechanics. The overall diagnostic scores for math, ELA, and reading are weighted averages of the strand scores.

The IXL Real-Time Diagnostic was collaboratively created and reviewed by a group of educators and mathematicians, and covers academic skills aligned to the Common Core and other important standards. As a rigorous assessment, the IXL Real-Time Diagnostic is widely used in schools and classrooms to evaluate student knowledge and track growth. These facts are indicators of the face validity and content validity of the IXL Real-Time Diagnostic.

To further examine the psychometric properties of the IXL Real-Time Diagnostic with respect to construct validity and predictive validity, IXL Learning conducted this study using student-level data from 12 public elementary schools in an Indiana school district. Students used the IXL Real-Time Diagnostic during the 2018-19 school year and took the Indiana Learning Evaluation Assessment Readiness Network (ILEARN) test at the end of that school year. Key findings include:

- **The IXL Real-Time Diagnostic in math and ELA showed a coherent internal structure and high reliability as a measure of math and ELA achievement, respectively.** The a-priori hypothesized one-factor models for math and ELA fit the data well, with all diagnostic strands having strong positive factor loadings onto their corresponding latent factor. As such, the results supported the construct validity of inferences drawn from IXL Real-Time Diagnostic scores with respect to math and ELA achievement.

- **The IXL Real-Time Diagnostic scores were a good predictor of subsequent performance using ILEARN as a criterion measure.** We found strong and positive correlations between IXL Real-Time Diagnostic scores and subsequent ILEARN scores. The overall correlation was .90 in math and .86 in ELA. These and individual strand correlations are shown in the figure below. Additionally, for both math and ELA, students whose IXL Real-Time Diagnostic scores were at or above their grade level were over 10 times more likely to reach proficiency on the ILEARN test.



ASSESSING THE CONSTRUCT VALIDITY AND THE PREDICTIVE VALIDITY OF THE IXL REAL-TIME DIAGNOSTIC

Study Design and Methodology

To further validate the IXL Real-Time Diagnostic, we examined data from 3,744 students in grades 3 through 6 attending 12 public elementary schools in an Indiana school district, who used IXL during the 2018-19 school year. We used two sources of data in this study: students' IXL Real-Time Diagnostic scores during the 2018-19 school year and their state assessment data in spring 2019.

IXL Real-Time Diagnostic scores during the 2018-19 school year were retrieved from the IXL database and included the overall math score, scores for the six math strands, overall ELA score, overall reading score, and scores for the four ELA strands. To get pinpointed diagnostic scores for a certain subject, a student needs to answer a certain number of questions. As a result, not all students in the sample had pinpointed diagnostic scores. Therefore, our validation analyses were conducted on data from 778 students who received pinpointed IXL Diagnostic math scores and 1,757 students who received pinpointed IXL Diagnostic ELA scores.

2019 ILEARN state assessment and demographic background data for students in grades 3 through 6 were provided by the district. Adopted in 2019, ILEARN is the state assessment of Indiana. It reports student achievement levels according to the Indiana Academic Standards. Students' proficiency levels were determined based on their scale scores. For example, for 6th graders, ILEARN math scores ranging from 6488 to 6544 indicate *Approaching Proficiency*, and ILEARN math scores ranging from 6545 to 6604 indicate the student is *At Proficiency* (Indiana Interpretive Guide for Statewide Assessments). Demographic information included student gender, race/ethnicity, economically disadvantaged status, English language learner status, and special education status (see Table A1 in Appendix A).

Research Questions

Using the data described above, in this study we focused on collecting two types of validity evidence by answering the following research questions:

1. **Construct validity:** Do the math strands in the IXL Real-Time Diagnostic math assessment and the language arts strands in the IXL Real-Time Diagnostic ELA assessment have a coherent factor structure? In other words, are the indicators making up the IXL Real-Time Diagnostic in math and ELA, respectively, related to one another and to the overall construct in theoretically expected ways? Given a coherent unidimensional factor structure, do the Diagnostic scores in math and ELA exhibit high internal consistency (i.e., reliability)?

2. **Predictive validity:** Do IXL Real-Time Diagnostic scores in math and ELA exhibit sufficient predictive validity? That is, do Real-Time Diagnostic scores in math and ELA predict how students performed on subsequent ILEARN math and ELA assessments? Do IXL Real-Time Diagnostic scores accurately predict proficiency on the ILEARN assessments for students at or above their grade level as measured by the Real-Time Diagnostic?

Analyses

Prior to conducting the substantive analyses to answer the research questions outlined above, we examined all descriptive statistics on the IXL Real-Time Diagnostic and ILEARN scores in math and ELA (see Appendix C). This was an important step, as it allowed us to (a) visually examine the correlation matrices which serve as input for each model and (b) assess the univariate and multivariate normality assumptions of the analytic techniques we discuss next.

To examine the construct validity of the IXL Real-Time Diagnostic, we conducted confirmatory factor analyses (CFA) under the structural equation modeling framework. In separate math and ELA analyses, a CFA model was constructed including a latent (i.e., unobserved) variable representing academic achievement in either math or ELA and all related strands representing its observed variables (see Figures 1 and 2 for visual depictions of the models). Similar to linear regression, the direct paths represented by arrows in the figures (also known as factor loadings) represent the relationship between the latent factor (i.e., construct being measured) and the observed indicators (i.e., manifestations of the latent construct in performance on the various strands). Before interpreting these coefficients, however, one must evaluate whether the theoretical model fits the data. The chi-square (χ^2) goodness-of-fit test and the following fit indices were used to evaluate model fit: the comparative fit index (CFI), the Tucker Lewis Index (TLI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA). As the chi-square is too sensitive in large samples (Hu & Bentler, 1998), we focus on the fit indices instead, using the following guidelines for good model-data fit: CFI >= .95, TLI >= .95, SRMR <= .08, and RMSEA <= .08 (Brown, 2006). Given a well-fitting unidimensional model in math and/or ELA, we would then compute McDonald's (1999) omega, a measure of internal consistency or reliability, based on each model's solution.

To examine the predictive validity of the Diagnostic, we examined the relationship between IXL Real-Time Diagnostic math/ELA scores and ILEARN math/ELA assessment scores using the Pearson product moment correlation (*r*). This correlation coefficient is a measure of the linear relationship between two variables and ranges from -1.00 to 1.00. An *r* value of 0 indicates no correlation, whereas *r* values greater than .70 indicate a strong positive relationship (Ratner, 2009). High positive correlations between IXL Real-Time Diagnostic math/ELA scores and corresponding ILEARN math/ELA assessment scores would provide strong predictive validity evidence for the IXL Real-Time Diagnostic with respect to future student academic achievement. In addition, we computed chi-square statistics and odds ratios using logistic regression to examine whether students whose IXL Real-Time Diagnostic scores were at or above their grade level were more likely to reach proficiency on the ILEARN assessments and vice versa.

Following What Works Clearinghouse (2020) guidelines, each analysis is accompanied by a test of statistical significance and a probability (*p*) value. The *p*-value is the probability of observing the current or more extreme data, assuming the effect is zero (Cohen, 1994). As such, the smaller the *p*-value, the less likely it is the result occurred at random, with .05, .01, and .001 commonly used as thresholds in research practice. Effects associated with *p*-values smaller than these thresholds are considered statistically significant at each of these significance levels.

Results

Construct Validity: the IXL Real-Time Diagnostic Math Assessment

To assess construct validity for the IXL Real-Time Diagnostic math assessment, we specified and evaluated the fit of the model shown in Figure 1. As the IXL Real-Time Diagnostic math assessment was developed to assess achievement in each of its strands, we hypothesized a one-factor model with strong positive relationships between the strands and the factor. Descriptive statistics are presented in Table C1, Appendix C. The model fit indices suggested good model-data fit: χ^2 = 25.549 (*df* = 9; *p* < .01), CFI = .997, TLI = .995, SRMR = .007, and RMSEA = .049. As expected, all math strands had high positive factor loadings on the latent factor (math achievement), with standardized coefficients ranging from .88 to .94 (see Table B1 in Appendix B for detailed results).

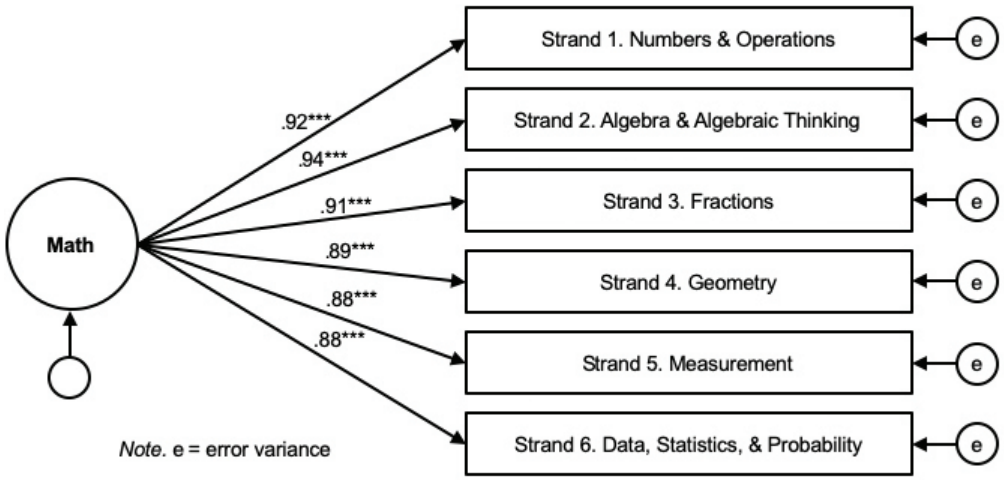


Figure 1. Standardized CFA model for the IXL Real-Time Diagnostic math assessment

Construct Validity: the IXL Real-Time Diagnostic ELA Assessment

Similar to the analysis above, we specified and tested a one-factor model to examine the factor structure of the IXL Real-Time Diagnostic ELA assessment (see Figure 2). Descriptive statistics are presented in Table C3, Appendix C. This model also fit the data well: $\chi^2 = 38.659$ ($df = 2$; $p < .001$), CFI = .995, TLI = .986, SRMR = .007, and RMSEA[1] = .102. Moreover, as hypothesized, all ELA strands had high positive factor loadings on the latent factor (ELA achievement), with standardized coefficients ranging from .90 to .93 (see Table B1 in Appendix B for detailed results).

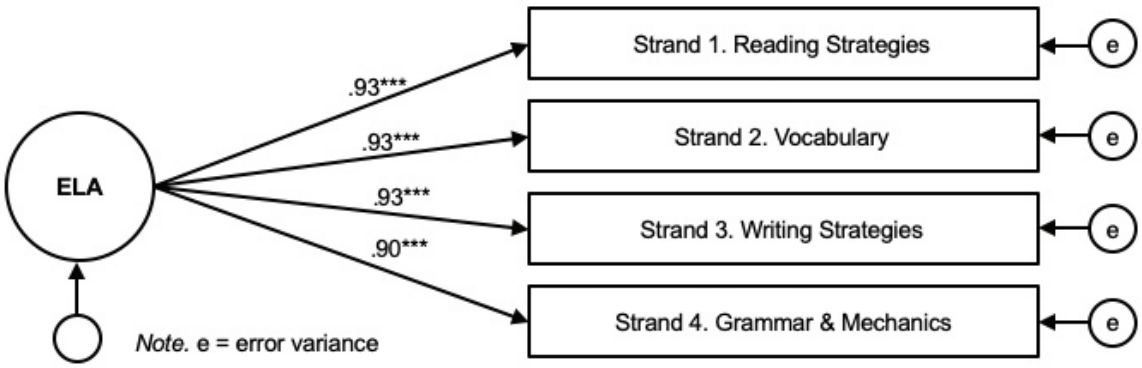


Figure 2. Standardized CFA model for the IXL Real-Time Diagnostic ELA assessment

Reliability: the IXL Real-Time Diagnostic Math and ELA Assessments

Given the favorable construct validity results above, we computed the reliability of the IXL Real-Time Diagnostic scores in math and ELA using the output from each model's unidimensional solution. Similar to other indices of reliability, McDonald's (1999) omega is a measure of internal consistency and ranges from 0 to 1 with values closer to 1 indicating adequate reliability. The omega values were .964 for math and .960 for ELA in the samples in our analyses. Reliability indices of this magnitude suggest that IXL Real-Time Diagnostic scores are highly reproducible.

Predictive Validity: the IXL Real-Time Diagnostic Math Assessment

The IXL Real-Time Diagnostic math assessment had a strong positive correlation with the ILEARN math assessment. The correlation between overall IXL Real-Time Diagnostic scores and ILEARN math scores was .90 ($p < .001$). This relationship is displayed via a scatterplot in Figure 3. The correlations between the six IXL Diagnostic math strand scores and ILEARN math scores were also high and positive (ranging from .80 - .86; all p values < .001). See Table C1 in Appendix C for the full correlation matrix and descriptive statistics.

[1] The value exceeded the criterion of RMSEA <= .08, but the RMSEA often falsely indicates poor model fit, as it penalizes models with few degrees of freedom (Kenny, Kaniskan, & McCoach, 2015), which was the case here (df = 2).

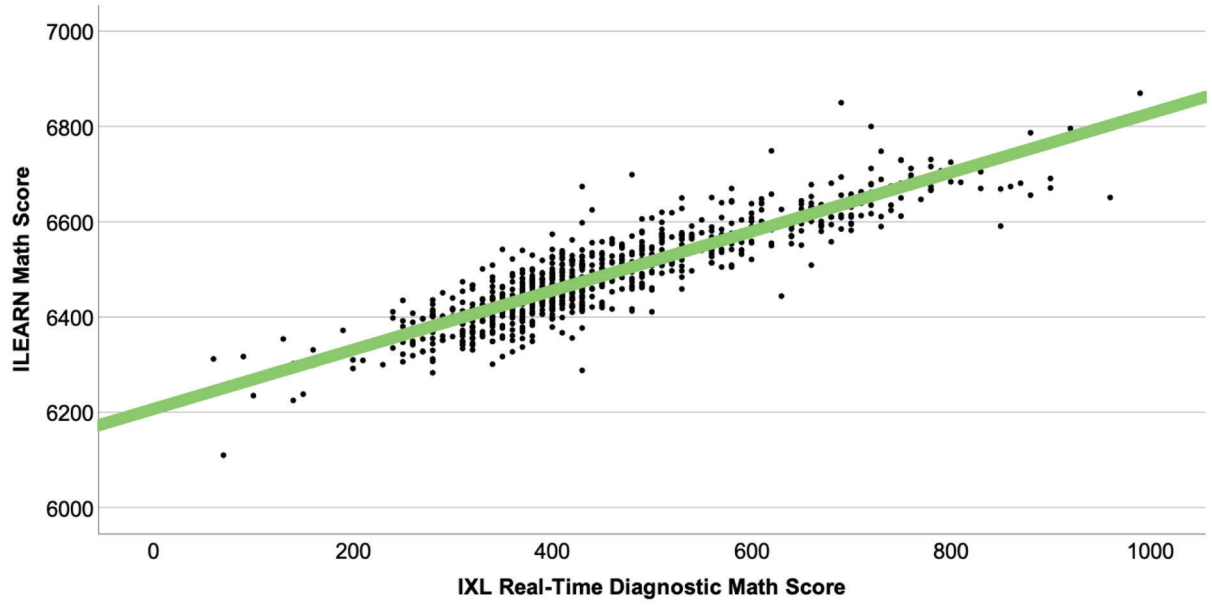


Figure 3. Correlation between IXL Real-Time Diagnostic math scores and ILEARN math scores

Additionally, we found that students whose IXL Real-Time Diagnostic math scores were at or above their grade level were more likely to reach ILEARN math proficiency ($\chi^2 = 243.18$, $p < .001$). Among students whose IXL Real-Time Diagnostic math scores were at or above grade level, 73.5% had ILEARN math scores at or above proficient. Among students whose IXL Real-Time Diagnostic math scores were below grade level, 84.4% had ILEARN math scores below proficiency. (see Figure 4 below and Table C2 in Appendix C).

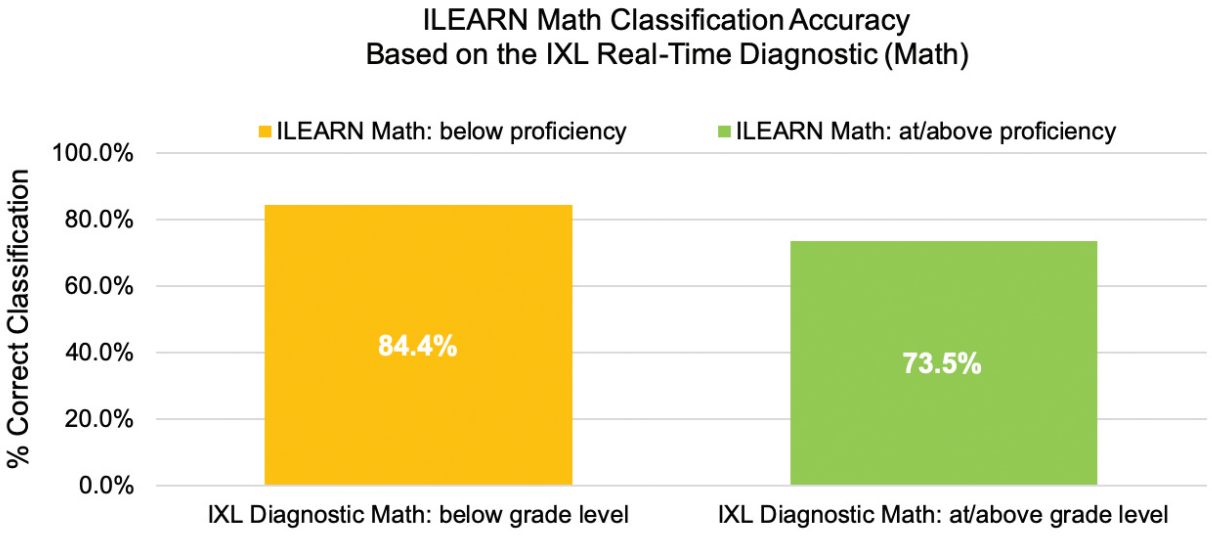


Figure 4. The relationship between IXL Real-Time Diagnostic math scores (at or above vs. below grade level) and the ILEARN math assessment scores (at or above proficient vs. below proficient)

We also examined the likelihood of reaching proficiency on the ILEARN math assessment for students whose IXL Real-Time Diagnostic math scores were at or above their grade level versus students below grade level. Logistic regression showed that students whose IXL Real-Time Diagnostic math scores were at or above their grade level were 11 times more likely to reach ILEARN proficiency in math than students whose IXL Real-Time Diagnostic math scores were below grade level, after accounting for gender, race/ethnicity, economically disadvantaged status, English language learner status, and special education status (see Table C2 in Appendix C for details). Therefore, not only is the IXL Real-Time Diagnostic math assessment a good predictor of future performance, but using it to keep students on track is very likely to result in achieving favorable results on state assessments as well.

Predictive Validity: the IXL Real-Time Diagnostic ELA Assessment

Similar to math, the IXL Real-Time Diagnostic ELA assessment had a positive and strong correlation with the ILEARN ELA assessment (r = .86, p < .001). The scatterplot presented in Figure 5 displays this relationship. Similarly, the correlation between IXL Diagnostic overall reading scores and the ILEARN ELA scores was .83 (p < .001). The correlations between the four IXL Diagnostic ELA strand scores and ILEARN ELA scores were also high and positive (ranging from .80 to .82, with p values < .001; see Table C3 in Appendix C for the correlation matrix and descriptive statistics).

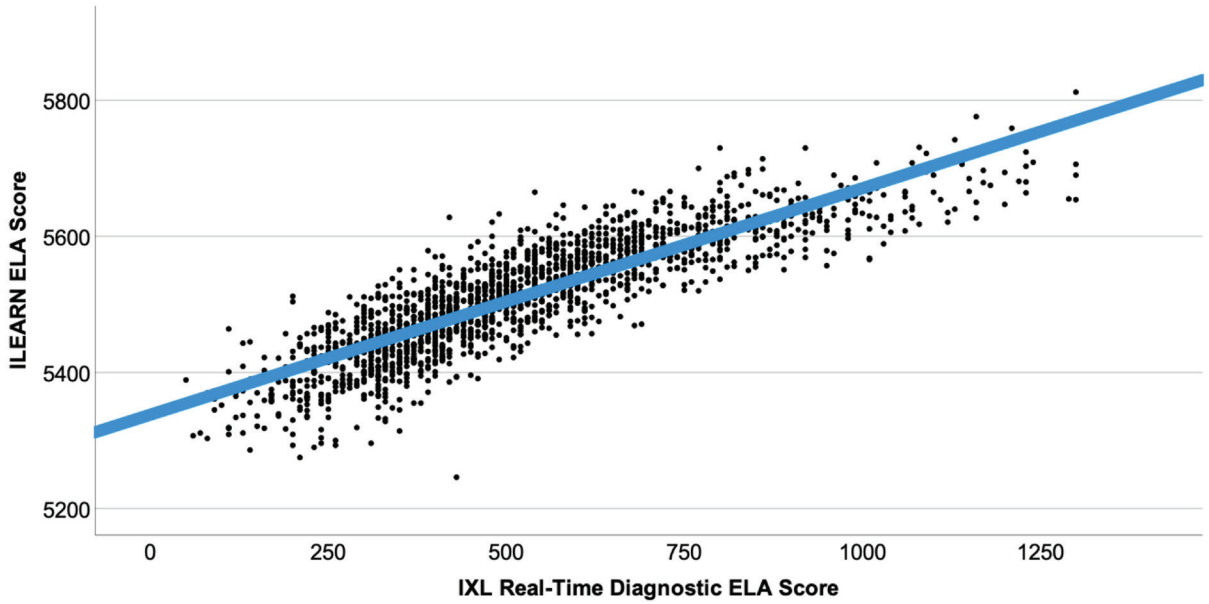


Figure 5. Correlation between IXL Real-Time Diagnostic ELA scores and ILEARN ELA scores

Additionally, students whose IXL Real-Time Diagnostic ELA scores were at or above their grade level were more likely to reach ILEARN ELA proficiency (χ^2 = 654.77, p < .001). Among students whose IXL Real-Time Diagnostic ELA scores were at or above their grade level, 77.3% had ILEARN ELA scores at or above proficiency. Among students whose IXL Real-Time Diagnostic ELA scores were below grade level, 87.8% had ILEARN math scores below proficiency (see Figure 6 below and Table C4 in Appendix C).

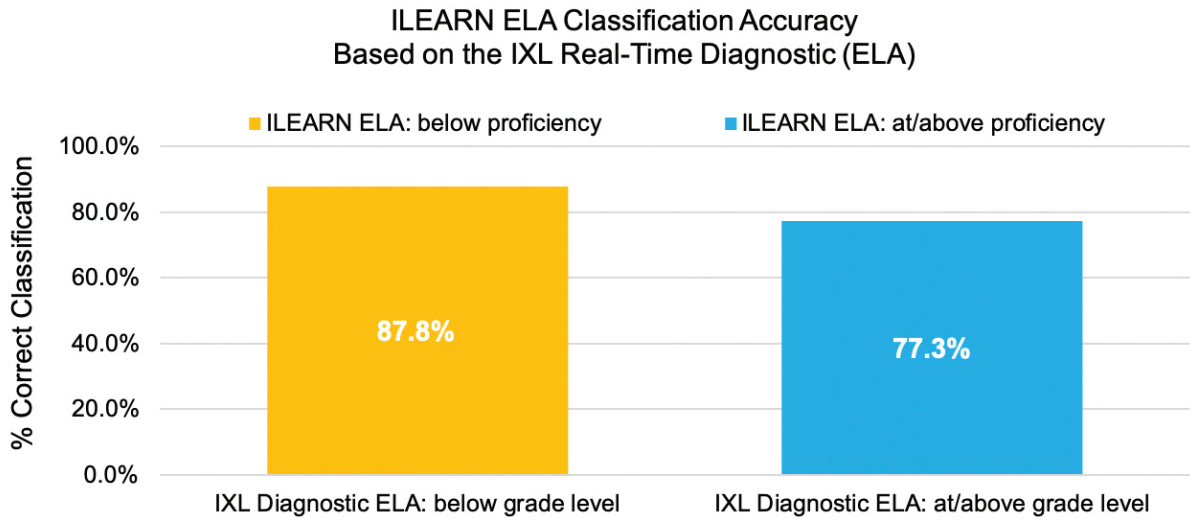


Figure 6. The relationship between IXL Real-Time Diagnostic ELA scores (at or above vs. below grade level) and the ILEARN ELA assessment scores (at or above proficient vs. below proficient)

We also examined the likelihood of reaching proficiency on the ILEARN ELA assessment for students whose IXL Real-Time Diagnostic ELA scores were at or above their grade level versus students below grade level. Logistic regression showed that students whose IXL Real-Time Diagnostic ELA scores were at or above their grade level were nearly 18 times more likely to reach ILEARN proficiency in ELA than students whose IXL Real-Time Diagnostic ELA scores were below grade level, after accounting for gender, race/ethnicity, economically disadvantaged status, English language learner status, and special education status (see Table C4 in Appendix C for details).

Conclusion

The IXL Real-Time Diagnostic exhibited excellent psychometric properties with respect to construct validity and reliability as a measure of math and ELA achievement, and with respect to predictive validity as a predictor of subsequent performance with the ILEARN used as a criterion. These findings provide ample evidence to support inferences based on the IXL Real-Time Diagnostic scores in regard to students' math and ELA achievement and future outcomes.

References

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*, 486-507.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum Associates.

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

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook* (Version 4.1). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

Appendix A. Demographic Background

Table A1. Demographic background information for the samples

	Students with IXL Real-Time Diagnostic math scores		Students with IXL Real-Time Diagnostic ELA scores	
# of students	778		1,757	
Gender:				
Female	389	50.0%	853	48.5%
Male	389	50.0%	904	51.5%
Race/Ethnicity:				
American Indian/Alaska Native	2	0.3%	6	0.3%
Asian	69	8.9%	174	9.9%
Black/African American	10	1.3%	25	1.4%
Hispanic	146	18.8%	299	17.0%
Native Hawaiian/Other Pacific Islander	1	0.1%	4	0.2%
White	515	66.2%	1,177	67.0%
Multiracial/Two or More Races	35	4.5%	72	4.1%
Status:				
Economically disadvantaged students	240	30.8%	530	30.2%
English language learners	125	16.1%	226	12.9%
Special education students	118	15.2%	230	13.1%
Grade levels:				
Grade 3	240	30.8%	465	26.5%
Grade 4	227	29.2%	522	29.7%
Grade 5	116	14.9%	370	21.1%
Grade 6	195	25.1%	400	22.8%

Appendix B. Construct Validity Results

Table B1. Unstandardized and standardized factor loadings for the CFA models

Observed variable: Strands	Latent variable: Academic achievement	*b*	β	*SE*	*p*
Numbers & Operations	Math	140.70	.92	4.20	< .001
Algebra & Algebraic Thinking	Math	168.58	.94	4.88	< .001
Fractions	Math	122.50	.91	3.69	< .001
Geometry	Math	132.08	.89	4.19	< .001
Measurement	Math	107.66	.88	3.47	< .001
Data, Statistics, & Probability	Math	145.24	.88	4.65	< .001
Reading Strategies	ELA	215.08	.93	4.19	< .001
Vocabulary	ELA	219.79	.93	4.26	< .001
Writing Strategies	ELA	238.14	.93	4.60	< .001
Grammar & Mechanics	ELA	185.75	.90	3.81	< .001

Appendix C. Predictive Validity Results

Table C1. Correlation matrix and descriptive statistics for mathematics

Correlation Coefficient (*r*)		1	2	3	4	5	6	7	8
1	ILEARN Math	-							
2	IXL Diagnostic Math	**.895**	-						
3	Strand 1. Numbers & Operations	**.838**	.948	-					
4	Strand 2. Algebra & Algebraic Thinking	**.861**	.947	.871	-				
5	Strand 3. Fractions	**.820**	.926	.838	.849	-			
6	Strand 4. Geometry	**.805**	.905	.820	.822	.819	-		
7	Strand 5. Measurement	**.811**	.880	.799	.816	.812	.775	-	
8	Strand 6. Data, Statistics, & Probability	**.838**	.885	.796	.831	.802	.782	.792	-
Descriptives									
M		6489.13	454.52	479.16	414.11	470.73	474.49	390.45	462.97
SD		101.12	145.87	153.12	180.32	134.15	149.03	122.58	164.80
Skewness		0.43	0.77	0.51	0.64	1.01	0.50	0.93	0.68
Kurtosis		0.24	0.56	0.47	0.14	0.41	1.38	0.73	-0.18

Note: all correlations were significant at α = .001 level; *n* = 778

Table C2. Chi-square test and logistic regression for mathematics

		ILEARN Math		χ^2	Odds ratio
		Below proficient	At/above proficient		
IXL Real-Time Diagnostic (Math)	Below grade level	243 (84.4%)	45 (15.6%)	243.179 ***	11.112 ***
	At/above grade level	130 (26.5%)	360 (73.5%)		

Note: ***p < .001; n = 778

Table C3. Correlation matrix and descriptive statistics for ELA

Correlation Coefficient (r)		1	2	3	4	5	6	7
1	ILEARN ELA							
2	IXL Diagnostic ELA	**.864**						
3	IXL Diagnostic Reading	**.828**	.967					
4	Strand 1. Reading Strategies	**.803**	.947	.983				
5	Strand 2. Vocabulary	**.816**	.950	.914	.864			
6	Strand 3. Writing Strategies	**.824**	.951	.896	.879	.867	-	
7	Strand 4. Grammar & Mechanics	**.824**	.929	.856	.830	.858	.839	-
Descriptives								
M		5512.27	523.86	535.67	525.82	573.09	531.99	467.54
SD		84.99	220.78	229.47	231.41	235.49	254.81	205.70
Skewness		-0.02	0.75	0.99	1.10	0.74	0.74	0.37
Kurtosis		-0.22	0.59	1.21	1.59	0.67	0.13	-0.07

Note: all correlations were significant at α = .001 level; n = 1,757

Table C4. Chi-square test and logistic regression for ELA

		ILEARN ELA		χ^2	Odds ratio
		Below proficient	At/above proficient		
IXL Real-Time Diagnostic (ELA)	Below grade level	488 (87.8%)	68 (12.2%)	654.771 ***	17.827 ***
	At/above grade level	273 (22.7%)	928 (77.3%)		

Note: ***p < .001; n = 1,757