RESEARCH REPORT

September 2021

# Demonstrating Grade-Level Predictive Validity of the IXL Real-Time Diagnostic Using the Virginia SOL as Criterion

Christina Schonberg, Ph.D.

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

*Peer Review: This study was peer-reviewed and presented at the 2022 annual meeting of the National Council on Measurement in Education in San Diego, CA.*

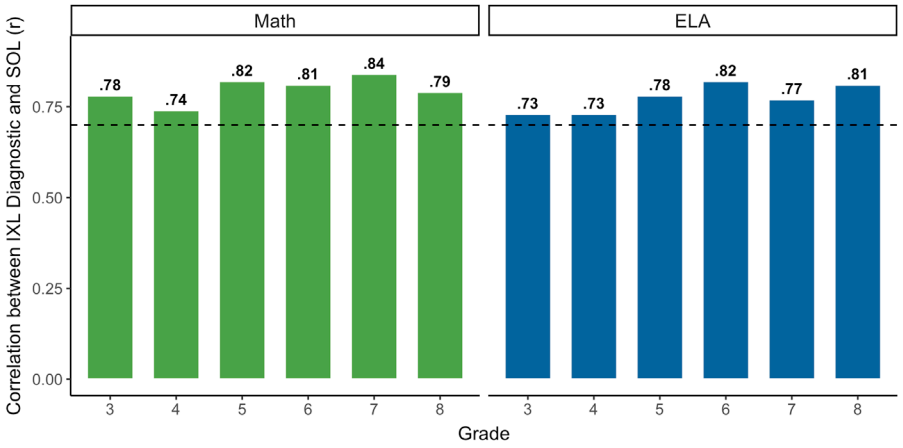
## Executive Summary

IXL is a personalized learning platform for students in grades PreK-12 and covers content in core subjects including mathematics and English language arts (ELA). A core component of IXL 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. Upon completing the diagnostic, students receive scores that correspond to their estimated grade-level proficiency as well as targeted recommendations for what skills to practice next to meet their learning goals.

IXL's Real-Time Diagnostic has been validated with other standardized assessments including the NWEA MAP Growth and ILEARN assessments (An, 2021; IXL Learning, 2020a, 2020b). The goals of this study were to assess the validity of IXL's Real-Time Diagnostic in both math and ELA using a new criterion measure—the Virginia state Standards of Learning, or SOL— at the grade level, and to examine the overall test-retest reliability of IXL's Diagnostic. We analyzed math and ELA data from students in grades 3-8 attending public schools in one Virginia district. Within each subject, we found:

- Strong, positive correlations between IXL's Diagnostic and SOL scores (all *r*s > .70)
- A high degree of overlap in student proficiency classifications by the IXL Diagnostic and SOL
- High test-retest reliability of IXL's Diagnostic, based on scores from the beginning and the end of the school year (*r*s > .90)

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



# Demonstrating Grade-Level Predictive Validity of the IXL Real-Time Diagnostic Using the Virginia SOL as Criterion

## Background

IXL is a powerful, flexible educational technology platform that provides personalized learning in subjects including mathematics and English language arts (ELA) for students in grades PreK-12; it is currently used by 1 in 5 students in the US and over 12 million students worldwide. A core component of IXL 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 the Common Core and other academic standards. The diagnostic allows students and educators to quickly identify current knowledge levels in key strands of math and ELA. Using Item Response Theory (Lord, 1980), the diagnostic analyzes student response patterns and provides personalized action plans that lay out clear next steps for students to reach their learning goals.

IXL's Diagnostic is especially valuable because it provides in-the-moment information about grade-level proficiency to students and educators. It is important, however, that the information provided by the diagnostic maps onto real-world outcomes such as performance on standardized assessments. As such, demonstrating the validity of 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, 2020a). Similarly, the IXL Diagnostic was shown to have high predictive validity with the Indiana Learning Evaluation Assessment Readiness Network (ILEARN) assessment (IXL Learning, 2020b). The primary goal of this study was to validate IXL's Diagnostic with a new external criterion measure: the Virginia state Standards of Learning (SOL).

Previous validity studies using other assessments have found strong, positive correlations between students' performance on IXL's Real-Time Diagnostic and subsequent standardized assessments (An, 2021; IXL Learning, 2020a, 2020b). These studies focused on overall predictive validity and construct validity (e.g., correlations by subject strand) but no research has yet examined the relationship between IXL's Diagnostic and standardized assessments by grade. This relationship is important to study as it may change across grades. Therefore, in addition to expanding the set of measures with which IXL's Diagnostic has been validated, this study also provides more precise validity evidence by examining correlations between IXL's Diagnostic and SOL assessments by grade.

In addition, many students in the district we studied had pinpointed diagnostic scores from several points throughout the school year. These data afforded us the opportunity to examine another important psychometric property of the diagnostic: test-retest reliability, or the correlation between students' scores at two different time points.

**RESEARCH QUESTIONS**

This study aimed to answer the following research questions for math and ELA:

1. How strong are the correlations between scores on the IXL Real-Time Diagnostic assessment and the Virginia SOL assessment in each grade from 3-8?
2. What is the degree of overlap in student grade-level proficiency classifications by IXL's Real-Time Diagnostic and the Virginia SOL? That is, to what extent do IXL's Diagnostic and the Virginia SOL identify the same students as being at or above (vs. below) grade level?
3. What is the test-retest reliability of IXL's Real-Time Diagnostic?

## Study Design and Methodology

**PARTICIPANTS**

To be included in the math or ELA validity analyses, students needed to have both a pinpointed IXL Real-Time Diagnostic score in math or ELA and an SOL assessment score in the corresponding subject. Furthermore, we specifically focused on students in grades 3-8. Using these criteria, we included 2,095 students in the math analyses and 2,058 students in the ELA analyses. Descriptive statistics for student demographic characteristics and performance on IXL's Diagnostic as well as the SOL assessment can be found in Appendix A.

To be included in the math or ELA test-retest reliability analyses, students needed to have two pinpointed IXL Real-Time Diagnostic scores in the subject of interest—one from the end of the school year and one from the beginning of the school year (at least 180 days before the end-of-year diagnostic score). Applying this additional criterion resulted in a sample size of 1,505 students in the math analysis and 1,395 students in the ELA analysis.

**DATA SOURCES**

IXL Real-Time Diagnostic data were obtained from IXL's internal 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. We obtained all of the available pinpointed diagnostic scores in math and ELA for students in this Virginia district across the 2020-2021 school year. Then, for the validity analyses, we selected each student's pinpointed end-of-year diagnostic score from the date closest to their SOL assessment date in each subject. For the test-retest reliability analyses, we also selected each student's earliest pinpointed diagnostic score within the school year, which had to be at least 180 days earlier than their end-of-year diagnostic.

To validate the IXL Real-Time Diagnostic with the Virginia SOL assessments, we obtained Spring 2021 student-level assessment data from 14 schools in one Virginia school district. The SOL is Virginia's statewide standardized assessment, and the mathematics and ELA SOL assessments are administered to students in grades 3-8. These assessments are computer adaptive, meaning that they are customized for each student through the student's responses to test items. Based on their performance, students are classified into one of four achievement levels in each subject: pass/advanced, pass/proficient, fail/basic, and fail/below basic. Additional information about these assessments can be found at https://www.doe.virginia.gov/testing/index.shtml.

## ANALYTIC APPROACH

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

We analyzed data from 14 elementary and middle schools in one Virginia school district where students completed IXL's Real-Time Diagnostic in math or ELA and SOL assessments in math or ELA. IXL's Diagnostic uses a continuous vertical scale across grades PreK-12 ranging from 0 to 1300, indicating grade-level proficiency. By contrast, SOL scores range from 0 to 600 within each grade, with 400 being the proficiency cutoff.

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 SOL scores would indicate that the IXL Real-Time Diagnostic is a valid measure of grade-level proficiency for students in grades 3-8 in the state of Virginia.

### *Research Question 2: Alignment of Student Proficiency Classifications Across Measures*

In addition to correlations between IXL Real-Time Diagnostic and SOL 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 grade level (i.e., passing) on the SOL (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 passing the SOL.

### *Research Question 3: Test-Retest Reliability*

To examine the test-retest reliability of the IXL Real-Time Diagnostic, we correlated students' pinpointed diagnostic scores from the beginning of the school year with their pinpointed diagnostic scores from the end of the school year. Given the adaptive nature of the IXL Real-Time Diagnostic, high correlations between two sets of diagnostic scores over 180 days apart would be considered not only coefficients of stability but also coefficients of equivalence. In this context, coefficients greater than .9 would indicate reasonably high test-retest and alternate form reliability (Bandalos, 2018, Chapter 8). Because this analysis correlates the same measure with itself and is not expected to differ by grade, we report test-retest reliability overall for math and ELA, respectively.

## Results

**MATH**

We found strong, statistically significant correlations between IXL Real-Time Diagnostic and SOL scores in each grade (smallest $r$ = .74, all $p$s < .001). See Figure 1 for scatterplots of these correlations by grade.

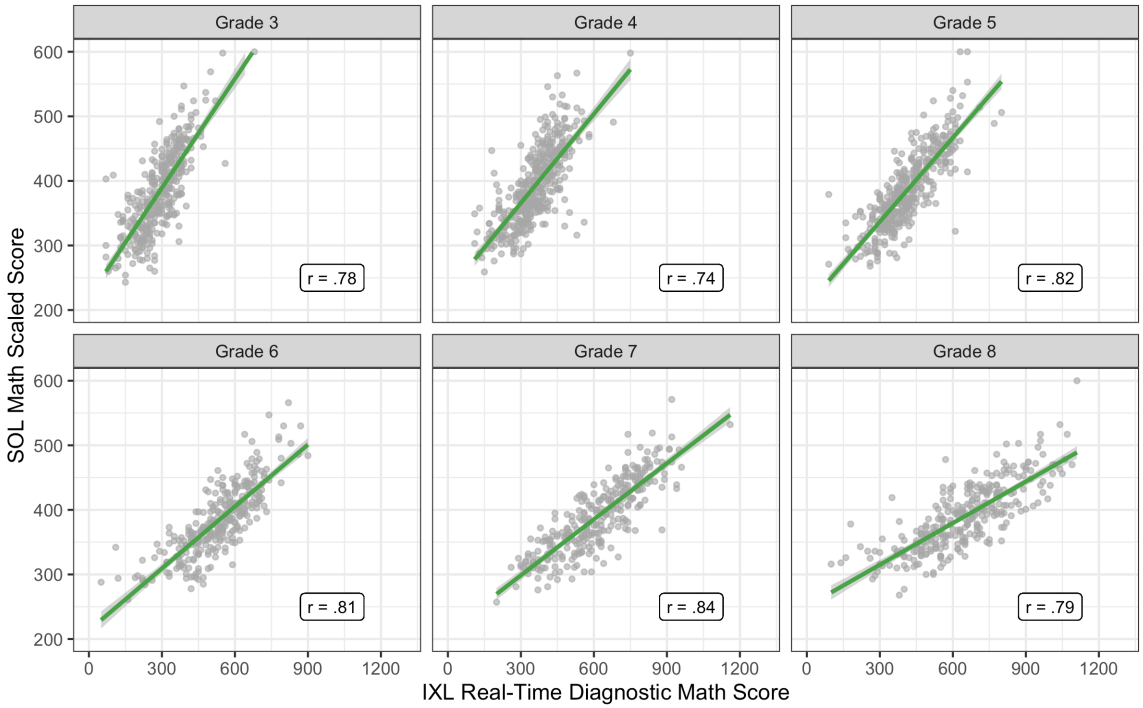



*Figure 1. Correlations between IXL Real-Time Diagnostic and SOL scores in math. All correlations are statistically significant at p < .001.*

Next, we examined the alignment of student proficiency classifications across the two measures. We found that IXL's Real-Time Diagnostic identified the majority of students performing at or above grade level (81%) as well as those below grade level (80%) based on SOL Math performance, $\chi^2(1) = 714.83$, $p < .001$ (Figure 2). In line with this finding, a logistic regression model predicting SOL proficiency status from IXL proficiency status showed that students who were classified as proficient by the IXL Real-Time Diagnostic were 17.09 times more likely to be proficient on the SOL than students who were not proficient based on IXL's Diagnostic (see Table B1).

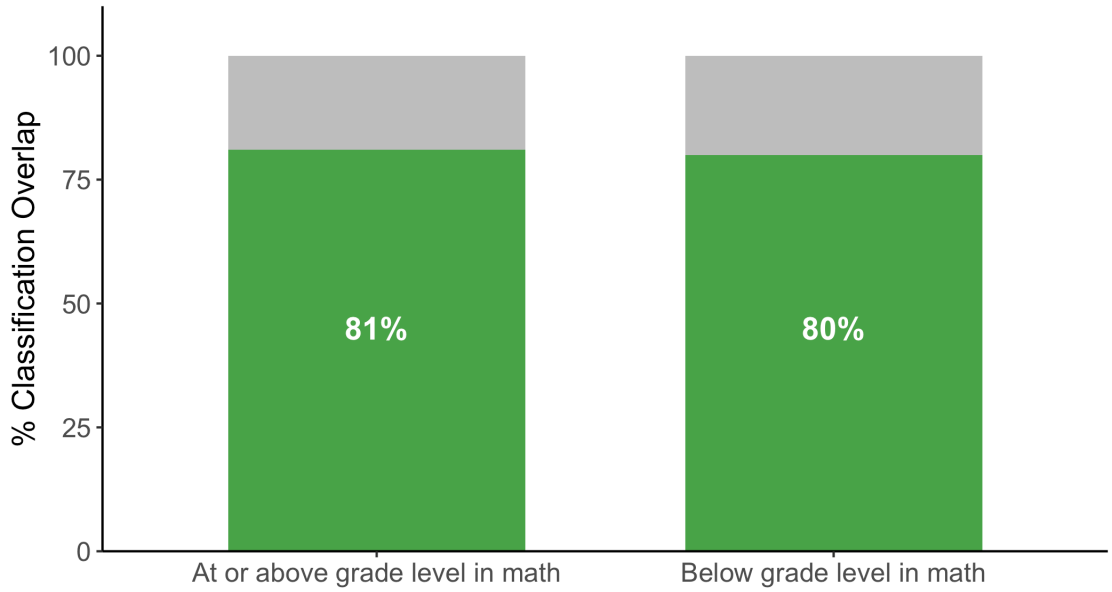



*Figure 2. IXL Real-Time Diagnostic Math classification alignment using the SOL Math assessment as criterion.*

Finally, we examined the test-retest reliability of IXL's Real-Time Diagnostic in math by correlating students' pinpointed diagnostic scores in math from the beginning of the school year with their pinpointed diagnostic scores in math from the end of the school year. The IXL Real-Time Diagnostic Math assessment exhibited a high coefficient of stability and equivalence between testing sessions, $r = .93$, $p < .001$.

**ELA**

As in math, we found strong, statistically significant correlations between IXL Real-Time Diagnostic and SOL scores in each grade for ELA (smallest $r = .73$, all $p$s $< .001$). See Figure 3 for scatterplots of these correlations by grade.

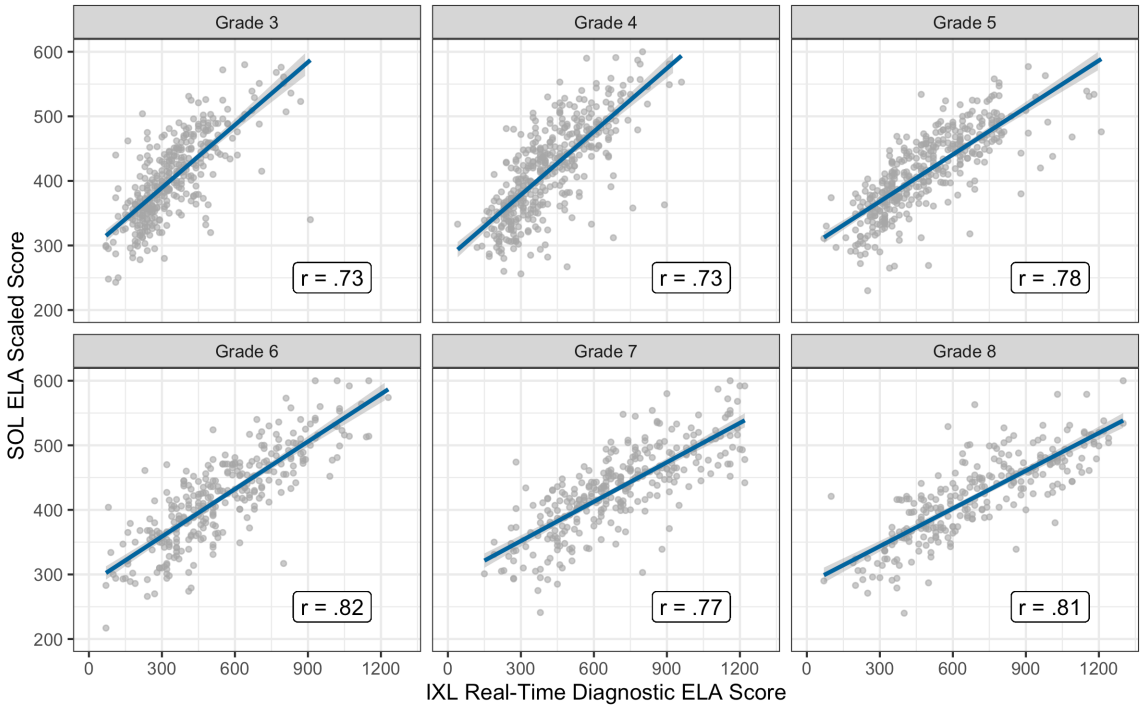



*Figure 3. Correlations between IXL Real-Time Diagnostic and SOL scores in ELA. All correlations are statistically significant at $p < .001$.*

Next, we examined the alignment of student proficiency classifications across measures. As in math, the IXL Real-Time Diagnostic in ELA identified the majority of students performing at or above grade level on the SOL ELA (89%) as well as those below grade level (65%), $\chi^2(1) = 629.13$, $p < .001$ (Figure 4). In addition, a logistic regression model predicting SOL proficiency status from IXL Real-Time Diagnostic proficiency status found that students classified as proficient by IXL's Diagnostic were 15.33 times more likely to be proficient on the SOL, compared to students who were not classified as proficient by the IXL Diagnostic (see Table B2).

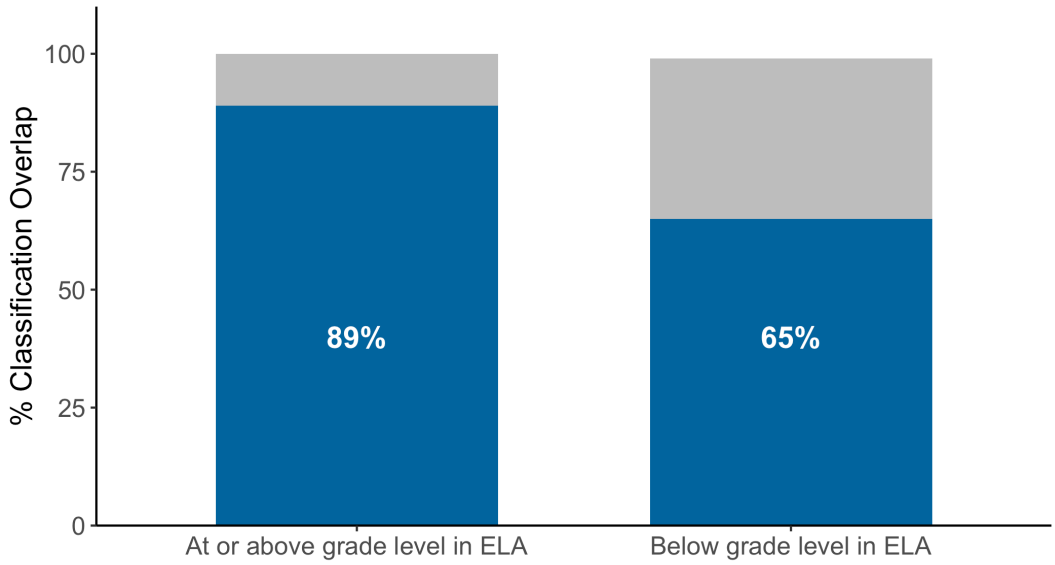



*Figure 4. IXL Real-Time Diagnostic ELA classification alignment using the SOL ELA assessment as criterion.*

Finally, we examined the test-retest reliability of IXL's Real-Time Diagnostic for ELA by correlating students' pinpointed diagnostic scores in ELA from the beginning of the school year with their pinpointed diagnostic scores in ELA from the end of the school year. As with math, the IXL Real-Time Diagnostic ELA assessment exhibited a high coefficient of stability and equivalence between testing sessions, $r = .95$, $p < .001$.

## Conclusion

In this study, we investigated the predictive validity of the IXL Real-Time Diagnostic with a novel set of state assessments, the Virginia SOL. We examined test-criterion relationships between the two assessments by grade, providing empirical evidence in each elementary and middle school grade level from grades 3 to 8. We found strong correlations between the IXL Real-Time Diagnostic and the Virginia SOL in each grade as well as a high degree of alignment in the two measures' classifications of student proficiency.

In addition, we found strong evidence of test-retest reliability for the IXL Real-Time Diagnostic, as measured by correlations between students' diagnostic scores across timepoints. One particular advantage of IXL's Diagnostic is that students answer different questions every time. This avoids the possibility of practice effects, which occur when students perform similarly or better on subsequent administrations of the same test because they have learned some of the answers for repeated items. Investigations of test-retest reliability are often critiqued due to the possibility of practice effects, because they may artificially inflate the correlation coefficients. Because students never answered the exact same questions across administrations of the diagnostic, we can be certain that the high test-retest reliability coefficients in this study indicate high degrees of measurement stability across time and equivalence across alternate forms of the assessment.

In sum, this study provides new evidence for the test-retest reliability of IXL's Real-Time Diagnostic, as well as the first evidence at the grade level that the diagnostic is a valid measure of students' proficiency in math and ELA. Coupled with prior studies of construct validity, internal consistency, and predictive validity (e.g., IXL Learning, 2020a, 2020b), our findings corroborate a strong program of reliability and validity for the IXL Real-Time Diagnostic.

## References

An, X. (2021). *Assessing the predictive validity of the IXL Real-Time Diagnostic among marginalized students*. 1-10.

Bandalos, D. (2018). Methods of Assessing Reliability. In *Measurement Theory and Applications for the Social Sciences* (pp. 172–209). Guilford Press.

IXL Learning. (2020a). *Validation study of the IXL Real-Time Diagnostic using MAP Growth Assessments*. 1-10. https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study.pdf

IXL Learning. (2020b). *Assessing the construct validity and the predictive validity of the IXL Real-Time Diagnostic*. 1-14. https://www.ixl.com/research/IXL-Real-Time-Diagnostic-Validation-Study-ILEARN.pdf

Lord, F. (1980). *Applications of Item Response Theory to practical testing problems*. Mahwah, NJ: Lawrence 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.

# Appendix A: Descriptive Statistics of Student Demographics and Performance

**Table A1.** Student demographic characteristics

| | Students with IXL Diagnostic and SOL math scores | | Students with IXL Diagnostic and SOL ELA scores | |
|---|---|---|---|---|
| # of students (total) | 2095 | | 2058 | |
| | N | % | N | % |
| **Gender** | | | | |
| Female | 1,075 | 51.31 | 1,049 | 50.97 |
| Male | 1,020 | 48.69 | 1,009 | 49.03 |
| **Ethnicity** | | | | |
| Hispanic | 116 | 5.54 | 117 | 5.69 |
| Non-Hispanic | 1,979 | 94.46 | 1,941 | 94.31 |
| **Race** | | | | |
| Asian | 41 | 1.96 | 43 | 2.09 |
| Black | 982 | 46.87 | 955 | 46.40 |
| Other | 228 | 10.88 | 220 | 10.69 |
| White | 844 | 40.29 | 840 | 40.82 |
| **Economically disadvantaged students** | 1,148 | 54.80 | 1,127 | 54.76 |
| **English language learners** | 35 | 1.67 | 36 | 1.75 |
| **Students with disabilities** | 351 | 16.75 | 356 | 17.30 |
| **Grade level** | | | | |
| Grade 3 | 366 | 17.47 | 364 | 17.69 |
| Grade 4 | 416 | 19.86 | 420 | 20.41 |
| Grade 5 | 399 | 19.04 | 398 | 19.34 |
| Grade 6 | 309 | 14.75 | 306 | 14.87 |
| Grade 7 | 300 | 14.32 | 309 | 15.01 |
| Grade 8 | 305 | 14.56 | 261 | 12.68 |

**Table A2.** Means (standard deviations) for math assessments

| | IXL Real-Time Diagnostic math assessment | SOL math assessment |
|---|---|---|
| Grade 3 | 291.01 (86.22) | 384.21 (62.41) |
| Grade 4 | 354.71 (92.47) | 390.94 (57.41) |
| Grade 5 | 414.74 (108.55) | 386.58 (57.34) |
| Grade 6 | 532.59 (133.37) | 383.64 (52.83) |
| Grade 7 | 616.87 (160.95) | 390.24 (55.21) |
| Grade 8 | 644.49 (185.08) | 389.03 (50.60) |

*Note. N* = 2,095. IXL Real-Time Diagnostic scores represent scores from the end of the school year.

Table A3. Means (standard deviations) for ELA assessments

| | IXL Real-Time Diagnostic ELA assessment | SOL ELA assessment |
|---|---|---|
| Grade 3 | 341.76 (142.43) | 403.03 (62.97) |
| Grade 4 | 415.88 (157.62) | 416.25 (70.07) |
| Grade 5 | 489.02 (197.87) | 413.95 (61.73) |
| Grade 6 | 548.37 (238.55) | 419.14 (72.19) |
| Grade 7 | 691.65 (247.07) | 431.40 (64.91) |
| Grade 8 | 692.72 (270.84) | 420.21 (65.16) |

*Note.* $N$ = 2,058. IXL Real-Time Diagnostic scores represent scores 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

| | | SOL Math | | $\chi^2$ | Odds ratio |
|---|---|---|---|---|---|
| | | Below proficient | At or above proficient | | |
| IXL Real-Time Diagnostic Math | Below grade level | 1,126 (80%) | 284 (20%) | 714.83*** | 17.09*** |
| | At or above grade level | 129 (19%) | 556 (81%) | | |

*Note.* $N$ = 2,095.

*** $p < .001$.

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

| | | SOL ELA | | $\chi^2$ | Odds ratio |
|---|---|---|---|---|---|
| | | Below proficient | At or above proficient | | |
| IXL Real-Time Diagnostic ELA | Below grade level | 722 (65%) | 384 (35%) | 629.13*** | 15.33*** |
| | At or above grade level | 104 (11%) | 848 (89%) | | |

*Note.* $N$ = 2,058.

*** $p < .001$.