

Predictive Validity and Forecasting Accuracy for the 2018–19 School Year: AzMERIT

by Christine Burnham, Ph.D. and Amanda Perino

OVERVIEW

Each year, Imagine Learning investigates Imagine Galileo K–12 assessments' predictive validity and the forecasting accuracy of risk levels once districts and charters upload their statewide assessment data for individual students into the Imagine Galileo database. Imagine Learning staff evaluates predictive validity by examining the correlation between student scores on each district/charter-wide assessment and student scores on the statewide assessment. Imagine Galileo then evaluates forecasting accuracy by examining how students classified at different risk levels ultimately performed on the statewide assessments.

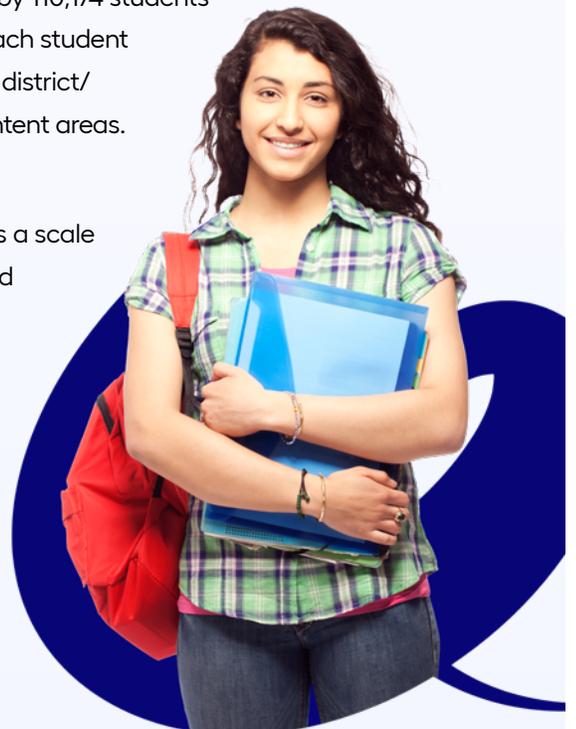
This document provides a comprehensive summary of the research on the predictive validity of Imagine Galileo assessments administered in Arizona during the 2018–19 school year and the forecasting accuracy of Imagine Galileo risk levels based on student performance on these assessments, with regard to their eventual performance on the AzMERIT assessment. For districts and charters who uploaded their 2019 AzMERIT assessment data, the results of the specific investigations for their administered assessments are available in Imagine Galileo's Forecast Report.

SAMPLE

This sample includes the first 12 Arizona districts and charters to provide Imagine Galileo with their 2019 AzMERIT data for individual students in grades 3 through high school. Subjects included are math and reading/English language arts. These districts/charters administered 1,038 district/charter-wide assessments in these grades and content areas, with the total sample including 580,388 scores recorded by 110,174 students from the 12 Arizona school districts and charter schools. On average, each student contributed 5.3 scores to the analyses, consistent with the average of 3 district/charter-wide assessments administered per year in each of the two content areas.

STUDENT PERFORMANCE MEASURES

The statewide assessment data uploaded by districts/charters contains a scale score for each student and an indication of whether the student passed the statewide assessment. For each district/charter-wide assessment administered, Imagine Galileo performs an Item Response Theory (IRT) analysis, which produces a scale score for each student, referred to as the Developmental Level (DL) score. Each student is further classified as to their level of risk of failing the statewide assessment based on their performance on all the district/charter-wide assessments they have taken within a given school year. The possible risk levels comprise "High Risk," "Moderate Risk," "Low Risk," and "On Course" in order of highest to lowest risk of failing the statewide assessment.



IMAGINE GALILEO ASSESSMENTS AND TIMEFRAMES

The 1,038 district/charter-wide assessments were not identical across all district/charters, nor were there constraints regarding the time of year during which the assessments were administered to students. Many of the assessments were standard Imagine Galileo Comprehensive Benchmark Assessment Series (CBAS) assessments, whereas others were district/charter curriculum-aligned assessments. Student DL scores on Imagine Galileo assessments within the same grade level and subject are on a common scale and are comparable, even if there are differences in the assessments' content. Therefore, the predictive validity and forecast accuracy analyses were conducted on the pool of student Imagine Galileo scores for each grade level and subject, regardless of the specific assessment.

However, it would be inappropriate to include student scores from the beginning, middle, and end of the year in a single correlation analysis; student DL scores are expected to increase as the year progresses, and this increase would obscure any correlation between student DL scores and end-of-year AzMERIT scores. Therefore, six separate predictive validity correlation analyses were conducted for each grade level and subject: one analysis for each of the six timeframes. Table 1 lists the date ranges for each student score timeframe.

Table 1. Date ranges for the six timeframes used in the correlation study

Timeframe	Start Date	End Date
T1	July 1	September 15
T2	September 16	November 30
T3	December 1	January 15
T4	January 16	February 15
T5	February 16	April 15
T6	April 16	June 30

PREDICTIVE VALIDITY ANALYSES

Predictive validity analyses examine the strength of the relationship between two measures of student performance: the student DL scores on an assessment in a given grade and content area, and the student scores on the statewide assessment in the same grade and content area. Predictive validity analyses can produce correlation statistics that range from -1 to +1, although typically only positive values are observed in this context. A positive correlation indicates a positive relationship; high scores on one measure are associated with high scores on the other measure. A negative correlation would indicate a negative relationship; high scores on one measure are associated with low scores on the other measure. A correlation of zero would indicate no relationship. Values of +1 or -1 indicate a perfect relationship between the two measures and are rare.

The predictive validity analysis for each grade level, subject, and timeframe was performed on the pooled set of student scores for all district/charter-wide assessments administered by the 12 districts/charters in the relevant grades and subjects during the 2018–19 school year.

PREDICTIVE VALIDITY RESULTS

Table 2 illustrates the correlation observed for the student Imagine Galileo DL scores on the assessments administered in each grade and content area during each timeframe. The chart shows correlations ranging, on average, from 0.66 for timeframe 1 to 0.71 for timeframe 6, with an overall mean of 0.76. The data demonstrate a tendency for correlation values to increase as the year progresses, reflecting that assessments taken early in the year can capture student ability levels before students acquire knowledge and skills in the content area.

This is especially true of high school math: A correlation between 0.7 and 0.9 indicates a high correlation between the two measures, while a correlation between 0.5 and 0.7 indicates a moderate correlation. Thus, the observed correlations suggest that student scores on the 2018–19 Imagine Galileo assessments were generally strongly related to student scores on the 2019 statewide assessment.

Table 2. Mean correlations between scores for the 2018–19 Galileo assessments and scores for the 2019 statewide assessment for each grade level, content area, and timeframe.

Subject	Grade	T1	T2	T3	T4	T5	T6	Average
ELA	3	0.67	0.75	0.80	0.70	0.79	0.81	0.75
ELA	4	0.80	0.83	0.82	0.84	0.80	0.81	0.82
ELA	5	0.80	0.82	0.84	0.85	0.71	0.84	0.83
ELA	6	0.72	0.71	0.81	0.70	0.80	0.83	0.76
ELA	7	0.71	0.68	0.81	0.71	0.80	0.86	0.74
ELA	8	0.75	0.77	0.80	0.79	0.81	0.83	0.79
ELA	9	0.71	0.74	0.78	0.77	0.83	0.77	0.77
ELA	10	0.73	0.76	0.76	0.78	0.68	0.76	0.74
ELA	11	0.60	0.68	0.69	0.64	0.80	0.63	0.67
Math	3	0.52	0.69	0.77	0.64	0.78	0.80	0.70
Math	4	0.73	0.80	0.81	0.86	0.85	0.82	0.71
Math	5	0.71	0.81	0.82	0.84	0.72	0.87	0.71
Math	6	0.62	0.72	0.84	0.66	0.83	0.88	0.76
Math	7	0.63	0.59	0.81	0.83	0.80	0.80	0.74
Math	8	0.65	0.78	0.80	0.82	0.83	0.80	0.78
Algebra I	8	0.54	0.72	NA	0.73	0.84	NA	0.71
Algebra I	HS	0.53	0.69	0.68	0.75	0.70	0.73	0.68
Algebra II	HS	0.60	0.78	0.77	0.85	0.76	0.75	0.75
Geometry	HS	0.62	0.76	0.69	0.80	0.81	0.81	0.75
Average of Correlations		0.66	0.74	0.78	0.77	0.80	0.79	0.76

FORECASTING ACCURACY ANALYSES

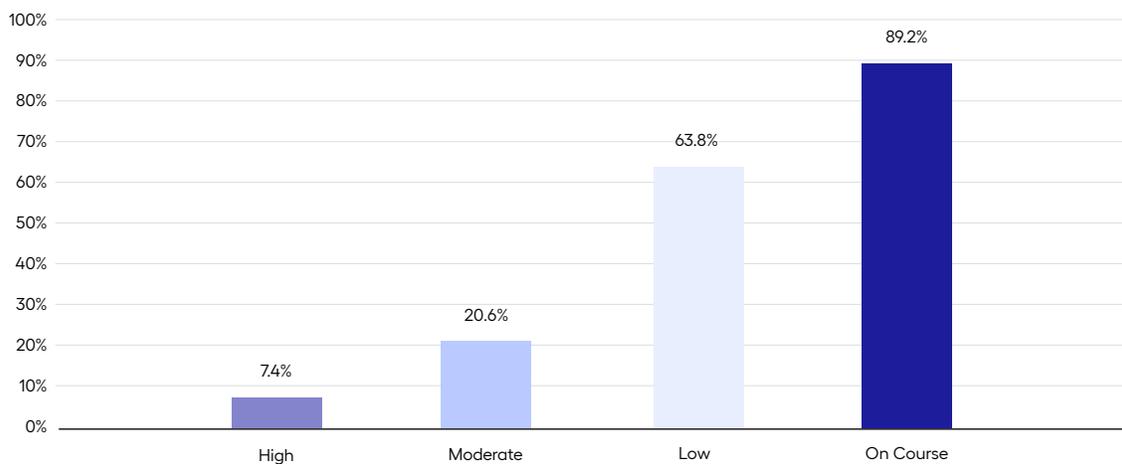
Forecasting accuracy analyses examine the accuracy with which Imagine Galileo risk levels for individual students predicted their ultimate performance on the relevant statewide assessment. Risk levels indicate the likelihood that a student is at risk of failing the statewide assessment. Although risk levels represent a continuum, students classified as “On Course” or “Low Risk” are predicted to pass the statewide assessment, while students classified as “Moderate Risk” or “High Risk” are predicted to fail it. Forecasting accuracy analyses were conducted for the group of 12 districts/charters described previously. The pooled student score data set from this group included risk level classifications and the corresponding AzMERIT classification for 110,174 student/content area records (most students contributed two records: one for math and one for ELA).

FORECASTING ACCURACY RESULTS

Figure 1 illustrates the percentage of students in each risk level who passed the statewide assessment. Figure 2 illustrates the overall forecasting accuracy, as well as the forecasting accuracy for each risk level. There are three critical aspects of the forecasting accuracy analysis to evaluate. First, as student risk level decreases, the likelihood of success on the statewide assessment should increase. This aspect is a prerequisite for accurate forecasting. As Figure 1 shows, most students classified as “On Course” based on their performance on the Imagine Galileo district/charter-wide assessments did pass the statewide assessment. Similarly, the majority of those classified as “High Risk” for not demonstrating mastery on the statewide assessment failed. The other two risk level groups also performed as expected.

Figure 1. Mean percentage of students passing the statewide assessment for each risk level.

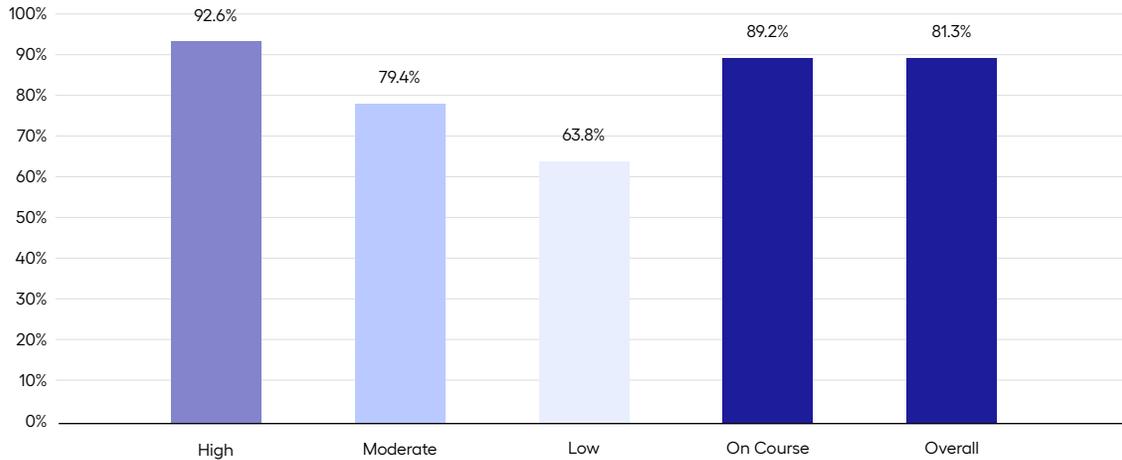
Percent of Students in Each 2018–19 Galileo Risk Group who passed Spring 2019 AzMERIT



Second, overall forecasting accuracy should be adequately high. Imagine Galileo considers forecasting accuracy to be adequate if a student’s risk level accurately predicted performance on the statewide assessment for at least 75 percent of students within a district/charter. As Figure 2 shows, the overall forecasting accuracy was relatively high, with statewide test performance accurately forecast, on average, for 81% percent of students.

Figure 2. Mean percentage of students passing the statewide assessment for each risk level..

Percent of Students for whom Spring 2019 AzMERIT Performance was Accurately Forecast



Third, forecasting accuracy should be highest in cases where student performance is most consistent. Students who consistently perform well on Imagine Galileo assessments and are thus classified as “On Course” should consistently pass the statewide assessment. Conversely, students who consistently perform poorly on Imagine Galileo assessments and are classified as “High Risk” should consistently fail to pass the statewide assessment. Students whose performance on Imagine Galileo assessments is more variable (i.e., the “bubble” students who sometimes perform well and sometimes don’t) should also display more variable performance on the statewide assessment. As Figure 2 shows, and as expected, forecasting accuracy was highest for students classified as “On Course” and “High Risk” and somewhat lower for students classified as “Low Risk” and “Moderate Risk.” If teachers and administrators are using the data provided by Imagine Galileo district/charter-wide assessments to implement effective interventions, many students classified as being at some risk of failing the statewide assessment should pass it instead, thereby reducing the accuracy of risk assessment forecasts for those groups. Imagine Galileo, therefore, considers a certain degree of inaccuracy in predictions of failure to be a sign of success.

CONCLUSION

The research presented in this document was conducted to evaluate predictive validity and forecasting accuracy for the 2018–19 school year. The results suggest that the 2018–19 Imagine Galileo assessments demonstrated adequate levels of predictive validity. The results also indicate that the 2018–19 Imagine Galileo risk levels displayed sufficient levels of accuracy in forecasting student performance on the statewide assessment. This research is consistent with similar research investigations performed in previous years, suggesting that Imagine Galileo assessments and risk levels continue to demonstrate adequate levels of predictive validity and forecasting accuracy.