Overview: ATI investigates the predictive validity of Galileo assessments and the forecasting accuracy of Galileo risk levels on an annual basis once districts'/schools' statewide assessment data for individual students has been uploaded into the Galileo database. ATI evaluates predictive validity by examining the correlation between student scores on each district-wide assessment and student scores on the statewide assessment. ATI evaluates forecasting accuracy by examining how students classified at different levels of risk ultimately performed on the statewide assessment. This document provides a comprehensive summary of the research on both the predictive validity of Galileo assessments administered in the 2011-12 school year and the forecasting accuracy of Galileo risk levels based on student performance on these assessments. For districts and schools who uploaded their 2012 statewide assessment data, the results of the specific investigations for their administered assessments are available in the online Galileo *Forecast Report*.

Sample: For the purpose of this brief, the first 26 districts to provide ATI with their 2012 statewide assessment data for individual students in grades three through high school in math, reading/English language arts, and science were included in the sample. The 26 districts included in the sample represent students in Arizona, Colorado, and Massachusetts and administered 1,105 district-wide assessments in these grades and content areas.

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

Predictive Validity Analyses: Predictive validity analyses examine the strength of the relationship between two measures of student performance, in this case 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, that is high scores on one measure are associated with high scores on the other measure. A negative correlation would indicate a negative relationship, that is high scores on one measure are associated with low scores on the other measure. A correlation of zero would indicate no relationship. Values of positive or negative one indicate a perfect relationship between the two measures and are rarely observed under real-world circumstances. Predictive validity analyses were performed for all of the district-wide assessments administered by the group of 26 districts in the relevant grades and subjects during the 2011-12 school year. Twenty-nine correlations were omitted from the summary of the results due to invalid statewide or district-wide assessment data.

Research Brief
Predictive Validity and Forecasting Accuracy
for the 2011-12 School Year
Assessment Technology Incorporated

1

Predictive Validity Results: Table 1 illustrates the mean correlation observed for the assessments administered in each grade and content area. As the chart shows, the mean correlations range from 0.69 to 0.78 across grades and content areas with an overall mean of 0.75. A correlation between 0.7 and 0.9 indicates a high correlation between the two measures. Thus, the observed correlations suggest that student scores on the 2011-12 Galileo district-wide assessments were strongly related to student scores on the 2012 statewide assessment.
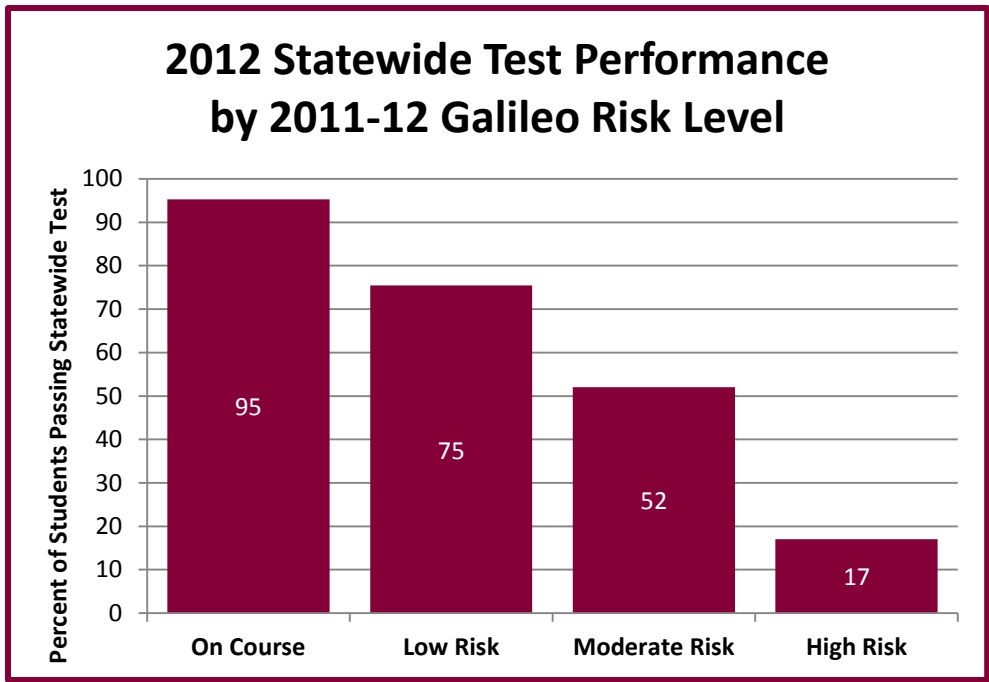
TABLE 1
Mean correlations for the 2011-12 Galileo assessments and the 2012 statewide assessment for each grade/content area.

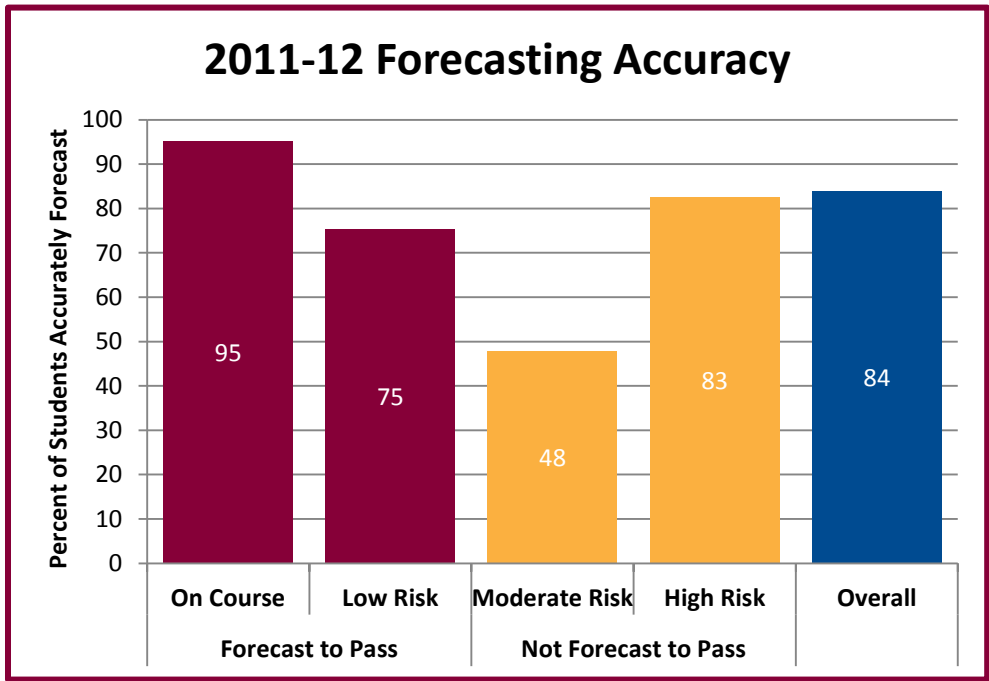| Correlation of 2011-12 Galileo Assessments with the 2012 Statewide Assessment					
		Content Area			
		Math	**Reading**	**Science**	**Overall**
Grade	3	0.77	0.76	N/A	**0.77**
	4	0.76	0.76	0.73	**0.76**
	5	0.78	0.76	N/A	**0.77**
	6	0.76	0.72	N/A	**0.74**
	7	0.73	0.71	N/A	**0.72**
	8	0.75	0.74	0.73	**0.74**
	10	0.69	0.71	0.76	**0.70**
	Overall	**0.76**	**0.74**	**0.73**	**0.75**
N/A= Data not available					

Forecasting Accuracy Analyses: Forecasting accuracy analyses examine the accuracy with which Galileo risk levels for individual students predicted their ultimate performance on the relevant statewide assessment. Risk levels provide an indication of the likelihood that a student is at risk to fail the statewide assessment. For the purpose of forecasting accuracy analyses, students who are classified as "On Course" or as "Low Risk" are predicted to pass the statewide assessment while students who are classified as "Moderate Risk" or "High Risk" are predicted to fail the statewide assessment. Forecasting accuracy analyses were conducted for the group of 26 districts described previously. Across these districts, there were 285 forecasting opportunities, where a forecasting opportunity is the student-level predictions made for the students within a given grade level in a specific district with regard to their performance on the statewide assessment in a given content area (i.e., one forecasting opportunity is for the third grade students in District A with regard to the statewide assessment in math). Forty-five forecasting opportunities were omitted from the summary of the results due to invalid statewide or district-wide assessment data.

Forecasting Accuracy Results: Figure 1 illustrates the mean percentage of students in each risk level who passed the statewide assessment. Figure 2 illustrates the mean overall forecasting accuracy as well as the mean forecasting accuracy for each risk level. There are three important aspects of the forecasting accuracy analysis to evaluate. First, as student risk level increases the likelihood of failure on the statewide assessment should increase. This is a prerequisite for accurate forecasting. As Figure 1 shows, the majority of students who were classified as being "On Course" based on their performance on the Galileo district-wide assessments did in fact pass the statewide assessment, while the majority of those who had been classified as being at "High Risk" of not demonstrating mastery on the statewide assessment did in fact fail. The other two risk level groups performed as expected as well.

Research Brief
Predictive Validity and Forecasting Accuracy
for the 2011-12 School Year
Assessment Technology Incorporated

2

1.800.367.4762
ati-online.com

2012 Statewide Test Performance by 2011-12 Galileo Risk Level

Percent of Students Passing Statewide Test

Risk Level	Percent
On Course	95
Low Risk	75
Moderate Risk	52
High Risk	17

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

Second, overall forecasting accuracy should be adequately high. ATI 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. As Figure 2 shows, the overall forecasting accuracy was quite high, with statewide test performance accurately forecast for 84 percent of students.

2011-12 Forecasting Accuracy

Percent of Students Accurately Forecast

	Forecast to Pass		Not Forecast to Pass		Overall
	On Course	Low Risk	Moderate Risk	High Risk	Overall
	95	75	48	83	84

Figure 2
Mean overall forecasting accuracy and mean forecasting accuracy for each risk level.

Research Brief
Predictive Validity and Forecasting Accuracy
for the 2011-12 School Year
Assessment Technology Incorporated

3

1.800.367.4762
ati-online.com

Third, forecasting accuracy should be highest in cases where student performance is most consistent. Students who consistently perform well on Galileo assessments and are thus classified as "On Course" should consistently pass the statewide assessment. Conversely, students who consistently perform poorly on Galileo assessments and are classified as "High Risk" should consistently fail to pass the statewide assessment. Students whose performance on 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, 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." It should be noted that, if teachers and administrators are using the data provided by Galileo district-wide assessments to implement effective interventions, many students who have been classified as being at some risk of failing the statewide assessment should pass it instead, thereby reducing the accuracy of risk assessment forecasts for the those student groups. ATI 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 2011-12 school year. The results suggest that the 2011-12 Galileo assessments demonstrated adequate levels of predictive validity. The results also suggest that the 2011-12 Galileo risk levels displayed adequate levels of accuracy in forecasting student performance on the statewide assessment. This research is consistent with similar research investigations performed in previous years and suggests that Galileo assessments and risk levels continue to demonstrate adequate levels of predictive validity and forecasting accuracy.