Comparison of Predictive Validity and Forecasting Accuracy for Assessments Aligned to Arizona's College and Career Ready Standards and Older Arizona Standards

by Sarah Callahan, Ph.D.
Assessment Technology Incorporated

Overview: Since the adoption of Arizona's College and Career Ready Standards (AZCCRS) in mathematics and English language arts by the Arizona State Board of Education in 2010, districts and charters have begun transitioning towards instruction and assessments aligned to the new standards. Assessment Technology, Incorporated (ATI) provides the Galileo® K-12 Online Instructional Improvement and Instructional Effectiveness System to 252 school districts and charter schools serving more than 500,000 students throughout the state of Arizona. On an annual basis, ATI investigates the predictive validity of Galileo assessments and the forecasting accuracy of Galileo risk levels with respect to the statewide assessment (e.g., Arizona's Instrument to Measures Standards [AIMS]). Currently, AIMS is designed to assess student mastery of older Arizona State Standards (older AZ standards) in mathematics (adopted 2008), English language arts (adopted 2003), and science (adopted 2005); however, in spring 2015, Arizona plans to administer statewide assessments in mathematics and English language arts aligned to the AZCCRS. During this transition period, one critical question is whether local assessments (e.g., benchmarks, pretests, posttests) administered by districts/charters to measure student progress towards mastery of AZCCRS can also provide accurate forecasts of student performance on the current AIMS assessments aligned to older AZ standards. This brief summarizes research evaluating the predictive validity and forecasting accuracy for Galileo assessments administered in the 2012-13 school year with respect to the spring 2013 AIMS assessments. A comparison is presented between Galileo assessments aligned to AZCCRS and those aligned to older AZ standards.

Sample: For the purpose of this brief, the first 36 Arizona districts/charters to provide ATI with their 2013 AIMS data for individual students in grades three through ten in mathematics, English language arts, and science were included in the sample. Based on the group of students who took the Galileo assessments and for which AIMS data was uploaded by the district/charter, the sample included 108,954 students for mathematics, 110,765 students for English language arts, and 18,775 students for science. The sample included districts and charters of various sizes from throughout the state of Arizona. On average, each district/charter tested 543 students in each grade and content area, with a minimum of 13 students and a maximum of 2,784 students. Within the sample, for most districts/charters, the set of standards tested varied across grades and content areas. As illustrated in Table 1, although most districts/charters were consistent in administering assessments aligned to a single set of standards (AZCC or older AZ) within a given grade and content area, a few districts/charters administered a mix of assessments aligned to different sets of standards within a grade and content area. Please note no new standards have been adopted in Arizona for science at this time.

Research Brief
Comparison of Predictive Validity and Forecasting Accuracy
for Assessments Aligned to AZCCRS and Older Arizona Standards
Assessment Technology Incorporated

1

TABLE 1

Number of districts/charters administering 2012-13 Galileo® assessments aligned to different sets of standards by grade and content area.

Number of Districts/Charters Administering 12-13 Galileo Assessments Aligned to Different Sets of Standards by Grade and Content Area				
Grade and Content Area		**All Older AZ Standards**	**All 2010 AZCCRS**	**Mix**
Math	3	21	11	1
	4	22	9	1
	5	22	9	1
	6	23	9	2
	7	19	9	3
	8	20	10	2
	10	6	2	0
Reading	3	18	14	1
	4	19	11	1
	5	20	11	1
	6	21	11	2
	7	17	12	2
	8	17	13	2
	10	4	2	0
Science	4	17	N/A	N/A
	8	18	N/A	N/A
	10	4	N/A	N/A
N/A = No data available. Arizona has not adopted new Science standards.				

Student Performance Measures: The AIMS data uploaded by districts/charters contains a scale score for each student as well as an indication of whether the student passed the statewide assessment. For each district/charter-wide Galileo 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 multiple district/charter-wide assessments 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 were conducted with respect to 1,580 district/charter-wide assessments administered by the 36 districts/charters described previously. This group of assessments included 550 assessments aligned to the 2010 AZCCRS in mathematics and English language arts, 910 assessments aligned to the 2008 Arizona standards in mathematics and the 2003 Arizona standards in English language arts, and 120 assessments aligned to the 2005 Arizona standards in science. Predictive validity analyses examine the strength of the relationship between two measures of student performance, in this case the student DL scores on a Galileo® 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 Results: Table 2 illustrates the mean correlation observed for the assessments aligned to each set of Arizona standards and administered in each grade and content area. As Table 2 shows, the mean correlations for assessments aligned to the older AZ standards in mathematics, English language arts, and science have an overall mean of 0.77 and range from 0.59 to 0.81 across grades and content areas. Similarly, the mean correlations for assessments aligned to the AZCCRS in mathematics and English language arts have an overall mean of 0.76 and range from 0.55 to 0.81. 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 2012-13 Galileo assessments were, on the whole, strongly related to student scores on the 2013 AIMS assessment. Moreover, correlations based on scores from Galileo® assessments aligned to older AZ standards and correlations based on scores from Galileo assessments aligned to the AZCCRS were similar.

TABLE 2
Correlation with the 2013 AIMS assessment for 2012-13 Galileo® assessments aligned to different sets of standards.

Correlation with the 2013 AIMS Assessment for 2012-13 Galileo Assessments Aligned to Different Sets of Standards				
	Older AZ Standards		**2010 AZCCRS**	
Grade and Content Area	**# of Assessments**	**Mean Correlation**	**# of Assessments**	**Mean Correlation**
ELA 3	66	0.78	51	0.79
4	69	0.76	41	0.80
5	73	0.79	52	0.77
6	76	0.78	55	0.75
7	63	0.77	57	0.75
8	63	0.75	51	0.76
10	14	0.59	9	0.71
Summary	**424**	**0.74**	**316**	**0.76**
Math 3	77	0.77	41	0.73
4	80	0.77	34	0.75
5	80	0.80	36	0.76
6	84	0.80	38	0.77
7	71	0.81	39	0.81
8	73	0.79	37	0.80
10	21	0.68	9	0.55
Summary	**486**	**0.77**	**234**	**0.74**
Science 4	51	0.74	N/A	N/A
8	56	0.74	N/A	N/A
10	13	0.71	N/A	N/A
Summary	**120**	**0.73**	**N/A**	**N/A**
Overall Summary	**1030**	**0.77**	**550**	**0.76**
N/A = No data available. Arizona has not adopted new Science standards.				

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. Although risk levels classify students along a continuum of increasing risk, for the purpose of evaluate the accuracy of forecasting, students who are classified as "On Course" or as "Low Risk" are considered to be predicted to pass the statewide assessment while students who are classified as "Moderate Risk" or "High Risk" are considered to be predicted to fail the statewide assessment. Forecasting accuracy analyses were conducted for the group of 36 districts/charters described previously. Across these districts/charters, there were 440 forecasting opportunities, where a forecasting opportunity is the student-level predictions made for the students within a given grade level in a specific district/charter 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 a given district/charter with regard to the statewide assessment in math).

Forecasting Accuracy Results: Figure 1 illustrates the mean percentage of students who passed AIMS for students classified into risk levels based on assessments aligned to different sets of standards. A prerequisite for accurate forecasting is that, as student risk level increases, the likelihood of failure on the statewide assessment increases. 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., "bubble" students who sometimes perform well and sometimes don't) should also display more variable performance on the statewide assessment. It should be noted that, if teachers and administrators are using the data provided by Galileo assessments to implement effective interventions, many students who have been classified as being at some risk of failing the statewide assessment may instead pass. As Figure 1 shows, the majority of students classified as being "On Course" based on their performance on the Galileo assessments did in fact pass the statewide assessment, while the majority of those classified as being at "High Risk" did in fact fail. The other two risk level groups performed as expected as well. Moreover, the observed relationship between risk level and pass rate was consistent regardless of whether the assessments used to generate the risk levels were all aligned to the AZCCRS, were all aligned to older AZ standards, or represented a mix.

**2013 AIMS Performance
for Students Classified into Galileo Risk Levels
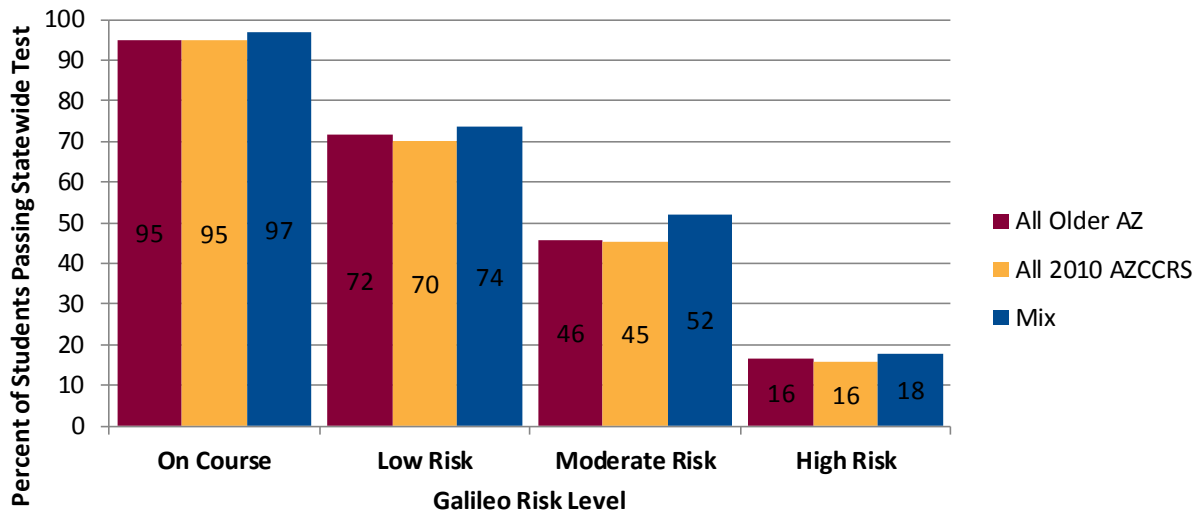Based on Assessments Aligned to Different Sets
of Standards**

Figure 1

Mean percentage of students passing AIMS for students classified into Galileo® risk level based on assessments aligned to different sets of standards.

Table 3 illustrates the mean overall forecasting accuracy for each grade and subject based on assessments aligned to different sets of standards. 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 Table 3 reveals, overall forecasting accuracy was also adequately high in a wide variety of grades and content areas. The mean overall forecasting accuracy was also consistently high regardless of the set of standards to which the assessments were aligned. Mean overall forecasting accuracy was 86 percent for risk levels based on assessments aligned to the older AZ standards, 85 percent for those based on assessments aligned to the AZCCRS, and 86 percent for those based on a mix of assessments aligned to different sets of standards.

Research Brief
Comparison of Predictive Validity and Forecasting Accuracy
for Assessments Aligned to AZCCRS and Older Arizona Standards
Assessment Technology Incorporated

5

1.800.367.4762
ati-online.com

TABLE 3
Mean overall forecasting accuracy for AIMS based on
2012-13 Galileo® assessments aligned to different sets of standards.

Mean Overall Forecasting Accuracy for AIMS Based on 12-13 Galileo Assessments Aligned to Different Sets of Standards

Grade and Content Area		All Older AZ Standards	All 2010 AZCCRS	Mix
Math	3	89	80	85
	4	85	82	87
	5	85	83	88
	6	84	84	85
	7	84	83	83
	8	83	82	85
	10	79	69	N/A
Reading	3	89	87	86
	4	88	87	88
	5	90	90	90
	6	89	90	91
	7	89	89	88
	8	84	84	79
	10	87	83	N/A
Science	4	84	N/A	N/A
	8	82	N/A	N/A
	10	78	N/A	N/A
Overall		**86**	**85**	**86**
N/A = No data available for this combination of factors.				

Conclusion: The research presented in this brief was conducted to evaluate predictive validity and forecasting accuracy for assessments aligned to different sets of standards and administered by districts/charters during the 2012-13 school year. Results suggest adequate levels of predictive validity for Galileo assessments aligned to 2010 AZCCRS as well as those aligned to older AZ standards. Results also suggest that 2012-13 Galileo risk levels displayed high levels of accuracy in forecasting student AIMS performance, both for risk levels based on assessments aligned to AZCCRS and those based on assessments aligned to older AZ standards. This research is consistent with investigations in previous years and suggests that Galileo assessments and risk levels continue to demonstrate adequate predictive validity and forecasting accuracy during this transition period from older AZ standards to AZCCRS. The results suggest that districts/charters can administer assessments aligned to either set of standards based on the curriculum currently being implemented in the district/charter as well as state recommendations related to the transition. Regardless of the set of standards selected, districts/charters can rely on Galileo® assessments to measure student progress towards standards mastery and to accurately forecast student performance on statewide assessments.