

Research Brief

**A Case Study Evaluation of a District's Galileo®  
Assessments in Kindergarten through Third Grade**  
by Sarah Callahan, Ph.D.  
Assessment Technology Incorporated



**Overview:** In recent years, states nationwide have focused on identifying and developing early childhood assessment systems that can help establish the strengths and weaknesses of students entering kindergarten, support differentiated instruction, guide early interventions, and track the progress of students throughout the early school years. State decision-making can be informed by research evaluating the approaches to early childhood assessment already being implemented by school districts and charter schools. The goal of this case study was to evaluate the reliability, predictive validity, and forecasting accuracy of the Galileo assessments administered by a large Arizona district in kindergarten through third grade. Since 2004, the district had been using the Galileo K-12 Online Instructional Improvement and Instructional Effectiveness System from Assessment Technology, Incorporated (ATI) to design, administer, and score periodic district-wide assessments in multiple grades and content areas. These assessments were designed to meet the district's goals by providing reliable, valid data including information about progress towards standards mastery and forecasts of likely performance on statewide assessments (i.e., Arizona's Instrument to Measure Standards [AIMS]).

**Student Performance Measures:** The assessments evaluated as part of this study included Galileo assessments in math and English language arts (ELA) administered by the district in kindergarten (2009-10), first grade (2010-11), second grade (2011-12), and third grade (2012-13). The average sample size for the Galileo assessments was 1755 students, with a minimum of 1513 students and a maximum of 1815 students. For each Galileo assessment administered, ATI performed an Item Response Theory (IRT) analysis that produced a marginal reliability estimate for the assessment and a scale score (i.e., Developmental Level [DL] score) for each student. As Table 1 illustrates, on the whole, marginal reliabilities for these assessments were adequate (mean=0.85). In addition, each student was classified as to their level of risk of failing the statewide assessment based on their performance on all the Galileo assessments within a given school year. In order of highest to lowest risk, the possible risk levels comprise "High Risk," "Moderate Risk," "Low Risk," and "On Course." In addition to risk level classifications, Galileo also provides empirically-based information about the mastery of individual standards and IRT-derived instructional recommendations supporting the reduction of student risk. For the 2013 AIMS assessments in math and ELA, each student received a scale score and also was classified into a performance category (i.e., Falls Far Below, Approaches, Meets, Exceeds).

**Predictive Validity of Galileo Developmental Level Scores:** For each school year, grade, and subject, a group of students were identified who had scores on both the Galileo assessments and the relevant 2013 third grade AIMS assessment. The number of students with scores on both measures was largest in third grade, where the vast majority of students with scores on the Galileo assessments also had AIMS scores. Due to student mobility, the size of the sample decreased as students were tracked back to kindergarten. However, even for kindergarten, the majority of students with scores on Galileo assessments also had AIMS scores. Correlation statistics were computed to evaluate the strength of the relationship between student DL scores on the Galileo assessments and student scale scores on the AIMS assessment. The correlation statistic ranges from -1 to +1 with the sign and magnitude of the correlation indicating the direction and strength of the relationship, respectively. A positive correlation indicates that high scores on one measure are associated with high scores on the other. A negative correlation indicates that high scores on one measure are associated with low scores on the other, although negative values are rarely observed in this context. Table 2 illustrates the correlation observed for each Galileo assessment administered in each school year, grade, and subject. As the table shows, on the whole, moderate to high positive correlations were observed. As

would be expected, the magnitude of the correlations typically increased as the time between the two measures decreased (i.e., from kindergarten to third grade).

**Forecasting Accuracy of Galileo® Risk Levels:** Forecasting accuracy analyses examine the accuracy with which Galileo risk levels predict the ultimate student performance on the relevant 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. As Table 3 shows, forecasting accuracy for the third grade AIMS assessment based on the Galileo risk levels in kindergarten, first, second, and third grade was uniformly high, ranging from 0.86 to 0.95 dependent on the grade and subject. In particular, it is noteworthy, that the forecasting accuracy for Galileo risk levels based on performance on kindergarten Galileo assessments was 0.86 in math and 0.87 in ELA even though the Galileo assessments were administered three years before the third grade AIMS assessment. Similarly, as Figures 1 and 2 illustrate, students who are classified at higher levels of risk based on their performance in kindergarten, first, second, and third grade are more likely to fail the third grade AIMS assessment. However, Figures 1 and 2 also suggest that even students who are classified as moderate or high risk early on can ultimately pass the third grade AIMS assessment. This provides evidence of the efficacy of district efforts to use the information provided by the Galileo assessments to reduce student risk throughout the early childhood years.

**Conclusion:** This case study demonstrated the reliability, predictive validity, and forecasting accuracy of the Galileo assessments administered by a large Arizona district throughout the early childhood years. The results provide strong evidence that the Galileo assessments administered by the district in kindergarten through third grade provided reliable, valid student DL scores that correlated well with student AIMS scale scores at the end of third grade. In addition, Galileo risk levels based on student performance in kindergarten, first, second, and third grade accurately predicted ultimate student performance on third grade AIMS assessments. The results of this case study also suggest that early intervention efforts can be effective in many cases in reducing student risk of failing to pass the statewide assessment. Therefore, it is critical that early childhood assessment systems provide not only classifications of student risk but also detailed information that can support districts and charters in implementing effective early interventions.

**Supplemental Tables and Figures:**

**TABLE 1**  
*Marginal reliabilities for Galileo assessments*

<b>Marginal Reliabilities for Galileo Assessments</b>				
<b>School Year, Grade, and Subject</b>	<b>Marginal Reliabilities</b>			
	<b>#1</b>	<b>#2</b>	<b>#3</b>	<b>#4</b>
<b>2009-10 KG Grade Math</b>	0.86	0.78	0.84	N/A
<b>2009-10 KG Grade ELA</b>	0.85	0.76	0.74	N/A
<b>2010-11 01st Grade Math</b>	0.78	0.73	0.67	0.76
<b>2010-11 01st Grade ELA</b>	0.88	0.84	0.84	N/A
<b>2011-12 02nd Grade Math</b>	0.93	0.89	0.87	0.88
<b>2011-12 02nd Grade ELA</b>	0.92	0.88	0.88	0.89
<b>2012-13 03rd Grade Math</b>	0.90	0.88	0.87	0.88
<b>2012-13 03rd Grade ELA</b>	0.91	0.88	0.87	0.90

**TABLE 2**

**Correlation between each Galileo® assessment and the 2013 third grade AIMS assessment in the relevant content area**

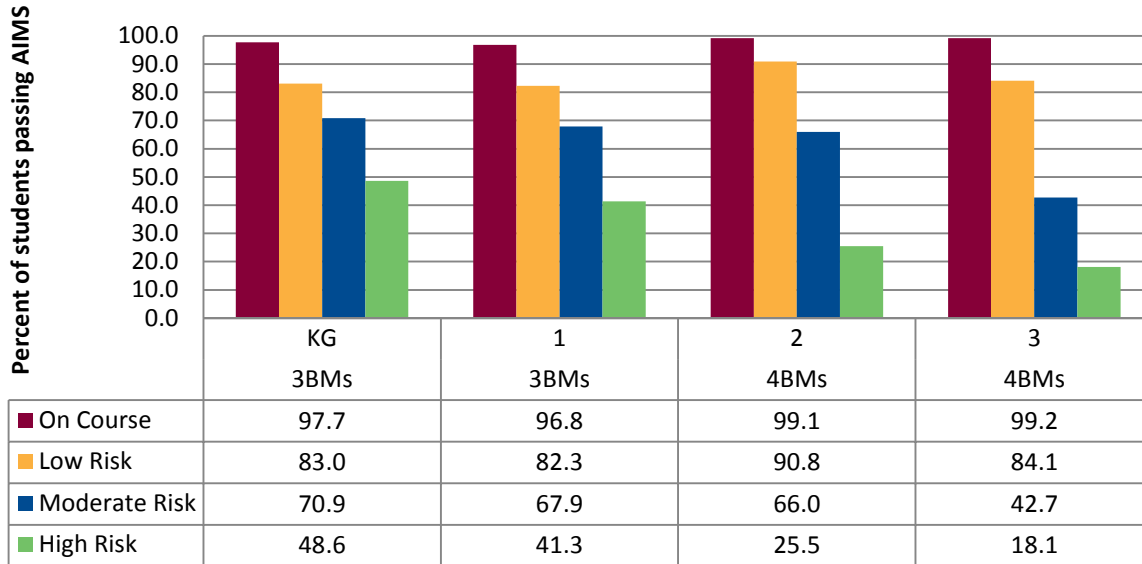
<b>Galileo Assessment Correlations with 2013 Third Grade AIMS</b>					
School Year, Grade, and Subject for Galileo Assessments	Students with Scores on Galileo Assessments and 2013 Grade 3 AIMS	Correlation between Galileo Assessment and 2013 Grade 3 AIMS			
2009-10 KG Grade Math	976	0.54	0.58	0.53	N/A
2009-10 KG Grade ELA	952	0.59	0.43	0.52	N/A
2010-11 01st Grade Math	1247	0.51	0.55	0.58	0.61
2010-11 01st Grade ELA	1317	0.57	0.62	0.56	N/A
2011-12 02nd Grade Math	1427	0.73	0.72	0.72	0.70
2011-12 02nd Grade ELA	1397	0.67	0.69	0.73	0.73
2012-13 03rd Grade Math	1649	0.79	0.75	0.78	0.81
2012-13 03rd Grade ELA	1667	0.76	0.77	0.77	0.80

**TABLE 3**

**Forecasting accuracy for the third grade AIMS assessment based on Galileo risk levels in kindergarten through third grade**

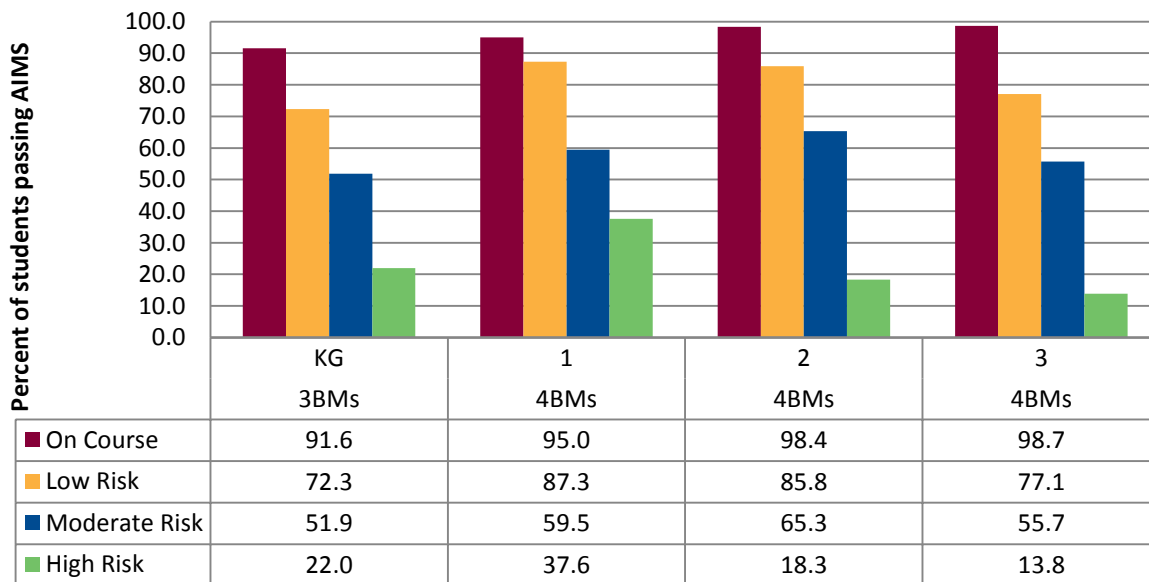
<b>Forecasting Accuracy for 2013 Third Grade AIMS</b>			
School Year, Grade, and Subject for Galileo Assessments	# of Galileo Assessments Administered	Students with Scores on Galileo Assessments and 2013 Grade 3 AIMS	Forecasting Accuracy for 2013 Grade 3 AIMS based on Galileo Risk Levels
2009-10 KG Grade Math	3	976	0.86
2009-10 KG Grade ELA	3	952	0.87
2010-11 01st Grade Math	4	1247	0.87
2010-11 01st Grade ELA	3	1317	0.89
2011-12 02nd Grade Math	4	1427	0.91
2011-12 02nd Grade ELA	4	1397	0.94
2012-13 03rd Grade Math	4	1649	0.92
2012-13 03rd Grade ELA	4	1667	0.95
<b>Average Forecasting Accuracy</b>			<b>0.90</b>

**Sample District:  
Percent of students in each risk group who passed the  
spring, 2013 AIMS assessment in ELA**



**Figure 1**  
*Percentage of students in each risk group who passed the third grade English language arts AIMS assessment*

**Sample District:  
Percent of students in each risk group who passed the  
spring, 2013 AIMS assessment in Math**



**Figure 2**  
*Percentage of students in each risk group who passed the third grade math AIMS assessment*