Assessment Instrument Table: Galileo

Element	Description	Assessment Instrument Information
Instrument Name	Name of specific instrument (more than vendor name).	Galileo® K-12 Online Pre-built and Customized Interim Benchmark Assessments
Vendor	Name of the company or organization that produces the instrument.	Assessment Technology Incorporated (ATI)
Purpose (Intended Use)	The described purpose and appropriate uses of the instrument.	Galileo K-12 Online includes an integrated comprehensive assessment system aligned to state standards in a variety of content areas including Common Core State Standards and Colorado Assessment Standards. Galileo assessments and item types (including technology-enhanced item types) are designed to reflect the guidelines released by statewide assessment consortia such as Partnership for Assessment of Readiness for College and Careers (PARCC). Galileo provides the capability to administer (online, offline, and via handheld wireless devices) multiple types of pre-built and customized district, school, and classroom assessments (e.g., interim benchmark, pretest/posttest, placement, end-of-course) in a wide variety of content areas (e.g., English language arts, math, writing, science, social studies, art, music). This description focuses on Galileo pre-built and customized interim benchmark assessments. Galileo interim benchmark assessments are designed to provide information about student standards mastery, to support the measurement of student growth, and to predict student performance on the statewide test.
Population	Who (which students) could be assessed using the instrument.	Students in grades kindergarten through 12.
When? How frequently?	How frequently the instrument can be administered in a school year, and recommended or required administration windows.	Galileo provides access to both pre-built and customized interim benchmark assessments. For pre-built assessments, five versions of the assessment are available to support administration multiple times throughout the school year. Typically, districts/charters implement between 3-5 interim benchmark assessments during the school year. Administration windows are determined by the district/charter. For customized assessments, Galileo's online Assessment Planner allows the district/charter to determine the number and timing of assessments.

Element	Description	Assessment Instrument Information
Content Area (s)	Content area or areas being assessed.	Galileo provides access to pre-built and customized interim benchmark assessments in math, English language arts (ELA), science, and writing. Districts/charters participating in ATI's Community Assessment and Item Banking initiative can also create customized pretests, posttests, and interim benchmark assessments using their own district/charter-created items in non-state-tested areas such as social studies, foreign languages, physical education, music, arts, career and technical education. As part of this initiative, ATI provides Item Response Theory (IRT) analyses and other statistical analyses designed to support the evaluation of student growth in these content areas.
Learning Objectives	Specific learning objectives being assessed, at as detailed a level as is provided. This may be "topics" or categories or may be actual learning objective statements.	Galileo provides access to pre-built and customized interim benchmark assessments in math, English language arts (ELA), science, and writing. Pre-built interim benchmark assessments comprehensively cover the Colorado Assessment Standards, focusing on the PARCC blueprint from the relevant grade level and content area (where applicable). Galileo also provides access to customized interim benchmark assessments aligned to the district/charter pacing guide and addressing the standards desired by the district/charter.
Individual Metrics	The scores provided at the individual (student) level.	**Raw Score:** The Raw Score for a test represents the points earned by the student (e.g., number or percent correct). **Developmental Level (DL Score):** The DL score is a scaled score similar to that provided by the statewide test. The DL score represents an Item Response Theory-derived estimate of student ability. **Performance Level:** The Performance Level indicates the student's performance relative to cut scores established to classify students in a similar way to the statewide test. **Risk Level:** The Risk Level is the best prediction of likely student performance on the statewide test because it takes into account student performance across one or more assessments throughout the school year. The Risk Level is established based on an initial assessment and refined as additional assessments are administered. **Observed Growth:** Observed Growth represents the change in a student's DL score across two assessments for which the scores have been placed on a common scale using IRT procedures. An

Element	Description	Assessment Instrument Information
		evaluation of Observed Growth relative to Expected Growth for the time period between the two assessments results in a categorization indicating whether or not the student has met Expected Growth.
		Standard Score: The Standard Score specifies position in a norm group in terms of standard deviation units.
		Percentile Rank: The Percentile Rank gives the percentage of scores in the norm group at or below a particular score.
		Normal Curve Equivalent (NCE) Score: The Normal Curve Equivalent score is a norm-referenced score with a mean of 50 and a standard deviation of 21.06.
Individual Comparison Points (cut scores)	Information provided regarding how good is good enough performance on the instrument. Comparison information should be available for every individual metric. This may be performance level ratings with specific cut scores.	ATI's in-house research staff use well-established statistical techniques (e.g., equipercentile equating) to establish cut scores for each pre-built and customized assessment that classifies students in a similar way to the relevant statewide test. These cut scores identify a raw score and a corresponding DL score that define the lower boundary for each Performance Level and identify the Performance Level that represents meeting the standard on the statewide test. A student's Risk Level is then established based on whether the student met the standard across one or more assessments.

ATI also conducts annual research across all ATI clients nationwide that provides districts/charters with estimates of the expected growth for students in a wide variety of grades and content areas. Based on this research, Galileo generates expected growth estimates that are customized to reflect the time period between two tests administered over a given time period. These expected growth estimates are compared to a student's Observed Growth to determine whether the student met the expected growth for the time period between two tests.

Galileo also provides information about comparison to a norm group via several norm-referenced scores (i.e., Standard Score, Percentile Rank, Normal Curve Equivalent Score). |
| **Aggregate Metrics** | Scores provided at the group level. The group | Galileo reports provide aggregated data for all of the individual metrics described previously (i.e., Raw Score, DL Score, Performance Level, Risk Level, Observed Growth, Standard Score, Percentile |

Element	Description	Assessment Instrument Information
	could be a grade level, school, district, or disaggregated groups (e.g. race/ethnicity, gender, IEP status, FRL status) Specify the group(s) and the score(s) provided.	Rank, Normal Curve Equivalent Score). Aggregated data are available by course, grade level, school or district/charter. Customized filters enable users to generate reports for disaggregated groups based on virtually any type of information available within Galileo (e.g., student demographics, intervention groups, teacher and program variables).
Aggregate Comparison Points (cut scores)	Information provided regarding how good is good enough performance at the group level.	As described previously, cut scores are established that identify a raw score and a corresponding DL score that define the boundary for each Performance Level and identify the Performance Level that represents meeting the standard on the statewide test. A student's Risk Level is then established based on whether the student met the standard across one or more assessments. In Galileo reports displaying aggregated data, districts/charters can evaluate group data relative to these cut scores. For example, Galileo includes reports that indicate the percentage and count of students in each Performance Level or Risk Level. Galileo also provides state-of-the-art statistical analyses evaluating aggregated Observed Growth data for a teacher or school relative to ATI's research-based expected growth estimates. Categorical Growth Analysis employs a well-established common statistical test, the repeated measures t-test, to evaluate whether the difference between observed growth and expected growth is significant for a particular group of students. A classification of Expected Growth Exceeded indicates that Observed Growth was significantly higher than expected growth while a classification of Expected Growth Not Maintained indicates that Observed Growth was significantly lower than expected growth. A classification of Expected Growth Maintained indicates that Observed Growth was not significantly different from expected growth.

Element	Description	Assessment Instrument Information
Aggregate Comparison Points (CDE Cut Scores)	Cut scores established for Requests to Reconsider	See tables below for aggregate metrics for which CDE established comparison points as part of the 2014-15 Request-to-Reconsider process. CDE provided comparison points for the percentage of students with Developmental Level scores categorized as proficient or advanced. **Note: The CDE comparison points for Galileo benchmark assessment scores for the 2016-17 Request-to-Reconsider process are currently being revised.**

50th Percentile of % of Students Proficient & Advanced on Post-Test				
Grade	Reading	Writing	Math	Science
K	71.65	53.52	70.89	N/A
1	71.65	53.52	70.89	N/A
2	71.65	53.52	70.89	47.53
3	71.65	53.52	70.89	47.53
4	71.65	53.52	70.89	47.53
5	71.65	53.52	70.89	47.53
6	71.43	57.77	52.48	48.00
7	71.43	57.77	52.48	48.00
8	71.43	57.77	52.48	48.00

Element	Description	Assessment Instrument Information
Data Reports	Description of data reports that are provided/available at the individual and aggregate level(s).	Galileo's reporting functionality displays actionable, real-time, reliable, and valid standards-aligned data on student learning and includes Dashboards accessible at the District, school, cohort (i.e., grade), and class level, as well as an overview page for students in the K-12 Student-Parent Center. A broad set of ready-made standards-aligned reports are available to Galileo users. In addition, with the use of Form Builder, Report Builder, and Custom Test Reports, users have the option of designing their own customized reports. Various formats are available for reports including PDF, Microsoft Excel, RTF, and comma-separated/delimited files. The following highlights selected examples of Galileo Dashboards and reports including examples of reports that provide access to the individual and aggregated metrics described previously. A wide variety of additional reports including curriculum monitoring reports, testing reports, and administrative reports are also available. **Teaching Dashboard:** Provides assessment results, recommendations for student groupings, targeted standards for instruction, intervention, and enrichment and links to online Instructional

Element	Description	Assessment Instrument Information

Dialogs. Dashboard information about student learning and progress includes: 1) information about student performance in the form of raw scores, DL scores, Performance Levels, and Risk Levels; 2) student Observed Growth; 3) student mastery of standards; and 5) student performance on specific items/skills.

School Performance Dashboard: Enables school- and district-level users to view aggregate data including assessment results in the form of Risk Levels, the results of analyses of student growth, recommendations for student groupings, targeted standards for instruction, intervention, and enrichment and links to online Instructional Dialogs.

Instructional Effectiveness Dashboard: Districts in Colorado and other states implementing initiatives related to instructional effectiveness (IE) may choose to make an additional Dashboard available. The Instructional Effectiveness Dashboard enables users to obtain all instructional-effectiveness-related information in one place. The IE Administration tab, available to staff with appropriate permissions, contains technology to manage a complete instructional effectiveness system including tools to access and/or build proficiency rating scales, administer proficiency rating scales, and import/view staff files. The Staff IE Results tab serves as a place for all users to access their evaluation results including proficiency rating results, current student growth and achievement results, and compiled evaluation data.

Student Growth and Achievement
Provides a graphical cross-classification of students based on their growth and achievement that facilitates the identification of student groups for intervention and enrichment (e.g., high performing students showing inadequate growth, low performing students showing exceptional growth). Also provides the results of a Categorical Growth Analysis evaluating whether, for the class as a whole, expected growth was exceeded, maintained, or not maintained.

Categorical Growth Summary
Presents a summary of the Categorical Growth Analyses for all teachers and schools. Growth classifications include Expected Growth Exceeded, Expected Growth Maintained, and Expected Growth Not Maintained.

Test Scores Reports

Element	Description	Assessment Instrument Information
		Present student performance on a test in terms of raw scores. **Development Summaries** Presents a series of standardized scores (e.g., Standard Score, Percentile Rank, Normal Curve Equivalent Score, DL Score) that summarize development at the individual student level or at various levels of aggregation. Ideal for monitoring progress in terms of the relative standing of a student, class, or school. **Aggregate Multi-Test Reports** Supports the monitoring of student progress and the forecasting of statewide test performance by providing Developmental Level scores for multiple tests and information about Performance Level or Risk Level. Aggregate views display the percentage and count of students in each classification. The user can drill down to identify groups of students at each level of risk as well as standards for which intervention is required. **Intervention Alert** Indicates the percent of students that have demonstrated mastery of each standard relative to district/charter-determined achievement level classifications. Standards where fewer than 75 percent of students have demonstrated mastery are highlighted in red.
Alignment	Information provided by the vendor about alignment of this instrument to other instruments, standards, etc.	Galileo includes access to pre-built interim benchmark assessments that comprehensively cover the Colorado Assessment Standards, focusing on the PARCC blueprint (where applicable). Districts/charters can also create customized interim benchmark assessments aligned to their pacing guide. All Galileo pre-built and customized assessments are created using items from the Galileo item banks. These item banks are among the largest secure and community (formative) item banks in the nation and contain more than 142,000 items including more than 8,500 technology-enhanced items. Item types include selected-response (e.g., multiple-choice, true/false, yes/no), constructed-response (e.g., essay, short answer), and technology-enhanced (e.g., selectable text, multi-part, expanded selected-response, sequencing, customized technology-enhanced) item as well as performance-based assessment tasks. ATI has extensive experience aligning Galileo items to state and national standards including Colorado Assessment Standards and Common Core State Standards. The item development process for assessments in Galileo has been carefully designed to produce high-quality items and includes the development of item specifications, item construction, and multi-tiered item review.

Element	Description	Assessment Instrument Information
		This item development process has ensured that ATI can adapt Galileo item banks in response to new and changing standards and statewide assessments. For example, item specifications may be designed to align very closely to guidelines and released items from assessment consortia (e.g., PARCC). The careful design of item specifications is especially critical to the development of technology-enhanced and constructed-response items similar to those students are likely to encounter on statewide assessments. Recent ATI expansion of the Galileo item banks has been focused on responding to the introduction of new standards in Colorado and many other states. Newly developed Galileo items address requirements related to the new standards such as the emphasis on the development of higher order thinking skills reflecting high Depth of Knowledge (DOK) levels. In the area of English language arts, there is a new emphasis on content-rich informational texts that require close reading and detailed analysis. In the area of mathematics, there is increased focus on understanding mathematical operations as well as fluently implementing mathematical operations. In both science and mathematics, there is increased emphasis on the use of physical models to inform mental models of processes of interest. The new standards also have required a shift from the heavy reliance on selected-response items characteristic of earlier assessments with a more balanced approach involving selected-response items, technology-enhanced items, and constructed-response items. This balanced approach results in an expansion of item types essential for assessment. In response, ATI has developed a wide variety of technology-enhanced and constructed-response items including those that are modeled on samples provided by PARCC. These items can be included on Galileo pre-built and customized assessments to provide students with practice prior to the statewide test.
Technical Quality	Information about the technical quality of the instrument. Reference to technical analysis if available electronically.	ATI research routinely indicates that Galileo pre-built and customized assessments are reliable, are valid, and are effective forecasters of student performance on statewide assessments. To help ensure adequate levels of reliability (i.e., 0.80 or higher), Galileo assessments typically contain between 35 and 50 items. A recent investigation of predictive validity revealed a high mean correlation (0.78) between scores on Galileo assessments and scores on the statewide test. Similarly, Galileo Risk Levels displayed a high level of accuracy (85 percent) in forecasting statewide test performance. For additional details on ATI reliability, validity, and forecasting accuracy research, see the Technical Manual (http://www.ati-online.com/pdfs/researchK12/K12TechManual.pdf, p. 20-55) as well as a recent Research Brief

Element	Description	Assessment Instrument Information
		evaluating the validity and forecasting accuracy of Galileo assessments (http://ati-online.com/pdfs/researchK12/PredictiveValidityandForecastingAccuracy2012-13.pdf, p. 1-4). Research evaluating Galileo assessments is undertaken annually.