

District Resource

# Building Reliable and Valid Benchmark Assessments

A Resource for District Leaders and Staff Responsible for  
Benchmark Planning, Construction, and Review

by  
John Richard Bergan, Ph.D.  
John Robert Bergan, Ph.D.  
Kathryn S. Bergan, Ph.D.  
Christine Guerrero Burnham, Ph.D.  
Scott A. Cunningham, B.S.  
Jason K. Feld, Ph.D.  
Karyn White, M.A.  
Kerridan A. Smith, B.A.



Assessment  
Technology  
Incorporated

**Assessment Technology, Incorporated**

6700 E. Speedway Boulevard

Tucson, Arizona 85710

Phone: 520.323.9033 • Fax: 520.323.9139

© 11/13/2006 and 9/15/2008 Assessment Technology, Incorporated



# Building Reliable and Valid Benchmark Assessments

*A Resource for District Leaders and Staff Responsible for Benchmark Planning, Construction, and Review*

## Table of Contents

- Introduction ..... 2
- 1. Overview of Benchmark Assessments ..... 3**
  - A. Benchmark Assessment Characteristics, Uses, and Psychometrics ..... 3
    - 1. Standards-Based Assessment ..... 3
    - 2. Assessment to Inform Instruction ..... 3
    - 3. Providing Multiple Assessments of Standards Mastery ..... 4
    - 4. Risk Assessment Analysis for Meeting Standards on Statewide Tests ..... 4
    - 5. Establishing and Sustaining Psychometric Integrity of Benchmarks ..... 6
      - a) Benchmark Reliability ..... 6
      - b) Benchmark Validity ..... 7
        - (1) Correlations with Statewide Tests and Between Benchmark Assessments ..... 7
        - (2) Accuracy of Forecasted State Classifications of Standards Mastery ..... 7
      - c) Benchmark Item Parameter Estimates ..... 8
        - (1) Discrimination Parameter ..... 8
        - (2) Difficulty Parameter ..... 8
        - (3) Guessing Parameter ..... 9
  - B. A Look under the Hood: ATI Benchmark Item Development and Certification Procedures ..... 10
    - 1. Guiding Standards for Developing Items ..... 10
      - a) Certified Items ..... 10
      - b) Suggestions for New Items ..... 10
    - 2. Item Specifications ..... 11
    - 3. Item Review and Certification ..... 13
  - C. Putting It All Together: Steps in Benchmark Assessment Planning, Construction, Review, and Delivery ..... 14
    - 1. Initial Benchmark Preparation Communications with ATI ..... 14
    - 2. Use of Benchmark Planner to Define Assessment Goals ..... 14
      - a) Assessment Delivery Date ..... 15
      - b) Number of Assessments per Year ..... 15
      - c) Standards Being Measured and Number of Items per Standard ..... 15
    - 3. Confirmation that Benchmark Planner is Complete ..... 16
    - 4. Development of Benchmark Draft by ATI ..... 17
    - 5. Guided Review of Benchmark Draft ..... 17
    - 6. Final Benchmark Assessment Delivered by ATI ..... 19
- 2. Navigation within Galileo Online Benchmark Planner ..... 20**
  - A. The Process of Designing Benchmark Assessments ..... 20
  - B. The Process of Creating Benchmark Assessments ..... 22
- 3. Navigation within Galileo Online Test Review ..... 24**
  - A. Set Initial and Final Reviewers ..... 24
  - B. Initial Reviews ..... 25
  - C. Final Review ..... 27

*Copyright © 11/13/2006 and 9/15/2008 by Assessment Technology, Inc.  
 All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the publisher.  
 Assessment Technology, Inc., Publishers  
 Tucson, Arizona, U.S.A, Printed in the United States of America. V3 Ap.*



## Introduction

---

As a part of our partnership with your district, ATI provides you with guidance and ongoing recommendations to assist in the planning, construction, and review of your benchmark assessments. The goal of this partnership is to help ensure that your Galileo district benchmark assessments and reporting tools help you to accomplish several fundamental objectives related to the use of benchmark assessments. Carefully designed district benchmarks can:

1. Provide reliable and valid information on student learning for teachers, administrators, specialists and parents to utilize in standards-based planning, instruction and intervention;
2. Ensure that benchmark assessments are aligned to state and district standards and tailored to reasonably accommodate district pacing guides;
3. Provide documentation throughout the school year on student progress and learning outcomes in meeting district and state standards;
4. Provide periodic updates on the extent to which students are on course for meeting statewide standards as measured by statewide tests.

To assist in your efforts to build well-designed benchmark assessments, we are providing you with this resource document. This document is designed for use by district leadership staff responsible for oversight and supervision of benchmark assessment planning, construction, and review activities. Throughout the year, this resource will be an important guide as ATI Educational Management Services staff and district staff work together on the construction of the district's benchmark assessments.

### ***Before you begin the benchmark development process, please:***

1. Read this document carefully, review it with staff involved in benchmark development activities, and contact ATI Educational Management Services with any questions.
2. Complete the “*Benchmark Contact Form*” and email it to Educational Management Services at [EdMgtSvs@ati-online.com](mailto:EdMgtSvs@ati-online.com). With the availability of the information on this form, ATI is better able to support district staff contributing to benchmark assessments.

This resource contains the following sections:

- 1) *Overview of ATI Galileo Benchmark Assessments: Purposes and Psychometrics*
  - a. *Benchmark Assessment Characteristics, Uses, and Psychometrics*
  - b. *A Look Under the Hood: ATI Item Development and Certification Procedures*
  - c. *Putting it All Together: Steps in Benchmark Assessment Planning, Construction, Review and Delivery*
- 2) *Navigating within Galileo Online Benchmark Planner*
- 3) *Navigating within Galileo Online Test Review*



## 1. Overview of ATI Galileo Benchmark Assessments

---

### A. Benchmark Assessment Characteristics, Uses, and Psychometrics

#### 1. Standards-Based Assessment

Benchmark assessments are locally relevant, district-wide assessments designed to measure student achievement of standards for the primary purpose of informing instruction. The first step in constructing a benchmark assessment is to identify the standards that the assessment will measure.

#### 2. Assessment to Inform Instruction

The fundamental purpose of benchmark assessments is **to provide information that can be used to guide instruction**. In this regard, Galileo benchmark assessments can measure student mastery of standards targeted for instruction. In so doing, they indicate what students have accomplished when given appropriate learning opportunities. Benchmark assessments also inform instruction in cases in which standards have not been mastered even though appropriate learning opportunities have been provided.

Benchmark assessments play a fundamental role in helping districts implement standards-based education-intervention initiatives. For example, an important component of these initiatives is to go beyond the basic responsibility of providing appropriate learning opportunities. These initiatives focus not only on whether skills have been taught, but also on whether or not the skills have been learned. Galileo benchmark assessments support these types of initiatives by pinpointing the specific skills that students need to acquire to master standards that have been targeted for instruction, but that have not yet been met.

Benchmark assessments provide information to guide instruction in a cyclical fashion. In some cases, initial instruction is preceded by a pretest designed to provide an overall picture of initial student mastery of standards. Initial instruction is followed by a benchmark assessment designed to assess mastery of standards covered during the initial instructional period. For example, a benchmark assessment may be administered after instruction implementing a set of unit plans has been completed. Teachers and administrators may use the results of that assessment to plan and implement interventions to address areas in which students may not have displayed mastery of standards measured on the test. For instance, a re-teaching intervention may be employed assisting students to master standards that they have not yet met. Short formative assessments built by using *Galileo Test Builder* or accompanying interventions administered through ATI Instructional Dialogs may then be used to ensure that standards not initially mastered have been mastered following re-teaching.



### **3. Providing Multiple Assessments of Standards Mastery**

Typically the cycle of teaching, assessment, and intervention with benchmark assessments supported by short formative assessments is implemented three or four times during the school year. Repetitions of the cycle provide an increasing body of information about student learning. This information, coupled with information on statewide test performance, provides the opportunity for a multi-test approach to the assessment of standards mastery. The multi-test approach has several benefits to the student and district stakeholders, including the following:

- Given the fact that any test, including statewide achievement tests used to make high-stakes decisions, has some degree of measurement error, use of the multi-test approach can reduce the impact of a single test on high-stakes decisions.
- The multi-test approach increases the likelihood that the assessments used to assess mastery cover the full range of content that has actually been taught.
- The multi-test approach increases timely access to assessment information that can be used in the overall determination of mastery.
- Multi-test information can be used in providing evidence of standards mastery in cases in which other available evidence is insufficient or subject to question.

### **4. Risk Assessment Analysis for Meeting Standards on Statewide Tests**

When a relationship has been established between performance on one or more benchmark assessments and performance on a statewide test, benchmark results can be used to assess the level of risk that a given student will not meet state standards as measured by the statewide test. The probability of accurately forecasting mastery of state standards will depend in part on the strength of the relationship between each benchmark assessment and the statewide test and in part on the mastery patterns evidenced by the students. For example, consider the situation in which performance on each of three benchmark assessments is correlated with performance on a statewide test. Suppose that a student has failed to meet standards on all three benchmark tests. The probability that the student will meet standards as measured by the statewide test will in all likelihood be substantially lower than will be the case for a student who has met standards on all three benchmark assessments. Galileo benchmark assessment results can provide risk forecasts for students, which are adjusted following subsequent administrations of benchmark assessments.

The table below illustrates the type of information that may be used to inform risk assessment. After data of the type shown in the table have been gathered, the results can be used to forecast risk given subsequent administrations of benchmark tests. As additional data becomes available, risk forecasts can be adjusted.



#### Risks of not Meeting State Standards Given Varying Patterns of Benchmark Standards Mastery

Assessments	District Risk Assessments - 5th Grade Math					
	Benchmark Mastery Patterns			Number of Students	AIMS Test	
	1	2	3		Met	Not Met
<b>Risk Assessment 1</b> <b>Benchmark 1</b>	Met			417	0.94	0.06
	Not Met			44	0.41	0.59
<b>Risk Assessment 2</b> <b>Benchmarks 1,2</b>	Met	Met		395	0.96	0.04
	Met	Not Met		22	0.59	0.41
	Not Met	Met		21	0.57	0.43
	Not Met	Not Met		23	0.26	0.74
<b>Risk Assessment 3</b> <b>Benchmark 1,2,3</b>	Met	Met	Met	375	0.98	0.02
	Met	Met	Not Met	20	0.50	0.50
	Met	Not Met	Met	14	0.79	0.21
	Met	Not Met	Not Met	8	0.25	0.75
	Not Met	Met	Met	13	0.77	0.23
	Not Met	Met	Not Met	8	0.25	0.75
	Not Met	Not Met	Met	5	0.80	0.20
	Not Met	Not Met	Not Met	18	0.11	0.89

The table summarizes data from three benchmark assessments and a statewide test, the Arizona Instrument to Measure Standards (AIMS). The first Benchmark Test was administered in the fall, the second in the winter, and the third in the spring. The AIMS test was administered toward the end of the school year following the three benchmark tests. The results of three risk assessments are shown in the table. The first risk assessment shown in the table examined the risk of not meeting the state standard as measured by AIMS given that the benchmark standard was met or not met on Benchmark Test 1. The second assessment examined the risk of not meeting the state standard given the various possible mastery classifications for Benchmark 1 and Benchmark 2. The third risk assessment utilized information from all three benchmark tests.

Results for the first assessment show that 417 students were classified as meeting the state standard and 44 students were classified as not meeting the standard based on their performance on Benchmark 1. The probability of meeting the standard on the statewide test given that the standard was met on Benchmark 1 was .94. The probability of not meeting the standard on AIMS was only .06. On the other hand, the probability of not meeting the standard on AIMS for students who did not meet the standard on Benchmark 1 was .59. Thus, the risk of not meeting the standard on AIMS was much higher for students who did not meet the benchmark standard than for those who did meet the benchmark standard. The results for risk assessment 2 and risk assessment 3 show that the risk of not meeting the state standard increases when students fail to meet the standard on consecutive benchmark tests.

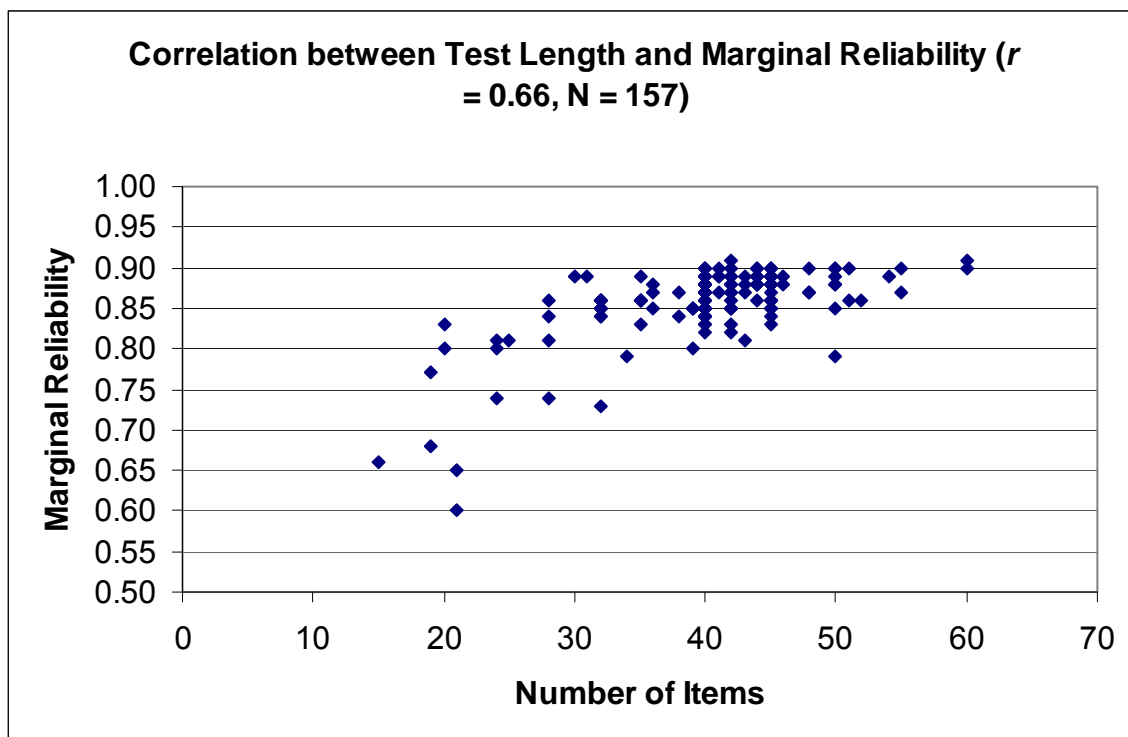


## 5. Establishing and Sustaining Psychometric Integrity of Benchmarks

- **Benchmark Reliability.**

Benchmark assessments must be reliable. Reliability has to do with the consistency of information provided by an assessment. A particularly important form of reliability for benchmark assessment is internal consistency. Measures of internal consistency provide information regarding the extent to which all of the items on a benchmark assessment are related to the underlying ability (Developmental Level) that the assessment is designed to measure. Benchmark assessments such as those provided to districts through Galileo are designed to correlate with other measures of student proficiency including statewide assessments. An assessment that lacks internal consistency does not correlate well even with itself. Therefore, it is unlikely that it would correlate well with other measures.

Reliability is directly affected by the length of the benchmark assessment. Longer assessments tend to be more reliable than shorter assessments. The illustration below plots the relationship between benchmark length and reliability for 156 Galileo benchmark assessments administered in multiple grades and multiple school districts. The data in the figure suggest that adequate levels of reliability can be achieved with benchmark assessments that are about 35 items to 40 items long. ATI recommends that benchmark assessments contain a **minimum of forty items** to ensure adequate reliability.



As shown above, benchmark assessments consistently begin to reach an acceptable level of reliability at a length of about 35 to 40 items.



- **Benchmark Validity.**

- **Correlations with Statewide Tests and Between Benchmark Assessments.**

Standards-based educational initiatives across the nation are targeting instruction toward the achievement of state and local standards. State standards provide a common set of goals for all school districts within the state. Statewide tests play a critical role in measuring the achievement of state standards. Benchmark assessments are also intended to measure the achievement of state standards. Accordingly, it is reasonable to expect significant correlations between benchmark tests in a particular state and the statewide test for that state. A finding revealing such correlations would provide important evidence of the validity of the benchmark assessments. Although significant correlations do support the validity of benchmark tests, it is important to recognize that benchmark tests differ from statewide tests in significant ways. The two forms of assessment serve different purposes. Statewide tests are typically administered toward the end of the school year to provide accountability information for the state and local education agencies and for the public. Benchmark tests are administered periodically during the school year to guide instruction. The skills assessed on a benchmark test are typically selected to match skills targeted for instruction in the curriculum at a particular time. This is not the case for statewide tests.

For these and other reasons, benchmark tests should not be thought of as replicas of statewide tests. Accordingly, although the two forms of assessment should correlate, the correlations should not be expected to be as high as the correlation between parallel forms of the same test. Multiple benchmark tests administered during the school year measure student achievement in the same or related knowledge areas. As a result, it is reasonable to expect benchmark tests to correlate well with each other. Thus, correlations among benchmark assessments provide evidence of the validity of benchmark assessments.

As one illustration of Galileo benchmark assessment validity, research conducted on the correlations between Galileo benchmark assessments and statewide tests such as AIMS (Arizona Instrument to Measure Standards) for grades 3 through 8 in the areas of Reading and Math showed correlations with a range of .71 to .83 and a mean of .77. Moreover, the relationships among the benchmark assessments were also adequate ranging from .57 to .80 with a mean of .74. Similar findings have been observed in other states.

- **Accuracy of Forecasted State Classifications of Standards Mastery.**

Determining whether or not students have mastered standards is, for all intents and purposes, a categorical decision (i.e., they did or they didn't). Consequently, the validity of a benchmark assessment program is informed not only by the correlations among benchmark assessments and statewide tests, but also by the accuracy of forecasted state classifications of standards mastery. Research conducted on the accuracy of forecasted classifications provides validity evidence for Galileo benchmark assessments. For example, one recent research study addressed data on the percentages of students correctly forecasted from Galileo district benchmark assessments to meet or not meet state standards as measured by their AIMS



performance. Three levels of forecasts were evaluated. The first was the forecast from Benchmark 1 to AIMS. The second was the forecast involving the combined mastery patterns in Benchmarks 1 and 2 and AIMS, and the third was the forecast including the combined mastery patterns for all three benchmark assessments and AIMS. The lowest levels of accuracy for the three forecasts were 84, 85, and 85 percent respectively. The highest accuracy levels were 93, 94, and 94 percent. The means were 89, 90, and 90 percent.

- **Benchmark Item Parameter Estimates.**

In addition to reliability and validity, an essential component of all Galileo benchmark assessments is establishing item parameter estimates of certified items in the ATI Item banks. ATI conducts this type of examination using Item Response Theory (IRT). IRT assumes that a student's response to a test item is determined by the student's ability and certain item parameters (i.e., characteristics of the item). For multiple-choice tests, ATI uses an IRT model that includes three item parameters: a discrimination parameter, a difficulty parameter, and a guessing parameter. The three parameters are described below.

- **Discrimination Parameter.**

Discrimination Parameter indicates the extent to which an item discriminates sharply between different levels of ability. Values approaching or exceeding 1.0 discriminate between levels of ability very well. Values close to 0.0 discriminate between different ability levels very poorly. The discrimination parameter indicates the relationship of the item to the underlying ability being measured divided by measurement error. Items with high discrimination values make a positive contribution to the test reliability.

- **Difficulty Parameter.**

Difficulty Parameter provides information on the relative difficulty of items in a benchmark assessment. In general, it is useful to construct benchmark assessments that include a broad range of difficulty levels. Benchmarks of this kind will be sensitive to a range of ability levels. Such benchmarks generally correlate higher with criterion measures (e.g., statewide assessments) than do benchmarks sensitive to a limited range of ability levels. Zero is the average, meaning that the item is about average level of difficulty. If item difficulty is above zero, the item is more difficult than the average. When item difficulty is negative, the item is less difficult than average.



#### o **Guessing Parameter.**

Guessing Parameter indicates the likelihood a student who does not know the answer to the question posed in a multiple-choice item will guess the correct answer. Given a multiple-choice item with four alternative choices, it would be reasonable to expect that the chances of guessing the correct answer would be about one in four, or .25. Sometimes this will be the case. However, sometimes the probability of guessing the correct answer will turn out to be lower than .25 and sometimes it will be higher than .25. Items that make it easy to guess the correct answer are less desirable than items that limit the likelihood of guessing the answer correctly.

Information regarding item parameter estimates for Galileo benchmark assessments is provided through an *Item Parameter Report* available in Galileo K-12 Online. Below is a sample *Item Parameter Report*.

<b>Item Parameter Report</b>			
<b>Test: 2005-06 Geometry Test 1</b>			
	<b>Discrimination</b>	<b>Difficulty</b>	<b>Guessing</b>
1. MHS-S4C1-01. Identify the attributes of special triangles. (isosceles, equilateral, right)	0.86	-1.08	0.13
2. MHS-S4C1-02. Identify the hierarchy of quadrilaterals.	0.62	-0.02	0.13
3. MHS-S4C1-06. Solve problems related to complementary, supplementary, or congruent angle concepts.	0.65	-1.14	0.13
4. MHS-S4C1-09. Solve problems using the triangle inequality property.	0.42	1.29	0.15
5. MHS-S4C1-11. Determine when triangles are congruent by applying SSS, ASA, AAS or SAS.	2.47	2.46	0.26
6. MHS-S4C1-13. Construct a triangle congruent to a given triangle.	1.09	-1.27	0.12
7. MHS-S4C3-01. Graph a quadratic equation with lead coefficient equal to one.	0.8	0.98	0.12
8. MHS-S4C3-02. Graph a linear equation in two variables.	1.14	0.34	0.21
9. MHS-S4C3-05. Determine the midpoint between two points in a coordinate system.	0.75	-0.36	0.13
10. MHS-S4C3-04. Determine the solution to a system of equations in two variables from a given graph.	1.57	2.19	0.24
11. MHS-S4C3-03. Graph a linear inequality in two variables.	0.68	1.59	0.17



## **B. A Look under the Hood: ATI Benchmark Item Development and Certification Procedures**

ATI's Benchmark Planner is used by districts to provide ATI Educational Management Services staff with information to help ensure that benchmark assessments built for the district meet district requests particularly as they relate to inclusion of specific standards, test length, and test delivery dates. The following information describes ATI's item development and certification procedures and is intended to help the reader understand these procedures that are used to ensure the quality of the items comprising ATI benchmark assessments.

### **1. Guiding Standards for Developing Items**

ATI's Assessment Department constructs and certifies new items for ATI item banks on a continuous basis. ATI's approach to item construction includes strict conformance to detailed **Item Specifications** for *consistent measurement* of standards. Continuous item development increases the variety of items available for inclusion in district assessments. This is essential not only to reflect the broad range of district needs, but also because standards are continually changing. As standards change, curriculum also changes. To keep pace with this rapidly changing educational landscape, ATI continually updates not only the content of its item banks, but also the psychometric analyses used to estimate item parameters including difficulty, discrimination, and guessing.

- ***Certified Items.***

Only certified items can become part of a test generated to meet the benchmark specifications of a district. ATI's certification procedures promote high standards for item quality, minimize the likelihood of errors in benchmark test items, and provide increased variety in item selection for districts. Item parameter estimates are saved to ATI item banks only for certified items following assessment with a significant number of students. The availability of item parameter estimates provides important information on item quality to be used in guiding test construction.

- ***Suggestions for New Items.***

ATI welcomes suggestions for new items. Such suggestions help to ensure that assessments will meet district needs. ATI currently has one of the largest item banks in the nation containing over sixty-thousand certified items. Moreover, as mentioned above, the banks are continuously growing. When making suggestions for new items, there are a number of considerations that the district may find useful. Suggestions for new items are likely to benefit the district most when they focus on the measurement of a specific skill or capability. If the district curriculum includes a particular capability that is not currently well represented in ATI item banks, suggesting the inclusion of additional items assessing the skill in question will be beneficial. Suggesting stylistic changes is generally less useful. ATI does develop items specifically aligned to the various styles of statewide assessments. Nonetheless, it is important to ensure that items on local benchmark and formative assessments not be limited to



a particular style. As items become closer in style to a statewide test, the danger that test outcomes will be contaminated by the phenomenon of “teaching to the test” increases. Benchmark and formative assessments are intended to measure the mastery of standards, not merely the ability to respond to items written in one particular style.

Finally, when making suggestions for new items, please keep in mind that in order to ensure that guiding standards for item construction, review, and certification are rigorously adhered to. ATI generally does not construct new items requested by individual districts during the formal planning, construction, and review of benchmark assessments for that district. This policy benefits the district in a number of ways. It maximizes the quality of items included on an assessment because the included items will already have been certified and typically used successfully on many assessments. It increases the likelihood that the assessment will be effective in forecasting performance on statewide assessments because the included items will be drawn from a pool previously used effectively in forecasting initiatives, and it eliminates scheduling uncertainties because it ensures that the planning and construction process can proceed following well established scheduling guidelines that have been proven to be effective for districts.

## **2. Item Specifications**

The item development process begins with the construction of ATI Item Specifications designed to guide item development for a particular standard. ATI Item Specifications include the following components:

- the standard that items conforming to the specification are designed to measure;
- a general description of the type of item to be covered by the specifications;
- the defining attributes of item components; and
- one or more sample items demonstrating the specification.

Item specifications serve as a consistent point of reference to guide and evaluate the overall quality of utility of each individual item. To ensure that specifications are always accessible during item development, ATI includes them in the online item development environment used by our Assessment Department. This is illustrated by the screenshots that follow.



Setup Assessment Curriculum Grades Student/Family Staff Reports

Test Planning | Test Construction | Bank Construction | Test Scheduling | Test Scoring | Printing

**Item Specification** District: K12 Online Test Item Development

Settings | Password | Tech Support | Site Map | Logout

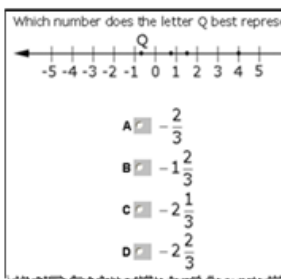
Subject: AZ-M08: 8th Grade Math  
 Standard: 8-1C1 UNDERSTAND AND APPLY NUMBERS, WAYS OF REPRESENTING NUMBERS, THE RELATIONSHIPS AMONG NUMBERS AND DIFFERENT NUMBER SYSTEMS  
 Objective: M08-81C1-01. Locate rational numbers on a number line.

Item Specification Title: number lines - identify a rational number

Font Size Color Symbols

**Description and Definitions:** This type provides a number line and requires students to identify specific points on the line.  
*Rational numbers* - numbers that can be expressed as the ratio of two integers.  
**Stimulus Attributes:** Graphic of a number line with labeled integers and letter labels on certain reference points.  
**Stem Attributes:** All items will provide the number line and ask students to identify a point or points on that line.  
**Response Attributes:** Each response is a point labeled on the number line, or a value identifying a given point on the number line.

**Prototype 01:**



Setup Assessment Curriculum Grades Student/Family Staff Reports

Test Planning | Test Construction | Bank Construction | Test Scheduling | Test Scoring | Printing

**Item Specification** District: K12 Online Test Item Development  
 School: TID 2004-05  
 Class: none

Settings | Password | Tech Support | Site Map | Logout

Subject: AZ-R08: 8th Grade Reading and Literature  
 Standard: 8-1C4 ACQUIRE AND USE NEW VOCABULARY IN RELEVANT CONTEXTS  
 Objective: R08-81C4-04. Determine the meaning of figurative language including similes, metaphors, personification, idioms, hyperbole, and technical language.

Item Specification Title: Type 01: Simile

Font Size Color Symbols

**Type One: Simile**

**Description:** Student is asked to identify the meaning of a simile.  
**Stimulus Attributes:** Grade-appropriate text that contains at least one example of simile.  
**Stem Attributes:** The student is asked to identify the correct meaning of the simile.  
**Response Attributes:** The responses are four sentences that describe possible meanings of the simile.

**Prototype:**

Read the sentence.

The students did poorly on the test because the directions were as clear as mud.

What does the simile "as clear as mud" mean?

A) The directions were easy to understand.  
 B) The directions were very lengthy.  
 C) The directions were very confusing.  
 D) The directions had not been included.

Correct answer: C



### 3. Item Review and Certification

Once an item is developed according to a specification, the item is subjected to an internal screening process designed to identify any problems that might be present in the item. The screening process involves looking closely at all aspects of the item including the fit of the specification to the standard, the fit of the item to the specification, the clarity and appropriateness of the language in the item components, the fairness of the content, the characteristics of the response options, the graphics and fonts in the item and the correct answer indication. If the item fails on any of the screening components, the item is returned for specified revision. The process is repeated until the item is “Accepted” and ready for certification, the final step in the item evaluation process. Certification involves a final item review in which alignment with standards and conformance of items to specifications guiding item development are again confirmed. The item certification process is completed online by education professionals in the Assessment Department.

The screenshot shows the 'Bank Editor' interface. At the top, it displays 'District: K12 Online Test Item Development', 'School: none', and 'Class: none'. There are navigation links for 'Settings | Password | Tech Support | Site Map | Logout'. Below this, there are dropdown menus for 'Library' (M06 District Assessment Bank), 'Test' (M06-S2C2-02C), and 'Specification' ([All Specifications]). A note instructs users to use the editor to edit text and to click the question number to edit in testbuilder. A 'Reset Formatting' button is present. A rich text editor toolbar shows 'Verdana', font size '3', and bold, italic, underline, list, and link icons. The main content area shows a question box for 'M06-S2C2-02C' with subject 'AZ-M06: 6th Grade Math', standard 'S2C2 UNDERSTAND AND APPLY THE BASIC CONCEPTS OF PROBABILITY', and objective 'M06-S2C2-02. Express probabilities of a single event as a decimal.' The question text is 'In a basket there are 10 purple markers, 9 blue markers and 6 green markers. What is the probability of pulling out a blue OR a green marker?' with four multiple-choice options: A) 0.24, B) 0.36, C) 0.40, and D) 0.60. A red arrow points to the question number '2.' which has a checked checkbox next to it. A callout box says 'This item has been certified.' A 'Save Changes' button is at the bottom.



## **C. Putting It All Together: Steps in Benchmark Assessment Planning, Construction, Review, and Delivery**

### **Step 1: Initial Benchmark Preparation Communications with ATI**

During the course of initial communications with ATI regarding your district's benchmark assessments, the following tasks need to be completed.

- The **Benchmark Contact Form** containing the names of the district contacts must be **completed** and **emailed to ATI Educational Management Services: EdMgtSvs@ati-online.com**.
- The District **Leader(s)** listed on the *Benchmark Contact Form* as **responsible for oversight** of benchmark planning, construction, and review begin the process by:
  - **Becoming familiar** with the **contents of this resource document**;
  - **Familiarizing district staff** involved in benchmark assessment development activities with the document contents; and
  - **Making phone contact with ATI Educational Management Services** via (520) 323-9033 or 1-800-367-4762.

During initial phone contact and in subsequent phone calls as needed, Educational Management Services staff will assist the district in the completion of the district's Benchmark Planner. The Benchmark Planner must be completed in order for ATI to begin the process of constructing your district benchmark assessments.

### **Step 2: Use of Benchmark Planner to Define Assessment Goals**

***Written instructions on navigation within the Benchmark Planner are in Section 2 of this document.***

The online Benchmark Planner makes it possible for your school district to provide ATI with specific input related to content area and the standards that are to be included in your local district benchmark assessments for each participating grade. The Planner helps to ensure that ATI constructed benchmark assessments are aligned with district selected state standards and local district benchmark goals. Specifically, your district uses the *Benchmark Planner* to communicate to ATI:

- a requested delivery date for each assessment;
- the number of district benchmark assessments planned;
- the specific standards to be measured on each benchmark;
- the number of items desired for measuring each standard on each assessment; and
- specific comments regarding your district benchmark needs.



During the course of completing Step 2, ATI Educational Management Services staff will be available to assist you with your questions related to completing the Benchmark Planner and to make recommendations as needed. As you complete your Benchmark Planner, it is important to keep the following in mind:

- **Assessment Delivery Date.**

It is important to keep in mind that the timeline for benchmark development starts 30 business days prior to the time of the **final delivery date** indicated on the Benchmark Planner. *The final delivery date **is the date the test is to be delivered to the district and this **is not** the date the district wants to administer the assessments.*** When selecting a delivery date for assessments that will be administered offline, districts need to be aware of how much time is needed to print the necessary number of test booklets at the district and any vacation days that will be occurring during that time frame. For example, if the district is administering an offline assessment on October 15<sup>th</sup> and it will take two weeks to print the assessments, October 1<sup>st</sup> should be selected as the final delivery date if district printers will be available during that full time period. When selecting a delivery date for assessments that will be delivered only online, the delivery date needs to take into consideration the time the district will need to schedule the assessment.

- **Number of Assessments per Year.**

Districts tend to prefer three to four benchmark assessments per year. Typically each of these assessments is designed to follow a period of instruction.

- **Standards Being Measured and Number of Items per Standard.**

An essential consideration in building reliable and informative benchmark assessments is the selection of the assessment content. Content selection is typically influenced by the district's approach to curriculum. Galileo technology allows the district to align benchmark assessments to district pacing guides or curriculum maps. Other curriculum related factors influencing the choice of standards to measure include:

- If the district's approach to curriculum includes provisions for revisiting previously taught skills, the district may wish to repeat selected standards over two or more benchmark assessments. For example, a curriculum may call for a spiral approach in which previously taught skills are revisited multiple times.
- If the curriculum is designed to be sensitive to students reflecting a broad range of abilities, the district may wish to include skills that have not yet been fully targeted for instruction. For example, a curriculum may support repeated assessments of the same standards and coverage of standards that while only being "introduced" are a topic of interest to the district for assessment purposes.



As mentioned above, another essential consideration in building reliable and informative benchmark assessments is test length. Tests with at least 40 items are considered important toward the goal of achieving reliability. A second length consideration is keeping the test short enough to be manageable for the students taking the test. Districts generally find that limiting Reading and Literature assessments to no more than 50 items and Math assessments to no more than 60 items is helpful in managing administration of assessments.

An additional consideration in building benchmark assessments is ensuring that the test reflects the full range of abilities that students evidence. Restrictions in the range of abilities assessed will reduce the magnitude of the relationship between the tests and other assessments including statewide tests. The consequences of range restrictions can be illustrated by considering the extreme case. If a test were so easy that every student got a perfect score, the correlation between that test and any other test would be zero.

The number of items selected per standard will need to vary depending upon the number of standards the district wishes to measure in a specific assessment. The larger the number of standards selected the smaller the number of items per standard will need to be if the test is going to meet the test length guidance provided above.

### ***Step 3: Confirmation that the Benchmark Planner is Complete and Obtaining Educational Management Services Feedback***

Once the online Benchmark Plan has been electronically submitted to ATI, Educational Management Services will send the District Leader(s) confirmation via phone and/or email. The confirmation will generally be followed by Educational Management Services contact that provides the district with feedback on its benchmark plans. The Educational Management Services review of the benchmark plans will be particularly attentive to:

- the length and content coverage in the planned assessments;
- assessment characteristics including the likelihood that the assessment can be administered within a reasonable amount of time.

Guidance from Educational Management Services staff with expertise in test development is provided to the district in order to help ensure that the district-designed benchmark assessments will have desired reliability, validity, and optimal utility for instructional planning, intervention, progress/outcome assessments and risk analysis related to student performance on state-wide tests.



#### **Step 4: Development of Benchmark Draft by ATI**

- Seven weeks from the specified delivery date for final versions of benchmark assessments, Educational Management Services will confirm with the District Leader(s) via phone and/or email that ATI is beginning the process of draft construction. At that time, **Educational Management Services will also confirm that the district has finalized its benchmark plans and remind the district that at six weeks from delivery no further changes in the Benchmark Planner can be accepted without changing the test delivery date. That means that no additional standards can be included once the test construction begins at six weeks.**
- Six weeks from the specified delivery date ATI staff begins construction of the benchmark assessment.

#### **Step 5: Guided Review of Benchmark Drafts**

***Written instructions on navigation within Galileo Online Test Review are in Section 3 of this document.***

Approximately four weeks from the specified delivery date, ATI will provide the district draft benchmark assessments for the district specified content areas (e.g., Math) and grade levels. Reviews of drafts are not required by ATI and are offered as an option for districts wishing to do reviews. For those districts choosing to review benchmark drafts, ATI provides specific guidelines for the District Leader(s) to implement during this process. See Section 3 of this document.

It is suggested that requests for changes in draft assessments be approached cautiously. When considering an item replacement, it is important to ensure that the replacement item does not adversely affect the range of difficulty desired for the assessment and that it contributes to the measurement of the standards selected for assessment in the benchmark plan. The items available for benchmark assessment have been reviewed and certified and are aligned with standards reflecting expected levels of student mastery. When a replacement is contemplated, it is important to ensure that the replacement item will increase the quality of the benchmark assessment as a measure of valued educational goals that it is reasonable to expect students to obtain.

#### ***Guidelines for District Review of Benchmark Drafts***

1. ***District Leaders supervise the review process and are the primary contact between the District and ATI Educational Management Services during the review.***



#### 2. ***District Leaders are responsible for the training of and guidance of District Reviewers.*** In this regard, please consider the following:

- a. **Providing reviewers with knowledge concerning item development, item certification, and benchmark construction will help to establish consistent criteria for test review, encourage review objectivity, and enhance the value of the review for the district.** ATI recommends that District Leaders request initial and final reviewers read and discuss this resource document prior to initiating the review process. Reviewers aware of the test construction, item construction and validating procedures have a knowledge base from which to operate during the test review process.
- b. Requesting changes in draft assessments should be approached cautiously. When requesting an item replacement, it is important to ensure that the replacement does not adversely affect the range of difficulty desired for the assessment and that it does contribute to the measurement of the standards selected for assessment in the district benchmark plan.

#### 3. ***District reviewers use Galileo online test review tools to provide feedback to ATI Educational Management Services.***

- a. During the review process, a district can accept the draft benchmark assessment as is or can request replacement of specific items (initial reviewers) or make item replacements (final reviewer only) in the draft.
- b. Final reviewers can make suggestions for item development in the comment box at the top of the Final Review interface. Please note carefully the material in Section 1.B.1 of this document for detailed guidance concerning requests for item development

#### 4. ***District personnel complete all steps within the Test Review process.***

- a. *Set the initial reviewers of the benchmark assessment.* These individuals will then be able to login to Galileo, see an electronic version of the benchmark assessment, and review the test. The initial reviews are submitted to the final reviewer for the District.
- b. *Set the final reviewer.* Only one person can be assigned this task for a specific assessment. This individual will look over staff comments and submit a final test review to ATI.
- c. *Complete initial reviews.* Those who were identified as initial reviewers will each complete a test review online.
- d. *Complete final review.* The final reviewer responsibilities include reviewing the work of the initial reviewers and making the final determination of content to be accepted as presented on the assessment draft and/or completing any item replacements. If the final reviewer has suggestions for items that the district wishes to have added to the item banks for future assessments, these suggestions can be included in the comment box on the Final Reviewer interface.



- e. *Save the final review.* ATI will then construct the final version of the test based on feedback from the final review.

#### **Step 6: Final Benchmark Assessment Delivered by ATI**

Following the interchange of district input and ATI recommendations, ATI produces a final benchmark assessment for administration according to the district's benchmark final delivery schedule. In order to maintain the integrity of the decisions reached between the district and ATI during the draft review process and ensure the timely delivery of benchmark assessments to the district, final benchmarks unlike drafts are not published for additional reviews and modifications.



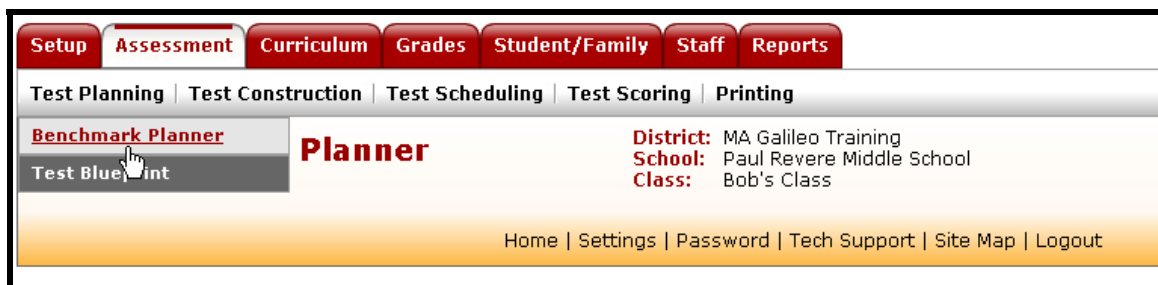
## 2. Navigation within Galileo Online Benchmark Planner

The ability to view the Benchmark Planner is available to all class, school and district level users. This allows all level of users to see a district's Benchmark Plan. The ability to make selections on the Benchmark Planner, though, is only available to district level users.

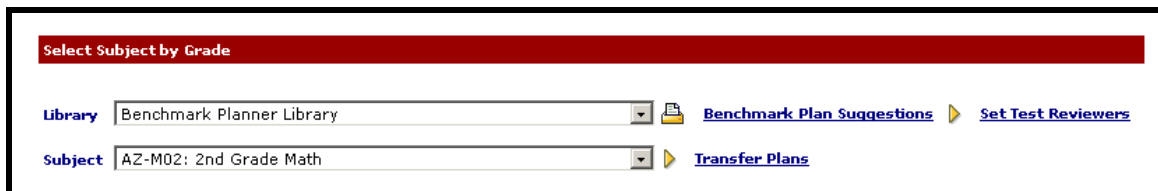
### A. The Process of Designing Benchmark Assessments

A district-level user will need to complete this process for each grade level they would like participating in the benchmark assessment planning.

1. When you first login, click on the **Assessment** link in the red menu bar at the top of the page.
2. Click on **Test Planning** in the gray menu bar at the top of your screen.
3. Click on the Benchmark Planner link.



4. After selecting Benchmark Planner, you will be taken to the Benchmark Planner page. On this page, there is a link that will provide you access to an electronic version of this document, Building Reliable and Valid Benchmark Assessments.
5. Select the **Library**.
6. Select a **Subject**. The screen will refresh.



7. Designate how many benchmarks your district will be administering —1, 2, 3, or 4.
8. Select the dates you would like each final version of the Benchmark Assessment delivered. Please note:
  - a. The delivery date of the final version is a date prior to the date of testing. The selected delivery date needs to allow time for printing tests that are to be administered offline and for scheduling tests that are to be delivered online.
  - b. The benchmark building process takes 6 weeks.



9. Select the number of items you would like for each objective:

Check the desired Benchmark Tests. Set the REQUESTED DELIVERY DATE for each selected Benchmark Test.

<input type="checkbox"/>	Benchmark Test 1	November	2	2005	Number of Items per Objective	1
<input type="checkbox"/>	Benchmark Test 2	November	3	2005	Number of Items per Objective	1
<input type="checkbox"/>	Benchmark Test 3	November	3	2005	Number of Items per Objective	1
<input type="checkbox"/>	Benchmark Test 4	November	3	2005	Number of Items per Objective	1

Save Dates

10. Click on the **Save Dates** button.

11. After you save the dates, a comment box will appear. If there are any special requests, this is where to place them.

Select objectives to be included in each Benchmark Test.

Save Plan

Comments

On assessment #1, please put Learning Standards 3.N.1,2,3 at the end.

12. Next, find each standard you would like to assess and place a checkmark next to the test on which you would like a test item measuring this standard to appear.

DATA ANALYSIS, STATISTICS, AND PROBABILITY STRAND

6.D.1 Describe and compare data sets using the concepts of median, mean, mode, maximum and minimum, and range.

Test 1 1  Test 2 1  Test 3 1  Test 4 1  Not Tested

6.D.2 Construct and interpret stem-and-leaf plots, line plots, and circle graphs.

Test 1 1  Test 2 1  Test 3 1  Test 4 1  Not Tested

6.D.3 Use tree diagrams and other models (e.g., lists and tables) to represent possible or actual outcomes of trials. Analyze the outcomes.

Test 1 5  Test 2 1  Test 3 1  Test 4 1  Not Tested

6.D.4 Predict the probability of outcomes of simple experiments (e.g., tossing a coin, rolling a die) and test the predictions. Use appropriate ratios between 0 and 1 to represent the probability of the outcome and associate the probability with the likelihood of the event.

Test 1 4  Test 2 1  Test 3 1  Test 4 1  Not Tested

Save Plan

13. Select the number of test items you would like assessing this objective. While you did this generally in step 9, you may adjust this as needed on specific objectives.

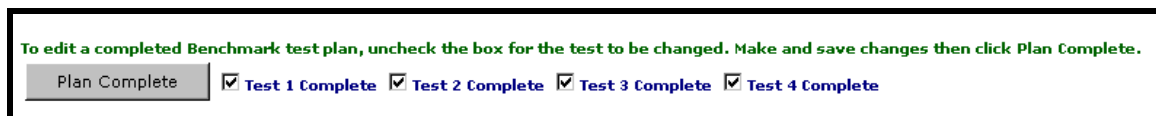
14. At any point in the process, you can leave and come back to it as long as you click on the **Save Plan** button.



- After you have saved the plan, you can confirm the size of the test by referring to the Total Test Items bar that will appear underneath the **Save Dates** button. At this point, if the number is too low or too high, you can go back and adjust any of your selections to modify the size of the test. After modifying your selection, hit the **Save Plan** button and the adjusted tally will appear on the Total Test Items bar.



- Once you have saved your plans and would like to submit them to ATI, click on the **Plan Complete** button. No assessments can be built unless the Plan Complete button has been selected.
- After you select Plan Complete, the plan will be locked. Next to the Plan complete button are a series of checked boxes that will confirm which assessment plans are locked.



- To unlock any particular assessment and to make a change, you need to uncheck the assessment in question and wait for the screen to refresh before you can make any changes. Once you have made the change, save the plan and then hit Plan Complete to again signal ATI that the plan is ready.

Note: No changes can be made if the delivery date on that particular assessment is within six weeks. The only way to make a change if the delivery date is within six weeks is to push the delivery date back.

## B. The Process of Creating a Benchmark Assessment

Following is a timeline that describes the assessment-development process. The timeline is controlled by the district-set final delivery date. Remember: The final delivery date is the date the district wishes to *receive* the final test and it takes into consideration the time needed by the district to print tests to be administered offline and/or the time needed to schedule tests to be administered online.

- Seven Weeks Out:** Educational Management Services will confirm with the district that the district has finalized its benchmark plans.



- **Six Weeks Out:** This is when ATI's Assessment Development Department starts constructing the assessments. For this reason, this is the deadline of when the Benchmark

Planner must be completed. At this time, the Benchmark Planner cannot be changed without extending the delivery date for the assessment. When the Benchmark Planner is completed close to the time that draft test construction begins, the district necessarily foregoes the benefit of Educational Management Services review unless the district wishes to extend the delivery date for the assessment.

Please Note: ***Construction of the test draft cannot begin without the completion of the Benchmark Planner.***

- **Four Weeks Out:** Two weeks after the assessment construction starts (four weeks out from final delivery date), a draft version of the assessment is delivered to the district to be reviewed online via Test Review.
  - **District Reviews:** ATI has an easy-to-use online Test Review tool that provides for different levels of review in the district and allows the final reviewer to send feedback to ATI's Test Development team.
- **Two Weeks Out:** It is expected that within two weeks of receipt of assessment drafts, the draft reviews will be complete and the final reviews will be saved and available to ATI with two weeks remaining before the specified delivery date.
- **Delivery Date:** After considering final reviewer comments and suggestions, the ATI assessment team prepares and delivers the final version of the assessment. A delivered assessment appears in the district's benchmark library and is ready to be scheduled.



## 3. Navigation within Galileo Online Test Review

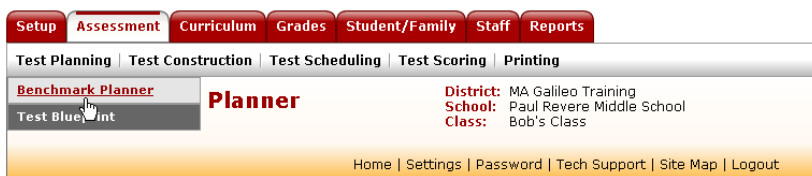
### Summary of Navigation Steps within the Test Review Process

- 1.) Set the initial reviewers of the benchmark.
- 2.) Set the final reviewer.
- 3.) Perform initial reviews.
- 4.) Perform final review.
- 5.) Save the final test review.

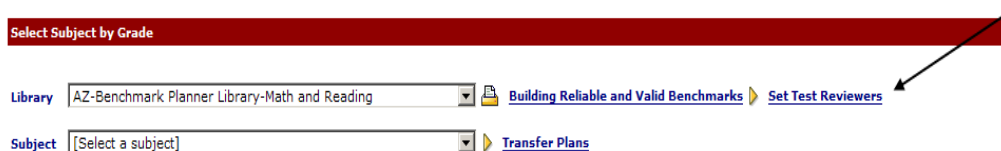
### A. Set Initial and Final Reviewers

You will first set the initial reviewers and save. Then you will repeat these steps to set up the final reviewer.

1. When you first login, click on the **Assessment** link in the red menu bar at the top of the page.
2. Click on **Test Planning** in the gray menu bar at the top of your screen.
3. Click on the Benchmark Planner link.



4. Click on the Set Test Reviewers link.



5. Under the **Review Level** heading select which type of reviewers you are about to set.

Select an item review level then select users and click save.  
To remove reviewers, select Remove then select users and click save.

Review Level	
<input type="radio"/> Final	<input checked="" type="radio"/> Initial
<input type="radio"/> Remove	

Select Users	
<input checked="" type="checkbox"/> Acronym, Amanda	<input type="checkbox"/> Airplane, Adrienne
<input type="checkbox"/> Andiron, Adam	<input type="checkbox"/> Anthropology, Allan



- Place a checkmark by the user you would like to be a reviewer. Please Note: To remove a person from the reviewer list, select the **Remove** option and then click on the name to be removed.

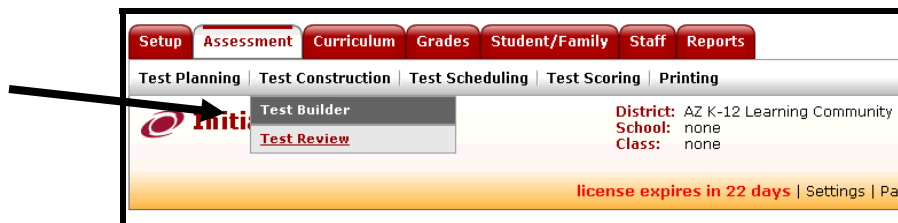
<input type="checkbox"/> Zithers, Zach	<input type="checkbox"/> Zoology, Zoe
<a href="#">Select All</a>   <a href="#">Deselect All</a>	
<input type="button" value="Save"/>	

- Click on the **Save** button. The **Save** button is also clicked when removing a reviewer from the list.
- Note that an individual cannot be an initial and a final reviewer at the same time.

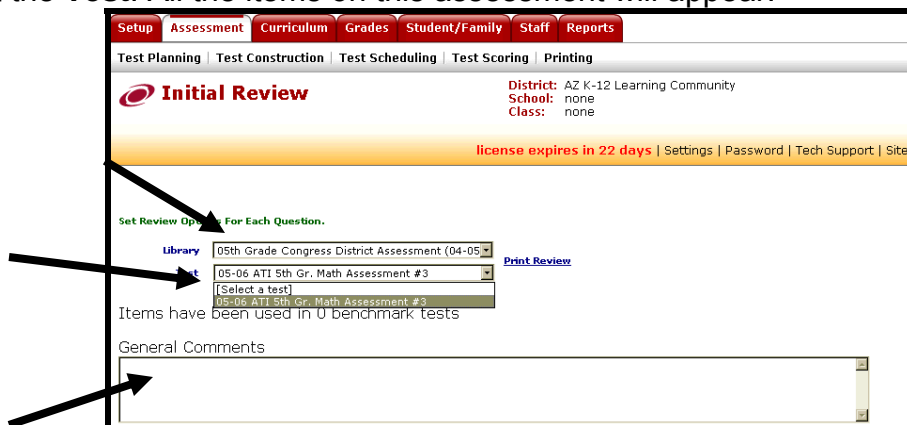
## B. Initial Reviews

Once a user has been set up as a reviewer, he or she will login to Galileo and have access to the Test Review tools. To access these tools and begin the reviewing process, he or she will follow these steps:

- When you first login, click on the **Assessment** link in the red menu bar at the top of the page.
- Click on **Test Construction** in the gray menu bar at the top of your screen.
- Click on the Test Review link.



- Select the **Library** that houses the benchmark assessment to be reviewed.
- Select the **Test**. All the items on this assessment will appear.






- At the top of the review, ATI has provided a box for the reviewer to make general comments about the assessment and make suggestions for future item development. Use of this box is optional.
- Read each test question and select a choice from the following: “Not Reviewed, Accept, Replace”. Please Note: It is suggested that requests for changes in draft assessments be approached cautiously. When suggesting an item replacement, it is important to ensure that the replacement item does not adversely affect the range of difficulty desired for the assessment and that it contributes to the measurement of the standards selected for assessment in the benchmark plan. When a replacement is contemplated, it is important to ensure that the replacement item will increase the quality of the benchmark assessment as a measure of valued educational goals that it is reasonable to expect students to obtain.

**Subject:** AZ-M05: 5th Grade Math  
**Standard:** S4C4 UNDERSTAND AND APPLY APPROPRIATE UNITS OF MEASURE, MEASUREMENT TECHNIQUES, AND FORMULAS TO DETERMINE MEASUREMENTS  
**Objective:** M05-S4C4-02. Draw two-dimensional figures to specifications using the appropriate tools. (e.g., draw a circle with a 2 inch radius)

29.  
Which of the following is true about the circle?



A  Its circumference is 15.  
B  Its radius is 15.  
C  Its diameter is 15.  
D  Its area is 15.

Initial Item Review

Not Reviewed  
 Accept  
 Replace (please comment)

Comments (please be specific)

- Please note: if selecting “Replace,” please provide the Final Reviewer clear comments concerning the request.

**Subject:** AZ-M05: 5th Grade Math  
**Standard:** S4C1 ANALYZE THE ATTRIBUTES AND PROPERTIES OF TWO AND THREE DIMENSIONAL SHAPES AND DEVELOP MATHEMATICAL ARGUMENTS ABOUT THEIR RELATIONSHIPS  
**Objective:** M05-S4C1-09. Identify the diameter, radius and circumference of a circle.

25.  
The radius is what part of the circle?

A  the distance across the circle and through the center  
B  the distance from the center of a circle to the circle itself  
C  the space on the inside of the circle  
D  the distance around the circle

Initial Item Review

Not Reviewed  
 Accept  
 Replace (please comment)

Comments (please be specific)

Please replace this item with on that that assesses knowledge of diameter instead of radius.



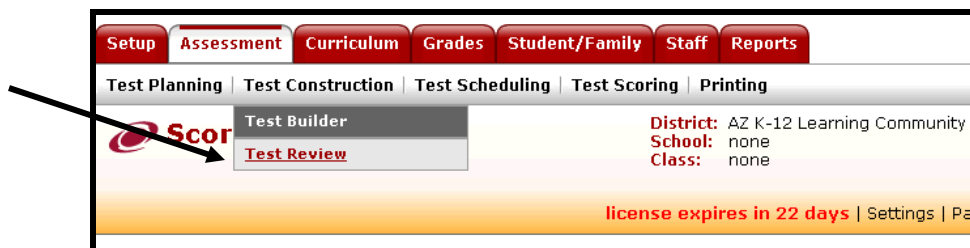
- When all items have been reviewed click on the **Save Review** button.

Once all initial reviewers have saved their input, the final reviewer may now login and do a final review of the items.

## C. Final Review

*The final reviewer considers the input of the initial reviewers*

- When you first login, click on the **Assessment** link in the red menu bar at the top of the page.
- Click on **Test Construction** in the gray menu bar at the top of your screen.
- Click on the Test Review link.



- Click on the View Initial Reviews link.

- Select the **Library** that houses the benchmark assessment under review.
- Select the **Test** under review.



7. Select the Reviewers whose comments you would like to see. You may view all comments by selecting the [All reviewers] option.

#### Set Item Review Options For Each Question.

Library: SA District Test

Test: ATI 5th Grade Math Test 1

Reviewer: [All Reviewers]

[All Reviewers]  
Acronym, Amanda  
Andiron, Adam

8. Determine how you would like to view the reviewed items.

Items

Show all reviewed items

Show reviewed items excluding accepted items

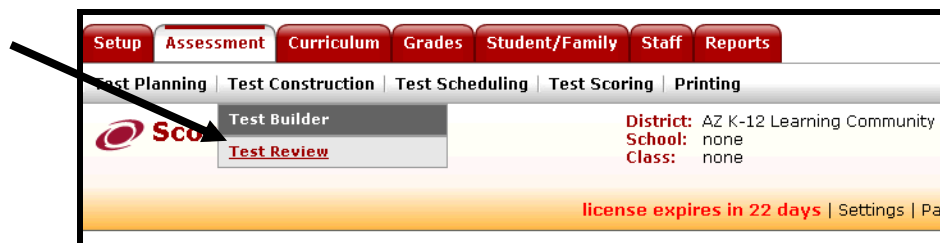
Show only rejected items

Show all items

9. The items will appear. Look over the input of initial reviewer(s).
10. Once you have considered this input you may now do a final review to submit to ATI.

#### ***The Final Reviewer will now complete the last and final review to submit to ATI.***

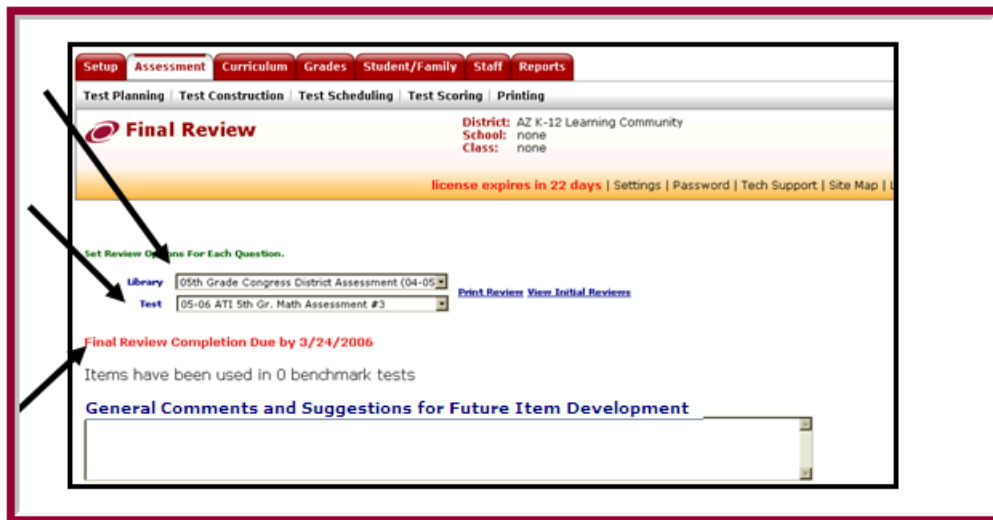
1. When you first login, click on the **Assessment** tab in the red menu bar at the top of the page.
2. Click on **Test Construction** in the gray menu bar at the top of your screen.
3. Click on the Test Review link.



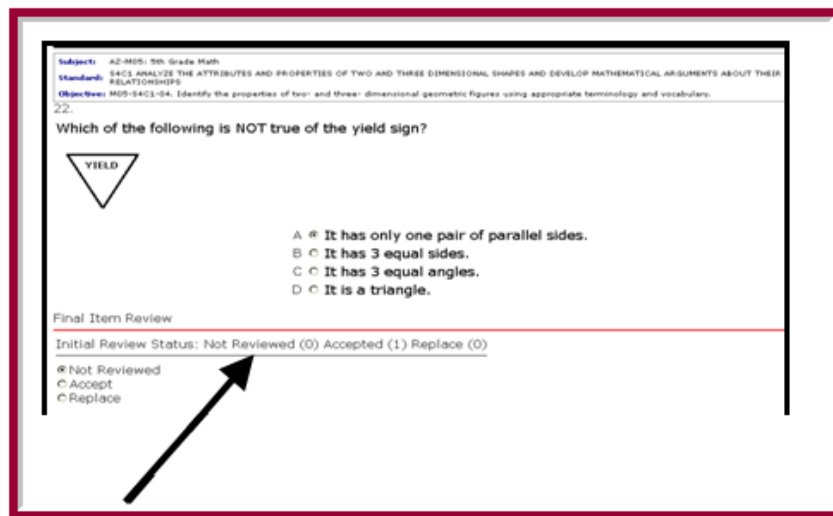
4. Select the **Library** that houses the benchmark assessment under review.



5. Select the **Test** under review. The complete assessment with each of its items will appear.



6. Please note that a Final Review Completion Due date will appear in red. The due date will usually be one week after the drafts have been delivered. If you want more time, that can be arranged, but it might impact the final delivery date.
7. At the top of the review, ATI has provided a box for the final reviewer to make general comments about the assessment and to provide suggestions for future item development. Use of this box is optional.



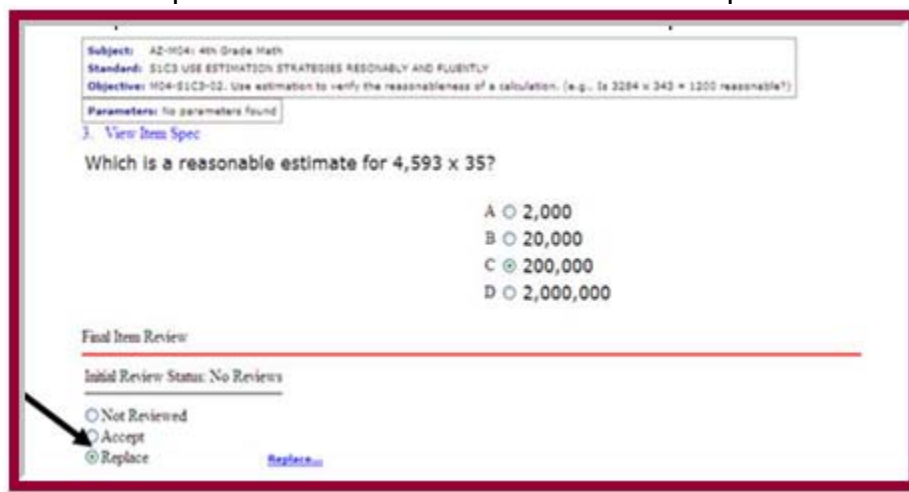
8. Please note that the screen will display, for the final reviewer only, the number of staff members who did not review this item, accepted the item, and asked that the item be replaced. This can help the final reviewer make his or her determination concerning the item.



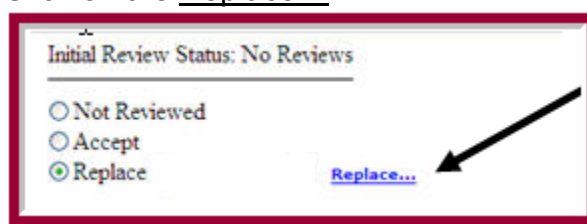
9. Select a choice from the following: “Not Reviewed, Accept, Replace.” Please Note: It is suggested that item replacements in draft assessments be approached cautiously. When considering an item replacement, it is important to ensure that the replacement item does not adversely affect the range of difficulty desired for the assessment and that it contributes to the measurement of the standards selected for assessment in the benchmark plan. When a replacement is contemplated, it is important to ensure that the replacement item will increase the quality of the benchmark assessment as a measure of valued educational goals that it is reasonable to expect students to obtain.

10. If you selected the “Replace” choice, questions may be replaced by following the steps below:

a. Locate the question that has been marked with “Replace.”



b. Click on the Replace... link.



c. All the questions in the bank that can replace the current question are displayed.



- d. Click on the radio button beside the question you would like to appear on the test. This question will replace the current question.

Which answer is a reasonable estimate for  $275 \div 7$ ?

A. 20  
 B. 30  
 C. 40  
 D. 50

Which answer is a reasonable estimate for  $178 \div 6$ ?

A. 30  
 B. 40  
 C. 50  
 D. 60

Replace Question

- e. Click on the **Replace Question** button. Please note, when doing replacements for reading assessments, the replacement question appears next to the text with which the question is referring. This changes the ordering of the questions.
- f. Change the question status to “Accept” and click **Save Review**.

Final Item Review

Initial Review Status: Not Reviewed (0) Accepted (1) Replace (0)

Not Reviewed  
 Accept  
 Replace

Save Review Delete Review

- g. Repeat these steps for each replacement question identified during the review process.
11. When you are finished with the review, select the “**Review Complete**” located at the bottom of the review. This will electronically **send a notice** to ATI that the review is now ready for our attention.

Click the button below to notify Assessment Technology, Inc. that your review is complete.

Review Complete

12. When **all** of the district’s final reviews for the Benchmark assessments under consideration are complete, contact ATI’s Education Management Services Department (1-800-367-4762 or e-mail [EdMgtSvs@ati-online.com](mailto:EdMgtSvs@ati-online.com)). When ATI knows that all of the final reviews are complete, ATI can proceed with the publication process as detailed below.



When the final review for an assessment is received and any interchanges between the District Leader(s) and ATI are completed concerning the content of the review, ATI will produce a final benchmark assessment to be delivered according to the district's benchmark final delivery schedule. In order to maintain the integrity of the decisions reached between the district and ATI during the draft review process and to ensure the timely delivery of benchmark assessments to the district, final benchmarks are submitted directly to the district for administration.