Achieve, Inc.

MARYLAND

TECHNICAL DOCUMENTATION FOR ACHIEVE ALIGNMENT STUDY: GEOMETRY RFP#3200770

FEBRUARY 2003

Table of Contents

Maryland State Department of Education

Maryland High School Assessment, May 2002
Goal 2: Geometry, Measurement, and Reasoning
Alignment of Assessment to Maryland's High School Core Learning Goal 2

Background

Achieve, Inc. contracted with the Maryland State Department of Education (MSDE) to carry out an external alignment evaluation of the state's high school geometry assessment to Maryland's High School Core Learning Goals—and in particular to High School Core Learning Goal 2, which defines student expectations with respect to geometry, measurement, and reasoning. MSDE seeks to provide the United States Department of Education (USDE) with information it needs to judge Maryland in compliance with the federal requirements of the No Child Left Behind Act (NCLB) of 2001 and the Elementary and Secondary Education Act (ESEA) of 1994. One important element required for compliance is an assessment for high school mathematics that aligns with the state's student expectations, as exemplified by the Maryland High School Core Learning Goals.

For purposes of this alignment analysis, a team of trained reviewers—experienced with mathematics teaching, learning, and assessment at the high school level—analyzed Form T of the Maryland High School Assessment (May 2002) for geometry. This team of reviewers was led by Achieve's senior associate for mathematics and included two additional experienced reviewers commissioned by Achieve. Two MSDE staff and a mathematics supervisor from a Maryland school district were provided training in use of Achieve's alignment protocol and also participated in the review to build internal capacity for the state in developing standards and assessments.

Maryland's high school geometry assessment was examined using Achieve's protocol, with reviewers systematically comparing test items to the indicators specified within Maryland's High School Core Learning Goal 2. Prior to meeting as a group to conduct the alignment analysis, individual team members had reviewed the Core Learning Goals and had taken the assessment themselves.

Maryland's High School Geometry Assessment

Maryland's high school geometry assessment consists of a variety of item types: selected response, student-produced response, brief constructed response, and extended constructed response. Selected-response items are multiple-choice items—presenting students with four item distracters from which to choose. Items that require student-produced responses call upon students to solve a problem and to grid in the answer on a response grid. Both the brief and extended constructed-response items require students to write—rather than select or bubble in—a response.

While the test consists of 47 items, only the first 37 items are operational items—i.e., items that will count toward a student score. The final 10 items of the test are field test items and were, therefore, not analyzed using Achieve's protocol. The 37 operational items examined present a solid mixture of items types---26 selected-response items, 6 student-produced response items, 2 brief constructed-response items, and 3 extended constructed-response items. The selected-response and student-produced response items are each worth 1 point, while the brief constructed-response items are worth 3 points and the extended constructed-response items are worth 4 points.

Maryland's high school geometry assessment is administered on a single day, and takes approximately 3 hours to administer, including time for students to listen as directions are read to them and to take a brief break between Sessions 1 and 2. Seventy-five minutes of testing time is allocated for each testing session. Different forms of the assessment (identifiable by different colors and letters on the cover) are given to students in each classroom.

Consistent with instructional practice in Maryland, students taking the geometry assessment are provided with a variety of tools including: a straightedge ruler with both metric and U.S. customary units, graph paper, a protractor, a calculator, tools for geometric construction, and a Formula Reference Sheet that also includes rubrics for the brief and extended constructed-response items and cues for students. The construction tools may vary by classroom but it is the expectation that students have access to the tools they use during regular classroom instruction. Such tools can include any or all of the following: straightedge, compass, patty paper, Mira™, and mirrors. With respect to calculators, students are expected to use what they typically use during classroom instruction. A scientific calculator with trigonometric functions is sufficient.

Methodology: Achieve's Protocol in Brief

Achieve's protocol considers five criteria to be central in determining the degree of alignment between an assessment and a state's student standards.

- **Confirmation of the test blueprint:** The test blueprint shows the intended match between each test item and the state's standards. Reviewers ask if the assignment of an item to a standard or indicator is justified, and make a "yes" or "no" judgment. This step is important since test blueprints are typically used by states as the basis for their score reports.

  Content centrality: This criterion analyzes the match between the content of each test question and the content of the related student expectation by examining the degree or quality of the match. Reviewers assign each item to one of four categories based on the degee of alignment. A score of 2 indicates clear consistency, a score of 1a indicates that the standard or indicator is not specific enough, a score of 1b indicates that an item assesses part of an objective, and a score of 0 identifies a lack of consistency between the test item and the standard it is purported to assess.

- **Performance centrality:** This criterion focuses on the degree of match between the type of performance (cognitive demand) presented by each test item and the type of performance described by the related standard. As with the criterion of content centrality, reviewers assign each item to one of four categories based on the degree of alignment with respect to performance demands. The categories are identical to those used for content centrality but are applied, in this case, to the degree of alignment with respect to performance expectations, rather than content.
- **Challenge:** This criterion is applied to a set of items to determine whether doing well on these items requires students to master challenging subject matter. Reviewers consider two aspects of challenge, at the item level, to determine whether sets of test items are sufficiently challenging: *source of challenge* and *level of demand.*

  Source of challenge attempts to uncover whether test items are difficult because of the knowledge and skills they target, or for other reasons not related to the subject matter, such as relying unfairly on students' background knowledge or the inclusion of unclear or inaccurate graphics in a test item. Reviewers rate each item as having an appropriate or inappropriate source of challenge. Any item judged to have an inappropriate source of challenge is not examined for level of demand.

  Level of demand focuses on the type and level of thinking and reasoning required by the student on a particular item. A Level 1 (recall) item requires students to recall information such as a fact, definition, term, or simple procedure, perform a simple algorithm, or apply a formula. A Level 2 (skill/concept) item calls for the engagement of some mental processing beyond a habitual response, with students required to make some decisions as to how to approach a problem. Level 3 (strategic thinking) items require students to reason, plan, or use evidence, and Level 4 (extended thinking) items require complex reasoning, planning, developing, and thinking---most likely over an extended period of time, which is typically not available in large-scale state assessments.

  Balance and range: *Balance* compares the emphasis of content supplied by an item set to the emphasis of content described by the standards. In addition to evaluating alignment, reviewers also judge whether the set of items emphasizes the more important aspects of content. Critique is provided in the form of a succinct summary of the balance of each item set. *Range*, on the other hand, is a measure of the coverage or breadth (the numerical proportion of all content addressed) of a test with respect to a state's standards. No one state assessment can measure the full range of knowledge and skills required by the state standards, but evaluation of both balance and range provides states with qualitative and quantitative information about the choices they are making.

Map of Test Items to the Maryland High School Core Learning Goals

The MSDE provided Achieve with a mapping of the 37 operational items on Form T of the high school geometry test to indicators from High School Core Learning Goal 2 (Geometry, Measurement, and Reasoning). This state-provided map was judged by Achieve's lead reviewer to be accurate and was also confirmed by the review team. It was, therefore, used as the basis for the alignment analysis. All items—with the exception of the three extended constructed-response items—are mapped to just one indicator; each of the extended constructed-response items maps to two indicators. In two instances (for Items 5 and 31), reviewers determined that one of the mappings was more prominent than the other, meaning that these two items are assigned a primary and a secondary map. Ratings are provided only for the primary mapping. In the case of Item 15—the third extended constructed-response item—reviewers judged that the item assesses two indicators in more or less equal proportions, so this item is assigned two primary mappings. Two sets of ratings were made for Item 15—one set with respect to Indicator 2.2.1 and one set with respect to Indicator 2.2.3. However, for purposes of summary statistics, the item is only counted once, with half of the weight of the item's rating attributed to each indicator.

Detailed Findings: Alignment of Maryland's High School Assessment in Geometry to High School Core Learning Goal 2

All of the 37 operational items on the Maryland High School Assessment for geometry map to indicators in High School Core Learning Goal 2. For purposes of this report, summary data and analyses will be reported by criterion, based on Achieve's Assessment-to-Standards Alignment Protocol.

Content Centrality

Maryland's high school geometry test received exemplary scores for content centrality. All 37 operational items received ratings of 2 for content centrality, indicating that they clearly and consistently assess content as specified in High School Core Learning Goal 2.

This speaks well not only for the test but also for the indicators in this Core Learning Goal—since a score of 2 is possible only when a student expectation is clear and specific and the test item clearly measures the content defined in that expectation. The indicators in Maryland's High School Core Learning Goals tend to be clearly and concisely, yet sometimes broadly, stated. However, each indicator is accompanied by assessment limits and skills statements that help to clarify the intent of the indicator. For example, while Indicator 2.1.1 addresses "properties of geometric figures," the assessment limits accompanying this indicator enumerate which properties, relationships, and geometric models students are expected to understand. This presentation format allowed reviewers to award an item such as Item 21—which assesses students' understanding of angle relationships with parallel lines—with a content centrality score of 2. Another formatting or language convention decision made when the High School Core Learning Goals were written also impacted reviewers as they rated items relative to the indicators. A number

of indicators use "and/or" notation when referencing content students are expected to know. For example, Indicator 2.3.2 references "perimeter, circumference, area, volume, and/or surface area." While it is unlikely that any one item on a large-scale assessment will address all of these constructs, the "and/or" notation allowed reviewers to award a content centrality score of 2 to an item such as Item 36, which addresses area. A simple change in language—to "and" as opposed to "and/or"—would have resulted in such an item receiving a rating of 1b, indicating that the item assesses only part of, and not necessarily the central part, of a compound standard. Further discussion about the test's ability to provide a balanced assessment of these compound expectations is provided later in this report.

Performance Centrality

Similarly, Maryland's high school geometry test received exemplary scores for performance centrality, with all 37 operational items receiving ratings of 2 for performance centrality. As was also the case for content centrality, the assessment limits and skill statements serve to clarify the intent of the indicators with respect to the performances expected of students. For example, Indicator 2.1.4 indicates that—among other things—students are expected to "validate properties of geometric figures using appropriate tools and technology." The assessment limits serve to clarify what validation of properties entails—specifically, in this case, the justification of solutions using definitions, mathematical principles, and/or measurement. Such clarification led reviewers to assign a rating of 2 for performance centrality when some may have advocated for a rating of 1a, without such clarification. This same indicator also provides student expectations with respect to constructing and drawing geometric figures. With respect to these student performances, the assessment limits also give parameters as to what kinds of tools and strategies are appropriate for students to use as they draw or construct geometric figures—making the performance expectation quite clear.

The use of "and/or" notation is perhaps even more evident with respect to student performance expectations than it was with respect to content. It is this notation that is largely responsible for 100% of the operational items on the geometry test receiving performance centrality scores of 2. For example, Indicator 2.3.2 indicates that students are expected to be able to "estimate, calculate, and/or compare" such attributes as perimeter, circumference, area, volume, and/or surface area. Again, it is unlikely that any one item on a state test will require students to engage in all three of these performances. Item 30, for example, requires students to calculate volume; it does not require them to estimate or compare. Yet this item—because of the "and/or" notation of the indicator—was able to receive a performance centrality rating of 2. If the conjunction connecting these performances had been "and," rather than "and/or," the item would have received a rating of 1b for performance centrality.

The fact that Maryland's geometry test includes a variety of items types helps ensure that students have the opportunity to demonstrate all of the performances defined in High School Core Learning Goal 2. Even though the "and/or" notation in the state's indicators makes it possible to achieve a performance centrality rating of 2 without an item necessarily calling upon students to demonstrate all of the performances enumerated in an

indicator, some indicators emphasize skills that can only be demonstrated by responding to a constructed-response item. For example, Indicator 2.1.4 defines three performance expectations, albeit joined by the conjunction "and/or." Each of these three performance expectations—constructing, drawing, and validating—calls upon skills that students cannot adequately display in an assessment composed solely of selected-response items. The mixture of item types is, therefore, a factor in the exemplary scores the Maryland geometry test received with respect to performance centrality. Later sections of this report on balance and level of challenge will address this issue further.

Source of Challenge

Items on Maryland's geometry assessment scored well for source of challenge. Thirty-six of the 37 operational items—or 97% of the test's items—received a score of 1, indicating that reviewers agree that the source of challenge is appropriate. These items appear to be fairly constructed and are not "trick" questions. The level of challenge comes from the content and the type of performance targeted. Only one item (Item 18)—mapping to Indicator 2.1.2—scored 0, indicating that reviewers agreed this item has an inappropriate source of challenge.

Item 18 requires students to use a graph to identify which line on the graph has a specified slope. While there does not appear to be a technical flaw with the item or the graph, reviews found the graph to be overly "busy" and hence potentially confusing to students—even students who have a strong understanding of slope. Reviewers expressed concern that the "cluttered" appearance of the graph—with four lines all labeled and portrayed in the first quadrant of the same graph—could prevent some students from selecting the correct response. They suggested that each line could be placed in a separate graph or in separate quadrants of the same graph to avoid this problem. A P-value of 0.43 confirms that students had more difficulty with this item than reviewers would have expected, given that slope is a concept students should have mastered by the time they complete coursework including basic concepts of high school algebra and geometry.

While reviewers did not identify any other items as having an inappropriate source of challenge, they did have considerable discussion about Item 31. This extended constructed-response item requires students to use geometric construction to solve a real-world problem, to explain the steps they used in the construction, and to mathematically justify their solution to the problem. The concern about this item is that students need to know to use angle bisectors to solve the problem. If they do not realize this, reviewers were concerned that there is no other entry point into the problem, meaning that students would likely miss all 4 of the points assigned to this item. Reviewers did conclude, however, that although the item is difficult, there is no inappropriate source of challenge. A low P-value of 0.34 and an omit rate of 0.34 confirm that this is indeed a difficult item for students.

Level of Demand

Item 18—which was judged to have an inappropriate source of challenge, as described above—was not scorable for level of demand using Achieve's protocol. Of the 36

scorable, operational items on the geometry test, the majority received ratings of Level 1 (recall) or Level 2 (skill/concept). Twenty items (or almost 56% of the items) were rated as Level 1, the least cognitively demanding type of item. About 38% of the items received Level 2 ratings.

The summary data charts provided with this report reflect an item count of 13.5 items receiving a score of Level 2 and 2.5 items (or almost 7% of the items) receiving a rating of Level 3 (strategic thinking). A fractional item count is possible since reviewers provided two sets of ratings for Item 15, one of the test's three extended constructed-response items. This was the only item identified by reviewers as having two primary mappings. Since Item 15 received identical scores for content centrality, performance centrality, and source of challenge, regardless of its mapping, fractional parts of items are not reflected in the summary data for these criteria. However, reviewers assigned different ratings for level of demand, rating Item 15 as Level 2 with respect to Indicator 2.2.1 and Level 3 with respect to Indicator 2.2.3. This particular item requires students to perform two discrete tasks. The one deemed to be Level 2 requires students to identify similar triangles in a real-world problem and then apply proportionality to arrive at a solution. The more cognitively demanding part of Item 15 requires students to construct a geometric proof, with no scaffolding provided. Hence, reviewers deemed a Level 3 rating to be appropriate, with respect to the mapping to Indicator 2.2.3.

The three items receiving scores of 3—either partially or in their entirety--for level of demand are all constructed-response items. Items 15 and 31 are both extended constructed-response items, while Item 10 is a brief constructed-response item. No selected-response or student-produced response items received a rating above Level 2. No items on Maryland's geometry test were scored as Level 4 (extended thinking), but this is to be expected on a large-scale, on-demand assessment that is administered in a specified time period under standardized conditions. It is rare indeed that Level 4 items are included on such tests, given the time constraints. It is possible—and even desirable, however—for a large scale assessment such as Maryland's to contain a higher proportion of Level 3 items. One of the challenges facing Maryland as they review and revise their high school geometry test is to include more Level 3 selected-response and student-produced response items.

Level of Challenge

In addition to scoring individual test items according to level of demand, reviewers were asked to evaluate the level of challenge posed by the sets of items mapping to the three Expectations in Maryland's High School Core Learning Goal 2. The three Expectations are as follows:

- Expectation 2.1: The student will represent and analyze two- and three-dimensional figures using tools and technology when appropriate.
- Expectation 2.2: The student will apply geometric properties and relationships to solve problems using tools and technology when appropriate.
- Expectation 2.3: The student will apply concepts of measurement using tools and technology when appropriate.

The reviewers' objective was to determine whether each of these sets of items spans the level of demand expressed by the indicators for that particular Expectation and whether the item set is appropriate for high school students. An ideal item set would have items ranging from simple through complex.

As discussed in the previous section, Maryland's geometry assessment includes items that span the levels of demand defined by Achieve's protocol. While the level of challenge of the assessment as a whole is enhanced by the inclusion of brief and extended constructed-response items and items requiring student-produced (i.e., gridded) responses, the overall level of challenge of the assessment could be further enhanced by including more demanding selected-response items. However, the test does a good job of assessing the range of content and performance expectations defined in High School Core Learning Goal 2 and is appropriate for high school students. In particular, the brief and extended constructed-response items provide students with the opportunity to be able to construct geometric figures and to explain and justify their answers—performances that are not possible in the context of a test consisting solely of selected-response items. Commentary specific to each of the three Expectations follows.

***Expectation 2.1*:** Eleven items—including 9 selected-response items, 1 brief constructed-response item and 1 extended constructed-response item—align with the four indicators associated with this expectation. This is an appropriate mix of item types and fairly reflects the content and performance expectations defined by Indicators 2.1.1, 2.1.2, 2.1.3, and 2.1.4.

The P-values for these items range from a low of 0.34 for Item 31, the extended constructed-response item, to a high of 0.76. The distribution of item difficulties spans a reasonable range, with 7 items having P-values greater than 0.50 and 4 items having P-values of less than .50. Both the brief and extended constructed-response items tend to present challenges to students, with P-values of 0.45 and 0.34 respectively.

Expectation 2.2: Fourteen items—including 10 selected-response items, 3 student-produced response items, and 1 extended-response item—align with the three indicators defined in this expectation. Again, the mixture of item types is robust and presents an appropriate level of challenge to students. Four of the fourteen items (or almost 29% of the items) require students to generate, rather than select, responses. In particular, reviewers commend the state for requiring students to demonstrate their geometric reasoning abilities by developing a proof. Inclusion of such an unscaffolded proof most definitely contributes to the level of challenge of this set of items.

The items in this set do a good job of spanning the level of demand of these indicators and are appropriate for high school students. The P-values for this set of items range from a low of 0.26 to a high of 0.95. The range, or spread, of item difficulties is greater for items aligning with Expectation 2.2 than it was for items

mapping to Expectation 2.1. Eight of the 14 items aligning with Expectation 2.2 have P-values greater than 0.50, while 6 of the items have P-values less than 0.50. As would be expected, the items which require students to generate responses tend to be difficult for students. The extended constructed-response item had a P-value of 0.38 while the 3 items requiring gridded responses on the part of students had P-values of 0.53, 0.40, and 0.26. Three selected-response items (Items 24, 28, and 32) also presented students with challenges, having P-values of 0.47, 0.49, and 0.48 respectively.

Expectation 2.3: Twelve items—including 7 selected-response items, 1 brief constructed-response item, 1 extended constructed-response item, and 3 items requiring students to grid their responses—align with the two indicators associated with this expectation. As was the case with Expectations 2.1 and 2.2, the item set mapping to Expectation 2.3 contains an appropriate mix of item types and tends to reflect the content and performance expectations defined in Indicators 2.3.1 and 2.3.2. The level of challenge of this set of items is appropriate for high school students. Five of the twelve items (or over 40% of the items) ask students to develop answers themselves, which tends to be more challenging to students than selecting a correct answer.

The P-values for this set of items range from a low of 0.37 to a high of 0.86. This range, or spread, of P-values more closely approximates that of the set of items mapping to Expectation 2.1 than it does with the set of items mapping to Expectation 2.2. The range of P-values for Expectations 2.1, 2.2, and 2.3 are 0.42, 0.69, and 0.49, respectively. Seven of the 12 items aligning with Expectation 2.3 (or almost 60% of the items) have P-values greater than 0.50, although 6 of these 7 items have P-values that cluster between 0.53 and 0.61. The remaining 5 items have P-values that range from 0.37 to 0.47. Interestingly, the items that require students to generate responses were not necessarily the most difficult for them. Two selected-response items—Items 34 and 35—tended to be among the most difficult for students with P-values of 0.43 and 0.37 respectively.

While, as pointed out earlier, Maryland's geometry test would benefit from the inclusion of more challenging selected-response items, the state has been successful in ensuring a consistent level of challenge across item sets assessing the three expectations defined for High School Core Learning Goal 2. The item sets mapping to these three expectations all contain a mix of item types, requiring students to generate at least some of their own answers, and reviewers were able to identify a Level 3 (strategic thinking) item associated with each expectation. While it is the case that Level 1 (recall) items make up the largest proportion of items aligned with each expectation, a comparable share of the items mapping to each expectation was identified by reviewers as Level 2 (skill/concept).

Balance

Reviewers also evaluated the degree to which each set of items mapping to an expectation reflects the balance of content delineated by the indicators defined for that expectation. The reviewers' objective was to determine whether the overall balance of a

set of items matches the emphases in terms of content and performance as defined by the indicators.

Prior to embarking upon a detailed analysis by Expectation, some comments about the overall balance of Maryland's high school geometry test are in order. Maryland has done an exemplary job of ensuring balance across Expectations 2.1, 2.2, and 2.3. Eleven items (or about 30% of the test's 37 operational items) assess indicators from Expectation 2.1. Fourteen items (or about 38% of the test's items) assess Expectation 2.2, while 12 items (or almost 33% of the items) assess Expectation 2.3. This is indicative of the painstaking care Maryland has taken to ensure fair and representative coverage in its geometry assessment. What follows is a more detailed commentary, organized by expectation.

> ***Expectation 2.1***: This expectation includes indicators that define what students should know and be able to do with respect to two- and three-dimensional figures. Specifically, the indicators address analysis of the properties of geometric figures (Indicator 2.1.1), identification and verification of properties of geometric figures using algebraic concepts and coordinate geometry (Indicator 2.1.2), the use of transformations (Indicator 2.1.3), and geometric constructions and drawings (Indicator 2.1.4). All four of these indicators are assessed by Form T of Maryland's Geometry test—with 3 items assessing Indicator 2.1.1, 3 items assessing Indicator 2.1.2, 4 items assessing Indicator 2.1.3, and 1 item assessing Indicator 2.1.4. One of the items assessing Indictor 2.1.3 is a 3-point brief constructed-response item, and the sole item assessing Indicator 2.1.4 is a 4-point extended constructed-response item. So the distribution of points across the four indicators is as follows: 3 points for Indicator 2.1.1, 3 points for Indicator 2.1.2, 6 points for Indicator 2.1.3, and 4 points for Indicator 2.1.4.
>
> The 11 items that assess this expectation generally provide good coverage of the content and performances specified in the indicators. The test includes a reasonable and balanced sampling of geometric properties, relationships and geometric figures—as defined by the details in the assessment limits that accompany these indicators. This particular form of the geometry test is especially strong with respect to transformations (Indicator 2.1.3), requiring students to apply the full range of transformations—reflections, rotations, translations, and dilations—to the solution of problems. As defined in Indicator 2.1.3, students are required to go beyond the mere identification of transformations, with Item 20—the brief constructed-response item—even requiring students to sketch the positions of figures and mathematically describe the transformations needed to create the needed configuration.
>
> Indicators 2.1.1 and 2.1.2 require students to analyze, identify, and/or verify the properties of a variety of types of geometric figures. Generally, this set of 11 items does a good job of addressing a variety of polygons (including triangles, rectangles, and rhombuses) and a variety of relationships (such as congruent, parallel, midpoint, and angle relationships). The items also adequately assess students' understanding of angle measurements, in a variety of contexts. For

example, some items (such as Items 1 and 8) deal with angle measurement in the context of polygons, while Item 21 deals with angle relationships associated with parallel lines. There is a strong, but appropriate emphasis, given to coordinate geometry, with Items 16, 17, 18, and 19 all being presented in such a context. The fact that students are provided with graph paper for this assessment ensures that they have the tools they need to solve such problems, if they are unable or choose not to use a conceptual or formula-based approach.

There are some areas that appear lacking, however. For example, none of the items aligning with this expectation address circles or the angle relationships involving circles (e.g., central and inscribed angles), and reviewers agreed that inclusion of one or more items addressing this content is important. Similarly, while some items involve parallel lines or right angles, none of the items require students to establish a perpendicular or parallel relationship. Plus, no items mapping to Indicator 2.1.1 or 2.1.2 assess an understanding of the properties of geometric solids. Reviewers are aware that items assessing this content may well be on other forms of the assessment and/or in the item bank used to create multiple forms of the assessment, yet no such items were included on Form T, the form provided to reviewers for purposes of this analysis.

Expectation 2.2: This Expectation includes indicators that define what students should know and be able to do with respect to applying geometric properties and relationships to solve problems. Specifically, the indicators address congruent and similar figures and the application of equality or proportionality to their corresponding parts (Indicator 2.2.1), use of two-dimensional figures and/or right triangle trigonometry to solve problems (Indicator 2.2.2), and the use of inductive and deductive reasoning (Indicator 2.2.3). All three of these indicators are assessed by Form T of Maryland's Geometry test—with 5 items assessing Indicator 2.2.1, 6 items assessing Indicator 2.2.2, and 4 items assessing Indicator 2.2.3. Item 15 assesses Indicators 2.2.1 and 2.2.3 so is counted twice in these tallies, although only 14 items in total align with Expectation 2.2. Item 15 is the only multi-point item aligning with Expectation 2.2, and if it is assumed that half of its total point value (4 points) is attributable to Indicator 2.2.1 and half to Indictor 2.2.3, then the distribution of points across the three indicators is as follows: 6 points for Indicator 2.2.1, 6 points for Indicator 2.2.2, and 5 points for Indicator 2.2.3.

The 14 items that align with this expectation generally provide good coverage, although there is a particularly strong emphasis on right-triangle trigonometry. Three items (Items 24, 27, and 28) specifically require the use of right-triangle trigonometry, and Item 37 can be solved either by the application of trigonometry or special right triangle relationships. The emphasis placed on trigonometry exceeds that in many state tests, and reviewers noted that, while the skill statement associated with Indicator 2.2.2 specifically mentions right-triangle trigonometry, the Pythagorean Theorem, and special right-triangle relationships, only one item on the test—Item 9—requires students to use the Pythagorean

Theorem. It is the case in Maryland, however, that the content standards for grade 8 also indicate that students need to be able to "use the Pythagorean theorem to solve problems by determining the missing side of a right triangle." So it may be a conscious decision on the part of the state to relegate emphasis on this to the grade 8 assessment.

The test does a good job of assessing students' understanding of congruent and similar figures (Indicator 2.2.1) in a reasonable and balanced way. Students are required to employ equality (in the case of congruence) and proportionality (in the case of similarity) as they solve these problems. Likewise, reviewers were favorably impressed that the assessment includes an extended constructed-response item that requires students to develop a geometric proof (focusing on similar triangles). It is frequently the case that state tests attempt to assess students' reasoning skills solely through selected-response items. There are definite limitations to such items, and Maryland is to be commended for including such an unscaffolded proof in its assessment.

Expectation 2.3: This Expectation includes indicators that define what students are expected to know and be able to do with respect to measurement. Specifically, the indicators address indirect measurement (Indicator 2.3.1) and the calculation and estimation of such attributes as perimeter, circumference, area, volume, and surface area (Indicator 2.3.2). Both of these indicators are assessed by Form T of Maryland's Geometry test—with greater emphasis being given to Indicator 2.3.2 than to Indicator 2.3.1. Four items assess the first indicator, while 8 items assess the second indicator. This emphasis is appropriate given the multitude of two-dimensional and three-dimensional shapes identified in the assessment limits for possible inclusion on the assessment—polygons, circle cubes, prisms, pyramids, cylinders, cones, spheres, and composite 2-dimension and 3-dimensional figures. The scope of Indicator 2.3.1 is substantially more limited, with emphases tending to be on such things as similarity, scaling, and proportional reasoning. Two multi-point items map to this Expectation—a 3-point brief constructed-response item (Item 10) to Indicator 2.3.1 and a 4-point extended constructed-response item (Item 5) to Indicator 2.3.2. The distribution of points, therefore, is such that 6 points are attributable to Indicator 2.3.1 and 11 points to Indicator 2.3.2.

The 12 items that align with this expectation provide good coverage of the concepts defined by these indicators. They also do a good job of representing a variety of geometric figures (including squares, rectangles, triangles, trapezoids, circles, spheres, cones, rectangular prisms, cylinders, and composite shapes) and such attributes as length, area, volume, and surface area.

While students are provided with a detailed Formula Reference Sheet, reviewers were favorably impressed by its presentation. Rather than simply providing students with rote formulas to apply, the sheet instead requires students to have a conceptual understanding of such derived attributes as area, volume, and surface

area. For example, rather than just telling students that the formula for finding the volume of a right circular cylinder is $V=\pi r^2h$, Maryland's Formula Reference Sheet provides a diagram of a right circular cylinder and a generic formula reminding students that volume is calculated by multiplying the area of the base time the height (V = Bh = area of base x height). Therefore, students need to reason that they must first find the area of the base of the cylinder—which is a circle—and multiply that by the height of the cylinder. This makes the solution of test items involving such attributes as volume and surface area more conceptually challenging—a step up from some other state tests where such calculations often turn out to be nothing more than simple evaluation.

Range

An examination of the range for the set of items that maps to each expectation gives a quantitative evaluation as to how well the items cover the indicators associated with that expectation. For the purpose of this calculation, range is considered to be the portion of indicators for a given expectation that is assessed by at least one item. Range is expressed as a decimal number, and its calculation includes all items that map to the expectation. As can be seen from the chart below, the range exhibited by Form T of Maryland's high school geometry test is exemplary, with 100% of the indicators associated with High School Core Learning Goal 2 being assessed.

Maryland High School Assessment, May 2002 Goal 2: Geometry, Measurement, and Reasoning	
Expectation	**Portion of Indictors Assessed**
2.1 Two- and Three-Dimensional Figures	4/4 or 1.0
2.2 Geometric Relationships and Properties	3/3 or 1.0
2.3 Concepts of Measurement	2/2 or 1.0
Total	9/9 or 1.0

Assessment-to-Standards Alignment Summary

- Items on the high school geometry test address the content specified in High School Core Learning Goal 2. This is evidenced by the fact that all of the items received scores of 2 for content centrality.
- Items on the high school geometry test address the performance expectations specified in High School Core Learning Goal 2. This is evidenced by the fact that all of the items received scores of 2 for performance centrality.
- Items on the high school geometry test score well for source of challenge, with reviewers expressing concern about only one item.
- The cognitive demand of items on the high school geometry test tends to be at the recall and basic problem solving (skill/concept) levels, with about 93% of the test items being rated at these two levels. Only about 7% of the test items were judged to require strategic thinking on the part of students, and all of these items were constructed-response items. The test could be made more challenging for students if more cognitively demanding items—particularly

selected-response and student-produced response items—could be infused into the test. While time constraints and scoring costs are certainly a consideration in a large-scale, on-demand assessment such as this, quick-response, machine-scorable items can be included that effectively measure strategic thinking.

- The level of challenge of the high school geometry test is appropriate for high school students having completed instruction is geometry—either in a course devoted to geometry or as part of a course sequence that addresses geometry as part of an integrated mathematics program. The test items do a good job of assessing the range of content and performance expectations defined in High School Core Learning Goal 2, and the level of challenge is generally consistent across the three expectations. In particular, the brief and extended constructed-response items provide students with challenging contexts for demonstrating their knowledge and skills in geometry. More demanding selected-response and student-produced response items would improve the level of challenge even more.
- Maryland has done an exemplary job of ensuring balance across its three expectations. Within each expectation, Maryland has also done a good job of ensuring that the item set generally matches the emphases in terms of content and performance as defined by the indicators. Coverage of transformations is thorough, and test items tend to address a wide variety of geometric figures, attributes, and relationships. Similarly, the test does a good job of assessing students' understandings of congruency and similarity—foundations of any strong preparation in geometry. Some topics that are key to geometry, such as properties of circles and angle relationships, were missing from the test, although it is may be that such items are included in other forms of the test and/or in the item pool used to generate the multiple forms of the test. A particularly strong emphasis on trigonometry was noted—one that exceeds what appears in many other state high school tests.
- The range, or coverage, of the high school geometry test relative to the indicators is exemplary. All nine geometry indicators are assessed by this test, and such thorough coverage is rare.

Item Mapping
Maryland High School Assessment, May 2002
Goal 2: Geometry, Measurement, and Reasoning

Item Number	MD Content Standard Primary Map	MD Content Standard Secondary Map
1	2.1.1	
2	2.1.3	
3	2.2.1	
4	2.3.1	
5	2.3.2	2.3.1
6	2.2.3	
7	2.2.3	
8	2.1.1	
9	2.2.2	
10	2.3.1	
11	2.2.1a	
12	2.3.1	
13	2.3.2	
14	2.2.2	
15	2.2.1, 2.2.3	
16	2.1.3	
17	2.1.2	
18	2.1.2	
19	2.1.2	
20	2.1.3	
21	2.1.1	
22	2.2.1	
23	2.3.2	
24	2.2.2	
25	2.3.1	
26	2.2.1	
27	2.2.2	
28	2.2.2	
29	2.3.2	
30	2.3.2	
31	2.1.4	2.1.1
32	2.2.3	
33	2.1.3	
34	2.3.2	
35	2.3.2	
36	2.3.2	
37	2.2.2	

Note: Only extended constructed response items (ECR) were deemed to assess more than one indicator. Item 15 was assigned two primary mappings, while Items 5 and 31 were each assigned one primary mapping and one secondary mapping.

Item Mappings and Ratings by Expectation and Indicator
Maryland High School Assessment, May 2002
Goal 2: Geometry, Measurement, and Reasoning

Expectation/ Indicator	Item	Item Ratings				Item Type and Other Comments
		Content Centrality (CC)	Performance Centrality (PC)	Source of Challenge (SoC)	Level of Demand (LoD)	
Expectation 2.1						
2.1.1	1	2	2	1	1	SR
	8	2	2	1	2	SR
	21	2	2	1	2	SR
2.1.2	17	2	2	1	1	SR
	18	2	2	0	Not scorable	SR
	19	2	2	1	2	SR
2.1.3	2	2	2	1	1	SR
	16	2	2	1	1	SR
	20	2	2	1	2	BCR
	33	2	2	1	1	SR
2.1.4	31	2	2	1	3	ECR
Expectation 2.2						
2.2.1	3	2	2	1	1	SR
	11	2	2	1	2	SPR – maps to Skill Statement 2.2.1.a
	15	2	2	1	2	ECR – also maps to Indicator 2.2.3
	22	2	2	1	1	SR
	26	2	2	1	1	SR
2.2.2	9	2	2	1	1	SR
	14	2	2	1	2	SPR
	24	2	2	1	1	SR
	27	2	2	1	1	SR
	28	2	2	1	1	SR
	37	2	2	1	1	SPR
2.2.3	6	2	2	1	2	SR
	7	2	2	1	2	SR
	15	2	2	1	3	ECR – also maps to Indicator 2.2.1
	32	2	2	1	2	SR
Expectation 2.3						
2.3.1	4	2	2	1	1	SR

Expectation/ Indicator	Item	Item Ratings				Item Type and Other Comments
		Content Centrality (CC)	Performance Centrality (PC)	Source of Challenge (SoC)	Level of Demand (LoD)	
2.3.1	10	2	2	1	3	BCR
	12	2	2	1	1	SPR
	25	2	2	1	1	SR
2.3.2	5	2	2	1	2	ECR
	13	2	2	1	1	SPR
	23	2	2	1	1	SR
	29	2	2	1	1	SR
	30	2	2	1	2	SR
	34	2	2	1	1	SR
	35	2	2	1	2	SR
	36	2	2	1	2	SPR

Note 1: Item types include selected response (SR), student-produced response (SPR), brief constructed response (BCR), and extended constructed response (ECR).
Note 2: Two sets of ratings are provided for Item 15, which was assigned two primary mappings by reviewers. This is the only item for which dual sets of ratings are assigned.

Maryland High School Assessment, May 2002
Goal 2: Geometry, Measurement and Reasoning

Test Format

Number of Operational Items	37
Selected Response	26
Student Produced/Gridded Response	6
Brief Constructed Response	2
Extended Constructed Response	3
Number of Field Test Items:	10
Total Number of Test Items	47

Item Coverage

Expectations and Indictors	# of Items	% of Operational Test Items
Expectation 2.1	**11**	**29.7%**
Indicator 2.1.1	*3*	
Indicator 2.1.2	*3*	
Indicator 2.1.3	*4*	
Indicator 2.1.4	*1*	
Expectation 2.2	**14**	**37.8%**
Indicator 2.2.1	*4.5*	
Indicator 2.2.2	*6*	
Indicator 2.2.3	*3.5*	
Expectation 2.3	**12**	**32.5%**
Indicator 2.3.1	*4*	
Indicator 2.3.2	*8*	

Note: Item 15 was mapped to two indicators (2.2.1 and 2.2.3). For purposes of this chart, half of Item 15 is counted as aligned to Indicator 2.2.1 and the other half of the item is counted as aligned to Indicator 2.2.3. Therefore, despite this double mapping of Item 15, the total item count remains at 37, which corresponds to the number of operational (i.e., non-field test) items on the assessment.

CONTENT CENTRALITY

Expectation	# of Items	2	1a	1b	0
2.1 Two- and Three-Dimensional Figures	11 (29.7% of operational items)	11	0	0	0
2.2 Geometric Relationships and Properties	14 (37.8% of operational items)	14	0	0	0
2.3 Concepts of Measurement	12 (32.5% of operational items)	12	0	0	0
Total	37 (100% of operational items)	37	0	0	0
		100% of all operational items received a 2	0% of all operational items received a 1a	0% of all operational items received a 1b	0% of all operational items received a 0

PERFORMANCE CENTRALITY

Expectation	# of Items	2	1a	1b	0
2.1 Two- and Three-Dimensional Figures	11 (29.7% of operational items)	11	0	0	0
2.2 Geometric Relationships and Properties	14 (37.8% of operational items)	14	0	0	0
2.3 Concepts of Measurement	12 (32.5% of operational items)	12	0	0	0
Total	37 (100% of operational items)	37	0	0	0
		100% of all operational items received a 2	0% of all operational items received a 1a	0% of all operational items received a 1b	0% of all operational items received a 0

SOURCE OF CHALLENGE

Expectation	# of Items	1	0	Not Scored
2.1 Two- and Three-Dimensional Figures	11 (29.7% of operational items)	10	1	0
2.2 Geometric Rclationships and Properties	14 (37.8% of operational items)	14	0	0
2.3 Concepts of Measurement	12 (32.5% of operational items)	12	0	0
Total	37 (100% of operational items)	36	1	0
		97.3% of operational items received a 1	2.7% of operational items received a 0	

LEVEL OF DEMAND

Expectation	# of Items	4	3	2	1	Not Scored
2.1 Two- and Three-Dimensional Figures	11 (29.7% of operational items)	0	1	4	5	1
2.2 Geometric Relationships and Properties	14 (37.8% of operational items)	0	.5	5.5	8	0
2.3 Concepts of Measurement	12 (32.5% of operational items)	0	1	4	7	0
Total	37 (100% of operational items)	0	2.5	13.5	20	1
		0% of scorable items received a 4	6.9% of scorable items received a 3	37.5% of scorable items received a 2	55.6% of scorable items received a 1	

Note 1: Item 18, which maps to Indicator 2.1.2, was identified as having an inappropriate source of challenge. It is, therefore, not scorable for Level of Demand.
Note 2: Item 15, which received 2 sets of ratings, received different ratings for Level of Demand Each set of ratings received a weighting of 0.5, resulting in fractional item counts in this table.

LEVEL OF CHALLENGE

Expectation	Level of Challenge of Item Set
2.1 Two- and Three-Dimensional Figures	Appropriate
2.2 Geometric Relationships and Properties	Appropriate
2.3 Concepts of Measurement	Appropriate
Total	Appropriate

BALANCE

Expectation	Balance of Item Set
2.1 Two- and Three-Dimensional Figures	Good
2.2 Geometric Relationships and Properties	Good
2.3 Concepts of Measurement	Good
Total	Good

RANGE

Expectation	Portion of Indictors Assessed
2.1 Two- and Three-Dimensional Figures	4/4 or 1.0
2.2 Geometric Relationships and Properties	3/3 or 1.0
2.3 Concepts of Measurement	2/2 or 1.0
Total	9/9 or 1.0

Summary Data Chart

Maryland High School Assessment, May 2002
Goal 2: Geometry, Measurement, and Reasoning

Indicator	*# items*	*Content Centrality*				*Performance Centrality*				*Source of challenge*		*Level of Demand*		
		2	*1b*	*1a*	*0*	*2*	*1b*	*1a*	*0*	*1*	*0*	*3*	*2*	*1*
Expectation 2.1: The student will represent and analyze two- and three-dimensional figures using tools and technology when appropriate.														
2.1.1	3	3				3				3			2	1
2.1.2	3	3				3				2	1		1	1
2.1.3	4	4				4				4			1	3
2.1.4	1	1				1				1		1		
Subtotal	*11*	*11*	*0*	*0*	*0*	*11*	*0*	*0*	*0*	*10*	*1*	*1*	*4*	*5*
Expectation 2.2: The student will apply geometric properties and relationships to solve problems using tools and technology when appropriate.														
2.2.1	4.5	4.5				4.5				4.5			1.5	3
2.2.2	6	6				6				6			1	5
2.2.3	3.5	3.5				3.5				3.5		.5	3	
Subtotal	*14*	*14*	*0*	*0*	*0*	*14*	*0*	*0*	*0*	*14*	*0*	*.5*	*5.5*	*8*
Expectation 2.3: The student will apply concepts of measurement using tools and technology when appropriate.														
2.3.1	4	4				4				4		1		3
2.3.2	8	8				8				8			4	4
Subtotal	*12*	*12*	*0*	*0*	*0*	*12*	*0*	*0*	*0*	*12*	*0*	*1*	*4*	*7*
TOTAL	*37*	*37*	*0*	*0*	*0*	*37*	*0*	*0*	*0*	*36*	*1*	*2.5*	*13.5*	*20*

Note 1: Item 15 was mapped to two indicators (2.2.1 and 2.2.3). The ratings for content centrality, performance centrality, and source of challenge were identical in both mappings (CC=2, PC=2, SC=1), but the item was rated Level 2 with respect to Indicator 2.2.1 and Level 3 with respect to 2.2.3. For purposes of this chart, half of Item 15 is counted as aligned to Indicator 2.2.1 and the other half of the item is counted as aligned to Indicator 2.2.3. Therefore, despite this double mapping of Item 15, the total item count remains at 37, which corresponds to the number of operational (i.e., non-field test) items on the assessment.
Note 2: Item 18, which maps to Indicator 2.1.2, was identified as having an inappropriate source of challenge. It is, therefore, not scorable for Level of Demand.