National Assessment Governing Board

Committee on Standards, Design and Methodology

May 14, 2010

COSDAM Attendees: Lou Fabrizio (Chair), Tonya Miles, James Popham, Andrew Porter (via teleconference), and Darvin Winick.
Other Governing Board Members: John Q. Easton (*Ex officio*), Director of the Institute of Education Sciences.
Governing Board Staff: Michelle Blair, Susan Loomis, Ray Fields, and Executive Director Cornelia Orr.
Other Attendees: NCES: Peggy Carr (Associate Commissioner of the National Center for Education Statistics) and Andrew Kolstad. AIR: George Bohrnstedt and Fran Stancavage. ACT: Nancy Petersen. CCSSO: Gene Wilhoit. ETS: Jay Campbell, Andreas Oranje, and Mary Pitoniak. Fulcrom IT: Scott Ferguson. HumRRO: Lauress Wise. McGraw Hill Education: Larry Snowhite. MetaMetrics: Heather Koons. Pearson: Brad Thayer.

Lou Fabrizio, Chair of the Committee on Standards, Design and Methodology (COSDAM), called the meeting to order at 10:15 a.m. and welcomed members and guests. Mr. Fabrizio noted that member Andrew Porter would participate in the meeting via conference call. Mr. Fabrizio announced that the COSDAM meeting includes a closed session beginning at 11:30 a.m. Neither Mr. Paine nor Ms. D'Amico can attend the meeting.

ACTION ITEM

1. **NAEP Grade 12 Mathematics Achievement Levels Descriptions**

   Lou Fabrizio, Chair, distributed the achievement level descriptions for the grade 12 mathematics National Assessment of Educational Progress (NAEP). He stated that the process used for developing these descriptions was the same as that for the reading descriptions that the Board approved at the March 2010 meeting. Public comments were solicited via a *Federal Register* notice, announcements on the Governing Board website, and invitations to stakeholder groups and other individuals likely to have an interest in mathematics achievement levels. The goal was to ensure that a wide array of participants had an opportunity to submit feedback. Susan Loomis offered to answer questions about the process for developing the achievement level descriptions.

Mr. Fabrizio presented the following **motion** to COSDAM:

> **The Committee on Standards, Design and Methodology approves the following achievement levels descriptions for grade 12 mathematics and recommends adoption of these descriptions for reporting performance for the National Assessment of Educational Progress for grade 12 mathematics beginning with the 2009 assessment.** (Achievement levels descriptions are attached to this report.)

Jim Popham made the motion, and Darv Winick seconded it. The motion was unanimously approved by voice vote.

John Easton asked a follow-up question about evidence to link the quality of the achievement levels descriptions (ALDs) to the outcomes of standard setting process. Specifically, he was interested in whether any research had been conducted to indicate that well-crafted ALDs facilitate their use by panelists and are associated with outcomes of the standard setting processes. Mr. Popham observed that there has been a shift in influence on standard setting panelists away from impact data and toward achievement levels descriptions. He stated that the panelists now are increasingly trained to become wedded to the ALDs which means that they are likely to pay little attention to the impact data.

2. Update on 12th Grade Preparedness Research

Mr. Fabrizio next asked Susan Loomis and Ray Fields of the Governing Board staff to update the Committee regarding the studies for grade 12 preparedness research. Ms. Loomis began by noting that two content alignment studies had been implemented since the March 2010 meeting: NAEP-SAT and NAEP-ACCUPLACER. Each study used the design documents approved by COSDAM last year. Each study included replicate panels for each subject alignment, so that four panels met concurrently for each study. Ray Fields reported that the NAEP-WorkKeys content alignment project is in the report-writing phrase. A meeting with contractors and staff will be convened later this month to plan reporting in order to maximize comparability across study reports. By the August COSDAM meeting, reports will be available for discussion of the content alignment studies.

Mr. Fields briefed the Committee on a new addition to the list of preparedness research studies. Through interactions with the Texas Higher Education Commissioner Raymund Paredes, who is also a member of the Governing Board's High School Achievement Commission, plans are now underway for a benchmarking study involving students entering two-year and four-year institutions in Texas. Samples for administration of the NAEP in reading and mathematics will include students placed into remedial courses and students placed into credit-bearing courses related to preparedness in these subjects, and the goal is to report the average score of each group of students for the NAEP in reading and mathematics. A small-scale pilot study with approximately nine institutions will be implemented in late summer 2010, prior to start of classes. The operational administration will be implemented in 2011.

Mr. Winick noted that Texas has been collecting longitudinal student data for 25 years, and Mr. Paredes is eager to have Texas serve as a test case for preparedness research using the longitudinal data.

Mr. Popham asked how this study fits into the preparedness research agenda and whether individual scores would be produced for the Texas students in the study. Mr. Fields responded that there will be no individual student scores; only aggregate averages will be reported for the preparedness research. Mr. Fields described the general approach of comparing results across studies to evaluate the extent to which study findings converge. He noted that we have made no assumptions regarding whether the results will converge on a single score range on the NAEP scale or on a few or on multiple points. Cornelia Orr added that data will be available for the Committee to discuss both results of the content alignment studies and the statistical analysis of scale score relationships with NAEP and other assessments. These comparisons will help to determine potential interpretations of the NAEP score scale in relation to preparedness.

Ms. Loomis reported that the technical advisors will be convened in June to continue discussions of design elements for the statistical analysis of NAEP data and data for other assessments. Data from the Florida Department of Education have been received and preliminary analysis work has begun. She noted that some estimated dates for completion of reports have been revised, and two checks appear in the update chart to indicate that the timelines have been modified for a study.

Exemplar jobs selection recommendations will be shared with COSDAM in August, as well. Ms. Loomis noted that the judgmental standard setting studies to set cut scores on the NAEP scales in reading and mathematics to represent preparedness for placement in job training programs for 5-7 exemplar jobs and the studies to set cut scores to represent placement for higher education coursework are somewhat behind schedule. The procurement process requires a lengthy period of time before the award can be made and work can begin.

The higher education survey study timeline is also behind schedule. Mr. Fields stated that this survey will create data on course placement that are not currently available, and this important point will help to promote the participation and increase the response rate among institutions.

Mr. Popham asked to have recommendations from the staff on how the pieces of information may fit together. Mr. Fields responded that a vetting process is being planned to help to guide the determination of how the various study findings fit together.

Ms. Orr noted that the 11 states that participated in the grade 12 state NAEP pilot are eager to know what they can say about the preparedness of their students. The states will be looking at this research for assistance.

3. **Common Core State Standards for College and Career Readiness and NAEP 12th Grade Preparedness Research Program**

Following his presentation to the Board Friday morning, Gene Wilhoit, Executive Director of the Council of Chief State School Officers (CCSSO), met with COSDAM to discuss Common Core State Standards (CCSS) and answer questions from COSDAM regarding the standards and how they may relate to the NAEP 12th grade preparedness program.

Mr. Wilhoit expects that NAEP will continue as the auditor/monitor of national progress. He cautioned that it is important to be very clear about how NAEP and CCSS relate.

Mr. Wilhoit reported that there will apparently be only two assessment consortia, and they are developing somewhat different approaches to assessing the Common Core State Standards. Each has affirmed a commitment to the goals of the CCSS program and assessment of the full breadth and depth of the standards. Mr. Winick suggested it is essential that unreliable measurement be excluded from the Common Core assessments. Mr. Wilhoit observed that this aspect of the program is playing out differently across the two assessment consortia, but this is an issue with which both consortia are consciously grappling.

Discussion next focused on whether the full depth and breadth of the Common Core Standards can and should be a goal of the assessments. Mr. Popham suggested caution about this goal for the assessments and expressed that this may be unrealistic. Mr. Wilhoit noted that the two assessment consortia seem to be developing somewhat different approaches and emphases, but each has agreed to the goals of common assessments. He noted that measurement depth has a higher priority than breadth, and he acknowledged that a deliberative discussion is needed. He supports having an essential assessment component that is common across assessments.

Tonya Miles asked whether consideration has been given to having the standards enacted without a common assessment: could the goals for state reform of curriculum and instruction be accomplished if the states continued to use their own assessments instead of having a common assessment? Mr. Wilhoit responded that this had been discussed at length, but the key objectives of comparability and transparency seem to require some common assessments, as well as common standards for curriculum and instruction. He stated that the state-NAEP mapping studies have been very instrumental in showing the need for common standards for curriculum and instruction, common assessments, and common performance standards.

Mr. Easton asked about the possibility of some type of common formative assessment across the consortia. The potential to conduct linking studies across the assessment consortia was discussed. Mr. Easton noted the need to be able to explain clearly to the public how students perform. Mr. Wilhoit expressed strong support for comparing results across the two assessment consortia and to have at least a portion of the assessments measure a common core to produce a common score. Despite differences in the approaches of the two assessment programs—one has a heavy emphasis on technology-based administration and

the other places a heavy emphasis on "authentic" assessments—there is a good faith effort to produce comparable score results. It is not yet clear what the final outcome will be.

Mr. Fabrizio mentioned that some changes in state assessment requirements will be needed in the ESEA reauthorization. Currently, states are required to assess all of their state standards. The plan for the CCSS includes an option for states to add state-specific standards, not to exceed 15% of the total. However, these state-specific standards will not be included in the common core assessments, so states could be non-compliant with the ESEA policies if the requirements are not modified. Mr. Wilhoit agreed that it is very important to avoid this sort of pitfall. The goal is to optimize outcomes for students. He noted that a key priority for Secretary Duncan is to work out a plan that is fair across states and that is productive for students. The goal is to make sure that the policies have a positive impact on students.

4. **2011 Pilot of NAEP Computerized Adaptive Test Design**

Andy Kolstad presented information about the first computerized adaptive testing study for NAEP to be implemented for the mathematics NAEP in 2011. A multi-stage test design is incorporated to help maintain the structure of NAEP blocks, to the extent feasible. A key goal of the adaptive testing effort is to match the cognitive demands of the assessment to the ability of the students to support better measurement for student performance on both ends of the scale—for both low-achieving and high-achieving students. Students in the middle of the range are currently measured quite well in NAEP, but having test items that challenge students at their actual performance level will support stronger measurement across the entire continuum of student performance. A second goal of this effort is to increase student engagement by increasing the likelihood that students can respond to more questions. Adaptive testing can create less frustration for students by reducing the potential for encountering many items that are either extremely easy or extremely difficult. A third goal of computerized adaptive testing is to produce more efficient assessments by reducing the time required for assessing students. The NAEP design already involves a relatively short testing time per student over a broad content domain, so this goal is not as important for NAEP as it may be for other testing programs.

Mr. Kolstad noted that the "routing" blocks, representing the initial testing stage, can include only multiple choice items because student performance on the first stage of testing must be scored immediately to route students to the block having the appropriate range of difficulty for the second stage of testing. In order to have the computerized adaptive test represent the framework requirements, it may be necessary to have mostly constructed response items in the second stage.

The Committee engaged in questions and answers about various aspects of the study design including the content restrictions posed by the framework requirements. Ms. Orr suggested COSDAM may want to consider recommending that the content restrictions be relaxed for this study to provide a better study design for informing future work.

In response to a question by Mr. Fabrizio regarding the need for better measurement at the lower end of the NAEP scale, Mr. Kolstad reported that studies are already underway to increase measurement precision at the lower end of the scale. Mr. Kolstad also noted that the study includes a non-adaptive component as a type of control group to evaluate the gain in measurement precision associated with the computerized adaptive design.

Mr. Popham questioned whether the assumption of unidimensionality for adaptive testing is met for the mathematics NAEP. Mr. Kolstad responded that while the assessment includes five separate subscales, they are highly intercorrelated.

Mr. Easton noted that greater measurement precision is stated as an advantage of computerized adaptive testing, but he asked whether an equity issue is raised in adaptive testing regarding the lack of equal access to the full array of assessment items by all examinees. Mr. Kolstad conjectured that a probabilistic model could ensure that every student has at least some probability of being assessed with items from any of the blocks, but the probability would depend on the student's performance on the first stage of the assessment. Mr. Popham said that the stronger equity argument may be that adaptive testing exposes students to an array of items, and students are likely to do well on at least some of them. It is important to communicate this effort to stakeholders.

5. Updates on On-Going COSDAM Issues and Recommendations for Future Topics

Mr. Kolstad shared two new developments in his update on the international benchmarking design. (1) The International Association for Educational Assessment (IAEA) gave approval for the NAEP-TIMSS linking study for 2011. (2) NCES has developed the criteria and procedure for identifying states and is ready to invite eight states to participate in the NAEP-TIMSS linking study. These states were selected and invited to participate in the linking study to provide validity information for the international linking design.

Ms. Loomis noted the need to issue a request for proposals (RFP) for achievement levels-setting for the 2011 writing NAEP. This will be the first completely computerized test for NAEP, but the pilot sample in 2010 was not large enough for conducting a scaling study to identify potential issues prior to beginning standard setting work. As noted by Andrew Porter in the May 7, 2010 COSDAM teleconference and reiterated at this meeting, a program of research should be conducted to evaluate the performance of items relative to student performance **before** panelists are convened to participate in the standard setting process. Therefore, Board staff will work with NCES staff to develop a program of research for this purpose. Ms. Loomis asked for questions and suggestions to be submitted privately in adherence with federal procurement regulations.

Mr. Fabrizio asked for suggestions for future topics for COSDAM discussion. Committee members noted that the Committee already has a full slate of issues for discussion and consideration. No new topics were recommended.

ACTION ITEM
CLOSED SESSION 11:30 a.m. – 12:20 p.m.
COSDAM Attendees: Lou Fabrizio (Chair), Tonya Miles, James Popham, Andrew Porter (via teleconference), and Darvin Winick.
Other Governing Board Members: John Q. Easton (*Ex officio*), Director of the Institute of Education Sciences.
Governing Board Staff: Michelle Blair, Susan Loomis, Ray Fields, and Executive Director Cornelia Orr.
Other Attendees: NCES: Arnold Goldstein and Andrew Kolstad. AIR: George Bohrnstedt and Fran Stancavage. ACT: Nancy Petersen. ETS: Jay Campbell, Andreas Oranje, and Mary Pitoniak. HumRRO: Lauress Wise.

In accordance with the provisions of exemption (9)(B) of Section 552b(c) of Title 5 U.S.C., the Committee on Standards, Design and Methodology met in closed session on May 14, 2010 from 11:30 a.m. to 12:20 p.m. in order to review and discuss reports including secure data and results of research conducted to set achievement levels cut scores for the National Assessment of Educational Progress in science.

6. Science Achievement Levels

Mr. Fabrizio summarized the discussion by COSDAM in the May 7, 2010 conference call. Ms. Loomis distributed updated and new materials and asked whether the committee had any further questions. Among the materials was a set of options with the advantages and disadvantages (pros and cons) enumerated for consideration by the Committee. Mr. Fabrizio clarified that the four options were for action at the May 2010 meeting.

1. Use the recommendations of the ALS panelists.
2. Use historical (normative) data to adjust the cut scores
3. Use data for the 2009 assessment to adjust ALS results
4. Delay action until August 2010 to provide time to collect more data and information to inform the decision and more time to deliberate the decision. (Staff recommendation)

With regard to option 1, Mr. Popham wondered about the extent to which panelists' judgments were emphasized as being of the utmost importance, thereby supporting their confidence in their judgment and lessening the likelihood of changing their judgments when confronted with impact data. Nancy Petersen responded that the expertise of panelists is not emphasized. Rather, the emphasis is on the achievement levels descriptions which are developed by the Governing Board to serve as the criteria for what students *should know and be able to do*. The achievement levels descriptions are held as *the* criteria for setting the cut scores. Mary Pitoniak, the ETS representative to the Technical Advisory Committee on Standard Setting, attended both the pilot study and the achievement levels-setting panel meetings. She recalled that that results for the 2005 science NAEP were presented to panelists at the start of the achievement levels-setting panel meeting, although the data were not juxtaposed with the panelists' final results.

Andy Porter noted that when the test is not well centered on student ability, as was the case for the 2009 NAEP science, the standard setting procedures are not likely to produce reasonable results. The policy-makers are then placed in a situation of needing to change the results to be more reasonable. He proposed that standard setting not be conducted unless there is assurance that the assessment data are reliable and provide good measurement precision. Ms. Loomis stated that his recommendation to this effect in the May 7, 2010 teleconference call had been clearly heard, and it has already been incorporated into planning for the upcoming writing NAEP achievement levels-setting process.

Ms. Orr reflected on her prior experiences with a state assessment program and noted that in her experience the exact recommendations of standard setting panelists were rarely implemented by the state: the cut scores were typically set by the policy board. The recommendations of panelists were considered as part of the decision-making process of the policy board.

Mr. Winick concluded that the process is fundamentally flawed because the scales are unanchored; thus, there is no way to judge if the results are too high or too low. In his opinion, there is no reason to believe that the process would produce the same results if a second set of panelists were convened. He acknowledged that there is currently no known better way to set the cut scores in the absence of empirical scale anchoring. He further observed that a thoughtful process should be developed to indicate when to stop spending money on standard setting.

Mr. Fabrizio reiterated that the process is judgmental; panel recommendations are just one part of the process. He stated that he and his state assessment staff are expected to make recommendations to the policy board in his state. Mr. Popham indicated that this is exactly what he wants for the science NAEP results: staff recommended cut scores. He emphasized that the achievement levels cut scores should reflect impact data and what students can do.

Mr. Easton stated that he was very sympathetic to the discussions, but he reminded the Committee that lower standards tend to be the result when policy boards make decisions.

Tonya Miles thanked staff for the options and the evaluation of pros and cons for each of them. She stated that option 4 enables the Board to be both *responsive* and *responsible*, which is of the utmost importance.

The Committee as a whole agreed that this option was the best choice. No further action was needed at the May 2010 meeting.

Mr. Popham asked that the Committee have the opportunity to discuss the recommendation developed by staff prior to the August 2010 meeting. Mr. Popham felt that no further studies were needed to support the recommended cut scores: if the policy board declares a cut score to represent the achievement level, then it represents the achievement level. No verification of the alignment of content to the performance is required. Ms. Orr confirmed that his message was understood.

The Committee meeting was opened at 12:20 p.m. No action was taken, and the Committee meeting was adjourned.

I certify the accuracy of this report.

Louis M. Fabrizio

Lou Fabrizio, Chair

May21, 2010

Date

National Assessment of Educational Progress
Grade 12 Mathematics Achievement Levels Descriptions

Adopted May 14, 2010

Basic

<u>Summary of Grade 12 *Basic* Level Achievement</u>: Twelfth-grade students performing at the *Basic* level should be able to solve mathematical problems that require the direct application of concepts and procedures in familiar mathematical and real-world settings.

Students performing at the *Basic* level should be able to compute, approximate, and estimate with real numbers, including common irrational numbers. They should be able to order and compare real numbers and be able to perform routine arithmetic calculations with and without a scientific calculator or spreadsheet. They should be able to use rates and proportions to solve numeric and geometric problems.

At this level, students should be able to interpret information about functions presented in various forms, including verbal, graphical, tabular, and symbolic. They should be able to evaluate polynomial functions and recognize the graphs of linear functions. Twelfth-grade students should also understand key aspects of linear functions, such as slope and intercepts.

These students should be able to extrapolate from sample results; calculate, interpret, and use measures of center; and compute simple probabilities.

Students at this level should be able to solve problems involving area and perimeter of plane figures, including regular and irregular polygons, and involving surface area and volume of solid figures. They should also be able to solve problems using the Pythagorean theorem and using scale drawings. Twelfth graders performing at the *Basic* level should be able to estimate, calculate, and compare measures, as well as to identify and compare properties of two- and three-dimensional figures. They should be able to solve routine problems using two-dimensional coordinate geometry, including calculating slope, distance, and midpoint. They should also be able to perform single translations or reflections of geometric figures in a plane.

Proficient

<u>Summary of Grade 12 *Proficient* Level Achievement</u>: Twelfth-grade students performing at the *Proficient* level should be able to recognize when particular concepts, procedures, and strategies are appropriate, and to select, integrate, and apply them to solve problems. They should also be able to test and validate geometric and algebraic conjectures using a variety of methods, including deductive reasoning and counterexamples.

Twelfth-grade students performing at the *Proficient* level should be able to compute, approximate, and estimate the values of numeric expressions using exponents (including fractional exponents), absolute value, order of magnitude, and ratios. They should be able to apply proportional reasoning, when necessary, to solve problems in nonroutine settings, and to understand the effects of changes in scale. They should be able to predict how transformations, including changes in scale, of one quantity affect related quantities.

These students should be able to write equivalent forms of algebraic expressions, including rational expressions, and use those forms to solve equations and systems of equations. They should be able to use graphing tools and to construct formulas for spreadsheets; to use function notation; and to evaluate quadratic, rational, piecewise-defined, power, and exponential functions. At this level students should be able to recognize the graphs and families of graphs of these functions and to recognize and perform transformations on the graphs of these functions. They should be able to use properties of these functions to model and solve problems in mathematical and real-world contexts, and they should understand the benefits and limits of mathematical modeling. Twelfth-graders performing at the *Proficient* level should also be able to translate between representations of functions, including verbal, graphical, tabular, and symbolic representations; to use appropriate representations to solve problems; and to use graphing tools and to construct formulas for spreadsheets.

Students performing at this level should be able to use technology to calculate summary statistics for distributions of data. They should be able to recognize and determine a method to select a simple random sample, identify a source of bias in a sample, use measures of center and spread of distributions to make decisions and predictions, describe the impact of linear transformations and outliers on measures of center, calculate combinations and permutations to solve problems, and understand the use of the normal distribution to describe real-world situations. Twelfth-grade students should be able to use theoretical probability to predict experimental outcomes involving multiple events.

These students should be able to solve problems involving right triangle trigonometry, use visualization in three dimensions, and perform successive transformations of a geometric figure in a plane. They should be able to understand the effects of transformations, including changes in scale, on corresponding measures and to apply slope, distance, and midpoint formulas to solve problems.

Advanced

<u>Summary of Grade 12 *Advanced* Level Achievement</u>: Twelfth-grade students performing at the *Advanced* level should demonstrate in-depth knowledge of and be able to reason about mathematical concepts and procedures. They should be able to integrate this knowledge to solve nonroutine and challenging problems, provide mathematical justifications for their solutions, and make generalizations and provide mathematical justifications for those generalizations. These students should reflect on their reasoning and they should understand the role of hypotheses, deductive reasoning, and conclusions in geometric proofs and algebraic arguments made by themselves and others. Students should also demonstrate this deep knowledge and level of awareness in solving problems, using appropriate mathematical language and notation.

Students at this level should be able to reason about functions as mathematical objects. They should be able to evaluate logarithmic and trigonometric functions and recognize the properties and graphs of these functions. They should be able to use properties of functions to analyze relationships and to determine and construct appropriate representations for solving problems, including the use of advanced features of graphing calculators and spreadsheets.

These students should be able to describe the impact of linear transformations and outliers on measures of spread (including standard deviation), analyze predictions based on multiple data sets, and apply probability and statistical reasoning to solve problems involving conditional probability and compound probability.

Twelfth grade students performing at the *Advanced* level should be able to solve problems and analyze properties of three-dimensional figures. They should be able to describe the effects of transformations of geometric figures in a plane or in three dimensions, to reason about geometric properties using coordinate geometry, and to do computations with vectors and to use vectors to represent magnitude and direction.