

National Assessment Governing Board

Committee on Standards, Design and Methodology

August 6, 2010

COSDAM Attendees: Lou Fabrizio (Chair), Steven Paine (Vice Chair), Tonya Miles, James Popham, Andrew Porter, and Darwin Winick.

Other Governing Board Members: John Q. Easton (*Ex officio*), Director of the Institute of Education Sciences.

Governing Board Staff: Susan Loomis and Ray Fields.

Other Attendees: NCES: Stuart Kerachsky (Deputy Commissioner of the National Center for Education Statistics), Jonathan Beard, Janis Brown, Samantha Burg, Steve Gorman, Elvie Hausken, Andrew Kolstad, Kashka Kubzdela, and Drew Malizio. AIR: George Bohrnstedt. ACT: Nancy Petersen. CRP, Inc.: Jasmine Fletcher and Kathy Smoot. Education Week: Catherine Gewertz. ETS: Steve Lazer, Andreas Oranje, and Mary Pitoniak. Hager Sharp: Joanne Lim. HumRRO: Laress Wise. MetaMetrics: Heather Koons. Pearson: Brad Thayer. Westat: Nancy Caldwell and Bob Patchen.

Lou Fabrizio, Chair of the Committee on Standards, Design and Methodology (COSDAM), called the meeting to order at 10:05 a.m. and welcomed members and guests. Mr. Fabrizio stated that Carol D'Amico has resigned from the Governing Board, and he expressed regret that she will no longer be a member of COSDAM. Mr. Fabrizio announced that the COSDAM meeting includes a closed session beginning at 11:30 a.m.

1. 12th Grade Preparedness Research Update

Mr. Fabrizio noted that the format for the preparedness update briefing was especially helpful, and complimented staff for that. He then asked Susan Loomis to report on the NAEP grade 12 preparedness research studies. For each type of study, the briefing included the background of the study and an update.

Susan Loomis reported that content alignment studies have been completed for NAEP in reading and in mathematics with WorkKeys, the SAT, and ACCUPLACER. These studies followed a design developed for NAEP by Dr. Norman Webb to promote standardization across studies to be implemented for different assessment programs by different contractors. Ms. Loomis noted that these studies were designed to evaluate the alignment of two assessments, rather than the alignment of an assessment to curriculum standards; and the studies were designed as bi-directional studies to maximize the information produced about the content overlap between the two assessments for each study. In addition to these studies for the 2009 NAEP, a content alignment study was designed and conducted by ACT, Inc. for the ACT and NAEP in reading and mathematics before the content alignment design document was developed. Development of

final reports is underway, and reports will be available for COSDAM in November 2010 after reviews by staff and the Technical Advisors for 12th Grade NAEP Preparedness Research.

Studies to establish statistical relationships for the national sample of grade 12 NAEP with SAT for reading and mathematics and with the state-representative sample of grade 12 NAEP for Florida students are also continuing. Mr. Fabrizio reminded the Committee that these are not equating studies, and Ms. Loomis stated that the strongest statistical relationship possible will be established with the available data. The statistical analyses are being conducted by ETS NAEP research analysts. Ms. Loomis reported that Florida WorkKeys data have been transferred by the state so that analysis can begin now. The first round of post-secondary updates will be added by the end of the calendar year to include information for the Florida sample about college course placement and grades for students who were 12th graders in 2009 and later entered college.

Ms. Loomis reported on some of the study findings and concerns that had been discussed by the Technical Advisors for 12th Grade Preparedness Research in June, and the Committee had several questions in response to the information about the statistical studies. In particular, Ms. Loomis noted that the lower correlation between the reading NAEP and other assessments of reading was a topic of discussion by the Technical Advisors and that further research to explore this finding are planned. Mr. Fabrizio clarified that the “unbalanced” design of the NAEP reading assessment refers to the fact that students respond to questions about two reading passages that may be of different types—informational, literary, and so forth. John Easton asked for more information about how the correlation studies are conducted without individual student scores for NAEP. Andreas Orange, an ETS NAEP research analyst for the project, stated that the average of the correlations across the five plausible values is used to compute the overall relationship of performance between NAEP and the other assessment. Research into those relationships will continue and focus on analysis of outliers, as well as further analysis of the impact of the “unbalanced” design of the NAEP reading assessment.

Ms. Loomis reported that a Request for Proposals has been issued for judgmental standard setting studies to set cut scores to represent preparedness for placement in college-level credit-bearing courses in mathematics and in reading and for entry in job training courses in up to five occupations. Using criteria recommended by the Technical Panel for 12th Grade NAEP Preparedness Research, a preliminary set of 20 occupations has been identified from which 5 occupations will be selected for the studies to set cut scores for the 2009 NAEP. A design document has been prepared to guide the work and provide for standardization across studies that may be conducted by different contractors.

Jim Popham cautioned that the standard setting studies for 12th grade preparedness may be quite different from the NAEP achievement levels-setting studies. He urged staff to anticipate the unanticipated.

Andrew Porter noted that the NAEP definition of preparedness is different from the definition used by Achieve and others in that NAEP is not saying that prepared for work is prepared for college. Ray Fields responded that the operational definition of preparedness was part of the policy adopted by the Governing Board in March 2009, and the Board policy does not assume that workplace preparedness and college preparedness are the same. Mr. Porter noted that the

type of careers included in the definition of workplace preparedness can determine the level of preparedness required. Mr. Fields added that the Board's focus is on eligibility for job training courses in specific occupations and not for direct placement in a job. Darv Winick commented that the level of investment that a company must make for training an employee is an important issue in hiring, and that training period and resource investment seem to be taken into consideration in the NAEP research plans.

Ray Fields reported on a survey of two-year and four-year post-secondary institutions to gather information regarding (1) the placement tests used and (2) the cut scores in reading and mathematics below which remedial reading and mathematics course placement results, and at or above which placement in credit-bearing entry level courses results. A pilot study of 100-120 postsecondary institutions is scheduled for administration within the next few weeks; the specific start date is contingent upon receiving Office of Management and Budget (OMB) clearance. After that, adjustments will be made to the survey and the operational survey will be administered to a nationally representative sample of institutions.

Committee members were very positive about the potential benefits and contributions of this study to the NAEP preparedness research and to the larger education community. In response to Mr. Fabrizio's question, Mr. Fields affirmed that the selectivity of institutions will be collected for analysis of responses. Mr. Fields noted that although the focus of the survey seems completely simple and straightforward, many questions have arisen regarding the survey: Who should be the respondent? How can we be sure that the correct person is the respondent? Who is the "target" student for which the cut score is to be applied? How does the 2-year institution differ from the 4-year? Mr. Fields expects to have preliminary information from the pilot study to share with COSDAM in November 2010.

The Governing Board and the National Center for Education Statistics (NCES) are collaborating on the implementation of a benchmarking study with entering students in Texas higher education institutions, both 2-year and 4-year. The study is a collaborative project of the Governing Board and NCES. Westat, the NCES sampling and administration contractor for NAEP, has conducted interviews to collect more information at each of the 9 institutions in Texas that have volunteered to participate in the pilot for this project. Mr. Fields reported that these institutions all seem very enthusiastic and cooperative. The opportunity for the study developed as a result of discussions with Texas Higher Education Commissioner Raymund Paredes who is a member of the Board's 12th Grade Preparedness Commission. If the results of the pilot study indicate that administering NAEP to college freshmen is feasible, a full-scale study will be implemented in the fall of 2011 on a much larger state-representative sample of institutions.

Tonya Miles noted that this study has the potential to identify learning needs and deficits that will help to develop strategies to better prepare students for college. Mr. Porter noted the importance of having administration procedures carefully planned and implemented to assure the ability to distinguish between placement in courses that contribute to fulfillment of general education requirements and remedial courses that are pre-requisite to such course placement.

Mr. Fields next presented a model for a preparedness validity framework and asked members to provide feedback. The framework consists of statements about preparedness that the Board may

wish to make, followed by propositions regarding relationships that would be needed to support the statement and evidence of the relationships. This is a model and the statements, propositions and evidence are all hypothetical at this time and presented only as examples of the type of information that might be presented in such a validity argument. COSDAM gave positive feedback to the overall approach and supported development of the framework. Mr. Porter asked if there is a way to support the statement that “X % of students are prepared” without having individual student scores. Mr. Popham recommended that we state in advance the outcomes that will be considered as evidence of “preparedness.” The discussion also led to the suggestion that the evidence evaluation be conducted by a third party. Staff will continue to work on this preparedness validity framework and provide more information to COSDAM in November 2010.

2. Writing NAEP Procurement and Plans for Preliminary Research

Ms. Loomis reported that a Request for Proposals (RFP) for writing NAEP achievement levels for grades 8 and 12 in 2011 and grade 4 in 2013 has been issued, and proposals are due August 10, 2010. Ms. Loomis noted that the key features of the statement of work for this procurement were shared with COSDAM before the release of the RFP. This is the first wholly computer-based assessment for NAEP, and this provides the opportunity to include computers in the achievement levels-setting (ALS) process. The plan is to conduct a small-scale field trial to test logistics of working with computers; a pilot of the process planned for achievement levels-setting; and operational achievement levels-setting study. Proposals are due August 10, 2010, and the award will be made by the close of the fiscal year September 30, 2010.

Ms. Loomis reported that there is neither a linking study to relate performance based on the current writing NAEP assessment framework to the new assessment framework nor a scaling analysis study to provide score data prior to the operational assessment. Thus, she requested that a team be formed to plan and conduct research to inform the writing ALS process to the extent possible using information from the pilot study for grades 8 and 12. She noted that the same conditions had held for the science 2009 NAEP, and she acknowledged Mr. Porter’s advice during the deliberations of science ALS results as a catalyst for organizing this program of research.

In addition to Ms. Loomis, the research team includes Mary Crovo of the Governing Board staff who directed development of the writing framework and oversees the development of the item pool, Andrew Kolstad and Steve Gorman of NCEES and the NAEP research analysts at ETS. Andrew Kolstad reported on the three types of studies that are currently planned. Study 1 includes analyses to compare 2010 writing pilot study data to data from the most recent writing assessment in 2007. Mr. Kolstad noted that preliminary information developed by ETS project staff indicates that performance on the 2010 writing assessment was lower in comparison to the 2007 writing assessment. For example, the modal response in 2010 was for score category 2, but for 2007 category 3 was the modal response. He cautioned that it is not possible to distinguish whether the 2010 assessment is harder, the rubrics more stringent, or if student achievement has declined. But, the pattern of data indicated that more student responses were scored in the lower rubric value ranges for the 2010 assessment than for the 2007 assessment in grades 8 and 12, and this pattern was especially clear for the grade 8 data.

Study 2 includes examination of information from the computer database collected from the student assessment. This information will be examined to identify any prompts that may have caused particular problems for students in terms of their interaction with the computer.

Study 3 involves creation of a data file to simulate a NAEP writing scale. Mr. Kolstad cautioned that this would require strong assumptions using data from the 2010 pilot, along with data from the 2007 assessment to develop the simulated scale. The simulated data will be used to evaluate alternative hypothetical outcomes of an ALS process, such as the impact of using the cut scores established in 1998 for writing, using 2007 impact data to derive cut scores, and using the percentage of total possible points associated with the 1998 cut scores.

Mr. Popham again advised that staff consider ways of having impact data introduced in the writing ALS process so that panelists will take it into account more fully. He recommended that research be incorporated in the writing ALS process to determine when and how to introduce data to have a greater impact. Mr. Porter noted his observations as a technical advisor to state assessment programs that successful standard setting procedures require that the assessments be centered on student ability. Having assessments that are centered at a higher level than student performance tends to result in cut scores that are deemed to be unreasonable.

CLOSED SESSION 11:30 a.m. – 12:20 p.m.

COSDAM Attendees: Lou Fabrizio (Chair), Steven Paine (Vice Chair), Tonya Miles, James Popham, Andrew Porter, and Darvin Winick.

Other Governing Board Members: John Q. Easton (*Ex officio*), Director of the Institute of Education Sciences.

Governing Board Staff: Cornelia Orr (Executive Director), Susan Loomis, and Michelle Blair.

Other Attendees: NCES: Stuart Kerachsky (Deputy Commissioner), Peggy Carr (Associate Commissioner), Jonathan Beard, Janis Brown, Steve Gorman, Elvie Hausken, Andrew Kolstad, and Drew Malizio. ACT: Nancy Petersen. AIR: George Bohrnstedt. ETS: Steve Lazer, Andreas Oranje, and Mary Pitoniak. Hager Sharp: Joanne Lim. HumRRO: Laress 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 August 6, 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.

Mr. Fabrizio asked Ms. Loomis to distribute the embargoed briefing materials. He noted that with the exception of one table, to which additional data were added, these were the same briefing materials that had been shared with COSDAM for a conference call on August 3. Mr. Fabrizio reviewed the decision by the Board in May 2010, based on the COSDAM recommendation, to delay action on the science achievement levels to provide more time for staff to collect additional information and continue evaluation of data. The Committee asked for a staff recommendation for science achievement levels for approval at the August 2010 meeting.

Mr. Fabrizio reviewed each table in the hand-out and provided the opportunity for Committee members to ask questions and make comments regarding the data. The first table included data

for the science 2009 cut scores for each grade and level, the percentage of total possible points associated with each cut score, and the percentage of students scoring at or above each level. The second table included data for the science NAEP for each assessment year, starting with the data for 1996 when achievement levels were set for science. The third table included impact data related to the cut scores for all NAEP subjects for which achievement levels have been set. (Table 3 had originally included data for mathematics and reading only.) The fourth table provided information for comparing NAEP science results to estimates of NAEP scores based on TIMSS performance and Massachusetts state assessment data. Finally, a listing of all sources of data collected as evidence of procedural validity, external validity, and other types of data was provided. Ms. Loomis had prepared copies of the Governing Board's policy on setting performance standards for review by the Committee, and Mr. Fabrizio brought the Committee's attention to the fact that the policy includes numerous references to the Board's role of setting cut scores based on the recommendations of the ALS panelists. In addition, references to the requirement for producing evidence of procedural validity were noted. Mr. Popham recommended that the term "procedural validity" be deleted from the policy and characterized it as an "unsanctioned construct."

The Committee discussed many aspects and considerations related to the decision for setting the NAEP science cut scores. They thanked staff for help in collecting information and for the staff recommendations.

Mr. Fabrizio announced the end of the Closed Session at 12:20 p.m.

ACTION

The meeting of the Committee on Standards, Design and Methodology continued in open session. Mr. Paine moved that the Committee approve the cut scores, achievement levels descriptions, and exemplar items recommended by staff for reporting the performance of 4th, 8th, and 12th grade students on the 2009 National Assessment of Educational Progress for science. Mr. Winick seconded the motion. There was no further discussion, and the motion carried unanimously.

The Committee adjourned at 12:30 p.m.

I certify the accuracy of this report.

Louis M. Fabrizio

Lou Fabrizio, Chair

August 11, 2010

Date