

National Assessment Governing Board

Committee on Standards, Design and Methodology

Report of August 5, 2011

COSDAM Attendees: Lou Fabrizio (Chair), Tonya Miles (Vice Chair), Governor Jack Markell, James Popham, and Andrew Porter.

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

Governing Board Staff: Executive Director Cornelia Orr, Susan Loomis, Michelle Blair, and Ray Fields.

Other Attendees: NCES: Jonathan Beard, Janis Brown, Eunice Greer, Andrew Kolstad, Drew Malizio, and Grady Wilburn. AIR: George Bohrnstedt. Kansas Department of Education and representative of CCSSO-NAGB Policy Task Force. Tom Foster. Office of Delaware Governor Markell: Jennifer Ranji. Data Recognition Corporation: Lara Milavickas. ETS: Steven Lazer, Rebecca Moran, and Andreas Oranje. Fulcrum IT: Jud Cole and Lori Rukus. Hager Sharp: Melissa Spade. HumRRO: Laress Wise. Measured Progress: Luz Bay. NAEP ESSI: Burhan Ogut. Pearson: Douglas Kuback and Connie Smith. U.S. Department of Education (retired): Jay Noell. Westat: Nancy Caldwell and Keith Rust.

Lou Fabrizio, Chair of the Committee on Standards, Design and Methodology (COSDAM), called the meeting to order at 10:30 a.m. and welcomed members and guests. Mr. Fabrizio noted that Leticia van de Putte had been called back to Texas for the birth of her grandson, and Blair Taylor had to cancel plans to attend the Board meeting, at the last minute, due to family illness. Special welcomes were extended to Andy Porter and Governor Markell who were unable to attend the May meeting.

1. ACTION ITEM

Writing Achievement Levels Descriptions for the National Assessment of Educational Progress (NAEP)

Mr. Fabrizio noted that the achievement levels descriptions require provisional approval of COSDAM for use in the achievement levels-setting studies, but official action by the Governing Board will not be taken until May 2012 when achievement levels for the 2011 NAEP for writing are presented for approval.

Susan Loomis provided a brief overview of the process for developing the writing achievement levels descriptions (ALDs) to be used for setting achievement levels for the 2011 NAEP for grades 8 and 12 and the 2013 NAEP for grade 4 in writing and for reporting results of the writing NAEP starting with the 2011 assessments. Ms. Loomis noted that the process included extensive reviews by writing content experts and edits by a small group of persons who helped develop the

NAEP writing framework. Ms. Loomis stated that the preamble was a recommendation of the panel of writing experts working with the contractor; and, she noted that the inclusion of a preamble is consistent with previous writing achievement levels descriptions.

Andrew Porter asked if the achievement levels descriptions for writing were evaluated relative to those of other subjects. Ms. Loomis said that this is not part of the process for developing NAEP achievement levels descriptions, although this can be added to the process. Mr. Porter recommended that this be added, and Jim Popham agreed. Mr. Fabrizio instructed that this suggestion be implemented in the future.

The following draft motion was distributed to members.

The Committee on Standards, Design and Methodology gives provisional approval for the use of the following achievement levels descriptions for the National Assessment of Educational Progress in Writing in procedures to set achievement levels for the 2011 assessments at grades 8 and 12 and the 2013 assessment at grade 4. Final approval will be determined after review of the analysis of results from the achievement levels-setting processes are completed and as part of the official approval of the achievement levels by the National Assessment Governing Board.

Mr. Popham moved approval of the motion, and Tonya Miles seconded the motion. There was unanimity in the voice vote of approval. A copy of the achievement levels descriptions, including the preamble, is attached to this report.

2. Update on 12th Grade Preparedness Research Program

The first part of the update on the Governing Board's 12th grade preparedness program of research was a focus on the statistical analysis of the relationship between NAEP and assessments that serve as indicators of preparedness at the national level and for the NAEP state-representative sample of Florida 12th graders in 2009. Mr. Fabrizio asked Andreas Oranje of ETS to present the report of the statistical analyses.

Mr. Oranje provided a brief review of the analyses of national level statistical relationships for NAEP and the SAT in reading and mathematics, presented to COSDAM in May 2011. He next turned to analyses focusing on the state-representative sample of Florida students. Florida was one of 11 states to participate in the first-ever state-level NAEP assessment of 12th graders in 2009. While the Florida data have not undergone all of the statistical treatments that will ultimately be investigated, Mr. Oranje noted that the correlations between NAEP and SAT performances for Florida students were remarkably close to the national level correlations for both mathematics and reading. As was the case for the national data, correlations for reading were substantially lower than for mathematics, and that was true for both the SAT-NAEP and ACT-NAEP data. While several hypotheses for the lower correlations for reading had been investigated, Ms. Loomis noted that no clear reason for this relationship had emerged in the discussions by the Technical Advisors for 12th Grade Preparedness Research of the analyses presented to date.

Benchmark scores to indicate college and career readiness by the ACT and first year success by the SAT showed relationships to Florida 12th graders' performance on NAEP indicating that these benchmarks are approximately the same as the grade 12 NAEP proficient cut scores for each subject. The performance of students in Florida at or above the Proficient level is slightly lower than the national level in each subject. Mr. Oranje noted that the demographic data for Florida differ from the national demographic profile. The percentage of both Hispanic and African-American students in Florida is higher than for the nation, and the average NAEP scores for these two demographic groupings of Florida students is higher than for the nation.

Of particular interest were the data presented for Florida 12th graders in 2009 who subsequently attended college. The Florida Department of Education has provided data to the Governing Board for students who attended public colleges and universities in Florida in 2009-10. Ms. Loomis anticipates adding data on attendance in private and out-of-state institutions in the future. Mr. Oranje reported that post-secondary data on first year course-taking were available for approximately 54% of the 2009 NAEP 12th grade sample of Florida students.

The average NAEP scores in both reading and mathematics for Florida students in the 2009 grade 12 NAEP sample who enrolled in college in 2009-10 was lower than the benchmark "readiness" score for the ACT and first year success score for the SAT: 153 for mathematics and 290 for reading. Average NAEP scores for the sample of students in the grade 12 mathematics NAEP who attended college in 2009-10 and took no remedial courses was 30 points higher than for students taking one remedial course and 33 points higher than for students taking two or more remedial courses in their first year of college. A similar pattern was found for the reading sample: the average grade 12 NAEP reading score for Florida students who took no remedial course in the first year of college was 27 points higher than the average for students taking one remedial course and 32 points higher than the average for students taking two or more remedial courses.

In general, the analyses with ACT and SAT data indicate the same performance score ranges on NAEP for the "readiness" and "success" benchmarks suggested by these two college admissions testing programs. The data for students in their first year of college are also consistent with the overall findings of NAEP preparedness.

Additional analyses are to be conducted for the Florida data and presented to a meeting of the Technical Advisors for 12th Grade Preparedness Research in October 2011. Further, Governing Board staff anticipate additional data from the Florida Department of Education (FLDOE) for the second post-secondary year for students in the 2009 12th grade NAEP samples. Ms. Loomis will contact the FLDOE to ascertain when data for students attending private and out-of-state post-secondary institutions will be available.

The second part of the 12th grade preparedness research update was for the recently-completed judgmental standard setting studies to set cut scores on the grade 12 NAEP in reading and mathematics. These studies are an important part of the studies for 2009 reporting in that they are the only source of information about performance on NAEP by students in job training programs. They involve judgments by job trainers of how students minimally prepared for entry in the job training programs would perform on NAEP.

Mr. Fabrizio stated that because an interim meeting of COSDAM via webinar is planned for November 2011 to review the complete results for the judgmental standard setting (JSS) studies, the 15 minutes cut from the Committee meeting by time-overruns of the general session of the Governing Board would be taken from the update on the JSS.

Ms. Loomis and Luz Bay, the project co-director from Measured Progress, provided information to COSDAM on these studies. The design of the JSS studies called for replicate panels, and both panelists and the item pool were divided to be as similar as possible and assigned to each paired group. A total of eight panels were convened concurrently for each JSS session: replicate panels for each subject, mathematics and reading, for each of two post-secondary activities at each session. JSS studies were implemented for job training programs in five different occupations (automotive master technicians; computer support specialists; heating, ventilation, and air conditioning technicians; licensed practical nurses; and pharmacy technicians) and for placement of college level credit bearing courses that fulfill general education requirements.

Ms. Loomis noted that the replication for the paired panels was not as substantial as anticipated. This was the first time that truly independent replicate panels had been used for NAEP standard setting, and there was no baseline for comparison of the level of replicability for these panels to other NAEP panels. The cut scores were mostly in the score range between the Basic and Proficient NAEP achievement levels cut scores. While panelists seemed to have a good overall understanding of the process and the tasks they were to perform, there was some concern about the content knowledge and requirements of the job training programs and the way the content is assessed in NAEP.

Ms. Loomis discussed the concerns raised by comments of panelists and observations by content facilitators, process facilitators, and technical advisors on site that the panelists often had difficulty in finding meaningful relationships between the requirements of their job training programs and the academic knowledge and skills assessed in grade 12 NAEP. One clear indicator of that difficulty was related to the need to change the procedures for producing descriptions of the knowledge, skills, and abilities (KSAs) required for correctly answering each item. These descriptions are a key part of panelists' training in a bookmark standard setting process. The computerization of this standard setting process provided easy access to the KSAs written by panelists, and it was possible to review the KSAs for each panelist. For the pilot study, it was apparent that changes to instructions for this task were needed because many panelists had simply described the items as "hard/easy" or "relevant/irrelevant." The process of writing KSAs was changed for each operational workshop to provide more and more content support to panelists. For the final operational workshop, the KSAs were provided to panelists for their consideration and use. The plan is to conduct an analysis of the KSA descriptions across the workshops, and to determine the extent to which the KSAs provided to panelists in the final workshop were modified by panelists or simply used verbatim. Ms. Loomis explained that the goal of providing the KSAs to the panelists was to assure that they had an accurate indication of the knowledge, skills, and abilities required by the items so that panelists would have a better informed judgment about the placement of their bookmarks.

The difficulties in producing KSAs and the anecdotal evidence collected across the workshops led to a decision to add a special study to the end of the final workshop. That study was to

collect information from panelists about the items that they judged to be irrelevant to minimal preparedness for their job training program as a technician in heating, ventilation and air conditioning (HVAC) or as computer support specialist. Items were arrayed on the item maps according to the content groupings of the framework (two subscales for reading and 4 for mathematics). Panelists were also asked to place a bookmark to represent minimal preparedness in each content area of the subject. Ms. Loomis stated that two facilitators failed to implement the study appropriately, but the data that are available will be analyzed to determine if there is a relationship between "irrelevant" items and bookmark placements and whether the "irrelevant" items impacted placement of the bookmarks.

Cornelia Orr noted that plans are underway to collect more information about the course content of the job training programs to help with evaluating the results of these studies.

Mr. Popham stated that in his opinion, having results that are not exactly consistent with what one expects or hopes to find lends more credibility to the overall findings.

Next, Mr. Fabrizio asked Ray Fields to report on the Higher Education Survey which will be distributed to nationally representative samples of 2-year and 4-year colleges in mid-August. Mr. Fields said that the survey distinguishes between policy for admissions to higher education institutions and placement into entry-level credit-bearing courses. He stressed that the survey is to collect information on course placement tests and the scores on those tests required for placement in college courses that are remedial and those that are credit-bearing courses. This purpose is in contrast to the statistical analysis studies to determine the score on NAEP representing the benchmark scores on ACT and SAT for college readiness/success, based on predicted first-year college performance. Sample stratification will provide the opportunity to evaluate placement scores by level of selectivity to determine the extent to which course *placement* requirements differ according to college *admissions* requirements. Preliminary survey results are expected in December 2011.

Mr. Fields next provided an update on the validity framework for preparedness research being developed to guide the evaluation of preparedness research study findings and to serve as the basis for structuring the validity argument for statements or inferences about preparedness that might be made in NAEP reports. Mr. Fields is taking the lead in development of this document. He reviewed the purpose of the validity framework and reported that he has received evaluations and recommendations from three individuals/groups already: Michael Kane, a leading expert in validity research; Chad Buckendahl, project director for the last NAEP evaluation study; and a sub-group of the NAEP Validity Studies Panel. Those comments will be studied further and modifications made to the framework, as appropriate. The plan is to submit the validity framework for broad review and vetting as well as to COSDAM and the Governing Board. Mr. Fields stressed that the goal of the Governing Board is to report preparedness on NAEP, but only if there is sufficient evidence in support of the validity of doing so. The Board's position has been clear from the start that the program of research is aimed at collecting validity evidence for reporting preparedness, and the statements or propositions about preparedness that will be reported will be those for which the evidence is judged to be clear and sufficient. The validity framework is being developed to share publically the statements that the Board would like to make about preparedness and the propositions and evidence that would be required to support

those statements. The statements in the document shared with COSDAM are illustrative of the statements to be included in the validity framework when it is finalized.

Mr. Porter had strong reservations about including both a statement about placement in remedial college courses and a statement about being prepared for placement in entry-level credit-bearing courses that lead to a BA degree. He suggests that only one of these two statements be put forward OR that something that he has used (red light, green light, or yellow light) be used to signal “clearly prepared,” “clearly needs remediation,” or “preparedness not clear.”

Ms. Orr noted the need for clear statements that can be understood by the general public and stressed that nuanced statements that might lead to confusion or inaccurate interpretations will be avoided for reporting results. Mr. Fabrizio asked staff to review the propositions and develop recommendations for reporting that will be clearly understood.

The final part of the 12th grade preparedness research update was a brief report by Michelle Blair on plans for developing and producing the preparedness reports. The procurement process is beginning for development of two reports on preparedness: (1) a report for the general public to summarize findings on preparedness and (2) a technical report on the research studies and evidence in support of reporting and of the interpretations of findings. Plans are for release of these reports in the early part of 2012.

3. COSDAM Recommendations for Future Topics

Mr. Fabrizio invited COSDAM members to recommend topics of interest for future COSDAM discussion. Mr. Popham expressed his interest in having COSDAM explore the possibility of research on whether NAEP is sensitive to instruction and how NAEP can be (more) instructionally relevant. Staff will distribute a report by Shepard and others on the NAEP Validity Study Panel that shows that the mathematics NAEP is sensitive to instruction. Mr. Porter noted a report by a former student that provides a methodology for evaluating evidence of whether items are instructionally sensitive.

Mr. Porter suggested that COSDAM should push forward the frontiers of psychometrics to do an *existence proof* of computer adaptive assessments for NAEP. He stated that NAEP should take the lead in implementing this technology with large-scale assessments and expressed his personal concerns about the ability of the two assessment consortia to lead the way in this area.

CLOSED SESSION

On Friday, August 5, 2011 the Committee on Standards, Design and Methodology met in closed session from 12 noon to 12:30 p.m. to receive a briefing on preliminary findings from the computer-based writing NAEP for grades 8 and 12 in 2011. The Committee was provided with embargoed results that could not be discussed in an open meeting prior to their official release. Premature disclosure of data would significantly impede implementation of the NAEP assessment, and is therefore protected by exemption 9(B) of section 552b(c) of Title 5 U.S.C.

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

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

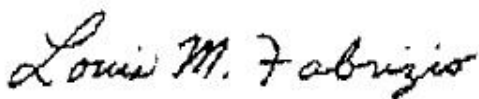
Governing Board Staff: Executive Director Cornelia Orr and Susan Loomis.

Other Attendees: NCES: Jonathan Beard, Janis Brown, Eunice Greer, Andrew Kolstad, Drew Malizio, and Grady Wilburn. AIR: George Bohrnstedt. Office of Delaware Governor Markell: Jennifer Ranji. ETS: Steven Lazer, Rebecca Moran, and Andreas Oranje. Fulcrum IT: Jud Cole and Lori Rukus. Hager Sharp: Melissa Spade. HumRRO: Laress Wise. Measured Progress: Luz Bay. Pearson: Douglas Kuback and Connie Smith. Westat: Nancy Caldwell and Keith Rust.

Andrew Kolstad, Senior Research Scientist of NCES, provided a report on the 2011 writing NAEP which is the first wholly computer-based assessment in NAEP. One focus of the report was universal design features, and Mr. Kolstad noted that these features reduced the incidence and costs associated with providing separate accommodations. He reported on the kinds of editing features used by students and indicators of how long students worked on producing their responses to each of two writing prompts.

The meeting was adjourned at 12:30 p.m.

I certify the accuracy of this report.



Lou Fabrizio, Chair

August 11, 2011

Date

Provisional Achievement Level Descriptions for the National Assessment of Educational Progress in Writing (2011/2013)

PREAMBLE

The 2011/2013 NAEP Writing Framework encourages students to draw from a wide variety of approaches to thinking and writing when deciding how to develop and organize their ideas in relation to audience and purpose. For example, students may choose to use narrative to support the presentation of an argument or analysis and synthesis to convey ideas in a narrative. Responses will be evaluated for their use of effective approaches in the development and organization of their ideas.

Grade 4

BASIC

Fourth grade students writing at the Basic level develop ideas with a few supporting details, reasons, or examples relevant to the topic, the writer's purpose, and the audience. Students may loosely organize their writing using one or more approaches, such as chronology, steps in a process, or main idea with some details. Word choice is generally clear and appropriate for the topic. Most sentences are simple and show little variety in structure. Students demonstrate some control of grammar, usage, and mechanics, including capitalization, punctuation, and spelling, although there may be distracting errors that occasionally impede readers' understanding.

PROFICIENT

Fourth grade students writing at the Proficient level develop ideas using some well-chosen details, reasons, or examples relevant to the topic, purpose, and audience. Students organize their writing using logical groupings and relationships among ideas, such as chronology, steps in a process, or main idea and details. The presence of some variety in word choice enhances the presentation of the writer's ideas and maintains the attention of the audience. Students vary some sentences in length or type, for example, by using exclamations or questions to engage the audience. Students demonstrate overall control of grammar, usage, and mechanics, including capitalization, punctuation, and spelling, but their writing may contain errors that detract from meaning.

ADVANCED

Fourth grade students writing at the Advanced level develop ideas using well-chosen details, reasons, or examples relevant to the topic, purpose, and audience. Their writing is organized effectively and shows an overall logical progression of ideas. Precise and varied word choice enhances the presentation of the writer's ideas and successfully sustains the attention of the audience. Students vary the length and complexity of sentences to engage the reader. Students' writing demonstrates control of grammar, usage, and mechanics, including capitalization, punctuation, and spelling, with relatively few errors.

Grade 8

BASIC

Eighth grade students writing at the Basic level demonstrate an ability to write a response that is appropriate to the topic, purpose, and audience. They develop their ideas with some relevant details, reasons, or examples. Writers at this level often create a loosely organized structure with some transitions. Their writing may contain some sentence variety, and clear word choices, as well as minimal evidence of voice and tone that may be appropriate for the topic, purpose, and audience. Students demonstrate an overall command of grammar, usage, and mechanics, including capitalization, punctuation, and spelling, although there may be errors, at least some of which impede readers' understanding.

PROFICIENT

Eighth grade students writing at the Proficient level demonstrate competence in writing a response appropriate to the topic, purpose, and audience. They develop their ideas with some well-chosen details, reasons, or examples. Writers at this level create an organized structure that shows a mostly logical, coherent progression of sentences and ideas. Their writing contains some sentence variety and precise word choices, and voice and tone are usually appropriate for the topic, purpose, and audience. These writers demonstrate an overall command of grammar, usage, and mechanics, including capitalization, punctuation, and spelling, although their writing may contain some errors.

ADVANCED

Eighth grade students writing at the Advanced level demonstrate strategic composing in addressing the topic, purpose, and audience. They develop their ideas with precise, well-chosen details, reasons, or examples that elaborate and extend those ideas. Writers at this level create a well-organized, effective piece of writing that shows a logical, coherent, and well-elaborated progression of sentences and ideas. Their writing contains varied and well-crafted sentences, precise word choices, and uses voice and tone that support the topic, purpose, and audience effectively. These writers demonstrate a strong command of grammar, usage, and mechanics, including capitalization, punctuation, and spelling, although there may be a few errors.

Grade 12

BASIC

Twelfth grade students writing at the Basic level demonstrate an ability to write a response that is appropriate to the topic, purpose, and audience. They develop their ideas with some details, reasons, and examples. Writers at this level may create a somewhat loose organizational structure with some transitions. The writing contains some sentence variety or complexity, though it may be inconsistently executed. Although word choice may be somewhat limited in variety and richness of expression, it generally conveys meaning clearly. Tone may be ineffective for the purpose and audience, fluctuating inconsistently in places. Writing at the Basic level reflects a general command of grammar, usage, mechanics, and spelling but may include a number of errors.

PROFICIENT

Twelfth grade students writing at the Proficient level produce a competent response to the topic that addresses a specific audience and states a clear purpose. They develop their ideas with some well-chosen details, reasons, and examples that clearly address the audience. Writers maintain focus on the topic and purpose, and make relationships among ideas clear. Overall, sentences are varied and structured to clarify the relationships among ideas, as well as to maintain the reader's interest. Precise word choices reinforce the purpose of the writing and are appropriate for the topic and audience. The tone is appropriate for both the purpose for writing and the audience. Writing at the Proficient level exhibits solid control of grammar, usage, mechanics, and spelling, although there may be some errors.

ADVANCED

Twelfth grade students at the Advanced level demonstrate flexible and strategic composing in addressing the topic, purpose, and audience. Well-chosen details, reasons, and specific examples support, extend, and deepen their ideas. Writers develop an organizational structure that supports a natural, logical, and coherent progression of ideas. Skillfully crafted sentences enhance the response, reinforce the purpose, and engage the audience. Word choices are purposeful, precise, and at times powerful. Appropriate tone clearly conveys the writer's position on the topic and is appropriate for the intended audience. Writing at the advanced level exhibits a strong control of grammar, usage, mechanics, and spelling that supports the presentation of ideas. Errors may be present, but they are few.