

National Assessment Governing Board Committee on Standards, Design and Methodology

May 17, 2013

JOINT MEETING WITH REPORTING AND DISSEMINATION COMMITTEE

Attendees

COSDAM Members: Chair Lou Fabrizio, Vice Chair Fielding Rolston, Andrew Ho, Terry Holliday, Tonya Miles, and James Popham.

Reporting and Dissemination Committee Members: Acting Chair David Driscoll (Chairman of the Governing Board), Anitere Flores, Rebecca Gagnon, Terry Mazany, and Father Joseph O'Keefe;

Governing Board Staff: Executive Director Cornelia Orr, Michelle Blair, Larry Feinberg, Ray Fields, and Stephaan Harris;

Other Attendees: John Easton, Director of the Institute of Education Sciences and Ex Officio Governing Board member; NCES – Commissioner Jack Buckley, Associate Commissioner Peggy Carr, Janis Brown, Arnold Goldstein, Andrew Kolstad, Dan McGrath, and Grady Wilburn; AIR – George Bohrnstedt, Cadelle Hemphill, and Fran Stancavage; CRP – Carolyn Rudd and Edward Wofford; ETS – Amy Dresher, Steve Lazer, Rebecca Moran, and Andreas Oranje; Hager Sharp Communications – David Hoff and Debra Silimeo; HumRRO – Steve Sellman and Laress Wise; Optimal Solutions Group – Robin Marion; Pearson – Connie Smith; Reingold Communications – Amy Buckley; Westat – Rima Zobayan; Widmeyer Communications – Jason Smith; California Department of Education – Julie Williams (NAEP state coordinator); WCER – Gary Cook; Dr. Albert Wilburn, MD.

NAEP Testing and Reporting on Students with Disabilities and English Language Learners

The Committees considered two sets of issues: (1) implementation of the Board policy that deals with testing of English language learners (ELL), and (2) reporting options for exclusions, participation rates, and the adjusted scores, termed full-population estimates, which try to take into account the differences in exclusion rates between the states and districts participating in NAEP.

Grady Wilburn, of NCES, gave an update to the committees on these two topics. He noted that under the policy adopted in 2010, the only ELL students that schools may exclude from NAEP are those who have been in U.S. schools for less than one academic year. Even students in this category should not be excluded if NAEP offers a translation in their home language. Students who speak Spanish now account for about 80 percent of ELLs nationwide. NAEP offers Spanish translations of all its tests (in bilingual booklets) except for reading and writing, which under the frameworks adopted by the Board are reading and writing in English.

Mr. Wilburn said the decision tree incorporating the Board policy on which ELL students to test and how to test them was implemented smoothly in the 2013 NAEP. He said exclusion rates went down.

On the reporting issues Mr. Wilburn noted that a joint meeting of the two committees in March had received a full report on implementation of the policy on SD students. Under this policy the only students that may be excluded from NAEP by school personnel are those with the most significant cognitive disabilities who take alternate state assessments with alternate achievement standards, expected to be about 1 percent of enrollment. For practical reasons NCES decided that schools could also continue to exclude students with an individualized education program (IEP) or 504 plan that provides for accommodations on state tests that NAEP does not allow. The non-allowable accommodations in nearly all cases have been read-aloud on the NAEP reading assessment or calculator use on all sections of NAEP math.

In 2013 for the first time NCES permitted students with IEPs requiring calculator use to take calculator-active blocks on NAEP even if they would have been assigned non-calculator blocks as part of the normal NAEP sampling. Mr. Wilburn said a study in 2011 indicated that this would have little impact on results, and, in any case, the number of students involved is small.

By law, student participation in NAEP is voluntary. Parents can refuse to have their children participate for any reason. Under the Board policy, “students refusing to take the assessment because a particular accommodation is not allowed should not be classified as exclusions but placed in the category of refusals under NAEP data analysis procedures.”

NCES has said doing this would break trends, depress reported scores, and contravene sound psychometric procedures. Under long-standing practice, excluded students are omitted from any calculations of NAEP results, and have no effect on state or district averages. Adjustments are made for refusals or absent students (a much larger group) by reweighting the scores of those with similar characteristics, which tends to lower state and district averages.

There is another analytic procedure, called full-population estimates (FPE), which NCES has used for about a decade to adjust state and district results by imputing scores for excluded SD and ELL students based on the performance of similar SD and ELL students who are tested. Data showing year-to-year changes in the full-population estimates are published on the NAEP website for participating states and districts, but these are given little prominence and do not include the adjusted scores themselves. The FPE scores were provided to the Board at this meeting. They showed most state averages to be about 3 to 6 points lower than reported. In only a few cases were year-to-year changes significantly different.

George Bohrnstedt, of AIR, chair of the NAEP Validity Studies Panel, said his group is concluding a study which shows that FPEs provide less biased results than the current NAEP analysis method, which overstates true scores considerably more.

Another proposal considered for reporting is to publish a total participation rate, based on all students in a sample divided into those tested and not tested for any reason whether excluded,

absent, or refused. At present the reported participation rates are calculated after excluded students are subtracted from the number in the sample.

After considerable discussion, the Committees endorsed the following in regard to the reporting of 2013 NAEP results:

- 1. Continue previous analysis procedures for exclusions and refusals.**
- 2. Give greater prominence and easier accessibility to full-population estimates as part of the information available online at the time of data release.**
- 3. Provide additional information online on the proportion of students excluded because they use an accommodation on state tests that is not allowed on NAEP.**
- 4. Issue a special report after the main data release with a full discussion of exclusion, participation, and refusal issues that includes data on the total participation rate for each jurisdiction in NAEP.**

COSDAM MEETING

Following the joint COSDAM/R&D session, R&D committee members adjourned to their separate meeting room. Lou Fabrizio, COSDAM Chair called the meeting to order at 11:15 a.m.

COSDAM Members: Lou Fabrizio (Chair), Fielding Rolston (Vice Chair), Andrew Ho, Terry Holliday, Tonya Miles, and Jim Popham.

Governing Board Staff: Cornelia Orr, Ray Fields, and Michelle Blair.

Other Attendees: John Q. Easton, Director of the Institute of Education Sciences and Ex Officio member of the Governing Board. NCES: Commissioner Jack Buckley, Associate Commissioner Peggy Carr, Janis Brown, and Andrew Kolstad. AIR: George Bohrndstedt and Fran Stancavage. ETS: Steve Lazer and Andreas Oranje. HumRRO: Laress Wise. Pearson: Connie Smith. Westat: Dianne Walsh. Widmeyer Communications: Jason Smith.

Preliminary Discussion on Setting NAEP Technology and Engineering Literacy (TEL) Achievement Levels

Mr. Fabrizio stated that setting achievement levels on a computer based, interactive and cross-disciplinary assessment such as TEL represents new terrain for educational assessment. Cornelia Orr noted that the Board was in a similar leadership position many years ago when the entire notion of achievement level setting for student assessments was a relatively new concept.

To determine the appropriate methodology to pursue in achievement level setting for the TEL assessment, the Board commissioned the development of an issues paper outlining what needs to be addressed. COSDAM's March 2013 meeting discussion was used to inform the drafting of this paper. The primary purpose of this meeting's discussion was to collect committee members'

perspectives on the issues outlined in the paper and to provide guidance to Board staff in preparing the associated procurement activities.

Mr. Fabrizio invited Andrew Ho to offer his comments in light of his recent related work. Mr. Ho applauded the paper's emphasis on the importance of providing plenty of feedback information to panelists in a clear way. Mr. Ho noted the briefing book method as a useful process to reference for this work.

Tonya Miles noted that the briefing material on evidence centered design (ECD) was not sufficiently clear in describing ECD in relation to TEL. Jack Buckley and Mr. Ho responded that the methodology associated with ECD essentially starts with claims (or inferences) to be reported and then rigorously cites the evidence from the assessment that will be used to support the claims, thereby making a more explicit connection between assessment development and eventual reporting. ECD represents an articulation of current best practices in the field of educational assessment.

The Committee discussed the applicability of the policy definitions of the NAEP achievement levels for the TEL assessment. Given the construct is labeled with the term "literacy" and there is not a unified presentation of TEL content in U.S. schools as a specific course or subject area, the Committee was not certain whether TEL achievement level definitions should be aspirational in nature with statements about what students *should know and be able to do* or whether TEL achievement levels should be more of a status description, providing the public with information about what students *do know and are able to do*. Relatedly, the Committee also echoed the difficulty of selecting appropriate panelists to participate in the achievement level setting, because there may be few, if any, instructors in TEL content. Steve Laser also noted it may be helpful to use a scale anchoring approach to inform the achievement level setting process given that this is a new construct. Mr. Ho noted that item maps are an important resource in the process.

Ms. Orr and Michelle Blair asked for Committee perspectives on the extent to which research should be incorporated into the achievement level setting activities. For example, there have been attempts to examine particular aspects of the process and the inputs to the process to help the Board determine whether the process was working well or should be enhanced in the future. Although conducting pilot or other research studies about the TEL standard setting methodology were not specifically addressed, other concerns about the standard setting process for TEL were discussed. Mr. Fabrizio noted that examining performance data from the assessment itself is critical to determining whether the process will produce strong defensible results. Mr. Ho said scaling analyses to investigate the unidimensionality of TEL are needed to support setting a cut score; Mr. Buckley noted the TEL assessment field test conducted in early 2013 was large in scope in order to provide early indications of any scaling issues.

The Committee agreed that more information is needed about how the assessment is functioning. In particular, the Committee expressed concern about whether scaling issues will challenge TEL achievement level setting efforts. The consensus of the Committee was to proceed with caution in light of these concerns.

Update on Evaluation of NAEP Achievement Levels Procurement

Mr. Fabrizio noted that Janis Brown, an NCES statistician, would provide the procurement update on the project to evaluate NAEP achievement levels. This is largely an information item; COSDAM last heard an update one year ago.

Ms. Brown noted that there have been different evaluation efforts in the past, which have each focused on different aspects of the NAEP program. This upcoming evaluation project will focus on NAEP reading and mathematics achievement levels. It is expected that U.S. Secretary of Education Arne Duncan will soon confirm whether the procurement activities relative to this evaluation project will be conducted within the Institute for Education Sciences (IES). In the current plan, key objectives of this evaluation effort are to:

- Propose how to operationalize the legislative mandate that achievement levels be “reasonable, valid and informative to the public”
- Identify, review and analyze extant evidence and gather additional evidence as needed
- Present collected evidence in a written report, discussing the strengths and weaknesses as well as the gaps in evidence
- Provide for an independent review of the draft report

A contract award is expected in September 2013, and the final report is expected in early 2015.

NAEP 12th Grade Academic Preparedness Research

Ms. Orr provided two brief updates on the Board’s preparedness research efforts. The Board just awarded a contract to review NAEP frameworks, primarily in relation to job training. This will help the Board determine how the frameworks may need to be updated or whether we need a new framework to address job training. Secondly, Ms. Orr announced that the final report from the Job Training Program Course Content Analysis is now available, and this should be a strong resource for the Board and the public. In particular, this research effort is more extensive than other similar efforts currently referenced in policymaking discussions.

Ms. Brown was available to respond to Committee questions on the research studies being conducted by NCES that are relevant to the Board’s preparedness initiative. These research studies include:

- (1) A NAEP Reading–Lexile® Linking Study to establish a link between the NAEP Reading assessment and the Lexile® scale, thereby developing a method to predict whether 8th-graders are on track for successful postsecondary outcomes at grade 12;
- (2) A NAEP–High School Longitudinal Study to create an overlap of NAEP and the 2009 High School Longitudinal Study (HSLs:09) samples, thereby enabling the background variable information from HSLs:09 to provide additional context for reporting of postsecondary outcomes; and
- (3) A Study Using Virginia Data to Examine the Relationship between NAEP Scores and Student Success in College and in the Labor Market.

The Committee indicated that the written briefing material did not prompt any questions at this time.

Interpreting NAEP Proficient Using Preparedness Research Findings

Continuing a discussion that began at the March 2013 COSDAM meeting, the Committee focused its attention on the draft validity argument developed to interpret the NAEP preparedness research findings prepared by Ray Fields. This draft validity argument proposes a specific interpretation of the 12th grade Proficient achievement levels for reading and mathematics relative to academic preparedness for college.

The structure of the validity argument is based on a model developed by Michael Kane, a renowned psychometrician and validity theorist. COSDAM earlier approved using the Kane model to organize the preparedness research findings and the statements supported by the findings. Mr. Kane is serving as an advisor to the Board in the development of the validity argument. The validity argument:

- presents the rationale for NAEP reporting on preparedness,
- stresses that this argument is for an interpretation of Proficient, not a new standard for preparedness,
- provides explicit limitations on interpretations, and
- addresses appropriate and inappropriate test uses and consequences

In the May 16, 2013 Executive Committee meeting, Mr. Fabrizio provided a detailed summary of the historical context leading up to the development of this validity argument and described the key aspects of the argument for the Board's examination. Mr. Fabrizio kicked off the COSDAM discussion by referring to these remarks, noting that COSDAM feedback will be used to refine the draft validity argument. Then the draft will be sent for independent external review by two technical experts. At the August 2013 Board meeting, the plan is for the validity argument to be presented for action by the full Board regarding its use in reporting grade 12 NAEP reading and mathematics results from the 2013 assessments.

Jim Popham asked what the purpose is of reporting NAEP results in relation to academic preparedness for college. Mr. Fields said that the nation uses various important indicators to monitor well-being and inform future policy. These include economic indicators, health indicators, air and water quality indicators, employment indicators, and the like. NAEP reporting on 12th grade academic preparedness will provide the nation a new indicator of the human capital potential of rising generations, a nation's most important resource.

Mr. Fabrizio and Mr. Buckley both commented that the caveats outlined in the validity argument may not receive sufficient attention, but Mr. Fabrizio was pleased that they were explicitly presented. Mr. Ho said there is a danger of people ignoring the caveats and misinterpreting the results. Mr. Fields responded that it is impossible to prevent all misinterpretation, but the Board can and must do all that it can to prevent misinterpretation.

Terry Holliday discussed how he would address questions about this topic at the state level. John Easton noted his earlier concerns, but expressed increased comfort with the way the validity argument is presented. Ms. Miles asked whether additional inferences might be added to the validity argument in the future. Mr. Fields said research findings from Phase 2 of the Board's Program of Preparedness Research might support revisions or additions to the inferences in the validity argument.

Ms. Orr noted that the Board's preparedness research findings provide external data and validity evidence about grade 12 achievement. She indicated that when the achievement levels were originally established by the Board, information like the findings of the preparedness research was not available for the Board to use in standard setting. Therefore, the achievement level descriptions were developed using the content of the NAEP frameworks and the professional judgment of informed educators. The key question for COSDAM is whether the preparedness research can be used to augment the meaning of the grade 12 achievement levels, especially the statements about NAEP Proficient.

Mr. Popham asked the committee members for their thoughts on whether plausibility was a sufficient criterion for evaluating the validity argument. Mr. Ho then asked for clarification on what exactly would be added in reporting performance results for the NAEP Proficient level. Mr. Fields responded by reiterating the inference presented in the validity argument:

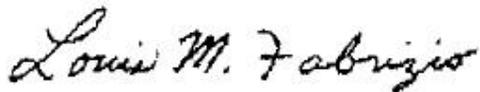
“12th grade students scoring at or above the Proficient achievement level on the 12th grade NAEP Reading or Mathematics Assessment are

- likely to be academically prepared for first year college courses,
- likely to have a first-year college GPA of B- or better, and
- not likely to need remedial/developmental courses in reading or mathematics in college”

Mr. Fabrizio commended the document for defining the target for the inference as “the typical student in the typical college” as a way of pre-empting the potential criticism that a wide range of selectivity exists in college admissions and placement.

Fielding Rolston said his primary concern was how these statements finally appear in the media, which may be different from how the statement appears in NAEP Report Cards. Mr. Fabrizio asked whether the Committee was comfortable with the overall direction of this validity argument and the prospective reporting which will be based on this validity argument. The Committee agreed that they were comfortable with the overall direction of the argument and the way in which caveats are outlined. However, the Committee expressed concern that the proposed inferences in the validity argument will be misinterpreted in secondary sources. The Committee was unanimous in supporting continued work on this topic and having the draft validity argument sent forward for additional review.

I certify the accuracy of these minutes.



Lou Fabrizio, Chair

6-14-2013

Date