National Assessment Governing Board

Committee on Standards, Design and Methodology

Report of August 2, 2013

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

Governing Board Staff: Cornelia Orr, Ray Fields, Sharyn Rosenberg, 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 and Patricia Etienne. AIR: Victor Bandeira de Mello, George Bohrnstedt, Young Yee Kim, Sami Kitmitto, and Fran Stancavage. CRP: Carolyn Rudd. DRC: Pat Roschewski. Education Week: Sarah Sparks. ETS: Steve Lazer and Andreas Oranje. Fulcrum: Saira Brenner. Hager Sharp: Joanne Lim. HumRRO: Sunny Becker. The Hunt Institute: Lucille Davy. Westat: Keith Rust. Widmeyer: Jason Smith.

Introductions and Welcome to Sharyn Rosenberg, the New Governing Board Assistant Director for Psychometrics

Lou Fabrizio, Chair of the Committee on Standards, Design and Methodology (COSDAM), called the meeting to order at 10:07 a.m. and welcomed members and guests. Committee members Andrew Ho and Leticia Van de Putte were unable to attend the meeting. Mr. Fabrizio introduced Sharyn Rosenberg, the new Governing Board Assistant Director for Psychometrics. Ms. Rosenberg gave some brief introductory remarks, noting that she attended several COSDAM meetings in her previous role as a NAEP contractor. She mentioned that former Board and COSDAM member Greg Cizek was not only her Ph.D. advisor, but that he first inspired her to pursue a degree in psychometrics.

Committee Questions on Information Items

Mr. Fabrizio asked whether there were any questions on either of the information items: 1) the update on evaluation of the NAEP achievement levels procurement (Attachment D); or 2) the NAEP 12th grade academic research: Phase 2 research updates (Attachment E). The Committee confirmed that there were no questions.

Interpreting NAEP Results Using Preparedness Research Findings (ACTION ITEM)

Mr. Fabrizio noted that COSDAM would be presenting a motion for full Board action on the inferences for interpreting NAEP results in terms of academic preparedness for college. He asked Governing Board staff Ray Fields to lead the discussion providing context for the motion.

Mr. Fields recognized the contributions of COSDAM, and in particular of Andrew Ho, who suggested a revision to the proposed inferences that were discussed by the full Board at the May 2013 meeting. Following the May 2013 meeting, the revised inferences suggested by Mr. Ho were reviewed by Michael Kane, advisor to the validity argument. Mr. Kane confirmed that the inferences were appropriate and considered them an important improvement.

In mid-July, COSDAM members engaged in email discussions to review the revised validity argument (a modification of the initial draft which was presented at the May 2013 meeting), reviews of the validity argument by independent technical experts Mark Reckase and Greg Cizek, and a staff-prepared prototype chapter for using NAEP as an indicator of academic preparedness in connection with the 2013 grade 12 results for reading and mathematics. Mr. Fields summarized these activities and invited the Committee to engage in additional discussion. In particular, Mr. Fields noted that the email exchanges had included Committee discussion on the use of "plausible" versus "reasonable," but that "plausible" was chosen because Mr. Ho expressed a preference for that term. Mr. Fields also reported that the validity argument and its accompanying external reviews would be finalized to include a table of contents, acknowledgment section, executive summary, and foreword, and that the independent reviews would be included as appendices.

Mr. Fields noted that Governing Board staff has been working with NCES since March 2013 to discuss options for reporting grade 12 results in terms of academic preparedness. With the parallels to the preparedness initiative, NCES had reminded Governing Board staff that the research and analyses related to inclusions and accommodations resulted in a chapter in the 2000 NAEP Reading Report Card about the reasons for pursuing a more inclusive NAEP, the research results, and the impact on score interpretations. The information in the draft prototype chapter on academic preparedness is intended as a concrete example of what might be included in the 2013 grade 12 NAEP Report Card and contains the following elements: the rationale for the Board's preparedness initiative, research results, policy context for preparedness, definition of preparedness, the proposed inferences, caveats, and limitations, and that this represents a transition in reporting. Mr. Fields noted that the prototype chapter provides context for the motion, but that the motion refers only to the inferences, not to the prototype chapter.

Jim Popham requested that the prototype chapter include several good examples of how the results are useful to the general public. He would like to see the grade 12 report address the questions, "Who cares?" and "So what?" when discussing academic preparedness. Real examples for the general public would be useful. Mr. Fields noted that Mr. Popham's previous suggestion at the May 2013 meeting had been incorporated into the current draft and that these questions would continue to be considered.

There was discussion about the use of "plausible" versus "reasonable." John Easton noted that "plausible" seems like a more statistical term, while "reasonable" seems to reflect rhetoric of argument. He also thought that "plausible" sounds like a probability term. Jim Popham stated that plausibility involves reason and is not just about numbers. Jack Buckley pointed out that "plausibility" comes from "worthy of applause." Cornelia Orr explained that "plausible" reflects the terminology used in the literature about validity arguments. The Committee was satisfied with this explanation of why "plausible" was chosen over "reasonable."

Mr. Fields asked whether the Committee had any other feedback to share on the prototype chapter. Mr. Fabrizio noted that the inferences in the prototype chapter should reflect the exact wording from the motion. Mr. Fields said that the prototype chapter would be changed to reflect the inferences as stated in the motion. Fielding Rolston asked about the timeline for the draft chapter; Mr. Fields clarified that the information would be incorporated into the Grade 12 NAEP Report Card, tentatively scheduled for release in March or April of 2014. Governing Board staff will work collaboratively with NCES on the information that will be incorporated into the main report or presented in a side publication. Tonya Miles noted that the prototype chapter was very comprehensive; she felt that it addressed the questions that others might have about the study conclusions.

Mr. Fabrizio raised the issue of choosing a scale score of 163 for mathematics, which is supported by the research but is lower than the Proficient cut score. Will this lead people to ask whether the cut score on grade 12 mathematics should be lowered?

Ms. Orr responded by discussing the desk side briefings on academic preparedness that she gave to policy leaders and organizations. She was asked whether the Board will lower the cut score for grade 12 mathematics, but this question assumes that the Proficient cut score is supposed to represent the college preparedness level, which it was not. More generally the briefings were quite positive; most people were impressed by the volume of research and expressed satisfaction in the content alignment of NAEP and SAT, even if NAEP is broader. Ms. Orr did receive requests for additional information, which varied by individuals and organizations. The comments received during the desk side briefings included an interest in including more longitudinal research with state partners in addition to Florida; a desire to know more about other exams linked to NAEP (such as the ACT) and especially placement exams; and the question of whether grade 12 students are motivated to try hard on NAEP.

In terms of student motivation, Ms. Orr noted that it is important to be aware of the tendency to question whether grade 12 results represent students' best efforts. Some people have a hard time believing that 12th-graders try hard on a test that does not count. On the other hand, TIMSS and PISA are at the secondary level and also do not count. Ms. Orr pointed out that there is some evidence that grade 12 students do take the test seriously, such as completion rates and completion of open-ended questions in particular. On the other hand, if an ERIC search was performed on the terms "NAEP" and "motivation," the search would likely yield studies that conclude students are not very motivated. Such a literature search would likely not turn up any evidence to indicate that students *are* motivated. Ms. Orr suggested that it is more important now than ever to make available the evidence that we do have to support the claim that students are motivated. Mr. Fields added that the research on NAEP and motivation that has been done and is often quoted has not been critiqued. Mr. Fields suggested that such research deserves to be critiqued; otherwise, the assertions made are left standing. One thing that has been discussed is that a literature review and critique of existing studies could be performed as part of the efforts on preparedness research.

The final topic that Ms. Orr reported from the desk side briefings was that people were really surprised by the lack of alignment and overlap between workforce training and the NAEP

reading and math assessments in the most recent studies that were performed. This was surprising to people given the rhetoric across the country of having the same standards and expectations for "college and career." This raises the question of what these results mean – should we not have the same high expectations for all students? The Board would say that all students should graduate high school prepared to pursue the path they choose, be it college or job training. Lastly, audiences wanted to know more about preparedness for the military, and people in career and technical education thought it would be helpful to know more about how the NAEP frameworks align with frameworks for career and technical education.

Mr. Popham asked whether there was going to be an effort to look at the appropriateness of the NAEP proficiency levels. He noted that the Smarter Balanced Assessment Consortium (SBAC) is formulating plans for setting performance standards. Several people think it is important to set the SBAC standards near the NAEP cut scores to ensure credibility. But one state representative from SBAC noted that if standards are set as high as NAEP to avoid political criticism, states may be in a position of having a standard that is too high to achieve for a long time. This caused Mr. Popham to wonder whether NAGB ever reexamines performance standards.

Mr. Fields responded that the NAEP legislation allows reconsideration of the performance standards. He also noted that it is very difficult to obtain external data for grades 4 and 8. There has been some work looking at NAEP achievement levels in relation to TIMSS (not performed by the Board), and this work has generally been confirming of the current performance standards for mathematics. Ms. Orr added that the state mapping studies could be used as disconfirming evidence to the NAEP achievement levels— but this begs the question of which performance standards are more appropriate.

Mr. Popham explained the source of his concern. If we assume that the NAEP achievement levels are a little too high, in the past this did not really matter too much. But now if the consortia adopt the same standards, and the tests are not instructionally facilitative or meaningfully diagnostic, it will be very tough to improve student performance on these assessments. If years go by with a consistent message of failure, this will not be helpful for American schooling. He asked Terry Holliday to provide his perspective on this issue.

Mr. Holliday explained that when Kentucky set performance standards a few years ago, technical advisors from colleges agreed on an academic preparedness cut score on the ACT. Then, the ACT EXPLORE data were taken all the way back to third grade, to set bands of proficiency levels benchmarked to that cut score. The inspiration for this approach came from Jack Buckley and his charts about confidence bands. Mr. Holliday feels that he can answer his many critics; NAEP Proficient falls within the band of state proficiency. Mr. Holliday suggested that SBAC and PARCC pay attention to the confidence bands; otherwise it will open the states to a lot of criticism, which will lead to pushback on NAEP. Mr. Holliday also noted that the grade 12 performance standards will only be relevant at the state level for the 13 states that participate in NAEP at the state level. Finally, Mr. Holliday noted that the Chief State School Officers were very receptive to the Board's academic preparedness work that was presented at their meeting in Wisconsin last week, but that they have not yet identified the implications of that work.

Mr. Rolston noted that Tennessee is very tied into the ACT – all students in Tennessee take the ACT, not only the students who intend to go to college. Having complete data on the ACT, along with grade 12 NAEP scores, will allow Tennessee to check some of these inferences about academic preparedness.

Mr. Popham asked whether the academic preparedness research results in reading are viewed as confirmatory evidence of the Proficient level. Mr. Fields responded, "So far," and added that Mr. Holliday and Mr. Rolston are supporting additional studies with their respective states, Kentucky and Tennessee, linking 2013 grade 8 NAEP scores (in both reading and math) to grade 8 ACT EXPLORE scores. The results of these studies will also help to inform this discussion for grade 8 performance standards.

Ms. Orr noted that we likely would not be having this conversation if the NAEP achievement levels had been numbered one through three rather than using the label of "Proficient." The initial achievement levels were clearly intended to be aspirational. Mr. Fields noted that the achievement levels came about during the same time as the National Education Goals were set in 1990, and that NAEP was intended to measure progress toward the goal for student achievement. Thus, the definition used in the National Education Goal for student achievement — "competency over challenging subject matter" — was embodied in the policy definition for the NAEP Proficient achievement level. This is clearly different from "performance at grade level." Mr. Fabrizio pointed out that the concern is with the states that use their current assessments to make promotion decisions. If standards are set really high and states continue to use these assessments to make promotion decisions, the public is not likely to accept 60 percent of students being retained in grades 3-8.

Mr. Buckley commented that COSDAM is a technical committee, but that this is really a policy discussion. The NAEP achievement levels are still in a trial status. Mr. Buckley would like to be able to resolve the trial status, but this is a difficult challenge. He added that NAEP was never intended to be a formative assessment even though the consortia assessments were conceptualized to provide more feedback. It is important to make this difference clear to policymakers.

Mr. Fabrizio presented the motion on the inferences to be included in the report of the 2013 grade 12 results. The motion was moved by Mr. Popham and seconded by Ms. Miles. There were no questions raised. COSDAM voted unanimously to pass the motion.

ACTION: COSDAM recommends approval by the Governing Board of the Motion on Reporting 12th Grade Academic Preparedness for College, which includes the inferences to be incorporated into the 2013 Grade 12 NAEP Report Cards for Mathematics and Reading. The motion is appended as Attachment 1 to this report.

CLOSED SESSION 11:20 a.m. – 12:25 p.m.
COSDAM Members: Lou Fabrizio (Chair), Terry Holliday, Tonya Miles, Jim Popham, and Fielding Rolston (Vice Chair).

Governing Board Staff: Cornelia Orr, Sharyn Rosenberg, 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 and Patricia Etienne. AIR: Victor Bandeira de Mello, George Bohrnstedt, Young Yee Kim, Sami Kitmitto, and Fran Stancavage. ETS: Steve Lazer and Andreas Oranje. Fulcrum: Saira Brenner. Hager Sharp: Joanne Lim. HumRRO: Sunny Becker. Westat: Keith Rust.

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 2, 2013 from 11:20 a.m. – 12:25 p.m. in order to discuss information regarding analyses of the TEL field trial, including secure data.

Discussion on Achievement Level Setting (ALS) on the 2014 NAEP Technology and Engineering Literacy (TEL) Assessment

Ms. Orr began the discussion on achievement level setting (ALS) for the 2014 NAEP Technology and Engineering Literacy (TEL) Assessment. She provided a summary of previous discussions related to this topic and noted that the TEL ALS poses some unique challenges due to being computer-based, using an evidence-centered design approach, and measuring a new construct that is not explicitly taught in most schools. Mr. Popham stated that the evidence-centered design process is great for assessment development, but he feels that it should not be incorporated into the standard setting process.

Andreas Oranje of ETS then provided a briefing on how the TEL achievement level setting (ALS) could be informed by current TEL data collections and analyses. His presentation addressed the following overarching questions:

1) Can we falsify (or fail to falsify) the hypothesis voiced by COSDAM members that TEL is a measure of general intelligence?
2) Is the TEL construct (either as a composite or a single scale) cohesive enough to warrant ALS?
3) Are there additional studies required to address these questions satisfactorily?

Mr. Oranje presented the currently available embargoed results from the TEL field trial and noted that there were some differences in the design and administration of the field trial and plans for the operational assessment in terms of timing, number of items, order of items, and spiraling of background questions. These differences are important when considering the use of field trial data for ALS pilot studies.

Mr. Oranje presented a schedule of completed and planned activities for the TEL field trial and the TEL operational administration. He noted that ETS had not yet examined correlations

between the assessment areas, but that this information was expected to be available by the December 2013 Board meeting. This presentation could also include performance patterns by subgroups. Results from the bi-factor model of the TEL field trial are expected to be available by the March 2014 Board meeting. For the TEL operational administration, results from the composite or univariate scale score distribution are expected to be available in August 2014, and results from the bi-factor model are expected to be available in October 2014.

Ms. Rosenberg asked for clarification on the schedule of analyses for the TEL operational administration. If composite or univariate scale scores are expected to be available in August 2014, and the bi-factor model results are expected to be available in October 2014, when will a decision be made about which model to use, and on what criteria will that decision be based? Mr. Oranje responded that this decision would likely be made a priori rather than on an empirical basis. It is safer to use the composite or univariate scale scores since this is the model that has been used in the past; the bi-factor model will likely be used to inform the extended reporting goals only.

Mr. Popham raised the concern of anticipatory validity. Given the assumption that most students have not received specific instruction in TEL, what kinds of factors will raise scores in the future? Mr. Fabrizio pointed out that the Assessment Development Committee has been having these conversations.

Two COSDAM members expressed concern about going forward with TEL, given that the budget constraints may undermine expansions in reporting preparedness. TEL is not a core area, but academic preparedness is heavily anticipated and needed in education at this time.

OPEN SESSION: 12:25 p.m. – 12:30 p.m.
(See list of attendees for the closed session. No additional attendees joined this session.)

Other Issues or Questions

In closing, Mr. Fabrizio invited COSDAM members to provide any additional comments or questions. Mr. Popham asked for an agenda item on instructional sensitivity. Mr. Easton noted that there is an IES grant on an assessment innovation for parsing out prior knowledge; COSDAM requested a briefing on this work. Mr. Holliday asked about the status of Dave Conley's report on course content analyses. Michelle Blair responded that the job training report has been released, but the work on colleges is ongoing. Content reviewers are examining course artifacts. A final report is expected by the March 2014 Board meeting.

I certify the accuracy of these minutes.

Louis M. Fabrizio

August 9, 2013

Lou Fabrizio, Chair

Date

Attachment 1

Motion on Reporting 12th Grade Academic Preparedness for College

The Committee on Standards, Design and Methodology (COSDAM) recommended adoption of the motion below.

The National Assessment Governing Board approved the motion on August 3, 2013.

Background

- COSDAM has responsibility for overseeing the Governing Board's program of research on 12th grade academic preparedness for college and job training.
- COSDAM has reviewed:
  - o the research results completed during 2010-2012,
  - o the staff-prepared validity argument developed in support of inferences proposed,
  - o the independent technical reviews of the validity argument, and
  - o a staff-prepared prototype intended to exemplify how the research on 12th grade academic preparedness for college would be reported.
- The inferences proposed in the motion below are intended as preliminary statements for reporting purposes and specifically not as performance standards for academic preparedness for college.
- Governing Board staff have worked and will continue to work collaboratively with staff of the National Center for Education Statistics to prepare the full explanatory text about the NAEP 12th grade academic preparedness initiative.
- The prototype document in the Board materials is a starting point for what may be said about the preparedness initiative in the report of the NAEP 12th grade reading and mathematics administered in 2013, with adjustments to the text and format to be made as needed.
- Further research to be conducted during 2013-2015 will be considered by COSDAM and the National Assessment Governing Board to make adjustments, as appropriate, to the statements in the motion below.

Motion

The National Assessment Governing Board approves the following statements for use in the reporting of the NAEP 12th Grade Reading and Mathematics assessments administered in 2013.

Reading: Given the design, content, and characteristics of the NAEP 12th grade reading assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness:

> **the percentage of students scoring at or above a score of 302 on Grade 12 NAEP in reading is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in reading that would make them academically prepared for college.**

A score of 302 corresponds to the cut-score for the Proficient achievement level in 12th grade reading.

In 2013, XX% of 12th graders nationally scored at or above 302 in reading.

Mathematics: Given the design, content, and characteristics of the NAEP 12th grade mathematics assessment, and the strength of relationships between NAEP scores and NAEP content to other relevant measures of college academic preparedness,

> **the percentage of students scoring at or above a score of 163 on the Grade 12 NAEP scale in mathematics is a plausible estimate of the percentage of students who possess the knowledge, skills, and abilities in mathematics that would make them academically prepared for college.**

A score of 163 in mathematics is between the cut-scores for the Basic and Proficient achievement levels in 12th grade mathematics.

In 2013, XX% of 12th graders nationally scored at or above 163 in mathematics.