

National Assessment Governing Board

Committee on Standards, Design and Methodology

August 3, 2012
10:00 a.m. - 12:30 p.m.

AGENDA

10:00 – 10:05 am	Welcome, Introductions, and Agenda Overview <i>Lou Fabrizio, COSDAM Chair</i>	
10:05– 10:40 am	Reporting 12 th Grade Preparedness Research: Validity Evidence for Reporting Preparedness on Reading and Mathematics NAEP <i>Ray Fields, and Susan Loomis Governing Board Staff</i>	
10:40 – 10:55 am	ACTION Trial Urban District Assessment (TUDA) Policy: Proposed Modifications in Eligibility Statements <i>Ray Fields</i>	Attachment A
10:55 – 11:20 am	Linking NAEP and PISA: Indirect Impacts <i>Taslina Rahman, NCES Staff</i>	Attachment B
11:20 am – 11:25 am	Recommendations of Future COSDAM Agenda Topics	
11:25 am – 12:30 pm	CLOSED SESSION	
11:30 am – 12:00 pm	Multi-Stage Adaptive Field Trial: Response Time Indicators of Engagement <i>Andreas Oranje, ETS Research Staff</i>	Attachment C
12:00 -12:30 pm	Evaluation of NAEP Achievement Levels <i>Jack Buckley, NCES Commissioner</i>	Attachment D

Clarifications to the Governing Board's Policy for The Trial Urban District Assessment Program

Ray Fields

The Trial Urban District Assessment (TUDA) program was first funded by Congress for assessments conducted in 2002. TUDA began with five volunteering districts; twenty-one volunteering districts participated in 2011 and have signed letters committing to participate in 2013.

The identification of the participating districts is the responsibility of the Governing Board, in consultation with the Executive Director of the Council of the Great City Schools (CGCS) and staff of the National Center for Education Statistics (NCES). The identification process is prescribed under the Governing Board Policy Statement entitled "Eligibility Criteria and Procedures for Selecting Districts for Participation in the National Assessment of Educational Progress: Trial Urban District Assessment."

The current TUDA policy statement was adopted by the Governing Board on March 3, 2007. Board staff periodically review extant policies to determine whether revisions may be needed. The TUDA policy statement was recently reviewed by staff. Through this review, a number of elements of the TUDA policy were identified that staff suggest would benefit from clarification, particularly the eligibility criteria. Governing Board staff have consulted with CGCS and NCES on the proposed changes.

Accordingly, the staff recommendations for clarifying revisions to the TUDA policy are on the following pages, along with a document showing the impact of the clarified eligibility criteria on the list of eligible districts.

COSDAM reviewed the proposed changes to the policy at the May 2012 Board meeting. A few minor edits have been made to the previous document and these are highlighted in yellow on the following pages.

The final page of this attachment compares the list of eligible districts under the current eligibility criteria and the clarified criteria. The impact of the clarifications is minimal. The first 21 districts in both lists are the current TUDA participants. The 15 districts highlighted in blue are also eligible under the current and the clarified criteria. Under the clarified criteria, 2 districts would be added, highlighted in yellow.

Governing Board staff recommend that COSDAM present the revised TUDA policy for full Board approval at the August 2012 meeting.

Adopted: March 3, 2007

Revised: DATE



National Assessment Governing Board

Eligibility Criteria and Procedures for Selecting Districts for Participation in the National Assessment of Educational Progress

Trial Urban District Assessment

Policy Statement

Purpose

To define the eligibility criteria and selection procedures for participation of urban school districts in the National Assessment of Educational Progress (NAEP) Trial Urban District Assessment (TUDA).

Guiding Principles

Principle 1

Participation in TUDA shall be voluntary.

Principle 2

A primary goal of TUDA is to ~~support promote education reform the improvement of student achievement in support of the large number of challenged populations enrolled in the schools of our nation's large largest~~ urban school districts and to focus attention on the specific challenges and accomplishments associated with urban education.

Principle 3

Districts participating in TUDA shall have the characteristics of large urban areas.

Principle 4

All ~~urban~~ districts that have participated ~~ed~~ in TUDA without interruption once included at the time the Governing Board adopts criteria and establishes a selection

~~process~~ shall be deemed eligible and permitted to continue to participate.

Principle 5

The eligibility criteria for participation in TUDA shall promote (1) inter-district homogeneity/comparability, — so that participating districts that are reasonably similar with respect to key demographics ~~across the districts~~ and (2) efficiency in resources required of the NAEP program.

Principle 6

The selection of any additional~~Increasing the total number of districts participating in for TUDA participation~~ shall be contingent on additional funding from Congress. ~~Current funding is sufficient to support ten (10) TUDA districts.~~

Principle 7

The Governing Board ~~may~~implements the selection procedures used to consider districts for participation in TUDA. ~~whenever sufficient funding is available to support the action.~~

Principle 8

Districts applying for participation in TUDA should be committed to long-term participation.

Eligibility Criteria

1. Only ~~large~~ cities having 250,000 or more population shall be represented in TUDA.
2. Districts participating in TUDA shall ~~be~~have a student enrollment large enough to support NAEP assessments in three subjects in each grade assessed ~~three subject assessment cycle for NAEP in grade levels included in the state assessment program.~~ The enrollment requirement is a minimum of approximately 1,500 students per subject per grade level assessed.
3. Districts participating in TUDA shall have an enrollment district-wide or in the grade levels assessed that meets majority (50% or more) of students meeting at least one of the following criteria:
 - a. 50% or more are minority students (i.e., Either African American, American Indian/Alaskan Native, Asian, Hispanic, Native Hawaiian/Other Pacific Islander, or and/or multi-racial).
 - b. 50% or more are eligible~~Eligible~~ for participation in the free and reduced-price lunch program (or other appropriate indicator of poverty status).

Districts that are very near to meeting a particular eligibility requirement may be considered eligible if they request to participate in the program and if funds are sufficient to permit participation. Eligibility data shall be updated and verified regularly/periodically.

Application and Selection Process/Procedures

To provide time for consultation, notification, and operational planning for the conduct of the Trial Urban District Assessments, the steps described below should be sequenced to conclude approximately 14 months prior to the start of testing.

1. Prior to the assessment year in which TUDA is to be conducted, the Governing Board Executive Director, in consultation with the National Center for Education Statistics (NCES), prepares a list of eligible districts and posts that list on the Governing Board website.
2. Prior to the assessment year in which TUDA is to be conducted, the Governing Board Executive Director sends a letter to each district that participated in the immediately preceding administration of TUDA to determine the district's interest in continuing as a participant in the upcoming administration of TUDA.
3. Based on funding from Congress and the decision of any previous TUDA participant not to continue, the Governing Board determines whether new districts can be considered for participation in the upcoming TUDA administration.
4. If the Governing Board determines that new districts can be considered for participation in the upcoming TUDA administration, the Governing Board Executive Director sends a letter notifying eligible districts of the opportunity to submit an application and the instructions for applying.
- ~~1.5.~~ Eligible districts seeking to participate in TUDA submit an ~~A~~ letter of application ~~from urban districts seeking to participate in TUDA should be submitted~~ to the Executive Director of the Governing Board. The application should be signed by the district superintendent or designee, include the most recent information documenting the district's enrollment and eligibility, and contain a commitment for long-term participation in TUDA if selected.
- ~~2.6.~~ The Executive Director of the Governing Board and appropriate staff of the Governing Board shall review applications in consultation with the Chairman of the Governing Board, the Chairman of the Board's Committee on Standards, Design and Methodology, staff of the National Center for Education Statistics, and the Executive Director of the Council of the Great City Schools.
7. The Executive Director of the Governing Board shall recommend new districts for participation in TUDA to the Governing Board for final action.
48. The Executive Director of the Governing Board shall send notification of the Board's decision regarding district participation in TUDA to the district and to the Commissioner of Education Statistics.

~~5. Districts must be accepted for participation at least 14 months prior to the first assessment cycle for their participation in TUDA.~~

Potential Pool of Eligible Districts

The list of eligible districts shall be posted on the website of the National Assessment Governing Board (www.nagb.org) and made publicly available through other appropriate means. The list of districts will change from time to time due to changes in the population of the district and the district setting.

District Eligibility for TUDA (Common Core of Data 2010-2011)

Current Participants Additional Eligibles

Current Criteria		Proposed Clarified Criteria	
ALBUQUERQUE PUBLIC SCHOOLS	NM	ALBUQUERQUE PUBLIC SCHOOLS	NM
ATLANTA PUBLIC SCHOOLS	GA	ATLANTA PUBLIC SCHOOLS	GA
AUSTIN ISD	TX	AUSTIN ISD	TX
BALTIMORE CITY PUBLIC SCHOOLS	MD	BALTIMORE CITY PUBLIC SCHOOLS	MD
BOSTON	MA	BOSTON	MA
CHARLOTTE-MECKLENBURG SCHOOLS	NC	CHARLOTTE-MECKLENBURG SCHOOLS	NC
CITY OF CHICAGO SD 299	IL	CITY OF CHICAGO SD 299	IL
CLEVELAND MUNICIPAL	OH	CLEVELAND MUNICIPAL	OH
DALLAS ISD	TX	DALLAS ISD	TX
DETROIT CITY SCHOOL DISTRICT	MI	DETROIT CITY SCHOOL DISTRICT	MI
DISTRICT OF COLUMBIA PUBLIC SCHOOLS	DC	DISTRICT OF COLUMBIA PUBLIC SCHOOLS	DC
FRESNO UNIFIED	CA	FRESNO UNIFIED	CA
HILLSBOROUGH	FL	HILLSBOROUGH	FL
HOUSTON ISD	TX	HOUSTON ISD	TX
JEFFERSON COUNTY	KY	JEFFERSON COUNTY	KY
LOS ANGELES UNIFIED	CA	LOS ANGELES UNIFIED	CA
MIAMI DADE PUBLIC SCHOOLS	FL	MIAMI DADE PUBLIC SCHOOLS	FL
MILWAUKEE SCHOOL DISTRICT	WI	MILWAUKEE SCHOOL DISTRICT	WI
NEW YORK CITY PUBLIC SCHOOLS	NY	NEW YORK CITY PUBLIC SCHOOLS	NY
PHILADELPHIA CITY SD	PA	PHILADELPHIA CITY SD	PA
SAN DIEGO UNIFIED	CA	SAN DIEGO UNIFIED	CA
CLARK COUNTY SCHOOL DISTRICT	NV	CLARK COUNTY SCHOOL DISTRICT	NV
CYPRESS-FAIRBANKS ISD	TX	CYPRESS-FAIRBANKS ISD	TX
DAVIDSON COUNTY	TN	DAVIDSON COUNTY	TN
DUVAL	FL	DUVAL	FL
EL PASO ISD	TX	EL PASO ISD	TX
ELK GROVE UNIFIED	CA	ELK GROVE UNIFIED	CA
FORT BEND ISD	TX	FORT BEND ISD	TX
FORT WORTH ISD	TX	FORT WORTH ISD	TX
GUILFORD COUNTY SCHOOLS	NC	GUILFORD COUNTY SCHOOLS	NC
LONG BEACH UNIFIED	CA	LONG BEACH UNIFIED	CA
MEMPHIS	TN	MEMPHIS	TN
MESA UNIFIED DISTRICT	AZ	MESA UNIFIED DISTRICT	AZ
NORTH EAST ISD	TX	NORTH EAST ISD	TX
NORTHSIDE ISD	TX	NORTHSIDE ISD	TX
SCHOOL DISTRICT NO. 1 IN THE COUNTY OF DENVER	CO	SCHOOL DISTRICT NO. 1 IN THE COUNTY OF DENVER	CO
		KATY ISD	TX
		WAKE COUNTY SCHOOLS	NC



Linking NAEP and PISA (the Program for International Student Assessment)

Why: PISA, the Program for International Student Assessment, measures reading, mathematics, and science literacy of 15-year-olds in over 60 countries. Although PISA scores are available for the U.S. as a country, its constituent states do not get PISA scores to compare their students with those of other countries in literacy of key learning subjects. The demand for state benchmarking data on the PISA scale is undeniable. To address the issue in this era of limited funds, NCES is embarking upon a Research & Development effort to link NAEP and PISA. If successful, U.S. states would get projected PISA scores without expending their resources.

When: NCES will conduct an exploratory study in the 2012–13 school year with the primary goal of projecting 2012 PISA mathematics score distributions for states from their 2013 NAEP performance. As of now, 13 states have agreed to participate in the 12th grade NAEP assessments in 2013.

How: A set of special NAEP mathematics booklets will be developed. These booklets will have one block of 8th grade math items and one block of 12th grade math items. These special booklets will be administered to an augmented sample of students drawn from grades 9, 10, and 11 from those high schools in the three states that agreed to participate in both NAEP's grade 12 assessment and in the fall 2012 PISA. In addition, some 8th and 12th graders sampled for NAEP will be administered those special booklets. The PISA/NAEP linking will be accomplished in two steps:

1. A cross-grade link will be developed that puts NAEP grades 8 and 12 onto a common scale (NAEP 8/12) scale. Since 2005, NAEP has reported mathematics results on two different scales: the range of the grade 8 scale is 0–500 and that of grade 12 is 0–300.
2. An equipercentile link will be established between the PISA scale and the NAEP 8/12 cross-grade scale based on common populations of 15-year-old students subsampled from NAEP's samples in grades 8, 9, 10, 11, and 12 assessed in the three states participating in both the NAEP and PISA assessments.

Design and Analysis Committee¹ (DAC) recommendations on the NAEP-PISA linking study. The DAC reviewed the preliminary design in June 2012 and offered the following recommendations:

1. Use the cross-grade NAEP 8/12 scale only to establish the statistical link between PISA and NAEP and not to use the cross-grade link as a vertical scale in the psychometric and content senses;

¹ The DAC is a technical advisory committee of the design, analysis, and reporting NAEP contractor.

2. Replicate the PISA/NAEP item alignment study (conducted in 2003) using the new NAEP and PISA frameworks and items;
3. Use a validation sample not used in the construction of the link itself (the training sample), perhaps through a split-half method; and
4. Carry out separate validation procedures using grade 8 only, grade 12 only, and a combination of grades 8 and 12 to produce a basis for comparisons between and among the resulting linked data.



Adaptive Testing and Response Time in NAEP's 2011 (MCBS) Study

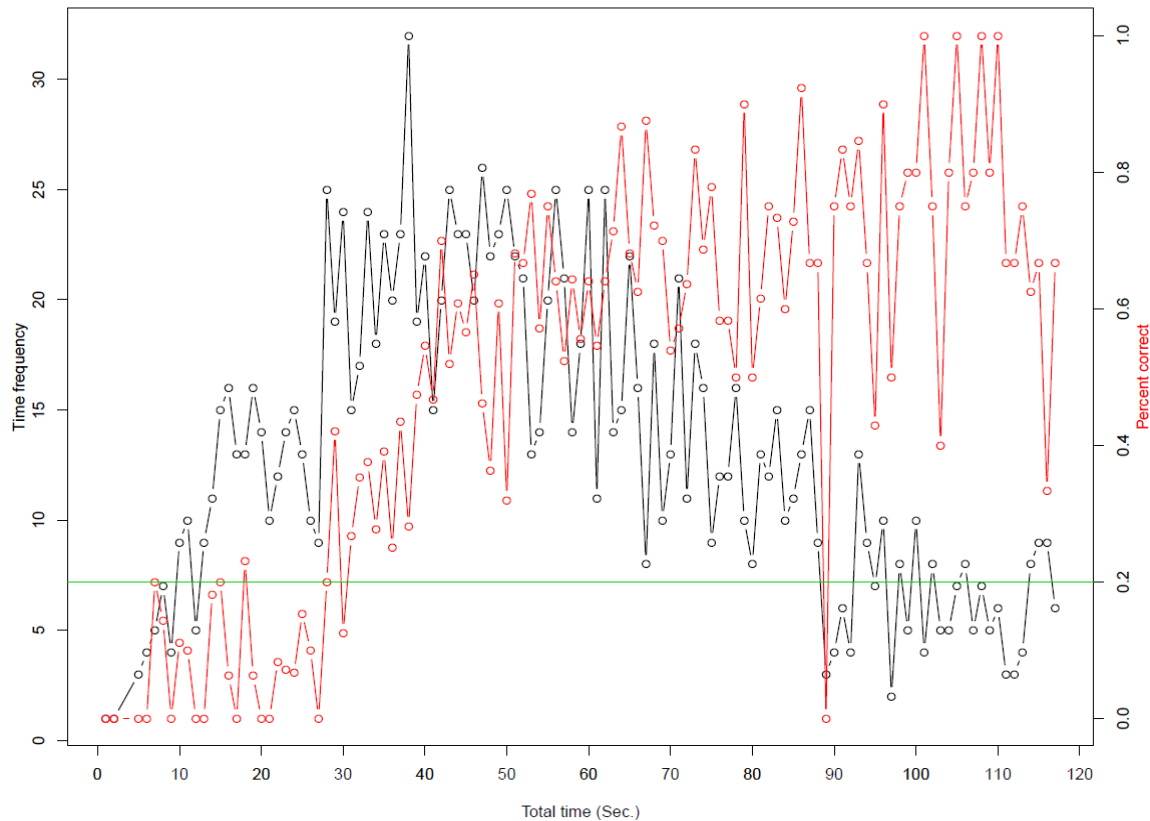
Adaptive testing, in which the cognitive demands of the assessment become matched to the ability of the student, offers both benefits and challenges to NAEP. Adaptive testing may offer NAEP an improvement in measurement precision across a wider range of proficiencies than its current design produces and could also improve student engagement compared to a test that is too hard or too easy. The NAEP Mathematics Computer Based Study (MCBS), a trial study of two-stage adaptive testing for grade 8 mathematics, was conducted in 2011. The primary goals of the study were to reduce measurement error and improve student engagement by administering items tailored to students' individual ability levels.

Presentation of MCBS results with respect to item response time. During this session, staff will present the results of analyzing the cognitive data collected from MCBS. Specific questions to be addressed include:

1. Based on various proxy measures, was there evidence that the multi-stage design improved students' engagement in the test?
2. Can response time be used to classify behaviors, to filter out cases of low student engagement, and to improve the quality of IRT parameter estimation?
3. Can response time be used to identify missing data in operational analysis?
4. Can results on an items' time demand assist test assembly?

Timing patterns on items like that shown in the figure on the next page were used to classify items as showing thresholds of 5, 10, or 20 seconds. The black line (higher on the left, lower on the right) shows the timing frequencies. Some items, like this one, show a bimodal timing frequency and the valley between the faster response mode (assumed to indicate disengagement and guessing) and the slower response mode (assumed to indicate engaged solution behavior) is used to estimate the threshold. Any student responding faster than this threshold on this item was coded as being disengaged on that item.

A 20 second threshold was used for the item in the figure on the next page, which is perhaps conservative as the valley is closer to 30 seconds. A second indication of disengagement is the average percent correct for the students' responses on this item in each total time bin, which is shown in the red line (lower on the left, higher on the right) in the figure. On this item, students did not consistently respond above chance until about 30 seconds. In other words, the timing and percent correct data converge on a plausible indication of disengagement if less than 20 (or 30) seconds is spent.



Design and Analysis Committee¹ (DAC) Recommendations on the analysis of item response time. The DAC is the technical advisory committee to the NAEP Design and Analysis Contractor. The DAC reviewed the preliminary evidence and offered several recommendations for further analysis:

1. Use the relationship between overall proficiency and percent correct on each item to perform a sensitivity analysis on threshold setting for item disengagement; that is, defining when Response Times become too short to provide informative responses.
2. Produce better statistics to support the claim that low performing students in the MCBS router block are less likely to indicate disengagement when routed to the easy second block than when routed to the medium or hard block.
3. Use Too-short Response Times to better classify missing responses to cognitive items into the “omit” and “not-reached” categories for analysis. It seems reasonable to pursue, but unlikely to have a large impact on the results.

¹ The DAC is a technical advisory committee of the design, analysis, and reporting NAEP contractor.

4. Develop a more focused investigation of the relationship between response time and ability for easier items and for better students. Rather than treating every response time the same, classify items by whether the relationship between percent correct and response time is positive or negative and analyze the relationship between the two separately.



Independent Evaluation of NAEP Achievement Levels

In his initial remarks to the National Assessment Governing Board, NCES Commissioner Jack Buckley indicated his interest in a new evaluation of the NAEP achievement levels. The achievement levels have been evaluated at various times in the past, but several changes have occurred since the last evaluation. The time is now appropriate for another look.

A solicitation was posted¹ for an independent, outside evaluation of the achievement levels. While the solicitation has been targeted for the National Academy of Sciences to perform the evaluation, the purpose of the announcement is to invite others to demonstrate their ability to do this work. Thus the possibility that the evaluation will be performed by a different organization is still open.

The draft Performance Work Statement acknowledges that since the time of the last achievement level review in 1999, a number of important changes concerning NAEP assessments have taken place. Among these is the widespread use of NAEP-like achievement levels in state testing, in which states set a bar for proficient performance for which their schools are held accountable. State standards are now compared to those in NAEP as an indicator of their rigor. In the near future, many states will be using performance standards under the Common Core State Standards assessments, which will almost certainly be compared with NAEP's achievement levels.

The focus of the new evaluation will be on the NAEP achievement levels for reading and mathematics. Except for 12th grade mathematics, these levels have been in place for a long period and performance data are available for several assessments. Achievement levels for other subjects may also be considered to provide a larger context.

The new study will address the following components:

- ***Reliability and validity.*** The draft Performance Work Statement makes reference to a Congressionally mandated evaluation of NAEP by the Buros Institute in 2007. Among other things, that audit was tasked to address “whether student achievement levels are reasonable, valid, reliable, and informative to the public.” Part of the new study will be a review of evidence for the reliability and validity of the achievement levels.

¹ Please see the solicitation at this location:

<https://www.fbo.gov/?s=opportunity&mode=form&id=c9e23849c5ee6b1ee388832e9d5d3a37&tab=core&cview=0>.

- Methodology. A second aspect of the study will be to examine the methodology used to arrive at the achievement level cut point recommendations. Current achievement levels were set using different procedures from many of those employed prior to the previous evaluations. The study will evaluate the procedures for setting the NAEP achievement levels.
- Policy Use. The study will examine those parts of the achievement level setting process that go beyond its technical aspects. The actual setting of cut points occurs within the context of several policy-level decisions. Each achievement level has a policy-level definition that the Governing Board has adopted. For example, “Proficient” is defined as mastery over challenging subject matter. Also, prior to setting the cut points, achievement level descriptions are adopted for each subject that describe what students should know and be able to do at each achievement level. Finally, after the technical committees recommend the cut points, the Governing Board makes the decision to accept or change them in light of the nation’s educational priorities and needs. The study will consider the achievement levels in the context of these policy-level decisions.
- Consequential validity. A fourth important aspect of the study will be to examine the evidence supporting the consequential validity of the achievement levels. The contractor will examine the levels’ impact on NAEP stakeholders, and any consequences, both intended and unintended, of setting the cut points at their particular locations on the NAEP scale. The study will also examine the usefulness of the achievement levels to stakeholders and the public.

To guide the study, the contractor will establish a committee of 10-12 members with a broad range of expertise related to assessment, statistics, social science, and education policy. The work statement calls for panel members to be impartial and objective, and to represent a variety of perspectives.