

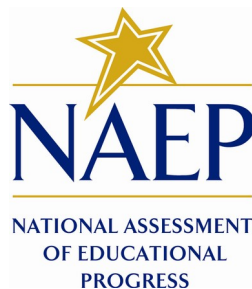
# ***NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS***

## ***Volume 1***

### ***SUPPORTING STATEMENT FOR WAVE 2 OF 2008 SUBMITTAL***

***(PART OF 2008-2010 SYSTEM CLEARANCE PROPOSAL  
OMB# 1850-0790)***

- Student Grade 4: Pilot Science; Reading Braided Study
- Student Grade 8: Pilot Science; Reading Braided Study
- Student Grade 12: Pilot Science, Reading Motivational Special Study
- Teacher Grade 4 (Background, Education, Training; Reading, Mathematics, Science)
- Teacher Grade 8 (Background, Education, Training; Reading)
- Teacher Grade 8 (Background, Education, Training; Mathematics)
- Teacher Grade 8 (Background, Education, Training; Science)
- School Grade 4 (School Characteristics & Policies; Reading, Mathematics, Science, Charter School)
- School Grade 8 (School Characteristics & Policies; Reading, Mathematics, Science, Charter School)
- School Grade 12 (School Characteristics & Policies; Reading, Mathematics, Science, Reading Motivational Special Study)



## **Explanation and Burden Information for This Submittal**

This document contains supplemental information pertaining to the 2008-2010 NAEP System Clearance proposal (submitted in January 2007). The terms of clearance for OMB approvals state that each subsequent submittal activity under the System Clearance is to be submitted to OMB.

This submittal contains burden information and the actual background questionnaires for the following components<sup>1</sup> of the 2008 assessments:

- Student Grade 4 (Science - both new 2008 questions and existing 2005 questions which are part of the bilingual book being administered; Reading Braided Study)
- Student Grade 8 - Science (both new 2008 questions and existing 2005 questions which are part of the bilingual book being administered; Reading Braided Study)
- Student Grade 12 (Science pilot, Reading Motivational Special Study)
- Teacher Grade 4 (Background, Education, Training; Reading; Mathematics; Science)
- Reading Teacher Grade 8 (Background, Education, Training; Reading)
- Mathematics Teacher Grade 8 (Background, Education, Training; Mathematics)
- Science Teacher Grade 8 (Background, Education, Training; Science)
- School Grade 4 (School Characteristics and Policies; Reading; Math; Science; Charter School)
- School Grade 8 (School Characteristics and Policies; Reading; Math; Science; Charter School)
- School Grade 12 (School Characteristics and Policies; Reading; Math; Science; Reading Motivational Special Study)

These specific questionnaires are the third group (Wave 2) of questionnaires submitted for approval for usage in 2008. The first groups of questions were submitted as part of the System Clearance submittal in January 2007 and included: student core questions (grades 4, 8, 12); reading and mathematics subject-specific background questions (grades 4 and

---

<sup>1</sup> Note: background questions that have already been approved in earlier 2008 submittals are not again included (i.e. 4<sup>th</sup> and 8<sup>th</sup> grade reading-specific background questions being used in the Reading Braided studies were previously approved).

8); long-term trend (LTT) core, reading and mathematics questions (ages 9, 13, 17). A subsequent (Wave 1) submittal included: grade 8 student questions (Arts); grade 12 student questions (pilot reading, pilot math); grade 8 school questionnaires for the Arts; and SD (Students with Disabilities) and ELL (English Language Learners) questionnaires for the LTT (long-term trend) assessments. There will be one final submittal (Wave 3) for 2008 NAEP which will include SD and ELL questionnaires. These questionnaires will be completed by school personnel for students participating in the operational, pre-calibration, and pilot 2008 assessments who are identified as SD and/or ELL.

**Estimated Respondent Burden for 2008 Assessments  
Contained in this Submittal  
By Grade**

2008

Part	Grade/ Component	# of Student s	Student Burden (in hours)	# of Teacher s	Teacher Burden (in hours)	# of School s	School Burden (in hours)	SD/ELL* (# of school personnel )	SD/ ELL* Burden (in hours)
1 of 8	4th Grade Student Science Reading Braid	5,000 10,000	1,250 2,500						
2 of 8	8th Grade Student Science Reading Braid	4,500 12,000	1,125 3,000						
3 of 8	12th Grade Student Science Reading Motivation	5,500 3,600	1,375 900						
4 of 8	4th Grade Teacher Science Math Reading			300 312 1,320	100 104 440				
5 of 8	8th Grade Teacher Science Math Reading			270 312 1,680	90 104 560				
6 of 8	4th Grade School Science Math Reading					100 104 440	50 52 220		
7 of 8	8th Grade School Science Math Reading					90 104 555	45 52 278		
8 of 8	12th Grade School Science Math Reading Reading Motivation					110 120 120 60	55 60 60 30		
		40,600	10,150	4,194	1,398	1,803	902		

\* SD-ELL burden estimates for operational, pre-calibration, and pilot for 2008 will be submitted in a Wave 3 submittal.

## **Overview of NAEP 2008-2010 Assessments**

The following broad overview of the 2008-2010 NAEP assessments was submitted as part of the initial systems clearance submittal. The National Assessment Governing Board determines NAEP policy and assessment schedule, and future Board decisions may result in changes to some aspects of an assessment (e.g., which subjects are assessed in which years). However, overall methodology and assessment process will remain constant. In the 2008 assessment year, questionnaires will be administered to students at grades 4, 8, and 12; to students at ages 9, 13, and 17 for LTT; to teachers at grades 4 and 8; and to school administrators at grades 4, 8, and 12.

The 2008 assessments consist of:

- National operational assessments in the Arts (Visual Arts and Music) at grade 8 and long-term trend assessments at ages 9, 13, and 17.
- Pre-calibration field test assessments for reading and mathematics at grades 4 and 8.
- Pilot assessments in science at grades 4, 8, and 12; pilot assessments in reading and mathematics at grade 12; LTT pilot assessments for mathematics at ages 9, 13, and 17; bridging ('braided') studies for reading at grades 4 and 8; and a special study on incentives for grade 12 reading.

## **How, by Whom, and for What Purpose the Data Will be Used**

In the original request for system clearance, NCES asked for approval of the instruments to be used to gather data from the 2008-2010 national and state assessments. This submittal applies to the third set of questionnaires (Science – Grades 4, 8, 12; Reading – Grades 4 & 8 Braided studies; teacher (grades 4 & 8) and school (grades 4, 8 12)) that will be submitted for the 2008 assessments. The first set of questions contained student core questions (grades 4, 8, and 12), reading and mathematics student subject-specific questions (grades 4 and 8), and student long-term trend (ages 9, 13, and 17). The second

set of questions (Wave 1) contained Arts - grade 8, Reading and Mathematics – grade 12, and SD-ELL for long-term trend questionnaires.

Given that the purpose of NAEP is to gather data on the achievement of students in the subject areas assessed for use in monitoring education progress, and because of the program's increasing visibility, it is incumbent on the program to develop the most reliable and valid instruments possible. To do so, NAEP employs four strategies:

- A. Small-scale pilot testing of new materials and test administration techniques;
- B. Pilot testing items to determine which items best measure the constructs under consideration;
- C. Field testing of operational assessments to accommodate the mandated six-month reporting; and
- D. Full-scale operations.

Questionnaire development follows the same pattern as that of cognitive item development, although we tend to pilot fewer items with less duplication and use the resulting data to refine the questions. Guidance for what is asked is provided by the National Assessment Governing Board. NCES develops the questionnaires, which the Governing Board then approves for submission to OMB in a two-stage process. The Governing Board approves the questionnaires prior to pilot testing, and then again after NCES and its contractors make selections for the operational assessment based on pilot data. The questions are designed to provide the information for disaggregating data according to categories specified in the legislation, to provide contextual information that is subject specific (e.g. reading, mathematics) and has an impact and known relationship to achievement, and to provide policy relevant information specified by the Governing Board.

## **Design Information for the 2008 Administrations Contained In This Submittal**

### **Pilot-test components.**

To support future assessments, NAEP pilot tests items to replace items that are released. In cases when new frameworks are being introduced (e.g., 2009 science at all grades and mathematics at grade 12), the majority of the items are developed and pilot tested. Pilot testing of items is intended to provide item-level data that is used for two purposes: (1) to revise and improve the items and (2) to select items for use in future operational assessments.

#### **i. Science pilot tests at grades 4, 8 and 12**

A new framework is set to be implemented with the 2009 science assessment. The 2008 science pilot will be used to provide information that will help revise, improve, and select the items and tasks that will be part of the 2009 operational assessment. Even though the 2009 science assessment will be the first under a new framework, NCES has decided to carry over some items from the 2005 science assessment that map to the new framework. Pilot tests in 2008 will be built using the common booklet model. However, because of the introduction of new hands-on-tasks (HOTs) and interactive computer tasks (ICTs), the test and sample design will vary somewhat from the traditional design.

### **Trend study components.**

NAEP will typically conduct trend studies to determine if significant changes in assessment conditions and/or procedures may impact differentially upon student performance. In 2008, NCES will conduct a “bridging” study in reading at grades 4 and 8 to facilitate a state-level trend study in 2009. The study will administer “braided” books to students, which contain one block from the old assessment and one block from the new assessment. Order and context effects would be examined using this information.



### **Teacher and School Components.**

**Teachers.** The teachers of fourth- and eighth-grade students participating in NAEP will be asked to complete questionnaires about their teaching background, education, training, classroom organization, and school community issues. Teacher questionnaire data will be collected at grades 4 and 8 for teachers whose students participate in the science pilot, reading and math pre-calibrations, and the reading braided studies.

**Principals/Administrators.** The school administrators in the sample schools will be asked to complete a questionnaire. As with the teacher questionnaires, the core questions are designed to measure school characteristics and policies that research has shown are highly correlated with student achievement. School questionnaire data will be collected at schools whose students participate in these administrations: arts operational (grade 8), science pilot (grades 4, 8, 12), reading and math pre-calibrations (grades 4 & 8), the reading braided studies (grades 4 & 8), and the reading and math pilots (grade 12).

### **Additional studies.**

NAEP frequently includes additional studies, as was discussed in the System Clearance submittal, in regular assessments to investigate content issues (e.g., Meaning Vocabulary Study in 2007), delivery options (e.g., various technology-based assessments), linking to other NCES surveys (e.g., Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K) Linking Study in 2007), or reporting variables (e.g., socio-economic status (SES) Indicator Study in 2007).

In 2008, a special study at grade 12 reading will be conducted to determine whether the outcomes for students sitting for the 12th grade NAEP reading assessment reasonably reflect their capabilities and, if not, what are the possible impacts on the many statistics reported by NAEP<sup>2</sup>. The study will investigate the issue of differential student engagement; that is, whether the phenomenon of reduced engagement and effort (if it exists) is more prevalent in certain subgroups of the student population. Some of the research questions addressed in the study will include:

- Do 12th grade students taking the NAEP reading assessment and offered “strong” performance incentives display greater levels of engagement and/or achieve

---

<sup>2</sup> For a more complete description of the background, purpose, and design of this study please refer to Appendix A.

higher scaled scores on average than comparable students who are not offered such incentives?

- Are there engagement or performance differences by treatment condition among students classified by ability level, by gender or by race/ethnicity?
- If there are differences in performance by treatment group, what is the likely impact on the statistics reported by NAEP, as well as other indicators that are constructed from NAEP data?
- Are there detectable differences between the control condition (fall administration) and the standard NAEP spring administration?

The study will be carried out in the fall of 2007. This will avoid any interference with the regular administration of 12<sup>th</sup> grade NAEP scheduled for the spring of 2008.

Under Section A, Question 9 in the 2008-10 NAEP Systems Clearance request submitted in January 2007, we indicated that NAEP does not typically offer any incentives to students, teachers or schools for participation. For this special study, NAEP will provide cash (i.e., gift cards) incentives of up to \$25.00 to each participating student.

# **APPENDIX A**

## **NAEP 12<sup>th</sup> GRADE INCENTIVE STUDY**

Henry Braun  
Boston College

Irwin Kirsch  
Educational Testing Service

## INTRODUCTION

As countries around the globe have come to appreciate the importance of human capital to long-term economic success, there has been a concomitant increase in attention to their education systems, with a particular focus on achieving greater progress with respect to the goals of access, quality and equity. In addition to mandated examinations that are used to determine student progress or graduation, large-scale assessments (LSAs) have come to play a significant role in informing public policy. In this country, examples include cross-sectional surveys such as the National Assessment of Educational Progress (NAEP), the National Assessment of Adult Literacy (NAAL) and longitudinal surveys such as High School and Beyond (HSB) and the Early Childhood Longitudinal Survey (ECLS-K). At the international level, examples include the Third International Mathematics and Science Study (TIMSS) series, the International Adult Literacy Survey (IALS) and the Program in International Student Assessment (PISA).

LSAs have faced greater pressure to establish the validity of their findings as those findings have achieved greater salience in the policy arena. This has meant addressing questions concerning both the nature of the raw data collected and the analytic procedures that generate the statistics reported. Clearly, high quality data is fundamental to the credibility of the enterprise. At the same time, as LSAs have grown in size and complexity, the task of fully documenting and validating their procedures, from data collection through reporting, has become a daunting one.

One of the concerns arising with low-stakes LSAs is whether student motivation and engagement is attenuated (in comparison with situations in which test results have consequences for the student), resulting in depressed performance. A thoughtful review of the literature on motivation and effort is provided by Baumert and Demmrich (2001). They also conducted an experimental study administering items from the PISA math literacy test to a small sample of German 9<sup>th</sup> graders under different incentive conditions. They did not find statistically significant differences with respect to either invested effort or to performance.

Concerns about data quality in large-scale data bases are certainly not limited to the education sector. In other policy arenas, such as economics and health care, such databases have also come to play increasingly important roles in monitoring current procedures and in establishing rationales for new initiatives. Efforts to evaluate and ameliorate the quality of these data bases have grown apace.

From its rather modest beginnings in 1969, NAEP has become central to many conversations about education in this country. With No Child Left Behind (NCLB), NAEP assumes greater prominence as an instrument for monitoring state-level results and as a basis for tracking the nation's progress over time. In this context, 12<sup>th</sup> grade NAEP holds a somewhat anomalous position in that it does not play a role in NCLB and it does not report results at the state level. However, this state of affairs is not permanent: There has been discussion at various levels concerning the possibility of transforming the

assessment into a “readiness” measure, as well as extending it by drawing representative samples for each state.

Inasmuch as such changes would require substantial outlays, policy-makers have to be convinced that the funds would be well spent, despite an ongoing concern with 12<sup>th</sup> grade NAEP’s data quality. Participation rates both at the school and student levels have been lower for 12<sup>th</sup> grade than for 4<sup>th</sup> and 8<sup>th</sup> grade. Moreover, the 12<sup>th</sup> grade rates have declined over time. A related issue is the level of engagement among students who do participate in the 12<sup>th</sup> grade assessment. Are they “doing their best” all through the assessment, working through part of it before slacking off, or just “going through the motions” for the entire period?

There are many plausible rationales for lower participation and engagement. In the spring of their senior year, most students have priorities other than sitting for a demanding assessment that is of no consequence to them personally. Indeed, given that the diversity of educational experiences and interests among 12<sup>th</sup> grade students is so much greater than in the earlier grades, devising an assessment that will be seen as meaningful and appropriate for all students is a daunting challenge.

The issues of engagement and effort are critical to the proper interpretation of the NAEP data. To the extent that the reported results underestimate (in some sense) what (at least some) American high school seniors know and can do at the end of their high school careers, policy-makers and the public at large can be misled. It is also the case that in international comparisons, the relative standing of American students declines with age and grade. It is not clear the extent to which this decline is due to national differences in motivation and engagement in the assessment process.

Clearly, both participation and engagement are critical issues. However, it is difficult to explore both in a single study because of the different strategies and designs that are required. We have chosen to focus on engagement as it is possible to carry out a substantively useful investigation with comparatively modest resources.

Accordingly, the primary goal of this study is to determine whether the outcomes for students sitting for the 12<sup>th</sup> grade NAEP reading assessment reasonably reflect their capabilities and, if not, what are the possible impacts on the many statistics reported by NAEP. The study will also investigate the issue of differential student engagement; that is, whether the phenomenon of reduced engagement and effort (if it exists) is more prevalent in certain subgroups of the student population.

To this end, we propose to carry out a randomized experiment in which students will either be placed in a control group or will be offered one of two incentives that are intended to motivate them to “do their best”. The size of the experiment will be such that there is substantial power to detect overall departures from the null hypothesis (of no difference among treatments) corresponding to a small (but practically meaningful) effect size. It will also be large enough to detect differential performance between large subgroups, if the difference corresponds to a moderate effect size.

Our intention is to place the student outcomes from the study on the NAEP scale so that the inferences with regard to differential performance can be presented in terms of NAEP score points. This will substantially enhance the value of the study and facilitate policy discussions concerning the meaning and import of the results. To accomplish this goal, we will work with the NAEP contractors responsible for data collection, processing and analysis. The procedures will mimic as closely as possible, given the budgetary constraints, those employed for NAEP operational work.

## RESEARCH QUESTIONS

1. Do 12<sup>th</sup> grade students taking the NAEP reading assessment and offered “strong” performance incentives display greater levels of engagement and/or achieve higher scaled scores on average than comparable students who are not offered such incentives?
2. Are there engagement or performance differences by treatment condition among students classified by ability level, by gender or by race/ethnicity?
3. If there are differences in performance by treatment group, what is the likely impact on the statistics reported by NAEP, as well as other indicators that are constructed from NAEP data?
4. Are there detectable differences between the control condition (fall administration) and the standard NAEP spring administration?

## INSTRUMENTATION AND DESIGN

1. A NAEP assessment is built from a pool of “item blocks”, with each block consisting of a number of items or questions. There is no overlap of items across blocks. Typically, about half the items in a reading block are in a multiple choice format and about half in a constructed response format, with a mixture of short- and extended-answer items. Pairs of blocks are assembled into booklets following a complex design pattern. Each booklet is expected to take about 50 minutes to complete.
2. We will select four blocks from the pool of NAEP 12<sup>th</sup> grade released reading blocks. Two blocks will have been classified as reading for literary experience (??) and two as reading for information. The blocks will be selected to reflect the constructs underlying the relevant reading subscales. Suppose the blocks for reading experience are denoted A and B, and the blocks for reading for information are denoted C and D. Then the blocks will be assembled into 8 booklets, and the booklets spiraled so that each booklet has an equal chance of being administered. The 8 booklets will conform to the following pattern:

Booklet	Block 1	Block 2
1	A	B
2	B	A
3	C	D
4	D	C
5	A	C
6	C	A
7	B	D
8	D	B

With this design, we will be able to

- Estimate and eliminate order effects
  - Establish each subscale
  - Estimate the covariance between subscales, which is essential to the construction of the composite NAEP scale.
3. We will administer the standard NAEP student questionnaire and ask that participating schools to fill out the standard NAEP school questionnaire and the administration schedule. Some of this information is needed for the conditioning model used to generate the plausible values for each student and some the information is required for contextual analyses.
  4. We will also administer a questionnaire to students to ascertain their general level of engagement in reading practices as well as their relative level of effort on this assessment. We intend to explore the possibility of obtaining student grades. That information will be helpful in interpreting the results of the study.
  5. The study will be carried out in the fall of 2007. This will avoid any interference with the regular administration of 12<sup>th</sup> grade NAEP scheduled for the spring of 2008. In a certain sense, a fall administration with no incentives can be considered as a type of treatment in comparison to the standard spring administration.

## SAMPLE SELECTION

The intent of this study is to obtain credible estimates of differences among treatment conditions overall and by student type, rather than national estimates of NAEP performance under each condition. Accordingly, sample selection is guided primarily by the need to enroll heterogeneous groups of students in each condition rather than nationally representative samples. The implication is that, consistent with cost constraints, the school sample should include as many schools as possible, in different locations with maximum diversity within and among schools with respect to student characteristics.

We will collaborate with the appropriate NAEP contractor to select a supplemental school sample to the national school sample for the spring 2008 12<sup>th</sup> grade NAEP

2008 Wave II Supporting Statement

administration. This will ensure that schools invited to participate in the study will not also be asked to participate in the operational assessment. It is expected that approximately 60 schools located in 6 or 7 states will be chosen.

We will work with both NAEP contractors and the NAEP state coordinators in the relevant states to obtain schools' agreement to participate in the study. The state coordinators are responsible, among other things, for school participation in operational NAEP. They are well-connected and experienced, and will be invaluable resource for this study. We also expect to present schools with a letter from the US Department of Education explaining the importance of the study. In addition, we will promise to make a donation to the school's senior class fund.

Within each participating school, approximately 60 students will be selected at random and invited to sit for the assessment. The students will be randomly assigned to one of three sessions of approximately 20 students each. The NAEP contractor will carry out the administration of the assessment and be responsible for collecting the booklets and shipping them to another NAEP contractor for processing and scoring.

## INCENTIVES

There will be three "arms" to the study:

1. Control condition. Students will be given standard NAEP instructions. Subsequent to completing the assessment, they will be given a debit card valued at \$5.
2. Incentive I. Students will be given standard NAEP instructions and told that at the conclusion of the session they will receive a debit card valued at \$20 in appreciation for their participation and applying their best efforts to answer each item. They will also be asked to indicate which of two debit cards they would like to have. The cards will be linked to different stores. It is hoped that the effect of the incentive will be enhanced by having the students actively make a choice in advance of the assessment.
3. Incentive II. Students will be given standard NAEP instructions and told that at the conclusion of the session they will receive a debit card valued at \$5. In addition, two questions will be selected at random from the booklet. The debit card will be increased by \$10 for each correct answer, so they can receive a maximum. They will also be asked to indicate which of two debit cards they would like to have. The cards will be linked to different stores. It is hoped that the effect of the incentive will be enhanced by having the students actively make a choice in advance of the assessment.