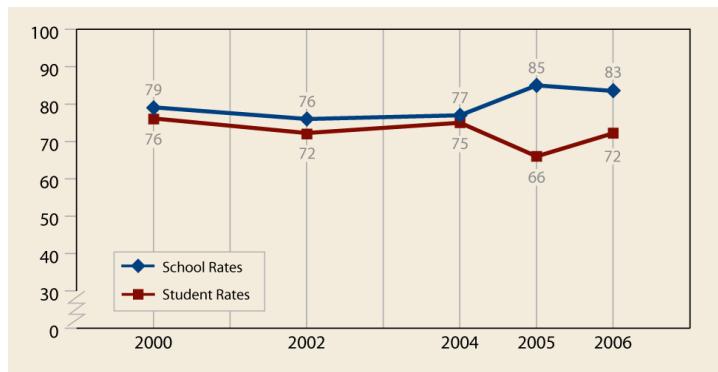


## Context for the 12<sup>th</sup> Grade Motivation Study

As early as 1998, NCES and the National Assessment Governing Board were concerned about school and student participation in NAEP. An AllStates 2000 Task Force was convened to address the issues and make recommendations for improving both school and student participation. The Task Force recognized that there were two related factors for 12<sup>th</sup> grade students that had to be addressed: participation and motivation.

In 2001, conversations began on the No Child Left Behind legislation (NCLB) and the requirement that all states receiving Title I funds participate in NAEP reading and mathematics assessments in 4<sup>th</sup> and 8<sup>th</sup> grades. In the rush to implement the NCLB provisions, the focus on 12<sup>th</sup> grade participation took a back seat. As early as 2003, all states and over 99 percent of the schools and more than 95 percent of the students (excluding some special needs students) participated in the NCLB NAEP assessments. However, the participation rate for 12<sup>th</sup> grade schools and students continued to decline as illustrated below.



Early in 2002, the Governing Board convened a National Commission on NAEP 12<sup>th</sup> Grade Assessment and Reporting followed in 2005 by an Ad Hoc Committee on 12<sup>th</sup> Grade Participation and Motivation. NCES began a series of panel discussions with principals, teachers, and district superintendents from the program/operational perspective. The conclusions from the groups were very consistent:

*the validity and credibility of the NAEP 12<sup>th</sup> grade results are severely compromised by the poor participation of both schools and students and by the perceived low motivation of students to try their best on a test that has no consequences for them or their schools.*

The educators went further in their analysis of the situation. They said, in effect, that

*while participation rates for both schools and students could be improved with some basic changes to operations, that low motivation is generally assumed to be a problem, is poorly defined and understood, and would be much more difficult to change.*

Their bottom line was that *if we are unable to improve the motivation of students, then the results will always be, at the least, suspect, and, at worst, misleading.*

The principals cautioned that if motivation of students to try on the NAEP assessments is not improved, it will, in the end, create a circular environment where schools will cease to try to help us resulting in participation rates going back down.

This prediction is now being borne out by the states. New York has, for the 2003 and 2005 assessment, refused to participate in 12<sup>th</sup> grade assessments because they perceived not only that the test is burdensome to their schools but that because of low motivation the results are misleading. New York, critical to providing a national score, was persuaded by US Department officials to let the NAEP contractor contact the schools and try to get them to participate. Other states such as Maryland, Rhode Island, and Nevada have just refused to permit NAEP 12<sup>th</sup> grade assessments in their schools on the principle that the results are not creditable.

The participation and motivation issues have been made more critical with the continued discussions in Congress to support a 10-state pilot study for a 12<sup>th</sup> Grade State NAEP.

Much has been done on improving the participation of both schools and students. In 2005, the new NAEP State Coordinators made a special effort to recruit schools bringing the rate to 85 percent, which is comparable to the participation rate of 4<sup>th</sup> and 8<sup>th</sup> grade schools when it was still voluntary. In 2006 and 2007, the focus was on improving student participation. The student participation rate grew from a low of 66 percent in 2005 to 72 percent in 2006, not good enough but moving in the right direction.

There has been much internal discussion about motivation over the last five years, but no concerted effort to influence student motivation, for NAEP as a program or for other NCES student surveys. NAEP could be criticized for not acting sooner to implement a concerted effort to address a serious threat to the program.

The proposed study, part of an emerging NAEP research agenda around student motivation, is a critical first step in understanding and addressing the issue. The results of this study are not envisioned to provide a definitive answer, but will help to define the magnitude of the problem and can provide insights into whether students can be incentivized to try harder on a low stakes assessment such as NAEP.

The use of monetary incentives in this study should not be construed as NCES' intention of offering 12<sup>th</sup> graders monetary incentives to participate in NAEP. Monetary incentives are used in this study primarily because it enables us to overcome the challenge of standardization of methodology across schools. The feedback from educators is that successful incentives vary by school or school district. Therefore, in an operational setting, non-monetary incentives will have to be tailored to the culture of the school/district. For example, while offering community service credit as an incentive would be allowable in one school or district, it might be forbidden in another.

The complete research agenda is evolving with the input of the NAEP contractors, the NAEP Validity Studies Panel, and the ETS Design and Analysis Committee, as well as our state partners. The next step is to talk to students. The design for a series of student focus groups is being completed now with a target for completion in the spring.

## **Treatment conditions and timeline for NAEP 12<sup>th</sup> Grade Motivation Study**

The following is a description of the three treatment conditions for the 12<sup>th</sup> Grade Motivation Study:

### **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 one of two different brands of debit cards they would like to have. 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.

### **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, upon completion of the administration, two questions will be selected at random from the booklet of each student. The debit card will be increased by \$15 for each correct answer provided to the two randomly selected questions, so that each student can receive a maximum of \$35. As in the first incentive group, they will also be asked to indicate which of two debit cards they would like to have.

### **Control**

Students will be given standard NAEP instructions. Subsequent to completing the assessment, each student will be given a debit card valued at \$35.

All students will receive an identical card value of \$35, at the conclusion of the study, regardless of condition. At that time, we will debrief students about the three treatment conditions, and explain that in order to be fair to all participating students, we have decided to give each student a \$35 card. Students who participated in the control will also be invited to select the debit card of their choice.

The following is the expected timeline for the study:

### **Schedule of activities for NAEP 12<sup>th</sup> Grade Motivation Study**

Week of Sept. 17	Train staff	Westat – H.O.
Week of Sept. 24	Contact schools to schedule assessments	Field staff
Week of Sept. 24	Receive materials from Pearson	Field staff
Oct 14 – Nov. 15	Conduct assessments	Field staff

Sample Letter from  
NAEP Coordinators to School Principal  
Draft 9-13-07

Dear [Principal Name]:

On behalf of the National Center for Education Statistics, Educational Testing Service, Princeton University, and Boston College, I want to thank you for <insert school's name> participation in the Grade 12 Motivation Study that is being conducted by the National Assessment of Educational Progress (NAEP). Data from recent Grade 12 NAEP assessments suggests that many 12<sup>th</sup> grade students do not do their best on NAEP tests. This study will examine how differences in student motivation are related to their performance on the test. Results of this study will be used to improve student engagement and participation in the future. This is particularly important, since NAEP results inform policy makers' decisions about education funding and allocation of resources.

*The assessment will take place between mid-October and mid-November on a day that is convenient for your school. A random sample of 75 12<sup>th</sup> graders will be selected to participate. The assessment will take approximately 90 minutes. Students will be divided into three separate groups and tested in different locations simultaneously. Students will complete a standard NAEP reading assessment and a short background questionnaire.*

In preparation for the assessment, please designate a School Coordinator who will serve as liaison for all study-related activities in the school. In the fall, several weeks prior to the assessment, a representative from Westat, the contractor responsible for administering the assessment, will contact your NAEP School Coordinator to discuss assessment details and finalize planning. Westat field staff will administer the assessment and provide support to the School Coordinator. At the conclusion of the study, participating schools will receive the study report and be invited to a web-based discussion of the results of the report. Schools participating in this unique study will receive a stipend of \$200.

*Participating students will receive a modest reward for their participation. A crucial aspect of the study is that the students not know that rewards will be involved until they are actually taking the assessment. Therefore, it is important that this information be kept confidential and not communicated to the students in advance.*

It is important that we achieve 100 percent participation of both schools and students. I know that we can count on your help in reaching this goal. Once again, thank you for your assistance with this very important study of our nation's high school seniors. If you have questions, please contact me at <insert telephone number> or <insert email address>.

Sincerely,

NAEP State Coordinator

# GRADE 12 Motivation Study

## Session M03

### INTRODUCTION

Good morning/afternoon! My name is (YOUR NAME). Today you will be participating in a test called the National Assessment of Educational Progress, also known as NAEP or The Nation's Report Card. NAEP is a way to show what students like you – from all around the country – know and can do in different subjects. You and your school were selected to represent other seniors and schools across (STATE NAME) and the United States.

Here are some things for you to keep in mind: the test takes about 90 minutes and you will be answering questions in reading along with questions about yourself and your experiences in and out of school. Do not write your name on the work you do; no one in the school will see your answers. Your answers will be combined with information from other seniors across (STATE NAME) and the United States. Because your responses are useful in showing our country's leaders and teachers what American high school students are learning, we ask that you try your very best. Thank you for your participation in NAEP.

**[The information below is an excerpt from the full session script]**

### GIFT CARD INSTRUCTIONS

Now remove the 3-ply form from the front cover of your booklet (HOLD UP FORM) and read along as I read out loud.

In appreciation of your agreement to participate in this study and in anticipation of your answering every question to the best of your ability, you will receive a \$5 gift card. In addition, 2 questions will be randomly selected from your booklet. You will receive an additional \$15 in gift cards for each of the questions you answer correctly. If you answer both of these

questions correctly you will receive gift cards totaling \$35. You may choose one of two brands of gift cards.

*After you have completed your booklet, you will be asked to write your birth date in the space provided on this form to indicate that you have both tried your best and have received the gift cards.*

## COLLECT MATERIALS

I'm now going to collect everything but the pencil from you. As I call your name, please come forward with your assessment booklet and the form.

### ***[Debriefing script to be inserted.]***

After you have received your \$35 gift card, please enter your birth date on the form to acknowledge this, and then return to your seat. I'll keep all copies of the form

- Read students' names in Administration Schedule order to collect booklets and forms.
- Give the student one \$35 gift card and have the student enter his or her birth date on the form. If a student objects to entering his or her birth date, ask him or her to mark an "X" instead.
- Collect the entire form. Students do not receive a copy.
- Verify that each student has returned an assessment booklet and that all unused booklets are accounted for.
- ***[Debriefing instructions to be inserted]***
- Thank the students and dismiss them according to school policy after all booklets and signed forms have been collected .

# **GRADE 12 Motivation Study**

## **Session MS1202**

### **INTRODUCTION**

Good morning/afternoon! My name is (YOUR NAME). Today you will be participating in a test called the National Assessment of Educational Progress, also known as NAEP or The Nation's Report Card. NAEP is a way to show what students like you – from all around the country – know and can do in different subjects. You and your school were selected to represent other seniors and schools across (STATE NAME) and the United States.

Here are some things for you to keep in mind: the test takes about 90 minutes and you will be answering questions in reading along with questions about yourself and your experiences in and out of school. Do not write your name on the work you do; no one in the school will see your answers. Your answers will be combined with information from other seniors across (STATE NAME) and the United States. Because your responses are useful in showing our country's leaders and teachers what American high school students are learning, we ask that you try your very best. Thank you for your participation in NAEP.

**[The information below is an excerpt from the full session script]**

### **GIFT CARD INSTRUCTIONS**

Now remove the 3-ply form from the front cover of your booklet (HOLD UP FORM) and read along as I read out loud.

In appreciation of your agreement to participate in this study and in anticipation of your answering every question to the best of your ability, you will receive a \$20 gift card. You may choose one of two brands of gift cards.

After you have completed your booklet, you will be asked to write your birth date in the space provided on this form to indicate that you have both tried your best and have received the gift cards.

## COLLECT MATERIALS

I'm now going to collect everything but the pencil from you. As I call your name, please come forward with your assessment booklet and the form.

### ***[Debriefing script to be inserted.]***

After you have received your \$35 gift card, please enter your birth date on the form to acknowledge this, and then return to your seat. I'll keep all copies of the form

- Read students' names in Administration Schedule order to collect booklets and forms.
- Give the student one \$35 gift card and have the student enter his or her birth date on the form. If a student objects to entering his or her birth date, ask him or her to mark an "X" instead.
- Collect the entire form. Students do not receive a copy.
- Verify that each student has returned an assessment booklet and that all unused booklets are accounted for.
- ***[Debriefing instructions to be inserted]***
- Thank the students and dismiss them according to school policy after all booklets and signed forms have been collected .

# GRADE 12 Motivation Study

## Session MS1201

### INTRODUCTION

Good morning/afternoon! My name is (YOUR NAME). Today you will be participating in a test called the National Assessment of Educational Progress, also known as NAEP or The Nation's Report Card. NAEP is a way to show what students like you – from all around the country – know and can do in different subjects. You and your school were selected to represent other seniors and schools across (STATE NAME) and the United States.

Here are some things for you to keep in mind: the test takes about 90 minutes and you will be answering questions in reading along with questions about yourself and your experiences in and out of school. Do not write your name on the work you do; no one in the school will see your answers. Your answers will be combined with information from other seniors across (STATE NAME) and the United States. Because your responses are useful in showing our country's leaders and teachers what American high school students are learning, we ask that you try your very best. Thank you for your participation in NAEP.

**[The information below is an excerpt from the full session script]**

### COLLECT MATERIALS

I'm now going to collect everything but the pencil from you. As I call your name, please come forward with your assessment booklet.

***[Debriefing script to be inserted.]***

After you have received your gift card, I'll ask you to enter your birth date on a form to acknowledge this, and then please return to your seat.

- Read students' names in Administration Schedule order to collect booklets and forms.
- Give the student one \$35 gift card and have them enter his or her birth date in Column O (Accommodation Booklet ID #) on the copy of the Administration Schedule (without names.) If a student objects to entering his or her birth date, ask him or her to mark an "X" instead.
- Verify that each student has returned an assessment booklet and that all unused booklets are accounted for.
- ***[Debriefing instructions to be inserted]***
- Thank the students and dismiss them according to school policy after all booklets and signed forms have been collected.

## **Grade 12 Motivation Study Debriefing Script**

Below are the objectives for development of the debriefing script. Once the official script is developed, it will be reviewed by the ETS Institutional Review Board

Objectives:

- Thank student for participating in the study and explain the purpose of the study
- Describe the three experimental conditions and explain why it was necessary to vary the incentives offered across the conditions
- Explain that, in appreciation of everyone's effort and in fairness to all, each student will receive the maximum \$35 regardless of the experimental condition.

## Responses to OMB Questions Regarding NAEP Wave 2 Clearance Package for the 2008 Assessment

### 4<sup>th</sup> Grade Science:

1. Please add PRA statement etc. to the front of this questionnaire.

Assuming this question is referring to the front of the background questions section that will appear in the grade 4 science bilingual booklet, the PRA statement approved for student booklets will be included at the beginning on this section (this statement also precedes the regular pilot student questionnaire section).

2. Please clarify the status of pretesting the new items. For example, #18, do 4<sup>th</sup> graders know what "hands on" means?

In the 2005 science student background questions section, the phrase "hands-on activities or projects" was used with grade 4 students. However, to be consistent with the wording used in questions 1-7 of the pilot science student background questions section and to avoid any confusion with the term "hands-on", we propose to reword questions 18 and 19 by replacing "hands-on activities" with "activities or projects."

18. In school this year, how often do you do activities or projects in science?

- A Never or hardly ever
- B Once or twice a month
- C Once or twice a week
- D Almost every day

19. In school this year, how often do you talk about measurements and results from your science activities or projects?

- A Never or hardly ever
- B Once or twice a month
- C Once or twice a week
- D Almost every day

3. #22, do most 4<sup>th</sup> graders "go" to a science class or is science a part of their day with a homeroom teacher, i.e., is the notion of "science class" clear?

Agreed. The word "class" will be deleted and question 22 will read:

22. I am sure I can understand whatever the teacher talks about in science.

- A This is not like me.
- B This is a little like me.
- C This is a lot like me.

### Special Study:

1. If the study is premised at least in part on the hypothesis that seniors are less engaged with graduation looming near, how will a fall test capture the same condition? We understand the desire to avoid complicating the spring operational test, but isn't there a way to do the study closer to graduation than the fall?

As indicated, the timing of the study in the fall was necessary so as not to interfere in any way with the regular administration of 12<sup>th</sup> grade NAEP that will take place in the spring of 2008. It should be noted, though, that the low-stakes nature of the assessment and the fact that seniors are generally pre-occupied with post-high school planning make a fall administration a feasible setting given the objectives of the study. Moreover, it affords an opportunity to compare results from the fall control condition to results from the regular spring administration. This is of independent interest as there has been interest in the impact of moving the regular grade 12 NAEP administration from the spring to the fall.

2. What evidence do you have that monetary incentives alter the level of engagement? The submission cites one study that did not show an effect.

There have been relatively few studies of the effect of extrinsic motivation on student performance on large-scale assessments. The initial study by O'Neill et al.(1992) was both underpowered and problematic because many students were “unaware” of the incentive treatment. The later study by O'Neill et al. (2001) had stronger and more memorable incentives but still suffered from low power. Both O'Neill studies focused on mathematics, while the current study focuses on reading. There is no extant research on NAEP 12<sup>th</sup> grade reading, so this study fills a gap in the literature. One can expect different dynamics in the two subjects: Improved performance on math requires both greater engagement and exposure to content; on the other hand, reading is less strongly related to curriculum at the 12<sup>th</sup> grade level. Brooks-Cooper and Bishop (1991) found an effect of monetary incentives on performance on literacy tasks. Segal (2006) found incentive effects on a coding speed test.

#### References

Brooks-Cooper, C. and Bishop, J. (1991). The effect of financial rewards on scores on NAEP tests. (Unpublished manuscript).

Sigel, C. (2006). Motivation, test scores and economic success. (Unpublished manuscript).

3. Can the study decouple "levels of engagement" and "achieve higher test scores" or will you have one result that will comprise both?

Indicators of level of engagement (i.e. situational motivation) will be constructed from patterns of items skipped and items not reached. These will be compared across the three groups in the study and with archival data on these same items. Information from the background questionnaire will be employed to estimate trait-level engagement and will be related to the patterns in the data. Indicators of achievement will be constructed from overall NAEP scale scores as well as performance by item type (i.e. multiple-choice, short constructed-response, extended constructed-response) and item position.

4. Please provide additional detail on power analysis used to determine number of schools needs and the subgroups of analytical interest.

The power analysis presented in the OMB clearance package demonstrates mathematically the trade-off between the magnitude of the effect sizes at different levels of aggregation and the degrees of freedom available. In other words, the fundamental problem is that, because of cost considerations, students must be sampled in clusters (i.e. by school) and there is a correlation in the scores among students in the same school. Thus, student scores cannot be treated as independent realizations from a common distribution. Accordingly, the appropriate method of estimation is to compute the difference in mean scores between treatment conditions within each school, yielding a one degree of freedom contrast. On the other hand, because the contrast is based on means, the effect size at this level of aggregation is (for the parameters presented) about 50% larger than the effect size at the student level. The required number of schools is then easily determined by traditional methods. Since there is also interest in evaluating the power for the comparison of subgroups, an illustrative calculation was presented under the assumption that only a quarter of the sample belonged to the subgroup of interest but that the student level effect size was twice as large. For the number of schools determined earlier, the power would then be the same. The point is that a study of the magnitude proposed will have substantial power (0.80) to detect small to moderate effect sizes. This power is considerably greater than that for studies in the literature.

5. Please explain why the control group is to be given an incentive (\$5) since the "normal" NAEP does not provide incentives.

Random sub-samples of students within each school will be allocated to each of the three treatment conditions. In order to reduce the possibility of ill-feelings within a school it was decided that each participating student should receive something of value. However, students in the control group are not informed before the administration that they will be given a \$5 gift card and the card is given to them only after the completion of the session. Thus, the control condition is not compromised.

The researchers also considered that students might perceive a degree of unfairness given that ultimately they will receive different dollar amounts. All students,

therefore, will receive an identical amount, at the conclusion of the study, regardless of condition. Specifications for the second incentive also have been revised in order to maximize effects. Please see below the revised description of the conditions.

6. Please explain in more detail what statistics this study will publish.

The study will publish for each treatment condition for all students and for selected student subgroups: (i) standard item statistics such as percent of students responding to the item, percent correct for dichotomous items and the score distribution for polytomous items; (ii) item characteristic curves; and (iii) scale score distributions. In addition, the study will publish inferential statistics derived from comparative analyses (e.g. analysis of covariance) of the three groups.

7. Related, how NCES hopes eventually to use the findings of this study, e.g., to develop models to adjust assessment scores for "non-engagement"?

The results of the study are intended to inform both the National Assessment Governing Board and the National Center for Education Statistics about the existence of reduced levels of engagement and sub-optimal performance -- and whether they result in material effects on the statistics routinely published in NAEP reports. If no effects are found, then the credibility of the reported statistics is enhanced. On the other hand, if main effects or interaction effects are observed, then policy makers should take these findings into account. In particular, if plans proceed to make fundamental changes in the purpose and design of 12<sup>th</sup> grade NAEP, then strategies to enhance situational motivation will need to be developed. However, the use of these results to make direct adjustments to NAEP results is not contemplated.

## 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 \$35.
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 \$15 for each correct answer, so they can receive a maximum of \$35. 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.

All students will receive an identical amount, \$35, at the conclusion of the study, regardless of condition.

Rev. 8/16/07

Answers to OMB Questions Regarding the Grade 12 Motivation Study

**1. Is the sample designed to be representative? This seems critical if the research objectives related to comparing a Fall versus Spring administration are to be valid.**

The sample is NOT designed to be nationally representative as in a regular NAEP administration. It is designed to be representative of the continuum of abilities found among 12th graders. The object is to study the effects of different incentives at various points along that continuum. By selecting diverse high schools in up to 7 states, we believe that we have achieved that objective.

It is important to recall that the key objective in our study is to estimate the impact of the incentive groups within season, so that the fall to spring differences don't enter directly. We did however make the point that we will be able to make statistical comparisons (not comparing mean scores) for our control group and those obtained in the regular spring administration. This will provide additional useful information to NCES and the National Assessment Governing Board.

*What do you find out by comparing fall to spring? Item parameters? See also response to question 6*

The fall-to-spring comparison would allow us to compare item parameters, such as difficulty and discrimination, and item characteristic curves between the two seasons. This item analysis would allow us, for example, to verify that the parameters that we have for these items, from previous spring administrations, are invariant when administered to students in the proposed fall study.

**2. Your concerns about ill will cause us to wonder about the potential for contamination. Please describe the conditions under which this experiment will be implemented. Specifically, it seems that if all 3 groups are not administered within a school virtually simultaneously, then contamination would be a serious concern.**

We have considered the possible issue of contamination. The plan is for the three sessions in each school to take place simultaneously whenever that is possible. In situations where this is not possible, the testing will take place in adjacent periods.

**3. Related, why did NCES choose to design the study with each of the 3 groups within a school, given concerns about contamination and ill will?**

By having all three conditions represented in each school, the estimated effect of the incentive will not be confounded with between school differences. Between school effects can be substantial and the size of the school sample is not large enough to insure that average between school differences will be negligible. Increasing school size would have added additional costs to the study that were beyond the monies available.

**4. Your description of your power analysis is useful, but it seems to presume that we've seen the actual analysis, which we haven't. Can you provide the actual numbers?**

The rationale and power analyses were described along with the actual numbers in the original proposal. A summary of the power analysis was included in the original OMB clearance package. The following is the more complete power analysis.

## POWER ANALYSIS

The power analysis will be carried out under the assumption that treatments will be approximately randomly assigned to students within schools. (Thus, average differences between schools will not contribute to estimates of treatment differences.) The calculation will be based on the comparison of a single treatment to a control, with the further assumption that the effect of the treatment is to add a fixed amount to each student's score. Formally,

$$y_{ij} = \mu + \alpha_i + \gamma I_{ij} + \varepsilon_{ij}$$

where

$y$  = NAEP score

$i$  indexes schools

$j$  indexes students within schools

$I_{ij}$  = 1 if student receives treatment and 0 otherwise.

Suppose that the  $\{\varepsilon_{ij}\}$  are identically distributed as random normal

deviates with zero mean and variance  $\sigma^2$ , and with correlation  $\rho$  between students in the same school but zero correlation between students in different schools.

Suppose further that within each school  $m$  students are assigned to each treatment condition and let

$y_{Ti}$  = mean for the treatment in school  $i$

$y_{Ci}$  = mean for the control in school  $i$ .

Then

$$\text{var}[y_{Ti} - y_{Ci}] = 2\sigma^2(1 - \rho)/m. \text{ Denote this variance by } \tau^2.$$

The effect size at the student level is  $\gamma/\sigma$  and the effect size at the school level is  $\gamma/\tau$ .

Denote these quantities by  $d$  and  $g$ , respectively. Then

$$d/g = \tau/\sigma = [2(1 - \rho)/m]^{1/2} = k(m, \rho).$$

For example, if  $d = 0.15$ ,  $\rho = 0.2$  and  $m = 16$ , then  $k \approx 0.32$  and  $g \approx 0.47$ . With these parameters, to obtain the number of schools to achieve a one-sided power of 0.8 at  $\alpha = 0.05$ , we enter the power tables (Cohen, 1988) with an effect size of 0.47, which yields 58 schools.

We also note that if  $d = 0.30$  and  $m = 4$ , then  $g = 0.47$ . Thus, if we are interested in detecting a treatment effect that comprises approximately 25% of the sample, the power is 0.8 for an effect size of 0.3.

## 5. What would the sample size difference be if you were to address the above concerns about contamination and representativeness?

We would likely have to double the school sample if we were to administer only one condition in each school. This would have the effect of increasing the budget by about 1/3.

6. **We would be interested in support for the assertion that "seniors are generally pre-occupied with post-high school planning...." in the Fall since NCES asserts that the Fall versus Spring timing doesn't really make a difference for this test.**

There is no documented research that we know of regarding this matter. However, it is empirically verifiable that many seniors in the fall are involved with taking admissions tests, completing college applications, visiting schools and the like. In the spring they have related concerns. Those students not planning for college also are concerned with post-high school planning throughout the year. It is important to recall that the key estimates are between incentive groups within season, so that the fall to spring differences don't enter directly. We have also made the point above that we will be able to make comparisons between the results for our control group and those obtained in the regular spring administration. While we will not be comparing actual means between groups, the kinds of statistical information we will gather (such as examining the stability of item parameters) will provide additional useful information to NCES and the Governing Board.

7. **Please justify the proposed incentive amount. OMB's recollection is that the \$20 amount used for the ELLs study was very successful in increasing 12<sup>th</sup> grade participation rates.**

This overall goal of this empirical study is concerned with level of engagement and effort and not participation. It is not clear that the cited results are relevant. Commissioner Schneider strongly recommended increasing the size of the second incentive condition and this was adopted.

*Provide clearer justification for incentive amount – link to participation such as results from ELS.*

The higher amount represents an attempt to increase the extrinsic motivation of the students. The literature shows that amounts offered to 12<sup>th</sup> grade students in similar studies have ranged as high as \$100. From that perspective, the \$35 proposed is rather modest in comparison. (See Harold O'Neil, 2004: Monetary Incentives for Low-stakes Tests). It is unlikely that we could create a similar incentive structure with a \$20 cap. Such a limit would not allow for sufficient differentiation between the two incentive groups, and would likely lessen the appeal of the incentives. The focus of the ELLs study was on student participation. The proposed study, however, focuses on student

8. **Has an IRB reviewed this study, particularly with the second set of incentives suggested? We are especially concerned about the ethical implications of misleading minors related to the incentive conditions and amounts.**

No subsequent IRB review has been done as the change was mandated by NCES at the time that the responses to OMB's first set of questions was being transmitted. NCES' concern was that after the administration, students would certainly share information and that discrepancies in the amounts received might cause ill-will. Therefore, it was felt that equalizing the final payments would be a preferred alternative even if it meant that students in the various groups would receive an amount greater than what had been promised.

*What kind of debriefing will be done? Is there a standard protocol for studies involving deception?*

We certainly appreciate the fact that a modest deception will take place. While we think that this is unlikely to cause undue distress, we will prepare a debriefing script that explains that we want to show equal appreciation to all the students who participated in the study. We will consult with social psychologists at ETS in the preparation of this script.

**9. Please provide copies of both cited unpublished studies.**

The Sigel article can be found at [http://www.people.hbs.edu/csegal/motivation\\_test\\_scores.pdf](http://www.people.hbs.edu/csegal/motivation_test_scores.pdf). Dr. Henry Braun, the principle researcher for the study, will provide a copy of the cited Brooks-Cooper and Bishop report. Unfortunately, Dr. Braun is traveling abroad and will be unable to forward the report until he returns to his ETS office in about a week.