

***NATIONAL ASSESSMENT OF  
EDUCATIONAL PROGRESS***

***Volume I  
Supporting Statement***

***Review of 2007 NAEP Mathematics Items  
Used in Puerto Rico***

***OMB# 1850-0803***

***(Generic Clearance for Cognitive, Pilot, and Field Test Studies)***



May 20, 2009  
(rev. June 3, 2009)

## Table of Contents

Table of Contents.....	2
1) Submittal-Related Information.....	2
2) Background.....	2
3) Design and Context.....	2
4) Item Rating and Teacher Interview Process.....	2
5) Consultations Outside the Agency.....	2
6) Assurance of Confidentiality.....	2
7) Justification for Sensitive Questions.....	2
8) Estimate of Hour Burden.....	2
9) Estimate of Costs for Recruiting and Paying Respondents.....	2
10) Cost to Federal Government.....	2
11) Schedule.....	2
Appendix A– Consent Form.....	2
Appendix B – Test Item Confidentiality Agreement.....	2
Appendix C- Thank You Letter to Teachers.....	2

## **1) Submittal-Related Information**

This material is being submitted under the generic Institute of Education Sciences (IES) clearance agreement (OMB #1850-0803 v.8) that was approved in July 2007. This generic clearance provides for the National Center of Educational Statistics (NCES) to conduct various procedures (field tests, cognitive interviews) to test new methodologies, question types, or delivery methods to improve survey and assessment instruments.

## **2) Background**

National Assessment of Educational Progress (NAEP) fourth- and eighth-grade mathematics assessments were administered to public school students in Puerto Rico in Spanish in 2003, 2005, and 2007. These assessments proved to be challenging in several ways. Compared to other jurisdictions, higher levels of missing data and fewer correct responses were observed at the item level. Moreover, the discrepancy between observed (empirical) and expected (model-based) responses was large, indicating that many of the mathematics items did not fit the assumptions guiding the NAEP Mathematics scale development. Consequently, NCES was not able to report the results of the 2007 Puerto Rico assessment on the NAEP reporting metric.

There are a number of potential explanations of the problems faced in NAEP's mathematics assessment in Puerto Rico. It is possible that the items in the assessment might not be adequately covering the ability range at which most Puerto Rican students perform. An alternative hypothesis is that the Puerto Rican students might be less motivated for a low-stakes assessment compared to students from the mainland. It is also possible that both low motivation and poor performance result from the differences between the curriculum Puerto Rican students follow and the NAEP assessment framework.

On the other hand, the potential issues listed above are not necessarily unique to Puerto Rico. For instance, there are other student groups that perform at the lower end of the achievement spectrum in the U.S. Problems such as item misfit and extremely low

percent correct values (such as values lower than 25% for multiple choice items with four response choices) do not appear to be as common in such student groups. NAEP is a low-stakes assessment for all students in the United States, not just for Puerto Rican students. In addition, differences between what is assessed by NAEP and what is taught in the classrooms exist in all jurisdictions since there is no national curriculum in the U.S.

There is, however, a challenge that is unique to Puerto Rico. NAEP's mathematics assessment is developed in English and then translated/adapted to Spanish for use in this jurisdiction. There is a substantial body of literature that illustrates how the difficulty and meaning of test items can be affected when they are administered in a different language to students with different learning experiences. In this study, we explore if this is the case for the NAEP mathematics items used in Puerto Rico. Using a survey tool, a number of Puerto Rican teachers will review selected 2007 NAEP mathematics items administered in Puerto Rico and rate the quality and the appropriateness of these items for the target Puerto Rican student population. Some of these items are secure, while some are released items. The main aspects the teachers will rate the items on are:

- Familiarity of the terms, visual representations (e.g. graphs), non-mathematical words/phrases used in the items,
- Familiarity of the context in which the problems are presented,
- Clarity of the language the items are presented in, and
- Complexity of the representations used in the items.

The teachers will also be asked in a group interview to comment in more detail on the problems they identified in these items, if any, and be asked to suggest revisions for the problematic items they identified. The study has the potential to not only explain the issues faced in NAEP's Puerto Rico assessment in 2007, but also inform item development, review, and translation procedures for future NAEP assessments in Puerto Rico.

### **3) Design and Context**

#### **Participants**

Thirty-nine teachers from Puerto Rico will participate in the study: 22 at grade four and 17 at grade eight. Teachers will be recruited with the help of the Puerto Rico Department of Education. The following qualifications are required for teachers to participate in the study:

- currently teaching mathematics at either fourth or eighth grade in Puerto Rican public schools;
- has taught at either fourth or eighth grade during the 2006-2007 school year; and
- has at least five years overall and at least two years of teaching experience at either fourth or eighth grade.

Although the study does not seek to attain a random sample of teachers, the 39 teachers will be selected from at least nine different schools. Further, not more than twelve of the fourth grade and six of the eighth grade teachers will be teaching in schools located within the boundary of San Juan.

#### **Items**

There are 163 fourth-grade and 166 eighth-grade items in NAEP's 2007 Puerto Rico mathematics assessment. We will select 60 items at each grade to be included in this study. Two groups of items will be used: one where the observed performance matches the model-based expected performance (items with best fit) and another group of items where the observed performance is lower than the model-based expected performance (items with worst fit). At each grade level, 30 items with the best fit and 30 items with the worst fit will be selected this way. Additionally, at grade four we will include 28 modified NAEP items (called 'accessible NAEP items') that are adapted in a way that they do not include construct-irrelevant aspects.

Items to be included in the study will be representative of both the content of the NAEP mathematics assessment (i.e., all five content areas: number properties and operations, measurement, geometry, data analysis and probability, and algebra) and of the different item types (i.e., multiple-choice and constructed-response).

Thirty-eight locally developed mathematics items at each grade level will also be selected by an Expert Panel (convened by Second Language Testing, Inc (SLTI) – see section 5 for more information on SLTI.) comprised of bilingual mathematics specialists for inclusion in the study. These items are developed in Spanish to assess Puerto Rican students. The experts will choose items from a larger pool obtained from the Puerto Rico Department of Education. They will select thirty-eight ‘local’ items that best match the NAEP items in terms of content and format.

As a result, 88 NAEP (30 items with good fit, 30 items with inadequate fit, and 28 ‘accessible’ items) and 38 ‘local’ items will be included in the study at grade four. Items will be assembled in booklets in a way that each booklet contains roughly equal numbers of items from different sources (NAEP items with good fit, NAEP items with bad fit, ‘accessible’ NAEP items, and local items). We will also make sure that the distribution of items in terms of item type and content are as similar as possible across booklets. The order of the items within each booklet will be completely randomized. Items will be assembled in booklets in such a way that each item is rated by nine teachers and each teacher rates 50 to 60 items.

Similarly, 60 NAEP (30 items with good fit, and 30 items with bad fit) and 38 ‘local’ items will be included in the study at grade eight. Items will be assembled following the same principles as described above. At grade eight each item will be rated by nine teachers and each teacher will be rating 50 to 60 items.

### **Item Review Tool**

Each teacher will rate 50 to 60 mathematics items using a survey tool (Item Review Tool). The Item Review Tool is contained in Volume II of this submittal (page 14). The tool features 11 statements about the mathematics item being reviewed. The Item Review Tool has been developed based on a comprehensive literature review. The statements included in the Item Review Tool ask teachers about, among other things,

- familiarity of the terms, visual representations (e.g. graphs), non-mathematical words/phrases used in the items,
- familiarity of the context in which the problems are presented,
- clarity of the language the items are presented in, and
- complexity of the representations used in the items.

The teachers will indicate their ratings using a Likert scale (1 to 4) on each statement for all items they review.

### **Item review and interview information**

The Item Review Tool and the data collection script (including interview questions) are contained in Volume II of this submittal (pages 4–17). The script includes:

- welcome/thank you/introductory remarks,
- consent and confidentiality forms,
- Item Review Tool,
- a sample mathematics item,
- a teacher background questionnaire,
- a questionnaire about reviewed mathematics items,
- interview questions following the rating process, and
- closing remarks/thanks.

Since some of the NAEP items the teachers will be reviewing are secure items, the teachers will be asked to sign a confidentiality form before they look at these items. The consent and confidentiality forms can be found in the Appendices of this Volume (pages 15-16). The teacher background questionnaire is contained in Volume II of this submittal (page 15). It includes questions about teacher’s gender, years of teaching experience, and school location among other relevant information.



The questionnaire about the reviewed items (found on pages 16-17 of Volume II) asks the teachers whether they recognized any of the items they reviewed. The purpose here is to assess the degree to which the item review process was truly blind. The same questionnaire also asks the teachers, given their ratings, to list the most problematic items they identified. This information will be used to select the problematic NAEP items, as identified by the teachers, to be discussed further in the group interview.

Once the most problematic items to be discussed are determined, the interviewer will ask the teachers two main questions: (1) what are the specific problems identified in the items, and (2) how would one modify the items to remove these problems.

### **Analysis Information**

ASPIRA, Inc. of Puerto Rico (<http://www.aspirapr.org/>) will facilitate the field activity, including recruitment of interviewers and data collection. Following the field activity, ASPIRA will enter data obtained from the Item Review Tool and teacher interviews in Excel spreadsheets. Data from teacher interviews will be translated into English by ASPIRA.

The quantitative data obtained from the Item Review Tool (ratings on a Likert scale) will be used to compare the mean ratings of NAEP and local items. As discussed above, two groups of NAEP items are used in this study: one where the observed performance matches the model-based expected performance (items with good statistical fit) and another group of items where the observed performance is lower than the model-based expected performance (items with bad statistical fit). The means ratings of NAEP items with good fit, those with bad fit, the accessible NAEP items (in the case of grade four), and local items will be computed and compared. The analyses will reveal on what aspect(s) the items from different sources are rated unfavorably by the teachers.

The information gathered from the teacher group interviews will be used to supplement the findings from the quantitative analyses mentioned above. We will feature example items (among the released NAEP items) in our study report to illustrate the kinds of

problems the teachers identified in NAEP items and the revisions or modifications suggested. NCES will share the final report and future related steps regarding this study with OMB.

### **Translation Information**

The following translation activities will occur as part of the study:

1. From English into Spanish, done at Second Language Testing, Inc.
  - consent and confidentiality forms,
  - Item Review Tool,
  - teacher background questionnaire,
  - questionnaire about reviewed mathematics items, and
  - data collection script.
2. From English into Spanish, done at ETS
  - accessible NAEP items.
3. From Spanish into English, done at ASPIRA
  - data collected in teacher interviews.

### **4) Item Rating and Teacher Interview Process**

ASPIRA, Inc. of Puerto Rico (<http://www.aspirapr.org/>) will facilitate the field activity, including recruitment of interviewers. ASPIRA is a well-respected national nonprofit organization supporting educational programs in Puerto Rico. It has the requisite infrastructure within the local educational community and the experience necessary to facilitate communication at the local level that will lend efficiency to the overall process.

Data collection will be conducted in about 7 sessions. Each session will include 5 or 6 teachers and at last 3 to 4 hours. In each session there will be two interviewers from ASPIRA. Interviewers will have at least an M.A. in Education, Social Sciences, or related area. In addition, there will be an American Institutes for Research (AIR) staff member in each session monitoring the quality of the data collection process. The script that the interviewers will follow is embedded in Volume II of this submittal (pages 4-9).

The data collection will be implemented following these steps:

1. The interviewers introduce themselves and the purpose of the study. Consent and confidentiality forms are signed and collected.
2. The interviewers give each teacher a booklet that includes mathematics items they will rate later. The teachers are asked to take about 30 minutes to get familiar with the mathematics items. They will be asked to read over the items and imagine how their students would approach these questions and answer them.
3. The interviewers distribute the Item Review Tool with the description (page 10 of Volume II of this submittal). The purpose is to make sure that all the teachers have the same understanding of each statement in the Item Review Tool.
4. The interviewers distribute the booklets that contain the math items to be reviewed and the Item Review Tool (page 14 of Volume II). Each booklet has a unique Booklet ID and also contains a Teacher Background Questionnaire (pages 15 of Volume II), which at the end asks basic information about the teachers, such as number of years of teaching experience and highest degree attained.

5. When all teachers are finished with their ratings, the interviewers hand out a short questionnaire (pages 16-17 of Volume II) that asks the teachers if they recognized any of the items they reviewed and to indicate the ID of such items and the source that they thought the item was coming from (local assessment, NAEP etc.). This information is recorded to assess the degree to which the rating process was truly blind. Next, the interviewers ask the teachers to identify four or more mathematics items that they found most problematic. The teachers record the IDs of those items on the same questionnaire and take a 10-minute break. During the break, the interviewers identify which of these items are NAEP items. After the break, the interviewers distribute a copy of the NAEP items identified as most problematic by the group. The interviewers ask the teachers to comment and discuss issues they found in these items. They are also asked to suggest modifications to these items if applicable. The interviewers record the summary of the discussion. At the end, the interviewers thank the teachers for their participation and distribute gift cards.

## **5) Consultations Outside the Agency**

The Puerto Rico Department of Education will recruit teachers to participate in the study. They will also provide ‘local’ items to be reviewed as part of the study.

Educational Testing Service (ETS), an NCES contractor, will provide translation from English to Spanish for the accessible cognitive items used in the study.

ASPIRA of Puerto Rico ([described](#) in Section 4) will facilitate the field activity, including recruitment of interviewers. ASPIRA will also deliver the data collected and translate the data captured in interviews. The executive director of ASPIRA is Adalexis Ríos.

Second Language Testing, Inc. (SLTI) is an independent agency that will provide translation services outlined on page 9 of this document. In addition, SLTI will convey an Expert Panel comprised of bilingual mathematics specialists to select local items to be

included in the study (as discussed on page 6 of this document). The president of SLTI is Charles Stansfield.

## **6) Assurance of Confidentiality**

Participation is voluntary. Written consent will be obtained from participating teachers before interviews are conducted. No personally identifiable information will be gathered from either schools or teachers. (See [Appendix A](#) for consent form.) Test security will be assured at the administrator, interviewer, and teacher levels. The interviewers and teacher participants must sign a confidentiality and test security agreement. (See [Appendix B](#) for the confidentiality and test security agreement.)

All participants will be provided with the following confidentiality pledge: The information you provide will be used for statistical purposes only. In accordance with the Confidential Information Protection provisions of Title V, Subtitle A, Public Law 107–347 and other applicable Federal laws, your responses will be kept confidential and will not be disclosed in identifiable form to anyone other than employees or agents. By law, every NCES employee as well as every agent, such as contractors and coordinators, has taken an oath and is subject to a jail term of up to 5 years, a fine of up to \$250,000, or both if he or she willfully discloses ANY identifiable information about you.

## **7) Justification for Sensitive Questions**

Throughout the interview protocol development process, effort has been made to avoid asking for information that might be considered sensitive or offensive. Reviewers have identified and eliminated potential bias in questions.

In addition, the cognitive mathematics item development process included sensitivity reviews before use in previously administered assessments.

## 8) Estimate of Hour Burden

Each interview is expected to take 3-4 hours. The estimated respondent burden follows:

Respondent	Hours per respondent	Number of respondents	Total Hours
Grades 4 teachers	4	22	88
Grades 8 teachers	4	17	68
Total			156

## 9) Estimate of Costs for Recruiting and Paying Respondents

The Puerto Rico Department of Education has agreed to recruit respondents.

Participating teachers will receive payments of \$120 in compensation for their time and effort. The study will take place outside the regular academic year. The monetary incentive is aimed at ensuring participation and motivation on behalf of the teachers.

Each teacher will spend about four hours in the study and we reasoned that paying \$30 per hour for their time and effort is reasonable.

## 10) Cost to Federal Government

For conducting the data collection, translation, and related travel, the costs are approximately \$238,555. The following table provides the overall project cost estimates:

<i>Review of 2007 NAEP Mathematics Items Used in Puerto Rico</i>	
1. Staff costs	104,331
2. Subcontractor costs	102,390
3. Other project materials (including cost of teacher remuneration)	31,834
<b>Total Cost of Task</b>	<b>238,555</b>

## 11) Schedule

<b>Activity</b>		<b>Dates</b>
Preparing data collection tools		
Item Review Tool		
	Prepare draft tool	March 16, 2009
	Submit draft for review by Expert Panel	March 16, 2009
	Feedback from Expert Panel due	March 24, 2009
	Revisions based on feedback	March 24-April 10, 2009
	Translate revised tool	May 20-May 27, 2009
Data collection script and interview protocols		
	Prepare draft data collection script and interview	March 16, 2009
	Submit draft for review by Expert Panel	March 16, 2009
	Feedback from Expert Panel due	March 24, 2009
	Make final revisions	April 10, 2009
	Translate data collection script and interview protocols	May 20-May 27, 2009
Item selection		
	Select NAEP items	February 27, 2009
	Translate accessible NAEP items	May 20-May 27, 2009
	Request local items from PRDE	March 23, 2009
	Puerto Rico Dept. of Education delivers items	March 24- May 21, 2009
	Expert panel chooses local items	May 29, 2009
OMB		
	Prepare and submit OMB package	March 2- April 27, 2009
Data collection		
	Recruit teachers and interviewers	June 1 - June 9, 2009
	Data collection	June 12-26, 2009
Data preparation		
	Build database for the item ratings	June 26-30, 2009
	Translate teacher interviews to English	June 26- July 1, 2009
Data analysis and report		
	Analysis of teacher ratings and interview data	July 1 -24, 2009
	Draft study report to NCES	July 31, 2009
	NCES feedback on draft report	August 10, 2009
	Final study report	August 28, 2009

Volume II of this submission includes the Item Review Tool and the data collection script.

## Appendix A– Consent Form

<DATE>

Dear Participant,

Thank you for agreeing to help refine mathematics questions. The purpose of this study is to improve grade 4 and 8 mathematics tests questions used to assess Puerto Rican students by having you evaluate the quality of questions and provide suggestions for improvement.

The study will last approximately 3 to 4 hours. During that time, you will be asked to review a small group of test questions. After your review, you will be asked a few questions and be given an opportunity to make comments.

All information obtained will be kept confidential and will only be used for the purposes of this study. We will not use your name or school name and will not attribute any quotes specifically to you.

The information you provide will be used for statistical purposes only. In accordance with the Confidential Information Protection provisions of Title V, Subtitle A, Public Law 107–347 and other applicable Federal laws, your responses will be kept confidential and will not be disclosed in identifiable form to anyone other than employees or agents. By law, every NCES employee as well as every agent, such as contractors and coordinators, has taken an oath and is subject to a jail term of up to 5 years, a fine of up to \$250,000, or both if he or she willfully discloses ANY identifiable information about you.

The Department of Education will not be responsible for any claim that could arise as a result of the study activities and of the information requested and provided by such. The Department of Education is not responsible of any damage or claim as a result of the procedural process or of the result of the investigation and such study is an independent investigation not sponsored by the Department of Education. The Department of Education is not necessarily in agreement with the results of the investigation.

We hope that you will give your consent to participate in the interview by signing this form. Without your signed consent, you will not be able to participate in the study.

Sincerely,

<ASPIRA NAME HERE>

Print

Name: \_\_\_\_\_

Signature: \_\_\_\_\_

Date: \_\_\_\_\_



*Appendix B – Test Item Confidentiality Agreement*

**CONFIDENTIALITY AGREEMENT:  
Test Materials Security Requirements**

Under this agreement, you will have access to secure and confidential test materials. These materials are confidential and may not be shared or discussed with any person who has not signed this confidentiality agreement.

*These materials may not be copied, published, announced, or in any other way made public.*

By signing this agreement, you acknowledge that the test materials constitute proprietary and confidential materials. You further understand that any disclosure, unauthorized use, or reproduction of these materials would damage the confidentiality of the assessment, is illegal, and can result in a felony charge. You agree to keep the test materials and data secure and confidential.

ACCEPTED AND AGREED TO:

Signature \_\_\_\_\_ Date \_\_\_\_\_

Full name (please print) \_\_\_\_\_

Title \_\_\_\_\_

Address \_\_\_\_\_  
\_\_\_\_\_

Witness \_\_\_\_\_

## Appendix C- Thank You Letter to Teachers

<DATE>

<NAME OF PARTICIPANT>  
<ADDRESS LINE 1>  
<ADDRESS LINE 2>

Dear <NAME>:

I would like to thank you for your participation in a study that will improve the quality of educational assessments in Puerto Rico. Because of the assistance of educators like you, researchers and test developers will be able to develop better assessments in future.

Our interviewers found the sessions extremely useful and productive. We thank you again for your time, effort, and valuable insights.

Best wishes for continued success to you and your students.

Sincerely,

Adalexis Ríos  
ASPIRA