

MEMORANDUM

OMB # 1850-0803 v.71

DATE: September 26, 2012

TO: Shelly Martinez
Office of Information and Regulatory Affairs, Office of Management and Budget

FROM: Patricia Etienne
National Center for Education Statistics

THROUGH: Kashka Kubzdela
National Center for Education Statistics

SUBJECT: Response to September 5 (NCES response in blue), September 14 (NCES response in green), September 24 (NCES response in pink), and September 26 (NCES response in purple) OMB passback on 2013 NAEP Read-Aloud Accommodations Study

Changes to the documents have been in response to these questions. In addition, in light of the passback, we have slightly modified the approach regarding gathering information about the composition of the classes, as shown in the revised documents.

Changes to the documents have also been in response to the follow-up questions. In addition, we have revised the race/ethnicity questions on the teacher questionnaire.

1. The memorandum mentions gaining information about feasibility as one of the purposes of this study. How will feasibility be assessed specifically?

An important part of the study is to assess any difficulties that were encountered in conducting the sessions. In addition, we want to assess feasibility in terms of the burden to students and schools. Feedback obtained from session debriefings with field staff will help inform what issues, difficulties, or concerns were encountered in these administrations to inform the feasibility of offering the read-aloud accommodation in reading.

We were hoping for some metrics here.

The main purpose of the study is to explore the use of the read-aloud accommodation in reading and evaluate the results to determine if there are any differences between the groups. In addition, we are interested in gaining information regarding the logistics and procedures used in implementing the accommodation. While we had originally described this as examining the “feasibility” of the study, we can see how the use of this word could be misleading. Therefore, to avoid confusion, we have removed this as one of the stated purposes of the study.

2. It appears that the research design (see figure 1) places students into 3 mutually exclusive categories. There is likely a fair amount overlap between the SD and ELL students in the schools participating in the study. How will this population be handled in terms of sampling and analysis?

Students with disabilities who are also English language learners will be included in the study. However, the sample will likely not be large enough to analyze this group separately. Rather, descriptive analyses will be done on this group. If a large enough sample is obtained, a fourth student category will be added to the analysis plan. We revised the Figure 1 categories and added a footnote to clarify this point.

(9/14) This answer somewhat clarifies analysis plans but does not speak directly to how this group will be sampled or included in analysis. Will the group be eligible to be included in either the ELL or the SD strata and be treated for analysis purposes as members of both groups (assuming insufficient sample to analyze as a stand-alone group)?

Given that the group of students who are both SD and ELL are often one of lowest performing subgroups, in an effort not to bias the results, we do not want to include them in only the SD group nor in only the ELL group. In addition, in order to perform certain analyses, independence of group membership is required and, as such, these students could not be in both groups. Therefore, students who are both SD and ELL would be excluded from statistical analysis of the different groups. However, as described earlier, descriptive analysis would be performed. And, if the sample proves to be large enough, we will create a fourth category for analysis purposes.

(9/24) We really believe that the dual EL/SD population is important to include in a study even if the study ultimately suggests that they can't be included in a production NAEP environment. Approximately 1/4 of CA students are ELs, and we would expect that the prevalence of disability in that population might be about the same as for the non-EL population. The number of EL students in the US is projected by some to double in the next 20 years, so this seems like a larger issue for NAEP going forward. Please consider how to include this group.

In order to obtain enough students who are both SD and ELL for meaningful analysis and interpretations of the data, the overall sample size would need to be nearly tripled. This conclusion is based on recent NAEP data. Specifically, in 2011, 3.73% and 3.59% of students in the NAEP sample from CA (for grades 4 and 8, respectively) were both SD and ELL. Based on the power analysis described in the response to Question #4, 130 students are required per group, per grade. Therefore, the overall sample would need to be 3485 and 3621 (for grades 4 and 8, respectively), which is nearly triple the proposed sample size of 1200 per grade. Increasing the sample size to this level would significantly increase the costs, as well as push the number of schools beyond that which we could expect to assess in the required timeframe. Given these considerations, we believe that the most efficient and cost effective method to obtain information on this important group is to perform descriptive analysis of the students who are both SD and ELL.

(9/26) Does this proposal involve including these students in the analysis of the single category groups ("just" EL and "just" SD) or excluding them all together? It was excluding them all together that had us concerned.

Upon reflection of the points raised by OMB, we suggest performing the analyses three different ways: 1) including the SD and ELL students in the ELL group, 2) including the SD and ELL students

in the SD group, and 3) not including the SD and ELL students. As such, the performance of these students can be fully considered and interpreted. We appreciate the advice offered by OMB in the design of this important aspect of the study.

3. Related, we understand that it would not be feasible or even advisable to have a large enough sample to disaggregate results by each disability category. However, there is some thought that understanding differences by severity of disability would be important. Specifically, this idea was shared with OMB by NCEE as part of their ongoing national study of students with disabilities (contact is Marsha Silverberg). Please consult with Marsha to see whether this idea could be incorporated into your study.

Typically, students with severe disabilities are considered to be the 1% of students with significant cognitive disabilities that require an alternate assessment. Students with disabilities of this severity would likely be excluded from participating in this study. The number of any such students included in the study would be too small to disaggregate for analysis. Other classifications of the severity of disabilities (e.g., mild, moderate, or severe) are not used consistently and are often based upon subjective judgment. We have, however, contacted Marsha to learn more about the NCEE study, and plan on discussing with her next week when she will be available.

We will wait to hear how that conversation goes before clearing.

We spoke with Marsha and reviewed our study design and reporting goals with her. Given that most students with severe disabilities would be excluded from the assessment, she agrees that the sample size would be too small to generalize. As such, and coupled with the purpose of the study, she does not feel that changes to the design are necessary.

4. Please provide your power analysis to justify the sample size.

The description of the power analysis has been added to the last paragraph of the [Population and Sample](#) section on page 6. It is also included here:

A power analysis was conducted to determine the sample size required for the read-aloud accommodations for the three groups of students in this study: (1) ELLs, (2) SDs, and (3) non-ELLs/non-SDs. The sample size was calculated based on a minimum detectible effect size (MDES) of 0.35 standard deviation with a Type-1 error rate of 0.05 (two-tailed) and statistical power of 0.80 (80%). Based on these parameters, the sample was calculated to be 130 subjects for each of the three subgroups of students with disabilities: (1) specific learning disability, (2) speech/language impairment, and (3) intellectual disability as well as for other subgroups of students (ELLs and non-ELLs/non-SDs)¹. While obtaining this sample size for each of the three groups of students with disabilities might be challenging, focusing efforts on schools/classes with larger number of students with disabilities in the three categories may yield the required sample size. However, if these minimum numbers are not secured, we have the option of combining students in the three categories

¹ The sample size in each subgroup could be reduced to 110 under a one-tailed test condition; however, a two-tailed test is preferred.

of disabilities to obtain the sample size required for analysis. For ELLs, we will have enough power to perform analyses by some subgroups of ELLs (such as ELLs at different levels of English proficiency). For non-ELLs/non-SDs, we will have a large enough sample that will allow analyses by levels of some of background variables such as gender, socioeconomic status (SES; as indicated by participation in the National School Lunch Program), and ethnicity.

Please discuss why this is the right MDE.

The MDES used in this study is between a small (0.2 standard deviation) and medium (0.5 standard deviation) effect size² and is appropriate to detect meaningful differences between the student groups. Therefore, combined with the other, traditional, parameters of a power analysis (i.e., 0.05 error rate, two-tailed test, and 0.80 statistical power), the necessary sample size was calculated. While the MDES could be decreased, it would require increasing the sample size beyond the number of schools that we could expect to assess in the required timeframe.

5. We are concerned that the teacher is not the most reliable reporter for some of the items on the class roster, whereas this information should be available via school administrative data. Please provide specific evidence that these teachers have ready access to the requested data elements, for example NSLP status and the parents' self-identified racial and ethnic designation (rather than a teacher observed one), or a proposal for an alternative procedure for collecting these variables.

We believe that the teachers are the best source for obtaining this information. Dr. Abedi has conducted several studies in which teachers were the main source of information for their students and provided student background data and student test scores. Some of the student-level data provided by teachers (test scores, race/ethnicity, NSLP (free/reduced price lunch program)) were cross-checked with the official sources and were found to be quite consistent. Please see:

Abedi, J., Courtney, M., Leon, S. Kao, J., and Azzam, T. (2006). *English Language Learners and Math Achievement: A Study of Opportunity to Learn and Language Accommodation (CSE Report 702, 2006)*. Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing. Download at: <http://www.cse.ucla.edu/products/reports/R702.pdf>.

Abedi, J., Baily, A., & Butler, F. (2005). *The Validity of Administering Large-Scale Content Assessments to English Language Learners: An Investigation From Three Perspectives (CSE Report 663, 2005)*. Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing. Download at: <http://www.cse.ucla.edu/products/reports/r663.pdf>.

The first report cites as its methodology: **“Student Background Data:** Collected English language development (ELD) information, gender, home language, and ethnicity from school records. Collected data on student language background characteristics from students.” This report seems to make our point – the data should come from school records, not from the teacher. If you are indicating that the teachers are to consult school records, the instructions would need to be clarified (eg, that is it inappropriate to use observation of race), but we do not understand why the NAEP study team doesn't take on that burden in working with the school offices rather than burdening the teachers.

² Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates. Please find a short discussion of this topic at: <http://effectsizefaq.com/2010/05/30/what-are-some-conventions-for-interpreting-different-effect-sizes/>

Based on OMB's feedback, we have revised the study design to include having a "school coordinator" from the school complete the class rosters (for all classes within the school). As such, the school coordinator (i.e., a staff member from the school office) will now provide the information and the burden will be removed from the teachers. The teachers will only be asked to complete the teacher survey. Please note that for confidentiality and security reasons, schools will not grant direct access of the school records to the NAEP field staff. Therefore, the school coordinator will be asked to provide the information requested on the class roster. We have revised the documents to reflect this change.

6. What is the purpose of questions 3 to 6 on the teacher survey?

In the teacher questionnaire for grade 4: Question 3 asks about access to instructional materials and other resources to see if a teacher's lack of resources may have an impact on students' accommodated performance; Question 4 asks about use of read-aloud in reading to find out if students' prior experience with the practice of read-aloud in the classroom may have an impact on their accommodated performance; and Question 5 and 6 ask teachers about their teaching experience to help us determine the level of impact of other factors (other than accommodations) on students' accommodated performance.

7. NCES indicates that "Other accommodations will not be provided in the study." It also says that standard NAEP assessment procedures will be followed. If a child would normally receive a specific accommodation in the NAEP reading assessment, does this mean that accommodation will not be provided here? Doesn't that seem likely to make interpreting the results of this study difficult to apply?

We have modified the design to offer extended time to all students, which is the most commonly required accommodation. In order to keep conditions as similar as possible across the groups, we are not offering other accommodations. In addition, we are collecting information regarding the accommodations that the participating students use on their state test, so as to account for this information in analysis and study conclusions.

8. We are not comfortable with the proposed incentive payment. First, the cited example provided as justification that "The study will take place outside the regular academic year." In particular, that study was essentially a four hour focus group during summer break (i.e., travel was necessary and the teacher was asked to participate in a group activity without the opportunity to identify a time most convenient for that individual). As NCES knows, we incentivize focus group participation in a different manner than teacher surveys given these considerations. Combined with the fact that the teachers do not appear to be the best reporters of much of the information on the roster, we suggests that NCES rethink the roster collection and associated incentive.

The teachers of the students involved in this study will play an important role in the success of the study. The completion of the rosters, for which we believe the teachers are the best source of this information, as indicated in the response to Question #5, will require about 3 hours of teacher time.

However, based on the concern noted by OMB, we have revised the teacher incentive to be \$25/hour for \$75 total. This amount was previously used as the incentive for school coordinators for similar

activities – completion of forms and questionnaires (e.g., 2011 TIMSS; OMB #1850-0645) and is comparable to incentives given to teachers for focus groups and cognitive interviews.

TIMMS did not provide a teacher incentive. Our previous note indicated that the incentive structure for focus groups and cognitive labs is completely different than for surveys, where the presumption begins that there will be no incentive. For NCEE studies, we often incentivize teachers when completing a survey in the context of a larger survey, somewhere in the \$20 to \$30 range. We are inclined open to something in that range, and again strongly suggest determining how to shift the burden for the non-survey pieces to the school office.

Related, we note that the burden estimate for the teacher survey is blank on the front of the questionnaire. It also is not broken out in the memo. Please provide.

Based on the creation of the school coordinator role (as described in the response to Question 5), we have modified the incentive plan. Teachers will now only be asked to respond to a 10 minute questionnaire (as noted in the burden chart in Table 1 and on the cover of the questionnaire). Therefore, teachers will not be provided a monetary incentive/token of appreciation for their time. The majority of the burden will now fall to staff from the school office (namely, the school coordinator). As a token of appreciation to the school for participating in the study, we are proposing to provide the school with \$200.

9. Why is the parent letter so specific (paragraph one)? This is unusual and may lead to lower salience and therefore participation of parents of non-disabled students. Again, this is a concern that NCEE has had to address in its study.

We have revised the parent letter by removing the detailed descriptive text from paragraph one.