Memorandum	**United States Department of Education**
	Institute of Education Sciences
	National Center for Education Statistics

DATE: June 23, 2017; revised July 6, 2017

TO: Robert Sivinski and E. Ann Carson, OMB

THROUGH: Kashka Kubzdela, OMB Liaison, NCES

FROM: Linda Hamilton, NCES

SUBJECT: National Assessment of Educational Progress (NAEP) 2018 and 2019 (OMB# 1850-0928 v.5) Responses to OMB Passback

This memorandum provides response to the OMB passback received on June 19th, 2017, requesting additional information regarding the NAEP 2018 and 2019 (OMB# 1850-0928 v.5) submission.

OMB Comment 1a:

Please provide more clarification/detail on the following two special studies in part A:

a. The Digitally Based Assessment (DBA) Bridge studies (page 13) needs a lot more detail. If the goal is to replicate a previously fielded instrument, why isn't that instrument discussed or included? What defines "randomly equivalent groups of students"? Age*sex*race*grade level? GPA?

Associated text (A.1.d, Digitally Based Assessment (DBA) Bridge Studies):

The term "bridge study" is used to describe a study conducted so that the interpretation of the assessment results remains constant over time. A bridge study involves administering two assessments: one that replicates the assessment given in the previous assessment year using the same questions and administration procedures (a bridge assessment), and one that represents the new design (a modified assessment). Comparing the results from the two assessments, given in the same year to randomly equivalent groups of students, provides an indication of whether there are any significant changes in results caused by the changes in the assessment. A statistical linking procedure can then be employed, if necessary, to adjust the scores so they are on the same metric, allowing trends to be reported. The following DBA bridge studies are planned:

- In 2018, DBA bridge studies are planned in U.S. history, civics, and geography;
- In 2019, DBA bridge studies are planned in science, reading, and mathematics in addition to the operational DBAs to confirm the findings from the 2015 initial bridge studies.

As described in A.1.c.5, NAEP is using a multi-step process designed to protect trend reporting to transition from PBA to DBA.

The PBAs will be administered to a representative sample, enabling the examination of the relationship between PBA and DBA performance. DBA bridge studies will be conducted for U.S. history, civics, and geography at grade 8 in 2018. Reading and mathematics at grade 12 and science at grades 4, 8, and 12 will be conducted in 2019. Given that the operational assessments of these subjects are at the national level, the DBA bridge study will be administered to a nationally representative sample for each of the 6 subjects. In 2018, the total sample size across the three subjects is 26,000. In 2019, the total sample size across the grades and subjects is 75,000. The size of the national sample is primarily driven by the need for sufficient numbers of student responses at item level to support IRT calibration.

NCES Response:

The school sampling procedure follows a specific sampling algorithm to generate a representative sample of schools, as described in section B.1.a of Part B.1.a and in Appendix C of this submission. The student sampling process ensures that all students within a selected school and grade have an equal probability of being selected, as described in section VIII of Appendix C. We revised the referenced text in Part A to the following:

"To support the transition from a paper based assessment (PBA) to a DBA, NAEP is conducting bridge studies that will compare student performance on paper versus digital platforms. The term "bridge study" is used to describe a study conducted so that the interpretation of the assessment results remains constant over time. A bridge study involves administering two assessments: one that replicates the assessment given in the previous assessment year using the

same questions and administration procedures (the bridge assessment; in NAEP 2018 and 2019 these are PBA), and one that represents the new design (the modified assessment; in NAEP 2018 and 2019 these are DBA). For example, in 2018 the same U.S. history content will be given to two groups of students, with one group taking a paper version and one group taking a digital version. Comparing the results from the two assessments, given in the same year to randomly equivalent groups of students (two distinct samples of students, each drawn from the same student population, and each using probability sampling methods that ensure that the sample is representative of that population, as described on pages 7-8 of the Appendix C in this submission), provides an indication of whether there are any significant changes in results caused by the changes in the mode of assessment. A statistical linking procedure can then be employed, if necessary, to adjust the scores so they are on the same metric, allowing trends to be reported. The following bridge studies are planned:

- In 2018, PBA-DBA NAEP bridge studies are planned in U.S. history, civics, and geography;
- In 2019, PBA-DBA NAEP bridge studies are planned in science, reading, and mathematics in addition to the operational DBAs to confirm the findings from the 2015 initial bridge studies.

As described in section A.1.c.5 of this document, NAEP is using a multi-step process designed to protect trend reporting to transition from PBA to DBA. The survey questionnaire and assessment content are the same for PBA and DBA. The survey questionnaire items are presented in the Appendix F library. As noted on page 2 of Appendix F, the item-level directions may differ between PBA and DBA versions. The final versions of the NAEP 2018 and 2019 questionnaires, which, as mentioned in section A.1.a of this document, will be submitted to OMB for approval as a non-substantive change request by October 2017 for NAEP 2018 and by October 2018 for NAEP 2019, as a future Appendix I, will include both the final PBA and DBA versions. While the PBA and DBA content is the same, the assessment items from PBA were converted to the DBA platform in order to support digital delivery. This conversion included adjustments such as adapting the visual layout, modifying answer selection mechanisms (e.g., selecting objects rather than circling them), and using digital tools to facilitate responding to the items (e.g., digital equation editor). The assessment items are not included in this request because they are not subject to the Paperwork Reduction Act.

The PBAs will be administered to a representative sample, enabling the examination of the relationship between PBA and DBA performance. DBA bridge studies will be conducted for U.S. history, civics, and geography at grade 8 in 2018. Reading and mathematics at grade 12 and science at grades 4, 8, and 12 will be conducted in 2019. Given that the operational assessments of these subjects are at the national level, the DBA bridge study will be administered to a nationally representative sample for each of the 6 subjects. In 2018, the total sample size across the three subjects is 26,000. In 2019, the total sample size across the grades and subjects is 75,000. The size of the national sample is primarily driven by the need for sufficient numbers of student responses at item level to support IRT calibration."

OMB Comment 1b:

Please provide more clarification/detail on the following two special studies in part A:

b. Oral Reading Fluency (page 14) – how large is the sample for this? Are all drawn from language and linguistic minority groups? What about students with disabilities that might show lower reading ability – will they also be included? (The sample size question is nominally answered in the table on page 28, and it looks like there will be no SD/ELL students, but this should have been spelled out in the text).

Associated text (A.1.d, Oral Reading Fluency (ORF)):

In this study, a sample of fourth-grade students will take the NAEP reading assessment followed by the ORF module. The ORF module will consist of a set of materials that students will read aloud in English after completing the NAEP reading assessment.

NCES Response:

We revised the text to the following:

"In this study, a sample of 2,000 fourth-grade students will take the NAEP reading assessment followed by the ORF module. The sample will be selected to be representative of all grade 4 public school students in the U.S. (including SD/ELL students). At the school level, schools with a relatively high proportion of students eligible for the National School Lunch Program (NSLP) will be oversampled. The ORF module will consist of a set of materials that students will read aloud in English after completing the NAEP reading assessment."

OMB Comment 2:

> Regarding the Automated scoring, section A.3, page 17 – what are the sample sizes for these assessments of automated versus human scoring?

Associated text (A.3):

One possible study involves using two different automated scoring engines and comparing the scores to those previously given by human scorers. This study would be conducted on items from the 2011 writing assessment, as well as some items from the 2015 DBA pilot. For each constructed response item, approximately two-thirds of responses would be used to develop the automated scoring model (the Training/Evaluation set) and the other third of responses would be used to test and validate the automated scoring model (the Test/Validation set).

NCES Response:

We added the following text to the end of the current paragraph:

"The sample will be selected from an estimated 2,000-2,500 responses to 22 different grade 8 prompts, plus 2,000-2,500 responses to 22 different grade 12 prompts. Total estimated responses to be scored using the automated system will be between 88,000 and 110,000. No new data collection or human scoring will be required."

OMB Comment 3:

> Part A, page 9, paragraph directly above the header "Development of Digitally Based Assessments" – NCES needs to specify exactly what about how students answer the question (changing answer, showing work used, etc.) will be captured and how these data will be used. I realize that agencies already capture paradata from digital surveys to look at improving their survey instruments (the time it takes to answer a question, whether an answer got switched etc.), but from what I've seen, these agencies always specify in OMB packages the data elements captured and assessed. The NAEP description on page 9 is very vague: "As such, NAEP will potentially uncover more information about which skills successful students use and where the skills of less successful students break down." This does not sound like the data will be used to improve survey questions, but rather will be used to assess the students in a way other than the final answers to questions. If that is an intended use of the data, it needs to be spelled out explicitly for OMB (along with exactly what will be collected).

Associated text (A.1.c.5):

These new item types and testing technologies may allow NAEP to capture information about students' problem solving processes and the strategies they use to answer items. For example, while PBAs would only yield the final responses in the test booklet, DBAs capture information about student use of the tools, whether students change their answer, etc. As such, NAEP will potentially uncover more information about which skills successful students use and where the skills of less successful students break down.

NCES Response:

We revised the text to the following:

"The digitally based assessment technology allows NAEP to capture information about what students do while attempting to answer questions. For example, while PBAs would only yield the final responses in the test booklet, DBAs capture actions students perform while interacting with the assessment tasks, as well the time at which students take these actions. These student interactions with the assessment interface are not used to assess students' knowledge and skills, but rather this information might be used to provide context for student performance. For example, it might be possible to detect a correlation between students who utilize the text highlighting tool in a reading passage and students who answer questions on the reading passage correctly. As such, NAEP will potentially uncover more information about which actions students use when they successfully (or unsuccessfully) answer specific questions on the assessment.

NAEP will capture the following actions in the DBA, although not all actions will be captured for all assessments:
- Student navigation (e.g. clicking back/next; clicking on the progress navigator; clicking to leave a section);
- Student use of tools (e.g. zooming; using text to speech; turning on scratchwork mode; using the highlighter tool; opening the calculator; using the equation editor; clicking the change language button);
- Student responses (e.g. clicking a choice; eliminating a choice; clearing an answer; keystroke log of student typed text);
- Writing interface (e.g. Expanding the response field; collapsing the prompt; using keyboard commands such as

CTRL+C to copy text; clicking buttons on the toolbar such as using the bold or undo button);

- Other student events (e.g. vertical and horizontal scrolling; media interaction such as playing an audio stimulus);
- Tutorial events (records student interactions with the tutorial such as correctly following the instructions of the tutorial; incorrectly following the instructions of the tutorial; or not interacting with the tutorial when prompted); and
- Scratchwork canvas (the system saves an image of the final scratchwork canvas for each item where the scratchwork tool is available)."

OMB Comment 4:

In the 2018 sampling plan (Appendix C, page 78), NCES says that it wants to oversample AIAN students. How will NCES classify AIAN students if they report additional races and/or Hispanic origin? AIAN is the most racially and ethnically admixed group….is NCES looking to oversample only non-Hispanic AIAN (single race alone)?

Associated text (Appendix C, Page 5):

In addition, we will implement three different kinds of oversampling of public schools. First, in order to increase the likelihood that the results for American Indian/Alaskan Native (AIAN) students can be reported for the operational samples, we will oversample high-AIAN public schools for Social Studies and TEL at grade 8. That is, a public school with over 5 percent AIAN enrollment will be given four times the chance of selection of a public school of the same size with a lower AIAN percentage. Recent research into oversampling schemes that could benefit AIAN students indicates that this approach should be effective in increasing the sample sizes of AIAN students, without inducing undesirably large design effects on the sample, either overall or for particular subgroups.

NCES Response:

We added endnote to Appendix C with the following text:

"As states, districts, and schools are only required to report race/ethnicity data at the 7-category level (and specifically because this is how the data are recorded on the Common Core of Data, used as the sampling frame), the data used to oversample high AIAN percentage schools are the percent of students who are non-Hispanic AIAN students with no other race category. This is also the basis for the primary reporting results of AIAN students for NAEP. Note that the oversampling is at the school level, so that students who report multiple races, including AIAN, who are in schools with a high percentage of AIAN students, will also be oversampled. However, as noted, current NAEP primary reporting practices will not report such students as AIAN."

Additional Revision by NCES:

To address OMB's verbal questions about respondent PII, given the use of the CIPSEA confidentiality pledge in NAEP, in the Supporting Statement Part A, section A.10 (p. 24), we added the following clarifying language about how PII is handled by NAEP Contractors:

"4.   At no point in time does any individual contactor have access to both the student name and student assessment and questionnaire responses. MDPS has access to both the student name and student assessment and questionnaire responses, but never at the same time. MDPS uses student PII to print Student Login Cards months in advance of the NAEP assessment window, and destroys the student PII file before MDPS begins to receive student assessment and questionnaire responses for scoring during the assessment window. SDC never has access to student responses, and no other contractor has access to Student PII. Going forward, MDPS will certify in writing to NCES, before MDPS begins to receive student assessment and questionnaire responses, that all files with student PII data have been securely destroyed."