Memorandum

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

DATE: July 29, 2016

TO: Robert Sivinski, OMB

THROUGH: Kashka Kubzdela, OMB Liaison, NCES

FROM: Pat Etienne, NCES

SUBJECT: NAEP Assessments for 2017-2019 (OMB# 1850-NEW v.1 (previous OMB# 1850-0790 v.43)) Responses to OMB Passback

This memorandum provides response to the OMB passback requesting additional information regarding NAEP Assessments for 2017-2019 3-Year Clearance. The follow-up passback comments are noted as **OMB Comment #a**.

OMB Comment 1: The supporting statements don't contain the required detail for the pilot and special studies. If details on specific research goals, sample size, sample design, items, survey administration procedures, and use of resulting data aren't available at this time, they'll need to be submitted in revisions with public comment.

Associated text (A.1): This submission requests OMB's approval for the following NAEP 2017-2019 assessments: operational, pilot, and special studies.

NCES Response: For pilot assessments, added the following text in Section A.1.c.6: "Pilot testing (also known as field testing) of cognitive and non-cognitive items is carried out in all subject areas. Pilot assessments are conducted in conjunction with operational assessments and use the same procedures as the operational assessments." In addition, added the following text in Section B.1.a: "samples sizes of approximately 3,000-12,000 for pilot assessments, depending on the size of the item pool[1]." For special studies, added clarification to the individual special study descriptions in Section A.1.d. (See OMB comment #9-14 for additional text related to this comment.)

OMB Comment 2: Please cite the relevant statute.

Associated text (A.1.c.1): NCES is responsible for developing the cognitive items and for selecting the final set of items.

NCES Response: Added the following text to Section A.1a: **"**NAEP is conducted by the National Center for Education Statistics (NCES) in the Institute of Education Sciences of the U.S. Department of Education. As such, NCES is responsible for designing and executing the assessment, including designing the assessment procedures and methodology, developing the assessment content, selecting the final assessment content, sampling schools and students, recruiting schools, administering the assessment, scoring student responses, determining the analysis procedures, analyzing the data, and reporting the results[2]."

OMB Comment 3: Please clarify 'policy-related questions'. Many of the items on the student questionnaires are used to inform policy.

Associated text (A.1.c.3): Policy-related questions are reserved for the teacher and school questionnaires.

NCES Response: The associated text sentence was removed.

OMB Comment 4: Why not? Without discussion this seems to conflict with the previous sentence.

Associated text (A.1.c.3): While completion of the questionnaire is voluntary, NAEP encourages teachers' participation since their responses improve the accuracy and completeness of the NAEP assessment. Teacher

[1] NAEP IRT scaling is conducted for most pilot assessments, requiring a minimum of 1,500-2,000 students per item in order to estimate stable item parameters. Therefore, pilot assessments with larger item pools have larger samples.
[2] The role of NCES, led by the Commissioner for Education Statistics, is defined in 20 U.S. Code Section 9622 (https://www.law.cornell.edu/uscode/text/20/9622) and OMB Statistical Policy Directives No. 1 and 4 (https://www.whitehouse.gov/omb/inforeg_statpolicy).

questionnaires are typically only given to teachers at grades 4 and 8; NAEP typically does not collect teacher information for grade 12.

NCES Response: Revised text to the following: **"**Teacher questionnaires are typically only given to teachers at grades 4 and 8; NAEP typically does not collect teacher information for grade 12. By grade 12, there is such variation in student course taking experiences that students cannot be matched to individual teachers for each tested subject. For example, a student may not be taking a mathematics class in grade 12, so he or she cannot be matched to a teacher. Conversely, a student could be taking two reading classes at grade 12 and have multiple teachers related to reading. Only an economics teacher questionnaire has been developed and administrated at grade 12. However, this data was not released (with either the 2006 or the 2012 results) due to a student-teacher match rate below statistical standards[3]."

OMB Comment 5: Softening this claim will make it more defensible.

Associated text (A.1.c.3): When developing the questionnaires, NAEP ensures through the pretesting process that the questions are not intrusive or sensitive, that they are grounded in educational research, and that the answers can provide information relevant to the subject being assessed.

NCES Response: Changed the sentence to now read: "When developing the questionnaires, NAEP uses a pretesting process so that the final questions are minimally intrusive or sensitive, are grounded in educational research, and the answers can provide information relevant to the subject being assessed."

OMB Comment 6: What portion? What's the research design? How will the results be used to ensure continuity? There were some slides presented to OMB on 5/23 that should be added here.

Associated text (A.1.c.5): Paper-based versions of the mathematics and reading assessments will be administered again in 2017 to a portion of the student sample within each state; the remainder will take the digital version.

NCES Response: Added the following text: "Inclusion of the paper-based component is designed to support a bridge study that both measures and potentially adjusts the metric in which results are reported for differences due to the change in mode. Details of the bridge study are presented in Section A.1.d." (See OMB comment #14 for additional text related to this comment.)

OMB Comment 7: Bear in mind that the entire package should be updated when revisions are submitted.

Associated text (A.1.d): The Governing Board determines NAEP policy and the assessment schedule, and future Governing Board decisions may result in changes to the plans represented here. Any changes will be presented in subsequent clearance packages or revisions to the current package.

NCES Response: This is understood and all revisions will be submitted as an entire package.

OMB Comment 8: These studies are lacking key descriptions of sample size and power, sampling methods and processes, recruitment, collection methods, analysis and estimation, and proposed uses of the results. These studies need the full level of detailed support that's required of any proposed collection

Associated text (A.1.d): The planned special studies are conducted in accordance with the assessment development, research, or additional reporting needs of NAEP. Additional details on each of the special studies are provided below.

NCES Response: The planned special studies have been expanded to provide additional detail. Please see responses to OMB Comments #9-14.

In addition, the following text has been added before the beginning of the special study descriptions: "With the exception of the High School Transcript Study and the National Indian Education Study, all data collection procedures are the same as those for operational and pilot NAEP assessments (as described in Part B.2). Additional details for the High School Transcript Study and the National Indian Education Study will be provided in 2018 (prior to these studies being conducted in 2019). At that point NCES will (a) publish on Regulations.gov an amendment to this package with all details for these special studies, (b) announce a 30-day public comment period on these details in the Federal Register, and (c) submit the amendment to OMB for review."

[3] The grade 12 economics teacher match rate was 56% in 2012. For comparison, the 2015 teacher match rates for grades 4 and 8 were approximately 94% and 86%, respectively.

Finally, Appendices D-20, D-21, and D-22 have been added to detail specifics for the writing comparability study, and Part B.2.a has been updated to introduce these new Appendices.

OMB Comment 9: Need a lot more detail.

Associated text (A.1.d): High School Transcript Study (HSTS).

NCES Response: Added the following text: "The 2019 HSTS study, will be conducted at approximately 800 schools, and will utilize similar methods as those used in previous years. Information related to the sampling, design, data collection methods, and analyses, as well as results from previous studies, can be found at http://nces.ed.gov/nationsreportcard/hsts/.

OMB Comment 9a: This language is unacceptably vague. By following the link provided I was able to locate this technical report: http://nces.ed.gov/nationsreportcard/pdf/studies/2011465.pdf

The report provides the level of detail expected in the supporting statements. Please copy and paste the relevant information from the most recent technical report directly into the supporting statements, making updates if necessary.

Associated text (A.1.d): The 2019 HSTS study, will be conducted at approximately 800 schools, and will utilize similar methods as those used in previous years.

NCES Response: While methods similar to previous years will be utilized, there will be some differences. As such, NCES plans to submit an amendment to this package identifying the specific methods for this study. This is noted on page 13 of Part A.1.d Overview of 2017-2019 NAEP Assessments, which reads: "Additional details for the High School Transcript Study and the National Indian Education Study will be provided in 2018 (prior to these studies being conducted in 2019). At that point NCES will (a) publish on Regulations.gov an amendment to this package with all details for these special studies, (b) announce a 30-day public comment period on these details in the Federal Register, and (c) submit the amendment to OMB for review."

We have also revised the text of the HSTS section in Part A to read: "The 2019 HSTS study, will be conducted at approximately 800 schools, and will utilize similar methods as those used in previous years. As noted above, an amendment to this package describing the study details will be submitted for approval prior to conducting the study."

OMB Comment 10: Need more detail.

Associated text (A.1.d): National Indian Education Survey (NIES).

NCES Response: Added the following text: "The 2019 NIES study will use similar methods as those used in previous years. Approximately 8,000 fourth-grade and 6,500 eighth-grade students will participate in the 2019 NIES study. Information related to the sampling, design, data collection methods, and analyses, as well as results from previous studies can be found at http://nces.ed.gov/nationsreportcard/nies/.

OMB Comment 10a: This language is unacceptably vague. The report provides the level of detail expected in the supporting statements. Please copy and paste the relevant information from the most recent technical report directly into the supporting statements, making updates if necessary.

Associated text (A.1.d): The 2019 NIES study will use similar methods as those used in previous years.

NCES Response: While methods similar to previous years will be utilized, there will be some differences. As such, NCES plans to submit an amendment to this package identifying the specific methods for this study. This is noted on page 13 of Part A.1.d Overview of 2017-2019 NAEP Assessments, which reads: "Additional details for the High School Transcript Study and the National Indian Education Study will be provided in 2018 (prior to these studies being conducted in 2019). At that point NCES will (a) publish on Regulations.gov an amendment to this package with all details for these special studies, (b) announce a 30-day public comment period on these details in the Federal Register, and (c) submit the amendment to OMB for review."

We have also revised the text of the NIES section in Part A to read: "The 2019 NIES study will use similar methods as those used in previous years. As noted above, an amendment to this package describing the study details will be submitted for approval prior to conducting the study.

OMB Comment 11: Needs more detail, it may be possible to point to supporting statements from the previous administration.

Associated text (A.1.d): Computer Access and Familiarity Study (CAFS).

OMB Comment 11a: How is the school sub-set sample selected? How is PBA/DBA assigned?

Associated text (A.1.d): The 2017 CAFS sample will be a nationally representative sample of 150 public schools participating in the reading and mathematics operational assessments at grades 4 and 8. All NAEP sampled students in the sub-set sample of schools will participate in the CAFS study. Some students will take NAEP using a paper-and-pencil administration mode (PBA), while others will take NAEP using tablets (DBA).

OMB Comment 11b: How was this sample size selected? What's the power/MDE? Wouldn't the recruitment materials and procedures be different? How can the sample be the same if the sample size is different? How can administration be the same if you're testing different modes of administration?

Associated text (A.1.d): The expected yield is approximately 3,000 DBA students per grade/subject and 750 PBA students per grade/subject.

NCES Response: Revised and added text to the CAFS section in Part A addressing these questions, which now reads: "The 2017 CAFS sample will be a nationally representative subsample of 150 public schools participating in the reading and mathematics operational assessments at grades 4 and 8. The sample will be stratified on characteristics such as census region, urban/rural, school race/ethnicity composition, and school enrollment size. All NAEP sampled students in the subsample of schools will participate in the CAFS study. Within a school selected for the NAEP reading and mathematics assessments, students will be randomly assigned to either DBA or PBA. The ratio of sample sizes for the two modes within each school will be approximately 4:1, with some variation depending upon the size of the school and the jurisdiction.

The expected yield is approximately 3,000 DBA students per grade/subject and 750 PBA students per grade/subject. Based on the results of the 2015 study, it was determined that a minimum sample size of 750 students were needed for each grade, subject, and mode. This sample size supports sufficient power in detecting an effect size of 0.1-0.16 and 0.2-0.32 for DBA and PBA, respectively, between students with low and high computer familiarity. This means that a sample of 150 schools per grade is needed to provide this sample size of 750 students per subject for PBA. These schools will also contain 3,000 students per subject who will be assessed using DBA. It is highly desirable from an operational perspective to have all NAEP students in a school complete the CAFS questionnaire, rather than a subset, and having the additional DBA sample will provide additional power for certain analyses. For the PBA sample, after the students complete their regular printed NAEP booklet, they will be given a separate booklet of CAFS questions. For the DBA sample, the CAFS questions will be an additional section of the student questionnaire, which is administered on the tablet.

Some analyses will be conducted combining the students in the different NAEP subjects, while other analyses will focus within subject only. Analyses, including factor analyses, IRT scaling, and correlational analyses, will examine the relationship between access and familiarity and performance on NAEP (overall and for certain subgroups), if the relationship varies by subject area or mode of administration, and if reliable composites related to computer access and familiarity can be constructed. The goal of the study is to inform the development and use of computer access and familiarity items in the questionnaires and reports for future NAEP assessment years."

OMB Comment 12: More detail, including a discussion of the 2011 study and why more research is needed.

Associated text (A.1.d): Multi-Stage Testing (MST) Study.

NCES Response: Added the following text: "The 2011 study was exploratory and a necessary first step to examine potential gains of MST for the NAEP program before a much more significant investment for operational deployment (in terms of resources, reputation, trend maintenance) could be considered. Gains were defined in terms of (conditional) standard errors, ability to meaningfully describe performance over a wider range of proficiency levels, and student engagement. A subset from the existing and pilot item pool containing predominantly multiple-choice, paper-based items, were transformed for computer-based assessment administered to an approximately national sample.

The current study is entirely geared towards preparing for operational deployment using a subset of items from the 2017 operational pool, the operational delivery system on tablets, and a nationally representative sample. The advisability to study an operational design before deploying operationally rests on the fact that, at the very core, the NAEP program is charged with maintaining trends. Therefore, any significant design changes require careful study and, in many cases, carefully designed bridge studies, in order not to interfere with the ability to maintain a robust

trend. Given that much of the previous research on MST design and implementation has been conducted on individual assessments and the psychometric and statistical parameters are very different for individual assessments than group-score assessments (such as NAEP), it is critical to study this major design change in the NAEP setting.

The 2017 MST study will be conducted at both grades 4 and 8 mathematics in conjunction with the operational assessments. As such, the same sampling, recruitment, and administration procedures as the operational assessments will be used. The only difference between this study and the operational assessment is how items are assigned to blocks and how blocks are assigned to students. In this study, students will first be randomly assigned to a 30-minute routing block, and then routed to a second 30-minute block targeted to ability level: easy, medium, or hard. The second stage (i.e., the target block) has different designs for the two grades. For grade 4, the design includes an adjacent routing component where some students are assigned to the adjacent targeted level rather than their intended level (i.e., some students routed to easy will be assigned a medium block). There will be no overlapping of items across blocks. For grade 8, on the other hand, blocks will be assembled with overlapping items among routers and between targeted levels. However, there is no adjacent routing component at grade 8 (therefore, all students routed to an easy block will be assigned an easy block). The analysis will evaluate the IRT item parameter estimates obtained from the MST designs in relationship to the item parameter estimates from the 2017 DBA operational assessments as the baseline. Consistency in parameter estimates between the 2017 MST study and the DBA operational assessment would be a positive outcome, indicating the MST design can be implemented in NAEP going forward.

The 2017 MST study will be administered to a national sample of 10,000 students at each grade. As with operational assessments, the sample size for this special study is primarily driven by the need for sufficient numbers of student responses at each item to support IRT calibration. For grade 4, the target sample size is approximately 3,000 per item for the first stage routing blocks, and approximately 1,100 to 2,700 per item for the second stage target blocks. For grade 8, the target sample size is approximately 3,300 and 800 per item for the first stage routing blocks and the second stage target blocks, respectively. The variation in sample sizes are functions of different numbers of blocks at each stage, as well as at each targeted level."

OMB Comment 13: This could be cleared under this submission with a little more detail. Clarify how the results are used to adjust or standardize the assessments and where the module is used (list all jurisdictions, number of students affected). How are jurisdictions selected?

Associated text (A.1.d): 'Knowledge and Skills Appropriate (KaSA) in Mathematics'.

NCES Response: Added the following text: "While the original KaSA instrument was designed to address a broader need to improve measurement precision on low-performing students, the KaSA special study has only been implemented in Puerto Rico as NAEP has had difficulties historically reporting scale scores for Puerto Rico. As the program moves to multi-stage testing design, KaSA items will be part of the MST instrument. And the selection of students (from all jurisdictions, including Puerto Rico) receiving KaSA items, as well as other targeted items, will be based on their performance on the routing items.

Currently, the KaSA special study serves as a bridge to enable NAEP to report on Puerto Rico similar to other jurisdictions. The 2017 KaSA study will be conducted at both grades 4 and 8 in conjunction with the operational mainland assessments. As such, the same sampling, recruitment, and administration procedures as the operational assessments will be used. For each administration mode (PBA and DBA) in 2017, the study design involves both Puerto Rico sample (3,000) and a nationally representative linking sample (3,000) receiving KaSA blocks in addition to the operational blocks. The sample sizes are primarily driven by the need for sufficient numbers of student responses per item to support IRT item calibration, as well as to support Puerto Rico jurisdiction-level reporting. During analysis, a statistical linking approach in IRT calibration is used to link the Puerto Rico student proficiency onto the operational reporting scale. Using this KaSA special study methodology, NAEP has been able to report scale scores for Puerto Rico since 2011."

OMB Comment 14: How? What's the power to discriminate differences? What actions will be taken if differences are detected? More detail was presented on 5/23.

Associated text (A.1.d): The term "bridge study" is used to describe a study conducted to ensure that the interpretation of the assessment results remains constant over time.

NCES Response: Replaced the description of the Digitally Based Assessment (DBA) Bridge Studies with the following text: "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. Three DBA bridge studies are planned:

- In 2017, PBA bridge studies are planned in reading and mathematics in addition to the operational DBAs to confirm the findings from the 2015 initial national-level bridge studies;
- In 2018, a PBA to DBA bridge study is planned in U.S. history, civics, and geography; and
- In 2017, a laptop to tablet DBA comparability bridge study is planned in writing at grade 8; it will be conducted after the regular NAEP administration window.

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. For reading and mathematics at grades 4 and 8, the 2015 PBAs will be re-administered at the state and TUDA level in 2017, along with the operational DBAs.

In 2017, the PBAs will be administered to a representative sample in each jurisdiction, enabling the examination of the relationship between PBA and DBA performance within each jurisdiction. The targeted PBA sample size is 500 students per state and TUDA, as well as 500 private school students for each subject within a grade. The sample sizes are driven by the need for sufficient numbers of student responses per item to support IRT item calibration, as well as to support evaluating mode effect at the state and TUDA level and for the private school population. The PBA will allow us to both measure and potentially adjust for differences due to the change in mode.

Similar PBA bridge studies will be conducted in 2018 for U.S. history, civics, and geography at grade 8. Given that the operational assessments of those three subjects are at the national-level, in 2018, the PBA will be administered to a nationally representative sample for each subject. The total sample size across the three subjects is 24,000. The size of national sample is primarily driven by the need for sufficient numbers of student responses at item level to support IRT calibration.

In addition to the PBA to DBA bridge studies mentioned above, NAEP will also study the transition from laptop-administration to tablet administration in writing. The first operational writing DBA was administered on laptop in 2011 at grades 8 and 12. The grade 8 writing assessment will shift delivery mode from laptop to tablet for the 2017 operational administration (note, grade 12 is not being administered in 2017). The goal of this study is to gather information about potential device effects on grade 8 student writing performance on tasks. Student writing performance on tasks on two devices—tablet vs. laptop—will be compared. This information will support better interpretation of trend results between 2017 and 2011.

A nationally representative sample of 3,000 students will participate in the study. Although the study will be administered during a separate window (from April to May 2017, as opposed to the rest of NAEP being administered from January to March 2017), the same recruitment, sample, and administration procedures will be used. Six of the writing tasks administered as part of the 2017 operational tablet-based grade 8 assessment will also be administered in this study. The task level summary statistics (e.g., average score on a writing task) will be compared, along with score distributions (ranging from 0 to 5). The comparison information will be used to inform interpretation of the trend results between 2017 and 2011.

While the sample size for most NAEP assessments is primarily driven by the need for sufficient numbers of student responses per item to support IRT item calibration, no item calibration is planned for the laptop-based sample. Therefore, the sample size for this study supports sufficient power (at least 0.8 with significant level of 0.05) in detecting a small effect size of 0.2 for average task score comparisons between the two devices."

OMB Comment 14a: Please explain this more. Wouldn't the recruitment materials and procedures be different? How can the sample be the same if the sample size is different? How can administration be the same if you're testing different modes of administration?

Associated text (A.1.d): A nationally representative sample of 3,000 students will participate in the study. Although the study will be administered during a separate window (from April to May 2017, as opposed to the rest of NAEP

being administered from January to March 2017), the same recruitment, sample, and administration procedures will be used.

NCES Response: The same assessment approach is utilized in both the writing comparability study and the main NAEP assessments. As such, many of the recruitment materials are the same; however Appendices D-20, D-21, and D-22 were added to reflect specifics for this study, such as the different mode of assessment, the purpose, and the assessment window (see Section B.2.a for a description of the materials). In order to draw a nationally representative sample, and as described in Section B.1.a., the sampling procedures will be the same as for the regular NAEP assessments. The sample sizes are different, based on the minimum sample needed for analysis and reporting purposes (as described in the response to Comment 14). Finally, the administration procedures for this study are essentially the same as the regular NAEP assessments, with the exception of the equipment. Regardless of mode or window, NAEP brings in all of the equipment for the study and, therefore, we are not asking the schools for anything different. In addition, the field staff directions to students for a laptop or tablet administration are essentially the same. The instructions provided to students by the test delivery system are specific to the device that is used by the students to respond to the assessment questions. Section B.2.c has been updated to indicate that NAEP field staff will bring in laptops for this study. The updated relevant text of Section B.2.c now reads "Trained NAEP field staff will set up and administer the assessment and provide all necessary equipment and assessment materials to the school, including paper booklets and pencils for the paper-and-pencil assessments; tablets with an attached keyboard, stylus, earbuds, and, for some subjects, mouse for the digitally based assessments; and laptops with a mouse for the writing comparability study. Internet access is not required for the digitally based assessments (DBAs). "

OMB Comment 15: What kind of investigation?

Associated text (A.3): With the increased use of technologies, the methodology and reliability of automated scoring (i.e., the scoring of constructed-response items using computer software) has advanced. While NAEP does not currently employ automated scoring methodologies, these may be investigated and ultimately employed during the assessment period of 2017-2019.

NCES Response: Added the following text: **"**One possible study involves using two different automated scoring engines and comparing the scores to those previously given by human scores. 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).

The Training/Evaluation set would be used to train, evaluate, and tune each scoring engine so as to produce the best possible scoring models for each constructed response item. The final scoring models would then be applied to the Test/Validation set producing a holistic score for each response.

Automated scoring performance is typically evaluated by comparison with human scoring performance. Evaluation criteria for the scoring models would include measures of inter-rater agreement such as correlation, quadratic-weighted kappa, exact and adjacent agreement, and standardized mean difference.[4] These measures would be computed for pairs of human ratings as well as for pairs of automated and human scores.

In addition to comparing how well each individual scoring engine agrees with human scorers, we would also compute how well the two scoring engines agree with each other. We would also evaluate how well the combination of the two engines (computed by averaging their scores) agrees with the human scorers. Results of these investigations would determine if automated scoring could be utilized for specific NAEP assessments or if additional investigations are required."

OMB Comment 16: This section needs a lot more detail in order to verify compliance with CIPSEA under the new collection protocols. Describe, step by step, how PII is created, how it's moved, and how it's destroyed.

Associated text (A.10): Students' names are submitted to the Sampling and Data Collection (SDC) contractor for selecting the student sample. This list also includes the month/year of birth, race/ethnicity, gender, and status codes for students with disabilities, English language learners, and participation in the National School Lunch Program.

[4] Evaluation criteria will be based on criteria advocated in Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practices, 31*(1), 2-13.

NCES Response: Added the following text: "In addition, the Sampling and Data Collection (SDC) contractor has obtained from the Department of Education's Chief Information Security Officer (CISO) a Security Authorization to Operate (ATO) at the FISMA Moderate level and adheres to and continuously monitors the security controls in said Authorization. Security controls include secure data processing centers and sites; properly vetted and cleared staff; and data sharing agreements."

"More specific information about how NAEP handles PII is provided in the table below:

PII is created in the following ways	1. Public and non-public school samples are released by the SDC contractor to NAEP State Coordinators (public schools only), NAEP TUDA Coordinators (public schools only), and SDC Gaining Cooperation Field Staff (non-public schools only) using the secure MyNAEP for Schools web site.
	2. Schools are recruited by SDC field staff for participation in NAEP.
	3. Participating schools need to submit a current roster of students for the sampled grade for student sampling.
	4. Rosters of students can be created by NAEP State Coordinators, NAEP TUDA Coordinators, or NAEP School Coordinators
	a. Rosters are submitted through the secure MyNAEP for Schools web site
	b. Rosters must be in Excel
	5. PII is contained in the roster files: student names, month/year of birth, race/ethnicity, gender, and status codes for students with disabilities, English language learners, and participation in the National School Lunch Program.
	6. PII is stored in the SDC contractor's secure data environments.
PII is moved in the following ways	1. Student names (PII) are moved to the Materials Preparation, Distribution, Processing and Scoring (MDPS) contractor via a secure FTP site. These names are used to print Student Login Cards
	2. Student Login Cards are only created for students taking DBAs so the student names for the PBA students are not moved
	3. Student PII data is available to the NAEP School Coordinators and the SDC contractor's Field Staff through the secure MyNAEP for Schools web site.
	a. NAEP School Coordinators can view and update PII for their own schools
	b. NAEP School Coordinators can print materials containing PII for their own schools
	c. NAEP School Coordinators store materials containing PII for their own schools in the "NAEP Secure Storage Envelope"
	d. SDC contractor Field Staff can update PII for schools within their assignment
	e. SDC contractor Field Staff can print materials containing PII for schools within their assignment
	f. SDC contractor Field Staff store materials containing PII for schools within their assignment in their "NAEP School Folders"
PII is destroyed in the following ways	1. MDPS contractor destroys the PII after printing the Student Login Cards
	2. School Coordinators destroy the materials containing PII on or before the end of the school year
	3. SDC contractor Field Staff destroy the materials containing PII after the school assessment has been completed. SDC contractor Field Staff return their NAEP School Folders to Westat Home Office for secure storage, and eventual secure destruction
	4. SDC contractor destroys student names after all weighting quality control checks have been completed, thereby making it impossible to link the responses to any directly identifiable PII. This activity is completed in August (approximately 175 days following the end of the administration).

"

OMB Comment 16a: Please identify the point in the process at which it would be impossible to link the responses to any PII, for example in the event of a subpoena.

Associated text (A.10): PII is destroyed in the following ways:

NCES Response: Revised the text in the last row in the table above, pertaining to "PII is destroyed in the following ways", to read as follows: "SDC contractor destroys student names after all weighting quality control checks have

been completed, thereby making it impossible to link the responses to any directly identifiable PII. This activity is completed in August (approximately 175 days following the end of the administration)."

OMB Comment 17: Is the destruction confirmed in any way?

Associated text (A.10): All paper-based student-specific materials linking Personally Identifiable Information (PII) to assessment materials are destroyed at the schools upon completion of the assessment. The field staff remove names from forms and place the student names in the school storage envelope. The school storage envelope contains all of the forms and materials with student names and is kept at the school until the end of the school year and then destroyed by school

NCES Response: Added the following text as a footnote: "In early May, schools receive an email from the MyNAEP system reminding them to securely destroy the contents of the NAEP storage envelope and confirm that they have done so. The confirmation is recorded in the system and tracked."

OMB Comment 18: Public comment indicates that the burden estimates may be inaccurate for some respondents. Describe the burden estimation process. Do the estimates include time spent retrieving records? What are ED's plans for using MyNAEP data to estimate burden?

Associated text (A.12): Exhibit 1 in Part A

NCES Response: Added the following text to the Student burden description: "Based on timing data collected from cognitive interviews and previous DBA assessments, 4th grade students can respond to approximately four non-cognitive items per minute, while 8th and 12th grade students can respond to approximately six non-cognitive items per minute. Using this information, the non-cognitive blocks are assembled so that most students can complete all items in the allocated amount of time. Each cognitive and non-cognitive block is timed so that the burden listed above is the maximum burden time for each student. The administrators and/or test delivery system will move students to the next section once the maximum amount of time is reached."

Added the following text to the Teachers burden description: "Based on timing data collected from cognitive interviews, adults can respond to approximately six non-cognitive items per minute. Using this information, the teacher questionnaires are assembled so that most teachers can complete the questionnaire in the estimated amount of time. For adult respondents, the burden listed is the estimated average burden."

Amended the text in the Principals/Administrators burden description to read: "The burden for school administrators is determined in the same manner as burden for teachers (see above) and is estimated to average 30 minutes per principal/administrator."

Added the following text to the Pre-Assessment and Assessment Activities burden description: "Based on information collected from previous years' use of MyNAEP, it is estimated that it will take three hours, on average, for school personnel to complete these activities, including looking up information to enter into the system. We will continue to use MyNAEP system data to learn more about participant response patterns and use this information to further refine the system to minimize school coordinator burden."

OMB Comment 19: Please break this out by year and major component.

Associated text (A.12): The estimated respondent burden across all these activities translates into an estimated total burden time cost of $17,106,770 for 1,262,051 hours.

NCES Response: The associated text was amended with additional text and a table as follows: "The estimated respondent burden across all these activities translates into an estimated total burden time cost of $17,106,770 for 1,262,051 hours, broken out by year and respondent group in the table below.

	Students		**Teachers and School Staff**		**Principals**		**Total**	
	Hours	**Cost**	**Hours**	**Cost**	**Hours**	**Cost**	**Hours**	**Cost**
2017	431,584	$3,128,984	135,892	$4,286,034	9,157	$386,334	**576,633**	**$7,801,352**
2018	68,500	$496,625	19,851	$626,101	1,439	$60,711	**89,790**	**$1,183,437**
2019	445,125	$3,227,156	136,610	$4,308,679	13,893	$586,146	**595,628**	**$8,121,981**
Total	**945,209**	**$6,852,765**	**292,353**	**$9,220,814**	**24,489**	**$1,033,191**	**1,262,051**	**$17,106,770**

“

OMB Comment 20: This section is lacking the required detail. The sampling, weighting, and estimation procedures should be thorough enough to allow replication. The attached sampling memo provides some additional info, but not enough to completely describe the process. The level of detail on sample design and estimation that OMB expects to see for a collection of this stature is reflected in the online tdw’s for NAEP collections, unfortunately it isn’t clear how up-to-date those materials are. For an example, see the page on 2008 school-level weight trimming: https://nces.ed.gov/nationsreportcard/tdw/weighting/2008/arts_weighting_2008_base_schtrim.aspx

Associated text (B.1.a): B.1.a. Sampling Procedures

NCES Response: Added the following text: “Additional information about the sampling procedures used in NAEP can be found in the technical documentation at: http://nces.ed.gov/nationsreportcard/tdw/sample_design/. Note, while the latest documentation that has been published (as of the drafting of this document) is 2008, the procedures have essentially remained the same. The 2007 procedures can be referenced for details on sampling of state-level assessments.”

NCES recognizes the importance of publishing technical documentation in a timely manner. However, given the lengthy NCES review process and limited resources, documentation for later assessment years has not been published at this time. In addition, we have updated Appendix C to now reflect the 2017 Sampling Memo (as opposed to the 2015 version). While NCES is aware that the sampling memo does not fill the need for complete technical documentation, it does reflect the plans for the 2017 assessment.

OMB Comment 20a: The relevant material should be copied and pasted into the supporting statements.

Associated text (B.1.a): Additional information about the sampling procedures used in NAEP can be found in the technical documentation at: http://nces.ed.gov/nationsreportcard/tdw/sample_design/. Note, while the latest documentation that has been published (as of the drafting of this document) is 2008, the procedures have essentially remained the same.

NCES Response: Added Appendix G to this submission and revised in B.1.a the first paragraph added in response to OMB Comment 20 (above) to read as follows: “Additional information about the sampling procedures used in NAEP can be found in the technical documentation at: http://nces.ed.gov/nationsreportcard/tdw/sample_design/. Note, while the latest documentation that has been published (as of the drafting of this document) is 2012, the procedures have essentially remained the same. The 2011 procedures can be referenced for details on sampling of state-level assessments. A summary of the sampling procedures are included below. Additional details (taken from the 2011 procedures on the technical documentation website) can be found in Appendix G.”

OMB Comment 21: I don’t believe it’s accurate to say that the schools represent demographic groups.

Associated text (B.1.a): This ensures that NAEP assesses students in schools that represent different demographic groups.

NCES Response: Thank you for pointing this out. This sentence has been deleted.

OMB Comment 22: Stratified PPS? What are the explicit strata? What’s the sample allocation across strata? What’s the MOS?

Associated text (B.1.a): 5. Select the school sample

NCES Response: Revised the description for step 5 (Select the school sample) to now read: “After schools are assigned a measure of size and grouped on an ordered list based on the characteristics that are referred to in previous steps, the sample is selected using stratified systematic sampling with probability proportional to the measure of size using a sampling interval. This procedure ensures that each school has the required selection probability. By proceeding systematically throughout the entire list, schools of different sizes and varying demographics are selected and a representative sample of students will be chosen for the assessment. Additional details regarding the selection of the school sample is included in the technical documentation (e.g., http://nces.ed.gov/nationsreportcard/tdw/sample_design/2007/sampdsgn_2007_state_schlsamp.aspx).”

OMB Comment 23: See AAPOR guidance on response rate calculation, and provide details on eligibility rates.

Associated text (B.1.a): 6. Confirm school eligibility

NCES Response: Added the following text: "Eligibility counts are included in the technical documentation (e.g., http://nces.ed.gov/nationsreportcard/tdw/sample_design/2007/sampdsgn_2007_state_inelg_table1.aspx). Information on response rates can be found in Section B.3.b."

Also, amended the last sentence of the first paragraph of Section B.3.b to read: "The NAEP response rates follow AAPOR (American Association for Public Opinion Research) guidelines. Response rates, in percentages, from the 2013 NAEP assessment are shown below and can be found in the technical documentation (for example, http://nces.ed.gov/nationsreportcard/tdw/sample_design/2007/sampdsgn_2007_state_schlresp.aspx)".

OMB Comment 24: Trimming? What are the rules?

Associated text (B.1.b): assignment of a "base" weight, the reciprocal of the overall initial probability of selection

OMB Comment 25: How? What variables are used? Raking? Poststratification?

Associated text (B.1.b): adjustments for school and student nonresponse

NCES Response: Added the following text to Section B.1.b: "Additional information about the weighting procedures used in NAEP can be found in the technical documentation at: http://nces.ed.gov/nationsreportcard/tdw/weighting/. Note, while the latest documentation that has been published (as of the drafting of this document) is 2008, the procedures have essentially remained the same." Also added: "School nonresponse adjustment cells are formed in part by census division, urbanicity, and race/ethnicity. Student nonresponse adjustment cells are formed in part by student with disabilities (SD)/English language learner (ELL) status, school nonresponse cell, age, gender, and race/ethnicity."

and: "The student weight trimming procedure uses a multiple median rule to detect excessively large student weights."

and: "Weighted estimates of population totals for student-level subgroups for a given grade will vary across subjects even though the student samples for each subject generally come from the same schools. These differences are the result of sampling error associated with the random assignment of subjects to students through a process known as spiraling. For state assessments in particular, any difference in demographic estimates between subjects, no matter how small, may raise concerns about data quality. To remove these random differences and potential data quality concerns, a new step was added to the NAEP weighting procedure starting in 2009. This step adjusts the student weights in such a way that the weighted sums of population totals for specific subgroups are the same across all subjects. It was implemented using a raking procedure and applied only to state-level assessments."

OMB Comment 25a: The relevant material should be copied and pasted into the supporting statements

Associated text (B.1.b): Additional information about the weighting procedures used in NAEP can be found in the technical documentation at: http://nces.ed.gov/nationsreportcard/tdw/weighting/. Note, while the latest documentation that has been published (as of the drafting of this document) is 2008, the procedures have essentially remained the same.

NCES Response: Added Appendix H (NAEP 2011 Weighting Procedures) to this submission and revised in B.1.b the first paragraph added in response to OMB Comment 25 (above) to read as follows:

"Additional information about the weighting procedures used in NAEP can be found in the technical documentation at: http://nces.ed.gov/nationsreportcard/tdw/weighting/. Note, while the latest documentation that has been published (as of the drafting of this document) is 2012, the procedures have essentially remained the same. The 2011 procedures can be referenced for details on sampling of state-level assessments. A summary of the sampling procedures are included below. Additional details (taken from the 2011 procedures on the technical documentation website) can be found in Appendix H."

OMB Comment 26: Which assessments? Why only some?

Associated text (B.1.b): adjustment of the student weights in state samples so that estimates for key student-level characteristics were in agreement across assessments in reading, math, and science.

NCES Response: In the last bullet point for the final weights assigned to each student as a result of the estimation procedures, replaced “adjustment of the student weights to reduce variability by benchmarking to known student counts obtained from independent sources, such as the Census Bureau (this procedure only applies to some NAEP assessments).” with “adjustment of the student weights in state samples so that estimates for key student-level characteristics were in agreement across assessments in reading, math, and science.”