

Institute of Education Sciences  
National Center for Education Statistics

## RELEASE OF PROCESS DATA TO RESEARCHERS

NISS Technical Expert Panel • March 2020

### EXECUTIVE SUMMARY

Now that most National Center for Education Statistics (NCES) assessments and surveys are conducted using electronic modes, electronic data capture means that in addition to basic background information and final responses, data includes documentation of the process of responding. These process data comprise a time-stamped, click-by-click record of each student's progress through the assessment or survey. To date these detailed process data have not been made available; however NCES is now preparing to establish a mechanism for release of process data for research purposes. Therefore NCES asked the National Institute of Statistical Sciences (NISS) to convene panels of technical experts to advise them on how to most efficiently and safely release the data in a useful form. The first panel was charged with making recommendations about how to minimize disclosure risks from such a data release. The second panel was asked to consider what data would be most useful to researchers, and how it should be preprocessed to allow a broad range of users able to use the data efficiently. Specifically, this panel was asked for guidance on the creation of new variables that will be useful to researchers in the fields of education sciences, test development and psychometrics, behavioral psychology and related fields. The goal for NCES is to make the research process more efficient for these data users by preventing duplicative effort in creating these variables.

The panel met via teleconferences with an in-person meeting at NCES on 11-12 March, 2020. This summary discusses the second panel's recommendations.

#### Primary Recommendations

1. An internationally used data exchange standard should be adopted for raw process files.
  - Such standardization will allow the exchange among researchers of robust, reliable processing scripts.
  - Such standardization will have the added benefit of bringing in additional researchers interested in data mining, machine learning, etc.
2. For researchers who do not desire to work with raw process files, but who do want access to micro-data, we recommend preparation of two pre-processed files for each content area. One should be focused on items and one on examinees. For both item and examinee files, summary information comes in two forms: background information and summary statistics computed from the process data itself.
  - The examinee file should include:
    - Background variables:
      - A student ID that allows for matching to other examinee files

- demographic variables, including SES variables
  - Disability information accommodation information
  - ELL status
  - Performance level and total points
  - School code
  - Complex sample design information necessary for population level estimation
  - Blocks the student took
  - Process summary data
    - Item path of the examinee's progress through the block, i.e., a block path file that maps the state of the user at each item, including time stamps, tools used, whether the question was answered/was correctly
      - For reading sections specifically, knowing the screen layout is critical - whether a student is viewing the prompt and the question (side by side or by page change), and the frequency of the change between views
    - # omitted, # not reached
    - Total time in block and %-ile of time for examinee
    - Examinee's answers to each item
    - Item visit count and total time in item
    - Item tool use for each type of tool
  - The item file should include:
    - Background variables
      - Item ID that allows for matching to examinee file
      - Position in block and assessment
      - Item parameters
      - Subscale
      - # words, # images, #tables in item
      - Tool indicator (presence or absence of each tool available to user)
      - Correct response
      - Depth of knowledge measure
    - Summary variables
      - Summary statistics for time spent on item over examinees
      - Summary statistics for # of visits to item over examinees
      - % answer changed, omitted, and not reached over examinees
      - % tool use for each tool
      - Responses summary over examinees (e.g., % of each multiple choice selection)
3. A tool similar to or included within the NAEP data explorer should be developed for the process data for users who do not have interest/access to micro data.

[Read the Full Report](#)