

The Evaluation Engine Data Request

I. Overview

The Evaluation Engine is a tool for conducting education impact evaluations. For any particular program under study, the tool will create an appropriate control group drawn from comprehensive student-level data on all students in the state. The tool will analyze the impact of the intervention on selected outcomes (such as achievement test scores, high school graduation, and possibly postsecondary education outcomes) and produce a report explaining the results.

Participating states are asked to submit:

- Student-level records for all public school students in grades K through 12 for each of the past 3 academic years: 2009-10, 2010-11, and 2011-12. (These records will be used to create longitudinal student-level records.)
- Data elements for all K12 students, as described in detail in the attached EE Data Request spreadsheet, that include the following main categories of data:
 1. Student identification.
 2. School and district identification.
 3. Student enrollment, exit/graduation, and GPA data.
 4. Student characteristics, including grade level and demographics.
 5. Achievement test results in English language arts/reading and mathematics .
 6. State-specific supplemental data.
- Background information on state tests (such as valid score ranges, proficiency cut points, and comparability over time) and grading scales.
- Postsecondary education data, if available in the state data system, for all postsecondary students in each of the past 3 academic years, for purposes of examining outcomes for high school students.
- An additional year of data, in the future, during Year 2 of the project, via an automated online submission process.

II. Goal and Objective

A goal of the project is to offer the Evaluation Engine to all states at little or no cost after project completion. To do so, we aim to automate data submission, transformation, and analysis to the extent possible. A primary objective during Year 1 of this project is, therefore, determining: 1) availability of requested data elements, 2) variability of data definitions and values across states, and 3) capacity of participating states to provide data elements in the requested formats.

III. Data Details: K-12

A. Data File Specifications

K-12 data files submitted to The Evaluation Engine should adhere to the following parameters:

- File format. The submitted data file should be in CSV (comma-separated value) format. (A sample CSV template containing variable names for the first five data categories, as well as the first four data elements of category 6, will be provided.)
- Header record. The data file should contain a header record (i.e., header row), using the exact variable names provided (see below). The variable name for a given column must correspond to the data in subsequent records for that column.
- Annual files. Data for each academic year should be submitted in a separate data file.

B. Data Element Specifications: Table

The attached “K-12 Data Detail” table contains the following information for each data element:

- Data Element. A brief description of each of the data elements to be collected.
- Definition. The definition of the data element along with any related instructions.
- Variable Name. The identifier for the data element in the returned data file. PLEASE NOTE: In order for the uploaded data to be correctly processed, variable names in the header record of the uploaded file should *exactly* match the variable names provided in the K-12 Data Detail.
- Values/Format. Formatting information and/or the list of available values for the data element. Common Education Data Standards (CEDS) coding, or another common data format, is used where possible.
- Missing Value Code. The value to provide if no value is available for the data element. (Typically, these are a string of 9’s or the entry “U”.)

C. Data Element Specifications: Additional Details

If a requested data element is unavailable in the state data system, the field should be coded with the appropriate “missing value code” listed on the data detail table. (Some omissions however, such as the state-assigned unique student ID, would preclude a state’s participation in Evaluation Engine or limit the potential analyses.)

The following notes each apply to multiple data elements:

- School and district (LEA) identifiers (section 2).
 - The first three data elements are district identifiers from your state’s *enrollment* data: (i) the Federal NCES ID, (ii) the state ID, and (iii) the district name.
 - The next three data elements are analogous school identifiers from your state’s enrollment data: (i) Federal/NCES ID, (ii) state ID, and (iii) school name.
 - The previous six data elements are then *repeated* with the distinction, however, that they are the district and school identifiers attached to the student in the state’s *assessment* data.
 - We ask for both sets of identifiers because students may be attributed to different schools at the enrollment census and at assessment time. If your state does not collect or cannot supply both sets of identifiers, the additional set may be omitted.
- Race/ethnicity data (sections 4 & 6). Race/ethnicity data are being requested in several formats.

- The race/ethnicity element (section 3) is meant to capture the single racial/ethnic category that best identifies the student.
- The six Y/N data elements (also section 3) are meant to allow a more nuanced description of race/ethnicity for each student; a “Y” in more than one category would therefore indicate a multi-ethnic student. If these separate data elements do not exist in the state data system, they should be omitted.
- State-defined race/ethnicity (section 6) is included to request that the state provide race/ethnicity data elements as they are natively stored in the state data system (only if these differ from the data elements provided in section 3).
- Free and reduced price lunch (section 4). Eligibility for free lunch or reduced price lunch may be reported together in a single data element, in two separate data elements, or both. Submit only the version(s) that are natively housed in the state data system.
- Dates: Diploma and withdrawal dates (section 3), date of birth (section 4) and test administration dates (sections 5 and 6). CCYMMDD indicates that date of birth should be provided as a string of eight digits, two each to indicate the century, year, month, and day of birth. All eight digits should be included. “May 9, 2012,” for example, would be reported as 20120509 (and not 120509 or 050912 or 5/9/12). If only year and month are available in the state data system, the “DD” portion of the entry should be coded as “01”.
- State achievement tests (section 5).
 - Data are being requested for up to four administrations of both ELA and math tests each year. (I.e., up to eight total tests.) Seven data elements are being collected for each administration of each test. These are:
 - a. Test name. (This is the test *series* name used by the state or the testing company – such as TCAP, FCAT, etc – rather than the test subject. Test names should be spelled consistently across student records.)
 - b. Exam version. (To indicate regular v. alternative assessments.)
 - c. Exam language. (To indicate English, Spanish, or other.)
 - d. Exam type. (Either grade-level, end-of-course, or high school exit.)
 - e. Specific test subject. (Note: For ELA the subject codes come from CEDS; for math they do not.)
 - f. Date of test administration. (See “date of birth,” above, for an explanation of the requested format.)
 - g. Scale score.
 - h. Proficiency level.
 - Additional background data are being requested to aid in the interpretation of state test data (see below).
- State-specific supplementary data (section 6).
 - The first four state-specific data elements request the district + school code for the district/school in which the student was enrolled at each enrollment census (or “snapshot”), starting with the last census of the *prior* academic year and continuing through the last census of the current academic year. Additional data elements may be provided if a census is taken more than three times during the academic year. A description should be provided to indicate which censuses are used in the state to deem a student as being enrolled for the full academic year.

- The next state-specific data element is “Age of student as of September 1”. This is being requested as an alternative for cases in which the state cannot provide birth month and year.
- For the next five state-specific data elements, a data element format is not being provided. The intent is to collect these data as they natively reside in the state data system. Each item may be provided as one or more data elements (fields). Variable names should reflect how these variables are identified in the state data system. A codebook should be provided to enable accurate interpretation.
- The remaining data elements are optional, to be provided at the state’s discretion.

IV. Data Details: State Tests Background Information

Separately, we will request detailed background information on the state’s K-12 testing/assessment program. (A web form is being developed for this purpose, though initially this detail might be collected via an Excel spreadsheet.)

V. Data Details: Postsecondary Education

For the Year 1 data submission, states that obtain postsecondary enrollment data from the National Student Clearinghouse (NSC) are encouraged to submit these data to The Evaluation Engine. So that these postsecondary data can be merged with K-12 data, records must contain the state-assigned unique student ID (in a field named STUDENTID_STATE). *Social security numbers must be removed* from the NSC data before submission to The Evaluation Engine. Student names, or at least initials, should be included if possible. (NSC data should be submitted separately from K-12 data.)

VI. Data Submission Procedure

An account will be provided to the state data manager for secure submission of data via the Evaluation Engine Data Submission Web Site. The account should be used by only this person and will be linked to their email address. The URL for the data submission site is <https://transfer.evaluationengine.org/>. Additional data submission instructions will be provided upon logging in to the account. A data validation report will be provided to the data manager after initial processing of the submitted data. If the validation report reveals errors, the data may be corrected and re-submitted. Data subsamples may be submitted, for testing/validation purposes, prior to submission of the complete data file.

VII. Contact Information

For any questions regarding this data request, please contact:

Roger Studley
Senior Research Associate
RTI International
510-665-8216
rstudley@rti.org