Maximizing ED_Facts_ Data Quality: A Comprehensive Approach

Background

The Data Quality Act of 2001 required that each federal agency, including the U.S. Department of Education (ED), issue guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated by the agency. In 2002, ED released the Information Quality Guidelines, which outlined the policy and procedures for reviewing and substantiating the quality of information it disseminates (e.g., reports, studies, and summaries). Then in 2006, ED established the ED_Facts_ Data Governance Board (EDGB), an intra-agency council with representatives from K–12 program offices across ED, that resolves issues and creates the policies needed to manage ED's asset of pre-kindergarten through grade 12 (preK–12) education data. One of the pillars of EDGB is validating the quality of ED_Facts_ data by working to ensure that data are usable and timely, accurate and valid for publication, and the same definitions are used across ED. EDGB adopted definitions and measures for five data quality attributes to operationalize the quality guidelines:

Attribute	Definition
Timeliness	Data are considered timely if submitted by the specified due (closing) date of the data collection.
Completeness	Data are considered complete if all of the required data are submitted, at each reporting level, for all education units, and all required category sets, subtotals, and totals. No data are missing and no placeholder data are submitted.
Accuracy	Data are considered accurate if they pass edits checks and data quality rules, contain no known errors, and are certified by the appropriate party.
Validity	Data are considered valid if they are reported in a consistent manner and measure what they were intended to measure.
Usability	Data are considered usable if the calculation/analyses are appropriate to the data and include explanations of anomalies.

This document describes a coordinated five-step data quality strategy developed by EDGB to review the quality of ED_Facts_ data. This strategy is described in the following section.

A note about data quality

Administrative standards for assessing data quality are important for ensuring data are fit for the intended uses by program offices. ED_Facts_ data are put through a quality review to ensure ED and other users have the appropriate understanding of and context for the usability of the data. The scope of a data quality review is associated with the purpose for which the program office collects the data and the intended use of the data for program planning, management, evaluation, or policy development. Administrative data quality standards vary from statistical standards for data quality. Though both sets of standards are used to ensure valid and accurate data, different rules and procedures are applied. Statistical standards reflect a more global set of norms and procedures set by the research community for the evaluation and correction of data errors in collecting, processing, analyzing, and reporting data for a generalizable research purpose. Applying statistical standards goes beyond assessing the quality of data fit for a designated administrative use. Understanding the different lens through which data quality is viewed and applied in administrative and statistical standards will help users of ED_Facts_ data make informed decisions about how to use the data.
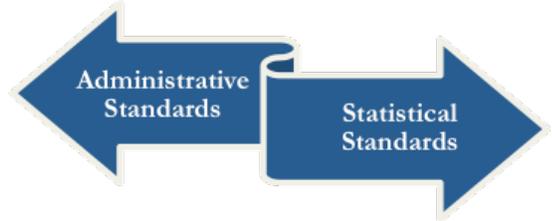
ED_Facts_ Data Quality Strategy

Data quality is reviewed throughout the cycle of submission, beginning with state efforts to collect data, state submission of data, acceptance of data into the submission system, post-submission review, and finally public release of data. This cycle is described in the figure and sections below.

1. **Efforts at the State**

   The foundation of the quality of data from administrative records is the capacity of the state education agency (SEA) to collect and store data from schools and/or school districts and submit the data. Many SEAs use longitudinal data systems to store data used for aggregation into ED*Facts* submission files.

   ED supports activities that improve the capacity of SEAs to manage data. For example, Common Education Data Standards (CEDS) is an education data management initiative that provides a common vocabulary for discussing data requirements, and the Statewide Longitudinal Data Systems Grant Program offers services and resources to assist both grantees and other states and territories with data system design, data governance, and data use.

2. **Data Quality at Data Acquisition Point**

   Data quality resources for data submitters

   ED provides a variety of tools and guidance to improve and support data submitters' capacity to manage and provide quality data, including:

   - Standardized reporting and data validation instructions for data submitters;
   - Standardized data dictionary;
   - Technical assistance through the ED*Facts* partner support center (PSC)—the primary interface between ED and the user community—who provide training, support, and a technical help desk;
   - Tools built into the submission system to check for errors and to review metadata about data sets; and
   - Guidance documents to assist with data submission:
     - o ED*Facts* Submission System (ESS) Users Guide—addresses the basic mechanics of system access and file transmission.
     - o ED*Facts* Workbook—provides information on how to submit files through the ESS for any given school year.
     - o ED*Facts* Submission System Release Notes—provide a description of the technical enhancements to each version of the ESS.
     - o File Specifications—provide technical instructions for building the files that are submitted through the ESS.

   Accurate and complete communication of data requirements

   The file specifications provide a description of the data to be submitted. The ED*Facts* PSC provides technical assistance related to submission and facilitates the responses to questions on the data requirements with the data stewards. The data stewards clarify the data requirements, using the approved paperwork clearance package.

   Central data collection system

   ED has a centralized, coordinated repository of state preK–12 education and performance data, and the ESS, an electronic data system capable of receiving data on more than 100 data groups at the state, district, and school levels. This centralized system facilitates data quality by ensuring consistent data management and standards across all data collected, reducing the potential for data inconsistencies.

3. **Data Submission Checks**

One key strategy at ED for enhancing the quality of ED_Facts_ data is to prevent errors and bad data from entering the system. ED has many automated edit checks programmed into each data submission system to prevent incorrectly formatted and inaccurate data from being accepted. Edit checks can also check for timeliness and completeness of the data submission.

Submission Business Rules

A set of business rules is applied to each ED_Facts_ file submission to check the overall quality of the data submitted. Data errors such as format and validation errors, submission errors, or match errors can be identified during the submission process. This process also identifies if data are consistently reported, unreported, misreported, or if expected relationships are present. During the submission process, data submitters can receive warning or error messages if the data fails one of the business rule edit checks. ED provides data submitters a Business Rules Guide that includes a comprehensive list of the business rules that apply to each file submission to help data providers prepare ESS submissions. In addition to helping users edit their data prior to submitting files, this guide is also a reference document to help submitters understand and correct any problems identified by the ESS. Business rules for data submitted to the ED_Facts_ Metadata and Process System (EMAPS) are included in the user guides.

4. **Post-submission Data Quality Reviews**

Some data stewards within ED conduct post-submission data quality reviews to further evaluate data anomalies or issues in the data. If a data steward conducts a review, the following are the steps that may be taken.

Standards for data quality review at ED

- Identify anomalies (outliers or unexpected values) in the data (both at macro and micro levels).
- Prioritize any anomalies identified for correction.
- Obtain feedback or context on the anomalies from the data submitters.
- Summarize the results for data stewards in easy-to-read and standard formats. Data stewards assess information to determine data quality.
- Data stewards decide if data are a fair representation and accurate enough to be accepted for program reporting and for use. They also decide what data to suppress if deemed to lack sufficient data quality.
- Provide data notes on quality for public releases.
- Implement a process of quality improvement for the next annual cycle of data reporting.

Data Quality Summary Reports

ED creates Data Quality Summary Reports to provide a high-level understanding of the timeliness, accuracy, and completeness across all ED_Facts_ data group files. These reports provide a universal format that

Who decides when data are valid? What level of error is acceptable?

Program office data stewards are responsible for assessing the validity of the data collected. With the assistance of the ED_Facts_ team, program offices decide on appropriate edit checks to evaluate data quality. If the data submitted fail an edit check, data stewards set a level of acceptable error and apply an appropriate threshold to determine data quality. This informs the decision about whether data are valid for the intended use and publication or whether the data should be suppressed due to data quality concerns.

program offices and data stewards can use as a starting point to better understand their data and follow-up with states as appropriate. There are two main types of data that are represented in these Data Quality Summary Reports: (1) metric files (most of which collect counts of students or teachers) and (2) status files (which are files that indicate whether [or not] a school or district satisfies criteria).

Data Quality Tracking List

ED developed a Data Quality Tracking list to track all ED_Facts_ data quality issues in a single location, tracking what actions ED took and what actions states took in response. This tracking list is reviewed at least annually to support development or revision of data quality checks throughout the data quality strategy cycle.

5. **Feedback and Public Data Release**

Out of the data quality review, feedback is provided to the data submitter, giving them the opportunity to explain data anomalies or correct data issues. This feedback and outreach step can be completed more than one time in a single cycle. Once the final data is received, program offices use the data to monitor implementation of the grant program. Program offices also publish data and engage the public to use the data. Data use is varied and products are designed to address the interests and data knowledge of intended audiences. Data uses include products such as summary statements in grant performance letters, annual reports to Congress, profiles or data display products, machine readable data files, and summary documents with key data results and visual data displays. In their use of the data, the public or researchers may ask questions, sparking additional reviews and checks of the data. By publicly highlighting aggregations of state-submitted data more frequently, ED hopes to increase the expectations for submitting quality data.

Once this cycle is complete on a school year of data, the cycle begins again at the state as they prepare to collect and submit data the next school year.