Career and Technical Education Data Validation

Strategies and Resources

Jon Boyette
jboyette@rti.org

Laura Rasmussen Foster
lrasmussen@rti.org

Prepared for the North Dakota Department of Career and Technical Education

June 2021
The Office of Career, Adult, and Technical Education (OCTAE) provides technical assistance (TA) each year to state and local career and technical education (CTE) leaders through the TA to States initiative. The TA helps states meet accountability requirements under the *Strengthening Career and Technical Education for the 21st Century Act (Perkins V)* and supports state and local CTE providers in using data to improve student and program outcomes.

Between October 2020 and May 2021, researchers at RTI International (“the TA team”) explored the following TA topics in partnership with key staff from the North Dakota Department of Career and Technical Education (NDCTE):

- What strategies can we take to ensure the reliability and accuracy of CTE data from secondary CTE providers?
- How can NDCTE streamline the data submission and review process for data submissions from tribal colleges?

RTI worked with staff from NDCTE—Paula Marschner, Education Data & Research Analyst; Wayde Sick, State Director; and Mark Wagner, Assistant State Director—along with Tracy Korsmo, Statewide Longitudinal Data System and BI Program Manager at the North Dakota Department of Information Technology.
BACKGROUND: TECHNICAL ASSISTANCE TO STATES (SLIDE 2)

To address North Dakota’s questions, the TA team

• reviewed national guidance and state CTE data validation practices in Colorado, the District of Columbia, Iowa, Kentucky, Nevada, and Utah;

• reviewed and updated NDCTE’s data collection workbook for tribal colleges with embedded data validation checks and routines; and

• facilitated a meeting between NDCTE and representatives of four tribal colleges to receive feedback on the data collection process and the revised data collection template.
This toolkit provides strategies that data analysts can use to review CTE student and program data for accuracy and identify recurring errors. It includes two sections:

• **Section 1: The data validation process** outlines common state approaches to data validation and offers strategies and recommendations for “manual” data validation.

• **Section 2: Tribal data collection template guidebook and checklist** describes the tribal data collection template and includes a checklist for template updates.
SECTION 1: DATA VALIDATION
The Data Validation Process: Overview

State CTE data validation strategies have four phases:
1. Training and guidance
2. Automated data review
3. Manual data review
4. Follow-up

NDCTE’s primary interest was the **manual data review** phase. The following slides include recommendations and examples of strategies for this phase.

### Purpose:
- **Training and guidance**: Establish shared understanding of CTE data concepts and requirements
- **Automated data review**: Ensure that the data are entered correctly
- **Manual data review**: Review data for data entry errors not caught by the automated review
- **Follow-up**: Raise and correct data concerns with local CTE providers
Establish a process to document data concerns and recurring issues

Formalize data validation sampling criteria (i.e., judgmental sample) and processes

Develop an analytical approach tailored to data concerns and recurring issues
Recommendations for North Dakota: Manual Data Review Strategies (slide 2)

- Establish a process to document data concerns and recurring issues
- Formalize data validation sampling criteria (i.e., judgmental sample) and processes
- Develop an analytical approach tailored to data concerns and recurring issues
Document Data Concerns and Recurring issues: Overview

**Key insights:**

- Manual CTE data validation, from sampling student records to analysis of potential errors, relies on the expertise of data analysts—not just analytical expertise, but expertise in specific state data systems.
- Different data systems provide different opportunities for errors to be introduced.
- Expertise in state-specific data systems and practices takes time and experience to develop and is therefore vulnerable to staff turnover.

**Cyclical review and documentation of data issues can improve data validation processes by**

- allowing data analysts to design their sampling methodology and analysis to review for known/historical data issues,
- identifying and addressing the points where data errors are more likely to arise, and
- preserving data analysts’ expertise by building it into data validation routines.
The template below documents data issues and potential explanations and resolutions, using examples from other states. A template like this can be used to review data issues with state and local staff, identify updates needed in the data collection and management process, and determine when data submitters need guidance.

<table>
<thead>
<tr>
<th>Issue</th>
<th>Explanation</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collapse in student CTE enrollment in a specific cluster</td>
<td>Misclassification of CTE cluster at local site</td>
<td>Recoded cluster</td>
</tr>
<tr>
<td>Sharp decline in percentage of CTE students reaching concentrator status</td>
<td>New program(s) at a specific site, leading to a higher number of participants relative to concentrators</td>
<td>Not an error, documented introduction of new program</td>
</tr>
<tr>
<td>Dramatic swings in CTE student enrollment</td>
<td>Lack of clarity around which students should be counted leading to differences in selection criteria from year to year</td>
<td>Developed/revised guidance for local providers</td>
</tr>
<tr>
<td>Dropped course data</td>
<td>Inconsistent course coding</td>
<td>(Short term) Recoding course data for match</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Multiyear) transition to common course numbering</td>
</tr>
</tbody>
</table>
Staff at NDCTE identified examples of recurring issues encountered in CTE data collection cycles, such as

- unexpected declines in CTE enrollment in 2019–20;
- declining enrollment in CTE programs at the postsecondary level, coupled with an increase in the number of concentrators in postsecondary CTE programs; and
- Large year-to-year changes in nontraditional enrollment.

NDCTE can begin the process of addressing these data concerns by

- documenting these issues for the next round of CTE data collection and validation, and
- investigating data collected for the 2021 Consolidated Annual Report (CAR) to see if these issues are present.
Recommendations for North Dakota: Manual Data Review Strategies—Overview

Establish a process to document data concerns and recurring issues

Formalize data validation sampling criteria (i.e., judgmental sample) and processes

Develop an analytical approach tailored to data concerns and recurring issues
State data analysts use a “judgmental” sample during the data validation process.

- A judgmental sample is designed to maximize the likelihood that student records will reveal errors introduced during data entry, data transfer, etc.
- It is not primarily designed to be statistically representative (though some states may begin their review with a representative sample).

The criteria used to select a judgmental sample are shaped by

- knowledge of the data system(s) involved and where errors are likely to be introduced
- errors encountered during previous CAR and CTE data collection
In Colorado, local CTE providers enter data into a dedicated CTE data collection portal using one of two options:

- **Batch upload of student data** in spreadsheets (favored by large districts)
- **Manual entry of individual student records** directly in the portal (favored by small districts)

In sampling student records for data validation, the Colorado Community College System includes data from very large and very small districts to increase the likelihood that the sample will include any errors introduced through batch upload (e.g., faulty data merges or misaligned columns in the source dataset) or manual entry (e.g., typos).
The Division of Community Colleges and Workforce Preparation (DCWP) in Iowa pulls CTE student data from the state’s K–12 Student Information Management System and reviews it at least twice.

- **Initial review for data entry errors:** The Department of Education’s IT Office, which manages the Student Information Management System, validates and reviews the data before sending to DCWP.

- **Stratified sampling for additional review:** The DCWP uses stratified sampling to draw a representative sample of student records for review, organizing students into strata based on district enrollment, number of CTE programs, CTE enrollment, and other factors.

DCWP data leads analyze the sample for statistically significant differences in student enrollment and concentration patterns over time and across districts and schools, before conducting a more targeted review of data where they have observed errors in data submitted in previous data runs.
Recommendations for North Dakota: Manual Data Review Strategies—Overview

- Establish a process to document data concerns and recurring issues
- Formalize data validation sampling criteria (i.e., judgmental sample) and processes
- Develop an analytical approach tailored to data concerns and recurring issues
Develop an Analytical Approach Tailored to Concerns and Recurring Issues: Overview

**Cross-sectional:** Analysis of student data drawn from a single period (typically, an academic year)
- **Process:** Comparison of student records across schools, districts, regions, or student groups
- **Purpose:** To catch errors that are systematically connected to location, data collection method, etc.
- **Example:** Comparison of special population enrollment across schools or districts to identify those who may not have included those data

**Trend analysis:** Analysis of student data drawn from multiple periods (e.g., multiple years)
- **Process:** Comparison of student enrollment or performance data for a similar set of schools, regions, or student groups from one year to the next
- **Purpose:** To identify errors that might result from changing staff or data collection and processing methods from year to year
- **Example:** Review of performance indicator denominators from year to year to identify dramatic changes in the number of students included
Develop an Analytical Approach: Sample Research Questions

Cross-sectional:

- Are there areas or districts with unusually high concentrations of students in special populations?
- Are there areas or districts with unusually high or low rates of CTE concentration (relative to participation)?

NDCTE-relevant examples for **trend analysis** include the following:

- Have we seen unexpected declines in CTE enrollment since last year?
- Have we seen seemingly opposing trends (e.g., declining enrollment in CTE programs at the postsecondary level coupled with an increase in the number of CTE concentrators in those programs)?
- Do we see dramatic changes in nontraditional enrollment at the secondary level?
## The Data Validation Process: Recommendations and Suggested Action Steps

| Establish a process to document data concerns and recurring issues | • Use the sample template included in the [appendix](#) to create a list of data issues encountered in previous data runs and document how those issues were resolved  
• Use the template to record issues that arise when preparing data for the 2021 CAR report  
• Review the template with the state and assistant CTE director and other key staff following the 2021 CAR report |
|---|---|
| Formalize data validation sampling criteria (i.e., judgmental sample) and processes | • Create a short document to summarize how data is sampled during the CTE data review process (currently, three strata of small, medium, and large school districts)  
• Identify potential entry points for data errors in the data collection process  
• Review explanations and/or resolutions for data issues identified in previous data runs  
• Where errors are systematically connected to student, school, district, or institutional characteristics, include those characteristics as sampling criteria in future data review |
| Develop an analytical approach tailored to data concerns and recurring issues | • Develop a list of research questions for cross-sectional and trend analyses  
• Incorporate previously identified data issues in the list, update the list to add questions as new issues are identified, and remove questions as issues are resolved |
SECTION 2: DATA VALIDATION TEMPLATE REFERENCE AND CHECKLIST
The TA team modified NDCTE’s tribal data collection template to identify basic formatting errors and reduce the need for repeated follow-up from NDCTE to the tribal colleges. Revisions include

- splitting data tables into individual, topic-specific data worksheets;

- adding explanatory notes to each worksheet and notes that appear when data analysts click on each column header to provide additional detail on the data requested; and

- adding macros that apply native Excel data validation to existing data to flag common data issues (see screenshot below).
Excel’s data validation function highlights cells that contain data that do not meet user-defined formatting rules.

However, this feature can be erased when data are pasted in, the method often used to transfer data from tribal college data systems to the data collection template.

To address this problem, the revised templates include macros (i.e., short computer programs) that apply data validation formulas after the data have been entered (see example at right). Unlike the data validation function, the macros are not erased when data are pasted in.
Updating the Template

NDCTE will need to update the template each year to
• ensure that program/reporting years are accurate, and
• verify that the range of possible values included in validation checks is up-to-date and reflect any definitional, program, or policy changes.

The following slide provides a checklist for updating the template. The checklist includes

• narrative updates, that is, updating text and instructions with the current reporting year and to reflect programmatic/definition changes, and

• data validation updates, for example, updating allowable values to reflect changes in data formatting rules, verify that correct date ranges were entered, etc.
Annual Updates to the Template: Checklist

On the “Home” tab
☐ Update the text to reflect the current reporting year.

In each data entry tab
☐ Update the “Who to include?” text box to reflect the current reporting year and academic years for which NDCTE is requesting data.

In the “Possible Values” tab
☐ Update the TERM and COMPLETED_TERM lists to reflect the time frame for which you are requesting student data.

☐ Update the Classification of Instructional Programs (CIP) code list to include CIP codes for approved CTE programs.

☐ Update the GENDER, RACE, LEVEL, and DEGREE lists as needed to reflect changes in policy, definitions, and demographic classifications.
APPENDIX
The following template can track issues encountered in each CAR reporting cycle (the Excel version was sent with this slide deck to NDCTE).

**Perkins V Data Validation: Issues and Observations**

<table>
<thead>
<tr>
<th>Analyst</th>
<th>[NAME OF ANALYST]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR Reporting Year</td>
<td>2021</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Issue</th>
<th>Explanation</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unexpected declines in CTE enrollment in 2019–20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Declining enrollment in CTE programs at the postsecondary level, coupled with an increase in the number of concentrators in postsecondary CTE programs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dramatic changes in nontraditional enrollment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>