Title III Data Clean-Up review

CGCC EASE Project
Prepared by Sara Mustonen, Project Director
2025

Activity 1: Strengthen the integrated approach of academic programming and support services to support student access and retention.

Activity 2: Facilitate an equitable student experience through the expansion of Tutoring and Mentoring services, the physical enhancement of teaching, learning, and gathering spaces across campus, the expansion of the HyFlex teaching method, and systematic professional development programming that supports diversity of talent acquisition and teaching strategies.

Contents

Data Clean-Up Project Overview

Data access and clarity plays a foundational role in enabling institutions funded by Title III Part A Strengthening Institutions Program (SIP) grants. Reporting timely and complete data to create capacity to meet grant goals and objectives enhances an institution's ability to provide quality education and support student success. In determining a need for enhancements in data quality in year one of the project, project management requested fund allocations under this project award specific to improving data outcomes. These approved funds provided the college with resources to perform data clean-up as an activity within the scope of the project. Clarity in data outcomes enhance an ability to report on specific metrics with higher rates of accuracy and identify student progress. In this case, data clean-up efforts served to broaden the n-group, or priority population identified as the target group as indicated by project design.

> For Data-Clean Up: The scope of this activity involves performing data clean-up tasks in the Student Information System (SIS) following an integration process during the Covid-19 pandemic in 2020 at the outset of the SIP grant award. The aim of the specific applied resource request is to ensure data accuracy, consistency, and completeness of the data migration within the SIS database.

Initial review of the outcomes from this data-clean up effort, the activity has been determined highly successful with reducing unknown race and ethnicity data markers from 39% (IPEDS DFR, 2023) at the initial phase of the project, to 7% (preliminary institutional data) at its completion in 2025. These positive outcomes from this activity are reflected in the National Student Clearinghouse IPEDS data reports and Postsecondary Data Partnership dashboards. These successful results have provided a foundation for improvements of systems and practice for institutional data governance and stewardship at Columbia Gorge Community College (CGCC).

Additional IPEDS DFR report references show zero unknown race and ethnicity data markers in 2020, and two percent unknown race and ethnicity markers in 2021. Variances in both the SIS and the application process have greatly influenced the ability to capture the information. By adding a "decline to respond" admissions option, it is unlikely the institution will see a reduction below the standard comparison group median of three percent (IPEDS DFR, 2024).

Objectives

The ability to collect, access, and report reliable data is necessary to support the grant objectives.

Three main goals of this activity include:
- Creating capacity to assess institutional progress;
- Monitor progress and outcomes; and
- Provide accountability through accurate reporting.

Those goals center tracking effectiveness of the initiatives funded by the Title III Part A SIP grant, measuring impact of these efforts, making data-driven decisions to adjust strategies as needed, and regularly reporting with clearer definitions on how Title III Part A funds are being utilized and the outcomes achieved.

Method

Challenges with the conversion of SIS systems during the Covid-19 pandemic were tied to logistical and resource constraints. Additionally, rapid enrollment decline attributed to staffing role changes and overall reduction in staff. For these reasons, data migration was incomplete at the outset of the start of the award performance period in 2021-22.

Discovery of gaps in data capture were evident and documented by the Title III Project Director at the first mid-term Interim Progress Report (IPR) report for the grant. Support to manage data recovery was unavailable and SIS configuration was prioritized to maintain enrollment operations.

Actions to improve needed evidence of data quality were determined through the following steps:
- Identify and rectify inconsistencies, errors, and duplications for integrated data.
- Standardize data formats, naming conventions,and categorizations for uniformity.
- Validate data integrity and ensure compliance with regulatory requirements.
- Enhance data quality to facilitate accurate reporting and analysis.
- Document processes and procedures for future reference and maintenance.

Subsequent Annual Performance Reports (APR) document the progress resolving the identified challenges. Additional internal and external evaluation and reporting demonstrate a high level of proficiency in performing the outlined work to completion. Particular to performance reporting is determining the success of the interventions applied to the Goals and Objectives of the SIP award.
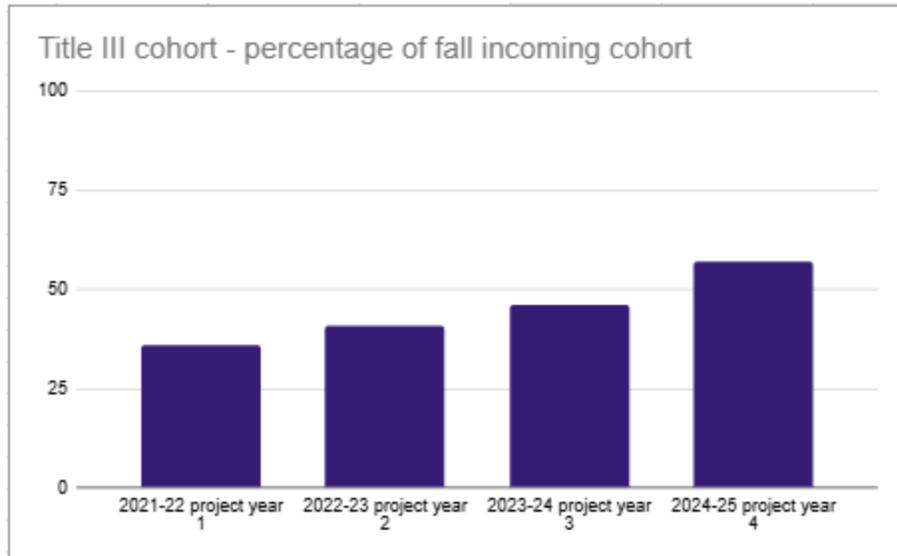

Figure 1. N value of CGCC Title III cohort increased year over year.

Quality of data in regards to sample size (Figure 1) improved because of two major activities conducted during the performance period. The Title III cohort measurements rely on particular markers tied to low-income and First-Generation, or First-Time Ever in College student status. Due to the challenges identified early on in the award performance period, very small cohort numbers limited evaluation of effectiveness of the planned interventions. In addition to data clean-up, resources were applied to increase student access to information regarding funding eligibility. These efforts combined provided a larger cohort to assess for intervention evaluation.

Data Assessment

For the scope of work, the college conducted a comprehensive analysis of the integrated data to identify issues such as duplicate student records, missing values, needs for corrections, and inconsistencies in informational fields within the SIS. For this work, teams met with Institutional Research to assess gaps. Report functions within the SIS were utilized to search for missing values. Admissions applications were reassessed and reviewed. Student enrollment forms

were reviewed to ensure all data was consistently collected. Errors and discrepancies were reported to National Student Clearinghouse (NSC) and resolved in a timely manner. Data quality metric assessments for accuracy, completeness, consistency, timeliness and validity were done in collaboration with Institutional Research to assess common gaps in data reporting and improvement of accuracy in data entry. Previous delayed reporting to NSC improved and continues to show progress and corrections for identified errors are complete.

Data Cleaning

Data improvement efforts included the development of data cleaning procedures to address identified issues systematically. The following initial steps were implemented:

- Evaluation of annual enrollment reporting.
- Reconfiguration of student record creation requirements.
- Alignment of application for admissions to required responses in all fields.
- Include "decline to answer" in fields to permit optional reporting where permissible.
- Collaboration to gather and verify complete information from attending students.
- Reviewed manual entries for accuracy and performed corrections.
- Increased digital security for admissions applications to reduce risk and threats from imposters.

For removing duplicate records and resolving conflicts by merging and updating information where necessary, these observations and actions were conducted:

- Implementation of logic to identify and contain duplicate records.
- Identification of common patterns of records that were created inconsistently.
- Collaboration with Title III Project Director and IT Department to contain imposter, or computer generated fraudulent accounts.
- Removed fraudulent accounts that were not generated by human student applicants.
- Initiated a process for identifying and merging duplicate student records.

To standardize data formats such as dates, naming conventions, and address structures, the following steps were actioned:

- Updated admissions application to perform consistent data collection.
- Installed force formatting in admissions application to ensure values are entered and captured correctly. (example: Date of Birth)

Throughout the process, review of data quality included validation of data entries against predefined criteria. Corrections of errors and discrepancies enhanced data outputs. Enrichment of data quality was verified internally and externally. Externally, evidence of effective activity completion is most evident in accurately capturing markers tied to race and ethnicity.
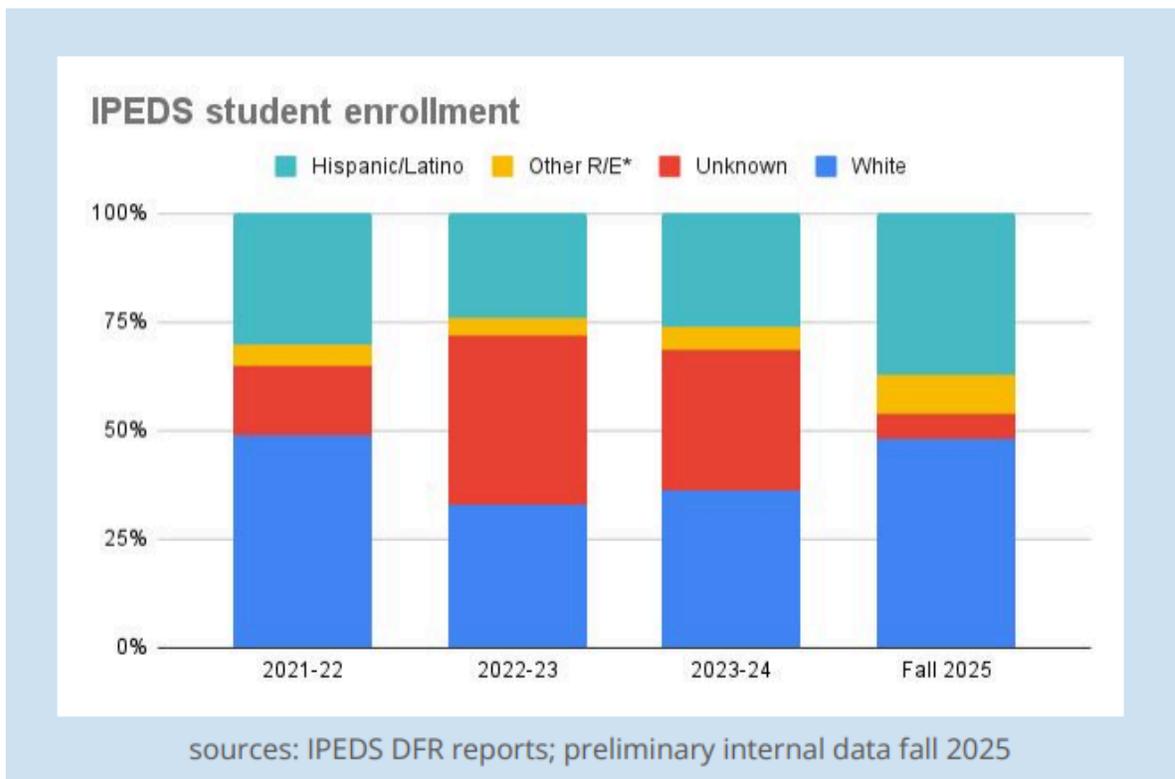

Figure 2. IPEDS DFR reporting combined with internal preliminary data from CGCC.

As a Hispanic Serving Institution, combined with high rates of low-income student enrollment, CGCC meets qualifications to be a Title III SIP award recipient. According to IPEDS DFR reports, at the time of the application and subsequent award, the gap of unknown student race and ethnicity markers was 16% of the total student population. In 2022, the unknown rate climbed to 39% due to the stall in data migration; and in 2023, it reduced to 32%. Preliminary internal data has determined less than 8% of student race and ethnicity markers remain unknown or are indicated with a "decline to respond" selection. At the time of this report, IPEDS has not produced a subsequent report to verify these findings (Figure 2).

Documentation

For documenting the data cleaning procedures, a digital **Registrar Quick Guide** was created to serve as a manual and process map for uniform procedural compliance. The guide includes steps for creating a student record and short recorded tutorials on basic SIS functions. Additional NSC trainings and tutorials are available in this comprehensive guide.

For uniformity in data input, the **Registrar Quick Guide** is available to all SIS users. This guide ensures consistency and enhanced quality of data outputs. Building consistency encouraged collaboration among SIS users across the Student Services Department from admissions application submission to student enrollment. Additionally, API process mapping is shared with CGCC IT to provide ongoing support to the data users and network.

Troubleshooting SIS usage included identifying where gaps in configuration resulted in abnormalities of desired data outputs. Additional challenges identified were related to historical knowledge and subject matter expertise in execution of systems configuration and maintenance. External technical support during systems implementation and availability of support for ongoing maintenance was inconsistent. Reporting accurate data outputs to NSC was screened and evaluated. Training ensured improved accuracy and timeliness of reporting.

Communication

Cross-collaborative efforts with internal and external stakeholders validated the improved quality of data outputs. Internally, quarterly updates to the CGCC Board of Education and President's Council provide information regarding decision making and resource allocation based on the data provided. Ongoing weekly communication with IT and SIS ensures the permanency of the process improvement. Regular collaboration with Institutional Research team members also serve to maintain efficiency in the process.

Deliverables

For the Title III project reporting, evaluative efforts were enriched by this lift and in turn, provided clearer information regarding opportunities for further interventions to mitigate gaps in student achievement. It also revealed success rate information to further inform pedagogical approach and instructional methods effective to improving student achievement for the priority population cohort (Figure 3).

In reviewing Early Momentum Metrics particular to college-level math course completion in the first year of attendance, students consistently performed well over the targets once the cohorts were clearly defined and coded to represent a larger sample of priority population student enrollment. Instructors were motivated by the feedback, and continued to seek alternative methods of instructional delivery, grading, and progress evaluation to push for further success in student achievement. Additional supplemental instructional supports were also identified and included in increasing the gains because of the quality of data reported for monitoring the interventions.
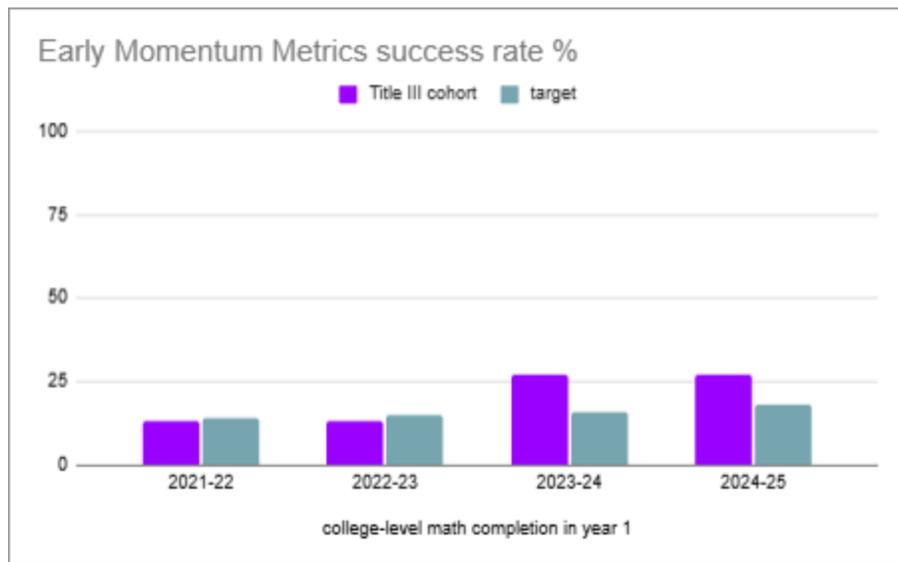

Figure 3. CGCC Early Momentum Metrics - college-level math.

As a participant of Excelencia and Rural Guided Pathways, CGCC is encouraged to regularly review disaggregated data to determine resource allocations and interventions to improve student success and rates of achievement for standard Early Momentum Metric milestones as well as monitoring student persistence, retention, and completion rates. As a part of the agreements with Excelencia, CGCC participates in the NSC Postsecondary Data Partnership (PDP). This network includes methods for sharing data as well as support for robust and ongoing student success evaluation.

Conclusion

Recommendations for further evaluation and inquiry remain ongoing, particularly in preparing for another SIS conversion in the near future due to current product expiration. Closer attention to established internal process and procedures have provided a foundation from which to

guide future conversions and an interdepartmental commitment to collaborative efforts for future data migration processes. In conclusion, the investments of this approved activity performed under the CGCC Title III grant fund award were highly successful in delivering excellent results which contribute to the overall goals and objectives of this project.

References

Data Source: Columbia Gorge Community College Institutional Research partners at Linn Benton Community College, Office of Data and Decision Support.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). (2021). *Data Feedback Report 2021* for Columbia Gorge Community College. Retrieved January 30, 2025, from https://nces.ed.gov/ipeds/dfr/2021/ReportHTML.aspx?unitId=420556.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). (2022). *Data Feedback Report 2022* for Columbia Gorge Community College. Retrieved January 30, 2025, from https://nces.ed.gov/ipeds/dfr/2022/ReportHTML.aspx?unitId=420556.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). (2023). *Data Feedback Report 2023* for Columbia Gorge Community College. Retrieved January 30, 2025, from https://nces.ed.gov/ipeds/dfr/2023/ReportHTML.aspx?unitId=420556.

U.S. Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). (2024). *Data Feedback Report 2024* for Columbia Gorge Community College. Retrieved January 30, 2025, from https://nces.ed.gov/ipeds/dfr/2024/ReportHTML.aspx?unitId=420556.