Identifying Carceral Enrollments: Navigating Challenges in Statewide Data Systems

Prepared by Research for Action | September 2026

Introduction

Researchers and advocates face persistent challenges when gathering and reporting data from institutions that serve incarcerated students, which create significant barriers to understanding postsecondary program impact, ensuring accountability, and advancing effective policy. Inconsistent, incomplete, or inaccessible data makes it difficult to assess whether incarcerated learners receive equitable educational opportunities, obscures our understanding of their educational pathways, and limits our ability to identify disparities or areas in need of improvement. These data limitations weaken advocacy efforts, hinder coordination between colleges and correctional facilities, and ultimately reinforce existing inequities, especially for a student population that is also disproportionately Black, Latine, and low-income. Structural factors such as fragmented data systems, varied reporting practices, and limited staff capacity, further compound these challenges, slowing progress toward statewide goals to expand high-quality college-in-prison programs, support reentry, and reduce recidivism.

Incarcerated students occupy an intersection between the justice system and the educational system and are typically tracked in one or both systems. However, these systems are commonly disconnected and may have conflicting missions; identifying educational enrollment, persistence, and completion often requires relying primarily on higher education data systems. But higher education data systems are not usually organized to support incarcerated learners, and Higher Education in Prison (HEP) program staff are commonly charged not only with supporting their students but also with managing data to effectively track students' progress through campus enrollment management systems. In our analysis, we analyzed two sources of data meant to identify the same incarcerated students and found they disagree on about 40% of students.

The purpose of this memo is to nuance the conversation about data collection for incarcerated students. In this memo, we outline how Research for Action (RFA) has approached identifying carceral enrollments in the California community colleges, and the challenges we encountered through this process. Specifically, we identified seven data challenges and show how these challenges might impact analytic goals. We have learned through this work that this data collection is not straightforward, and that obtaining reliable estimates for enrollment and completion among this population is more difficult than one might expect. We hope that through our efforts,

researchers studying incarcerated students may anticipate such challenges and traverse the landscape with more transparency.

The California Context

California has an extensive array of postsecondary opportunities available to individuals incarcerated in the state's adult prisons, provided by community colleges and public four-year institutions.[1] As is common across the country, most of the postsecondary education provided to currently incarcerated students in California is at the community college level. California has the largest community college system in the country, with 116 institutions, and has one of the nation's largest incarcerated populations.

As part of a larger study to systematically document how California's recent statewide developmental education reforms (i.e., AB 705 and 1705) have been implemented in colleges that serve incarcerated learners and how reforms have been experienced by the faculty and community college administrators supporting those learners, RFA sought to understand trends in course offerings, enrollment, and completion, both within and across institutions. To do this, we sought data on math and English course offerings, enrollment, and completion for students who experience incarceration and those who do not.

Students enrolling in community college while incarcerated, just like those who enroll on campus or online, are tracked in the California Community College Chancellor's Office Management Information System (COMIS) data. The COMIS data holds systemwide longitudinal information on student college attendance, demographics, course enrollment, and outcomes, as well as data on courses and sections offered. Included in these data are students experiencing incarceration who attend courses through carceral institutions' partnerships with California community colleges. Students who are enrolled in higher education while incarcerated are also tracked by the California Department of Corrections and Rehabilitation (CDCR), although those academic records are more limited.

Challenges with COMIS data

In early conversations with individuals working with COMIS data, we learned that researchers and advocates often encounter a disconnect between carceral enrollments at colleges and data available in COMIS, limiting their ability to use COMIS to understand enrollment and completion trends among incarcerated students. Specifically, the indicator sg04 that identifies carceral status of students in COMIS has not always been consistently populated with fidelity, leading to

[1] While many California community colleges also provide educational services to juvenile detention facilities and jails, we focus in this brief on the educational opportunities provided to individuals incarcerated in state and federal prisons.

inconsistent and potentially inaccurate estimates of incarcerated learners across the system and over time.[2]

We conducted an initial exploration of this claim by examining the enrollment of incarcerated students in Fall 2023 for those institutions identified as providing services to this population in the publicly available COMIS data, as well as additional institutions identified in the state's Rising Scholars Network.[3] Our exploration revealed two types of issues: data entry lags and inconsistencies in COMIS data across domains.

Data Entry Timelines and Snapshot

We initially ran this data query in May 2024, finding that, among institutions identified in the Rising Scholars Network as providing education to incarcerated learners, four showed zero carceral enrollment according to publicly available COMIS data. However, when we reran the same data query in August 2026, all institutions had carceral enrollments in the system. While most colleges had the same number of students at both timepoints, several had gained or lost students between the two data queries. In most cases colleges gained students, with a net enrollment growth of 1,960 students across the system between timepoints.

Challenge #1: Data entry timing hampers real-time enrollment estimates

This exercise helped us understand that there can sometimes be a perceived data entry delay from the perspective of a data user, with institutions finalizing enrollment numbers in subsequent semesters. Individual institutions follow unique data reporting schedules, meaning a snapshot download of publicly available data may be incomplete at any given time, despite reporting institutions being completely within the expected bounds of regular reporting schedules[4]. Considering these data real-time estimates of incarcerated student enrollments is not fully accurate; these figures can be unreliable because reporting institutions rectify reporting with post-census enrollment, limiting researchers to retrospective analysis of enrollment and completion trends. Data users should inquire with technical assistance at the Chancellor's Office about their reporting timelines before treating data snapshots as final.

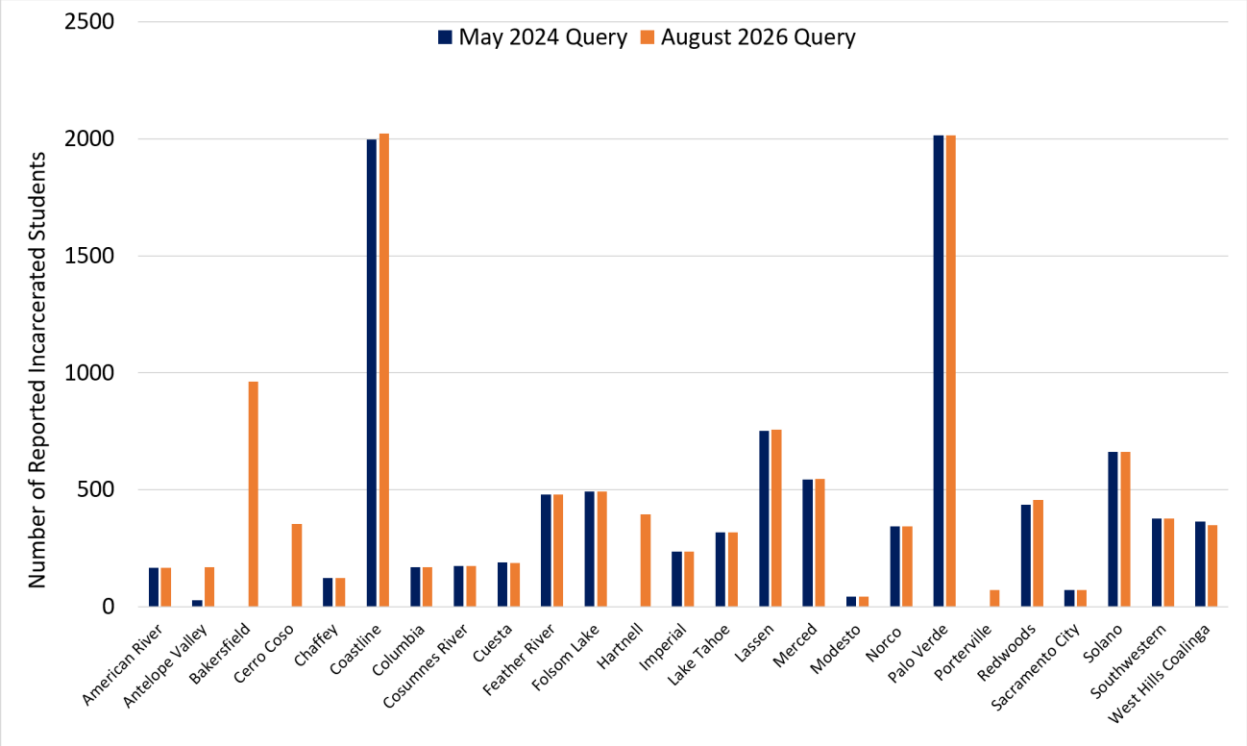
---

[2] This phenomenon is further explored in Challenge #3 later in this memo.

[3] The Rising Scholars Network is a statewide strategy to serve incarcerated students. https://risingscholarsnetwork.org/about-us/

[4] Reporting censes can be delayed for appropriate reasons, for example facility lockdowns impacting enrollment recording.

Figure 1| Fall 2023 Enrollments of Formerly and Currently Incarcerated Students in California Community Colleges, queried in May 2024 and August 2026



Note: This data was sourced from queries made using California Community College Chancellor's Office Management Information Systems (COMIS) Data Mart: https://datamart.cccco.edu/Services/Special_Pop_Count.aspx

Major discrepancies across COMIS domains

Data from COMIS are organized under a set of domains, each of which holds a different set of variables at different levels of observation.[5,6] To conduct analysis of trends in access to courses, course enrollment, and course completion among incarcerated learners compared to their non-incarcerated peers, data across several domains are needed, including from the "Student College Enrollment" and "Student Special Populations" domains. The former holds college enrollment records and the latter holds student characteristics, including carceral status. These two domains *should* successfully merge using student ID, institution ID, and term, with little to no unmatched observations remaining. However, when our team attempted to merge domains, several challenges were encountered, each of which presents a major limitation for analysis.

[5] We are using anonymized student-level data to calculate unique enrollment in and completion of course sections. RFA has access to this student-level data through an existing data use agreement with COMIS.

[6] Details on the COMIS domains relevant to this memo can be found in Section IV of Appendix B.

Challenge #2: Incarcerated students often lack complete enrollment data

First, over four million student records in the special populations domain file don't appear in the enrollment file, or in the financial aid, goals, or demographic domain files.[7] These may be active students or students only partly recorded, and this can undermine accurate counts of incarcerated students. This may be due in part to the many changes over time in the administrative process for coding these students, in response to legislative requirements. Researchers treating the college enrollment domain as the primary source on which to build the analytic sample would reasonably drop all records that did not appear in the enrollment domain. Doing so excludes both incarcerated and non-incarcerated learners at roughly the same magnitude. However, it brings into question the accurate recording of incarcerated learners across all domains within the COMIS central database, calling into question the reliability of systemwide data to accurately represent incarcerated students. This, in turn, introduces serious limits to research that can be conducted on this population. For example, without consistent demographic data on these students, it is impossible to compare student outcomes across student subpopulations, limiting our ability to explore equity gaps. Treating the college enrollment domain as the primary source on which to build our analytic sample, we moved forward by dropping all records that did not appear in the enrollment domain.

Challenge #3: Incarcerated students may be inconsistently flagged in the data

Furthermore, when we attached each student's incarceration flag to their section enrollments, 2645 sections across the seven years included both flagged and unflagged students, representing about 1% of our section sample. These mixed sections involved 7,385 students, about 2% of our student sample. Carceral sections are generally either closed to the general student population or held at a location and time inaccessible to non-incarcerated students making a mixed-status section improbable, suggesting that the discrepancy reflects errors in the data. This may be related to inconsistent practices at the institution level. Some colleges flag carceral status for each student individually, whereas others – likely those with limited program staff capacity – assign carceral status based on whether they are enrolled in a designated carceral sections, without necessarily updating each student's individual flag. The 7,385-student discrepancy is most likely due to under-identification of incarcerated students rather than misclassification of non-incarcerated students.[8]

Since COMIS flags students, not sections, and since there were issues with sections with mixed enrollments, we had to decide when to count a section as carceral. We counted a section as carceral if at least one enrolled student in that section was flagged as incarcerated. This rule assumes that unflagged students in those sections are incarcerated but were missed, as discussed above. Because

[7] Details on the misalignment found between the domains and patterns seen in those misalignments can be found in Section I of Appendix A.

[8] More details on the extent of this discrepancy can be found in Section I of Appendix A.

a single flagged student is enough, this rule likely produces more carceral sections than a stricter one would. This is a small but clear discrepancy that may have implications for our subsequent analyses in that **we are assuming that students not flagged as incarcerated that are enrolled in sections with incarcerated students are actually incarcerated but were incorrectly flagged.**

A solution that surfaced additional limitations with COMIS

As a second way to identify carceral enrollments, we asked colleges for the course sections they offered in prisons and matched those sections to COMIS enrollment records. This approach assumes that students enrolled in a prison section were incarcerated at the time, and that colleges' lists of prison sections are complete. Specifically, we sought to collect course- and section-level data from individual institutions on all community college course sections available to incarcerated students from Fall 2017 through Spring 2025.

We contacted Rising Scholars program staff at the 24 community colleges that partner with state and federal prisons.[9] In our initial outreach, we learned that the most efficient and least burdensome approach to this data collection was to request section identifiers directly from the Rising Scholars program staff. We requested course names, course identification numbers, section identification numbers, and term data for all courses offered to incarcerated students in state or federal prisons.[10] No student-level data were requested or shared by colleges.

RFA received all requested years of prison section data from most, but not all, colleges.[11] To connect the prison section data from colleges with administrative course section- and student-level COMIS data, we prepared the data for a successful merge, including initial reformatting and standardizing course and section identifiers, and generating source-specific flags for carceral sections.

After exhausting all matching strategies (described in Section III of Appendix B), the research team was able to match 97% of all sections identified in the institution-provided data to sections in the COMIS data.[12] The remaining 315 sections (3%) did not appear in COMIS after accounting for

[9] The directory used can be found here: https://risingscholarsnetwork.org/program-directory/

[10] For details on how the research team contacted and communicated with institutions to receive requested data, please see Section I in Appendix B.

[11] For details on the data received by the research team, please see Section II in Appendix B.

[12] For details on the matching process, please see Section III in Appendix B.

formatting differences. We don't know why these sections are missing from COMIS, and we excluded them from all later analyses.[13]

Upon merging data files, several additional limitations and discrepancies emerged that are consequential for data analysis.

Challenge #4. State systems may underreport a small share of carceral sections.

The undermatching surfaces a small, but clear discrepancy between the two data sources prior to investigating carceral status. Institutions provided us with details of sections that, according to the COMIS data, did not exist, despite RFA's best attempts to account for formatting differences between the two sources. This means, according to RFA's research, that **the number of course sections being offered to incarcerated students may be underreported to the state's primary data collection system.** For analysis, this would mean that we lack complete information on the number and content of course offerings currently being offered to incarcerated students.

Challenge #5. Some incarcerated student credits may not be formally recorded in state data systems.

Further, this issue points to a very real possibility **that incarcerated students may be participating in coursework that is not being formally recorded in the community college system's centralized database,** which calls into question whether credits attempted and earned by incarcerated students are being accurately counted.

Challenge #6. Estimates of course offerings differ depending on the data source used

The number of course sections identified as taking place in a carceral institution was slightly smaller according to the Institution-Provided data, compared to the COMIS data; however, these differences are minimal and stable across years (Table 1). Across both data sources, we can see a gradual increase in the number of carceral sections identified, though this is likely due to the variation in the number of colleges included in each year (Table 2). Over 98% of all sections were successfully identified as either carceral (3.9%) or not carceral (94.4%) across both data sources, with discrepancies only existing in approximately 1.6% of all sections (Table 1).

[13] Unmatched sections, which represent sections in Institution-Provided data that did not exist in the systemwide COMIS data, were unable to be used and dropped from all subsequent analyses. It is not clear why these courses would be absent from the COMIS data.

Table 1. Section carceral status discrepancies across data sources

		Institution-Provided Data	
		Carceral Section	Non-Carceral Section
COMIS data	Carceral Section	8,754 (3.9%)	3,198 (1.4%)
	Non-Carceral Section	357 (0.2%)	209,379 (94.4%)

Table 2. Number of course sections identified as taking place in a carceral institution, by year and data source

	# Carceral Sections		
School Year	Institution-Provided Data	COMIS data	Difference
2017-18	1,040	1,269	229
2018-19	1,141	1,535	394
2019-20	1,252	1,765	513
2020-21	1,408	1,785	377
2021-22	1,470	1,899	429
2022-23	1,396	1,796	400
2023-24	1,404	1,903	499

Table 1 shows that over 98% of sections were identified as either carceral or non-carceral in both datasets. 0.2% of sections identified as carceral in the COMIS data were identified as non-carceral in the Institution-Provided data, and 1.4% of sections identified as carceral in the Institution-Provided data were identified as non-carceral in the COMIS data. However, out of all sections identified as carceral in at least one of the two sources (12,309), only about 71% (8,754) were identified by both sources as carceral. The remaining 29% of sections (3,555) were marked as carceral in one of the datasets, but not the other.

These sections, without a definitive carceral status, cause estimates of course offerings to differ depending on the data source used. Because of this, the choice in which source to use can notably affect what carceral course offerings look like to researchers and practitioners.

Challenge #7. Estimates of incarcerated students differ depending on the data source used

We also see slightly elevated counts of incarcerated students in the COMIS data compared to the Institution-Provided data in each year of data except 2017-18, though these differences are once again minimal. 96.7% of all students were consistently identified as either incarcerated or non-incarcerated across both data sources, with discrepancies affecting only 3.3% of the sample. Contrary to the stable accuracy of the section-level identification, we see a clear increase in the accuracy to which both sources report the same students as incarcerated or non-incarcerated over time, increasing from 92.7% in 2017-18 to 98.8% in 2023-24. Table 3 below shows the number of unique students per year flagged as currently incarcerated according to each data source.

Table 3 | Number of students identified as incarcerated learners, by year and data source

School Year	# Incarcerated Students According to Institution-Provided Data	# Incarcerated Students According to COMIS data	Difference
2017-18	3,279	3,271	-8
2018-19	2,963	3,793	830
2019-20	3,324	4,042	718
2020-21	2,893	3,151	258
2021-22	2,942	3,191	249
2022-23	2,968	3,282	314
2023-24	3,269	3,443	174

Note: "COMIS data" in this table refers to the analytic sample constructed by merging student special populations data with student college enrollment data, and dropping students who did not appear in the college enrollment data. This is discussed in more detail in Challenge #2.

Table 4 below shows how the carceral status of students identified by COMIS data differs from the carceral status identified by Institution-Provided data.

Table 4 | Student incarceration status discrepancies across data sources

		Institution-Provided Data	
		Incarcerated	Non-Incarcerated
COMIS data	Incarcerated	16,871 (4.6%)	7,302 (2.0%)
	Non-Incarcerated	4,767 (1.3%)	340,871 (92.2%)

Note: “COMIS data” in this table refers to the analytic sample constructed by merging student special populations data with student college enrollment data and dropping students who did not appear in the college enrollment data. This is discussed in more detail in Challenge #2.

In Table 4, we see that nearly 97% of students were identified as either incarcerated or non-incarcerated in both datasets. 2% of students identified as incarcerated in the COMIS data were identified as non-incarcerated in the Institution-Provided data, and 1.3% of students identified as incarcerated in the Institution-Provided data were identified as non-incarcerated in the COMIS data. However, in looking at all students who were flagged as incarcerated by at least one data source (28,940 students), only 58.3% showed matching carceral statuses across both data sets. This leaves 41.7% of these students flagged as incarcerated in one data source, but not the other. This discrepancy is very large and troubling, especially when thinking about subsequent analyses of trends among incarcerated students. **If nearly 42% of students flagged as incarcerated are not established as such across both data sources, this can lead to major differences in estimates of enrollment and completion depending on the source used.** The result of this discrepancy is **two data sources purporting to identify the same population of incarcerated students, but which actually capture different samples of these students.**

Discussion

In this memo, RFA sought to capture the level of accuracy to which incarcerated learners are being recorded and tracked as they interact with the California Community College System. The quality of this data is not only necessary for administrative purposes, but to ensure rigor in research conducted on the experiences and outcomes of incarcerated learners. As RFA seeks to answer questions about incarcerated learners’ educational experiences in the context of AB 705, it was important for us to also assess our ability to do so based on the data available to us.

One of the early factors that brought this work into existence was a data query of publicly available COMIS data conducted by RFA in May 2024. The results showed four institutions in the Rising Scholars Network appeared to have zero incarcerated students enrolled. When this same query was done again over two years later in August 2026, all Rising Scholars Network institutions had carceral enrollments. Additionally, enrollment figures across the system changed between timepoints, with the system adding 1,960 incarcerated students between timepoints. This was a key exercise in uncovering the existence of a data lag in the COMIS data system for the research team.

Matching course sections across data sources was a challenge due to inconsistencies in section numbering. Most differences in section numbers across sources were able to be reconciled, however some sections in the institution-provided prison section data did not exist in the COMIS data, suggesting to us that the number of course sections offered to incarcerated students may be

underreported to the state's primary data system. This would mean that incarcerated students could be enrolled in course sections but those course sections are not being formally recorded in the COMIS database.

A close examination of the COMIS data found that incarcerated students whose records appear in the Student Special Population domain did not appear in domains holding student financial aid status, educational goal, or demographics, bringing the accurate reporting of all incarcerated students into question across the data system. This lack of data coverage for incarcerated students could suggest that data for those students are not being recorded completely in COMIS. It could also mean the records don't belong to currently enrolled students. The accurate and reliable records of incarcerated students in the COMIS data are in question, which could have implications for institutional administrators and policymakers seeking to develop policy related to serving incarcerated students or future research examining education for incarcerated students. Based on the data that are currently available, we cannot determine whether these are active students. Further, if they were to be active students, there would not be demographic information linked to those incarcerated students. Incarcerated students are already a uniquely marginalized population; if a substantial portion of them do not have complete and accurate demographic information recorded, data cannot be disaggregated to examine issues of equity for those students. Our understanding of the COMIS data points towards the most likely cause of the discrepancy to be under-identification of incarcerated students rather than misclassification of non-incarcerated students. A discrepancy like this could mean the state not does have an accurate and timely understanding of incarcerated enrollment.

When looking across datasets, just under 30% of course sections were indicated as being offered to incarcerated students in one data source, but not the other. Without a definitive indication of carceral status at the section-level, the estimates of course offerings for incarcerated learners can differ based on the source used. This finding is particularly important for policymakers, researchers, and institutional administrators because the source of the data can have notable effects on the conclusions a user makes on carceral course offerings; the choice of data source can lead to differing – and possibly inaccurate – conclusions about carceral course offerings.

There is a large discrepancy between data sources in how incarcerated students are flagged; 41.7% of students are flagged as incarcerated in one source and not the other. A major discrepancy like this would not only have implications for stakeholders seeking to understand course offerings, but it would lead to significant differences in estimates of enrollment and completion depending on the source used. Put plainly, both data sources are intended to measure the same student population, but neither can be reliably assumed to hold an accurate count of incarcerated students. Further, we found cases of incarcerated students showing enrollment records in the same section as non-incarcerated students. This only affects 1.9% of our student sample but clearly indicates either over- or under-identification of incarcerated students within the data.

Overall, the question of which data source more accurately identifies and tracks incarcerated students within the California Community College system remains. Despite not being the central focus of this analysis, the within-source discrepancies emerge as a more pressing threat to data validity and likely contribute to the across-source discrepancies. Though the limited ability of complete data from these institutions will impact our assessment of student outcomes, it has little impact on our assessment of between-source discrepancies, as evidenced by the 97% match rate. The COMIS database is the primary source of information on course offerings, student enrollment, and student outcomes trends and is widely utilized by administrators, researchers, and policymakers, therefore any systematic or structural issue with the data's integrity or accuracy, no matter how small, has large implications for the field.

Conclusion

The research is clear on the value of postsecondary education for incarcerated learners. However, obtaining national estimates on the academic outcomes of incarcerated students is notoriously challenging, in part because these students are often tracked by multiple mission-misaligned systems, and because programs serving this population are chronically understaffed. Our team approached this task leveraging data provided by institutions as well as the state administrative data, finding significant misalignment which challenges our ability to state with confidence that we have reliable estimates of carceral enrollment. Without being able to rely on enrollment estimates, we are unable to track students through the higher education system, to see how they fare in their coursework, where they may need additional support, and observe their retention and completion patterns. We also found that students experiencing incarceration may not be fully captured in demographic data, hampering our ability to consistently identify patterns among student subpopulations or equity gaps.

Educating incarcerated learners is an equity mandate. We must invest adequately in data infrastructure, training, and personnel to ensure that state investments and partnerships supporting these students have what they need to reliably document student enrollment and progress. California has built an impressive network of programs serving this population, and has invested state funds to support these programs. In a time of unprecedented growth, however, these programs remain underfunded and understaffed, straining to consistently deliver education to an expanding number of students, which makes accurate record-keeping more challenging in an already-challenging environment.

Appendix A. Details on Findings

Section I. Within-Source Discrepancies

Table A1. COMIS discrepancy between student college enrollment domain and student characteristics domain

	In College Enrollment Domain Only	In Special Populations Domain Only	In Both Domains
Incarcerated Students	X	92,357	95,231
Non-Incarcerated Students	X	4,081,667	5,194,710
Unknown Carceral Status	18,001,867	X	X

Approximately 188,000 unique students in the "Student Special Populations" domain were flagged as incarcerated learners, but only approximately 95,000 of those students were also found in the "Student College Enrollment" domain (shown in Table A1 above). We know from our previous work with the COMIS data that some domains may hold records of inactive students, which result in observations from those domains not appearing in the "Student College Enrollment" domain. However, the absence of nearly half of all students flagged as incarcerated learners in the "Student College Enrollment" domain is too large a discrepancy to be explained by this alone. It is important to note that we see the same pattern with students not flagged as incarcerated learners, in which approximately half were not found in the "Student College Enrollment" domain, suggesting that this discrepancy affects incarcerated and non-incarcerated students at a similar magnitude. We know that COMIS domains holding student characteristics do not always represent all enrolled students, sometimes only containing students who satisfy the characteristics being reported on. Therefore, we look primarily to the unmerged observations from the "Student Special Populations" domain to investigate patterns in this discrepancy. Though the discrepancy is seen across all terms and institutions, we can see some patterns:[14]

1. By academic year, the number of unmerged observations from the "Student Special Populations" domain rises steadily across all eight terms, from about 329,000 to 708,000. This increase can be largely attributed to the growth in total enrollment over the same period rather than a worsening discrepancy, so the rate of unmatched student characteristics observations stays relatively stable across terms even as the raw count grows.
2. By institution, the share of each college's "Student Special Populations" records that failed to match to "Student College Enrollment" records varies widely across colleges, from about 2% to 38%. High-discrepancy colleges include Cosumnes River (38%), Sacramento

[14] The discrepancies outlined below are reported in full detail in Tables A1 and A2 in the Appendix.

City (37%), and American River (35%). Low-discrepancy colleges include Modesto (2%), and Chaffey (5%). This pattern points to a variation in institution-level reporting accuracy rather than a systematic or time-based cause for this discrepancy.

One structural explanation for this gap is that the incarceration flag is not necessarily re-set each term. Our understanding is that once a student is coded as incarcerated in the "Student Special Populations" domain, the flag persists with the student across terms, including terms in which the student is not actively enrolled. Students in carceral settings transfer between facilities, and subsequently between the colleges serving those facilities, more frequently and unpredictably than the general student population. A student who is transferred mid-term, moved between multiple institutions, or experiencing fragmented access to educational services at any point over the course of their sentence may therefore appear in the "Student Special Populations" domain for a given term without a corresponding "Student College Enrollment" record for that same term. This also may explain some of the fluctuation in percentage of students without "Student College Enrollment" data across institutions: colleges serving facilities with higher rates of student transfer would be expected to show a larger share of records that appear in the "Student Special Populations" domain only.

This partially supports our hypothesis of student characteristics domains, such as the "Student Special Populations" domain, being cumulative rather than term specific. Our understanding is that the COMIS Incarcerated Student Flag is tied to a student's current incarceration status and is only updated once a student is released, not updated based on whether a student is enrolled in college coursework during each term.

Table A2. COMIS discrepancies between students' carceral statuses within a single course section

All Incarcerated Students	Incarcerated Students in Sections with Non-Incarcerated Students	All Non-Incarcerated Students	Non-Incarcerated Students in Sections with Incarcerated Students	All Sections	Sections Containing Both Incarcerated and Non-Incarcerated Students
17,860	2,546 (14.3%)	284,751	2,379 (0.84%)	241,862	2,645 (1.09%)

Appendix B. Details on Data Collection and Processing

Section I. Institutional Data Requests

Emails were sent to colleges requesting the prison course section data. Initial data requests were sent to colleges in April 2025. Follow-up outreach to obtain data included emails, phone calls, and voicemail messages. The range in intensity of outreach required was varied. Table B1 below summarizes the methods of outreach undertaken by the research team.

Only four colleges shared the data after just one email request; RFA sent at most ten email requests to one college. A total of 100 email requests (i.e., initial outreach and follow-up requests) were sent across all 24 colleges to collect the data identifying course sections offered to incarcerated learners. On average, 4.2 emails were sent to colleges before data were received. The average time from initial request to data receipt was 3 months.

Table B1. Prison section data request outreach

Email correspondence	
Total data request emails sent:	100
Average number of data request emails sent per college:	4.2
Phone correspondence	
Total data request phone calls/voicemails:	13
Timeline to data submission	
Overall data request timeline:	April 2025 – August 2025
Average timeline from initial request to data submission	3 months

Section II. Data Collection Details

RFA received all requested years of prison section data from most, but not all, colleges. Table B2 below summarizes the institutions from which we received any data, and Table B3 summarizes the years of data we received from colleges that *did not* provide all years of requested data.

Table B2. Results of Institution-Provided Data Collection

Sampled College	RFA received data from this college	RFA did not receive data from this college
American River College	X	
Antelope Valley College		X
Bakersfield College	X*	
Cerro Coso Community College	X	

Chaffey College	X	
Coalinga College	X	
Coastline College	X	
College of the Redwoods	X	
Columbia College	X	
Cosumnes River College	X	
Cuesta College	X	
Feather River College	X	
Folsom Lake College	X	
Hartnell College		X
Imperial Valley College	X	
Lake Tahoe Community College	X	
Lassen College	X	
Merced College	X	
Modesto Junior College	X	
Norco College	X	
Palo Verde College	X	
Sacramento City College	X	
Solano Community College	X	
Southwestern College	X	

Note: Data was received from Bakersfield College, however additional verification from the college was required to determine which sections were offered in prisons. Several unsuccessful attempts were made to verify the data RFA received.

Table B3. Prison section data received from institutions, by term, from institutions that did not provide all years of data

College	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23	2023-24
American River					X	X	X
Chaffey	X	X	X	X	X		
Cosumnes River				X	X	X	X
Feather River					X	X	X
Lake Tahoe	X	X	X	X	X		
Modesto						X	
Sacramento City				X	X	X	X

Most colleges provided the requested data. Colleges shared the data directly with RFA via email. Of the 24 colleges which received data requests, we obtained usable data from 21 colleges. Some included additional data, for example, instructional modality and information about the carceral facilities wherein courses were offered. However, those indicators were not consistently captured in the data requests across each college.

The data RFA received was generally of high quality and fidelity to the parameters of the data request. In some cases, additional correspondence was needed to confirm the contents of data files. For example, some colleges repeated course section numbers over terms while most others do not. In cases of repeating course section numbers, additional data verification steps were conducted with the colleges to confirm the accuracy of the course section data before combining it with COMIS data.

College contacts typically compiled prison section data using internal collection systems and provided this data to us in the form of spreadsheets. However, some college contacts provided data in other formats or provided incomplete data. For example:

- **Prison section data were shared as screenshots.** A staff person at one college provided a screenshot of their internal view of the college's web-based course catalog, including the course and section identification data of courses offered to incarcerated students. RFA transposed data from that screenshot into spreadsheet format for processing.
- **Prison section data from another college was shared via PDF.** The data were limited to courses offered to incarcerated learners and formatted as a course catalog. These data were then transposed for processing and analysis by RFA.
- Staff at a third college reported that **limited capacity was a barrier** to fulfilling the data sharing request. RFA only received prison section data from Fall 2017 through Summer 2022 in weekly installments.

Section III. Data Manipulation and Merging Details

We treated the format of the COMIS data as the standard, adapting the prison section data to reflect this format. The following steps outline the initial reformatting done to all prison section data to prepare it for merging with the COMIS data:

Adding college ID codes to prison section data: The prison section data included the colleges' names as identifiers; however, COMIS data uses colleges' systemwide ID. Using a crosswalk document that specifies each college's corresponding ID[15], we created a new variable in the data that held this information. This allows us to merge with the COMIS data using the college identification number as an identifier.

Limitations of data completeness for some analyses: Analyses linking student-level COMIS data with prison section data are key to documenting the courses offered to incarcerated students. The share of

[15] College ID crosswalk can be found in the Appendix.

colleges that did not provide data for all requested years (see Table 2 above) will need to be excluded from trend analyses requiring such a link, that show, for example, the arc of course offerings, enrollment, and completion within and across colleges for incarcerated students before and after AB 705.

Reformatting term data to match with COMIS format: Information on the term in which a section was offered came from colleges in various formats (e.g. "FA17", "Fall 2017", "147"). We chose to reformat these values to match the COMIS data's system for recording term (e.g. 147 = Fall 2014, 153 = Spring 2015, 157 = Fall 2015, etc.).

In our attempt to merge the prison section data to COMIS data, we also came across several college-specific discrepancies between the two datasets. The following section describes the steps taken to rectify these discrepancies when combining datasets.

- Several colleges had **systematic differences in how the course identification number was formatted** in one dataset compared to the other; manually aligning these values would have been cumbersome. For colleges where many or all of their course numbers were formatted differently in the prison section data than the COMIS data, the decision was made to merge these two datasets based on course section number and term only. Before moving forward with this strategy, it was confirmed that these colleges' section IDs were unique at the course and term level, so they were sufficiently capable of identifying individual sections offered. The following colleges were merged using this strategy: American River, Consumnes River, Sacramento City, Folsom Lake, and Cerro Coso.
- The prison section data we received from Cuesta College had **incorrect term values**, but we confirmed that their sections could be uniquely identified in the COMIS data using only course and section numbers. Therefore, the prison section data and COMIS data for Cuesta College were merged only using section and course numbers, after confirming that these values identified individual course/section offerings in the COMIS data.
- In their prison section data, Imperial Valley College and Modesto College provided **course numbers only containing numeric characters** (e.g. "100" instead of "ENGL100"). For these colleges, we dropped the alphabetic characters from the COMIS dataset's course numbers so they could be successfully merged with the prison section data. We confirmed that the numeric course numbers could still identify individual courses prior to making this edit.
- Chaffey College's prison section data included **section identification numbers that did not contain any leading or trailing zeros**; however, their section numbers did contain these zeros in the COMIS data. We chose to drop the leading and trailing zeros from the COMIS values to ensure a successful merge with the prison section data.
- In the COMIS dataset, Lake Tahoe College's **term values were even** (e.g., 178 instead of 177 for Fall 2017). To allow for this data to merge with the prison section data, which contained odd term values, we added 1 to the term values in the prison section data. We are confident that the adjusted term values still reflect the same term as the COMIS data, because we know that, for example, 178 is also a fall semester for some colleges. Additionally, Lake Tahoe College's section

number values were “Y1”/”Y2”/”Y3” in their prison section data, but “01”, “02”, “03” in the COMIS dataset. These values were edited to allow for one dataset to merge with the other.

- English course identification numbers for Feather River College were **named slightly differently between the two datasets**. In the COMIS data they were structured as “ENGLXXXX”, whereas the format in the prison section data was “ENGXXXX”. Observation-level edits to the course identification numbers in the prison section data were implemented to allow the data to successfully merge.

Table B4 below shows the results of the research team’s attempts to match sections from the institution-provided data to sections in the COMIS data. The remaining unmatched sections are only a concern for the following four colleges, in which less than 94% of sections merged:

- Chaffey (79.2% merged)
- Cuesta (62.8% merged)
- Imperial Valley (68% merged)
- Modesto (60% merged)

Table B4. Prison section and COMIS data merge results

College	Number of Prison Sections	Number of Merged Prison Sections	Percent of Prison Sections Merged
American River	48	48	100.0%
Coastline	546	546	100.0%
Columbia	355	355	100.0%
Cosumnes River	39	39	100.0%
Lassen	300	300	100.0%
Norco	414	414	100.0%
Sacramento City	24	24	100.0%
Lake Tahoe	605	604	99.8%
Palo Verde	3,394	3,383	99.7%
Southwestern	287	286	99.7%
Solano	543	539	99.3%
College of the Redwoods	521	517	99.2%
Feather River	121	120	99.2%
Merced	649	639	98.5%
Folsom Lake	249	240	96.4%
Cerro Coso	1,433	1,374	95.9%
Coalinga	310	291	93.9%
Chaffey	258	194	75.2%
Imperial Valley	356	242	68.0%
Cuesta	43	27	62.8%

Modesto	5	3	60.0%
Total	10,500	10,185	97%

Section IV. Data Discrepancies Details

The COMIS data domains relevant to this analysis are as follows:

1. "Student College Enrollment" domain: This domain contains one observation per student per year of enrollment in a California community college. We use this domain as our basis for constructing our analytic sample of enrolled students.
2. "Student Course Enrollment" domain: This domain contains one observation per student per course section they are enrolled in. We use this domain to identify which course sections students were enrolled in for each year of college attendance.
3. "Course Section" domain: This domain contains one observation per course section offered at California community colleges. We use this domain to connect Institution-Provided data to student-level records, as well as to pull descriptive characteristics from courses such as transfer status and subject in subsequent analyses.
4. "Student Special Populations" domain: This domain contains one observation per student per year and holds flags for certain student characteristics such as economically disadvantaged status, special education status, English language learner status, and carceral status. We use this domain to capture carceral status and add it to our analytic sample built from the "Student College Enrollment" domain.

Section V. Across-Source Discrepancy Details

Table B5 below displays the final analytic dataset after combining both data sources. The number of institutions included each year varies due to incomplete Institution-Provided data, which in turn varies the number of sections and students included in each year.

Table B5. Final analytic sample of colleges serving incarcerated students, by year

School Year	# Colleges	# Sections	# Students
2017-18	16	28,305	48,083
2018-19	16	27,596	47,040
2019-20	16	27,982	46,123
2020-21	18	30,898	49,302
2021-22	20	35,625	55,248
2022-23	18	32,412	52,790
2023-24	19	38,870	71,225