

The Research and Planning Group for California Community Colleges (RP Group) is an independent nonprofit in service to California's community colleges, students, and our institutional research, planning, and effectiveness (IRPE) colleagues. Our work is based on a commitment to equitable student success. As we work to document the impacts of our student success reforms (e.g., AB705), it is critical to ensure such reforms are equitably impactful. However, such data analysis should never come at the expense of protecting students' right to anonymity and privacy.

## Purpose

One of the five critical areas in the [RP Group's Code of Ethics](#) is **Integrity & Authenticity**. In the code, we state:

*Our role as data stewards is one of great responsibility, ensuring trust and a positive culture of inquiry. Within this area is a promise to maintain transparency of data reporting and collection processes while ensuring the confidentiality and privacy rights of students and other participants.*

This area is aligned with the tenants of the Family Educational Rights and Privacy Act ([FERPA](#)), a federal law that, among other things, affords students ages 18+<sup>1</sup> *the right to have some control over the disclosure of personally identifiable information from their education records*. While writing the Code of Ethics, it became clear that the IRPE field would welcome guidance on effective practices for adhering to this promise. As such, we wanted to provide guidance to the IRPE field on to manage data in a way that meets FERPA requirements — notably, how to address data suppression in ways that provide actionable data that can be used to drive decision-making and assess progress towards meeting goals (e.g. Vision for Success), while still adhering to the privacy rights of students and other participants.

The following guidance sets in writing a set of recommended effective practices and principles for the external sharing of data and results that reflect the link between the RP Group's mission, vision, and values, and the day-to-day decisions and behaviors of our staff, Board, and consultants. Our hope is that IRPE professionals throughout the system will find this information useful as they attempt to navigate the issues around privacy and usefulness.

## The Importance of Data Suppression

The IRPE field recognizes the critical importance of providing actionable, meaningful data, yet also understands that, first and foremost, we must do so in a way that protects the privacy rights of the subjects involved. A failure of the latter would result in an inability to do the former. While there are likely many reasons a researcher may opt to suppress data, as it relates to the work IRPE professionals do, **protecting participant anonymity** and **promoting reliability** clearly rise to the top.

- **Protecting Privacy** - While we typically think of data results as being anonymous if it does not include a singular, unique identifier (e.g., name, student ID, social security number), when enough variables are presented together and/or a sample size is small enough, data that would typically not be considered as identifying suddenly becomes so. For example, in a course with 30 students being surveyed on satisfaction, a student's "anonymous" response is no longer anonymous when results are parsed out by race and gender (e.g., data for Latinas aged 18-24).

---

<sup>1</sup> And parents of students under age 18

- **Promoting Reliability** - Unless a survey sample is incredibly large, results presented are typically descriptive in nature (e.g., percents, n-counts). The onus is on the reader to determine whether a given result is noteworthy/action-worthy and there is a degree of subjectivity in the interpretation (e.g., one individual may think a five percentage point difference between groups is noteworthy while someone else may not). However, when the n-count of a given group presented in a result is incredibly small, most researchers would agree that the results should not be interpreted in a way to compare across groups as a statistical test would almost definitely not reveal significance. Suppressing those data prevents the reader from having to make that decision for themselves.

## Recommended Practices in Data Suppression

The following recommendations are based on a combination of FERPA<sup>2</sup> regulations, California *State ESSA Plans*, and our own experience working with data across the CCC. These recommendations are put forth in response to common issues we have seen in researchers' attempts at dealing with small sample sizes.

### RECOMMENDATIONS FOR PROTECTING PRIVACY

- **Recommendation 1:** Suppress n-counts and results for groups where there are 10 or fewer individuals<sup>3</sup> AND<sup>4</sup>
- **Recommendation 2:** When suppressing data in a given table cell, one must also suppress the totals from any row or column, including a suppressed value to ensure that a reader cannot calculate the individual cell numbers, thus breaking the anonymity that Recommendation 1 is designed to protect.<sup>5</sup>

### RECOMMENDATION FOR PROMOTING RELIABILITY

- **Recommendation 3:** When there are substantially different sized groups, yet cell sizes are large enough to not need suppression (> 10), include a standard footnote encouraging the reader to use caution in interpreting findings.

---

<sup>2</sup>The Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99) is a Federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education.

<sup>3</sup>Though see Q&A section below for alternatives to suppression when looking to ensure all voices are presented in the data

<sup>4</sup>It is important to note that Recommendations 1 and 2 must be used in conjunction with one another as anonymity can be violated if the two methods are used independently.

<sup>5</sup>Complimentary suppression (suppression of all values in the row) is another option, though it further limits the conclusions that can be drawn from the data as a whole.

Table 1 presents examples of each recommendation, color-coded as indicated above.

**TABLE 1. DISTRIBUTION OF TRANSFER STATUS BY RACE/ETHNICITY**

Race/Ethnicity	Near the Gate		At the Gate		Transfer Achievers		Total
	%	#	%	#	%	#	#
African American	12.2%	130	6.8%	65	81.2%	920	1,115
Asian American	8.7%	94	11.2%	102	82.1%	890	1,086
Filipino	7.1%	180	10.4%	255	82.7%	1,951	2,386
Hispanic/Latino	12.7%	544	14.1%	589	73.7%	3,165	4,298
Native American and Alaska Native	s	<11	s	<11	61.7%	21	<b>Total s</b>
Pacific Islander	8.7%	14	10.6%	17	80.7%	130	161
Two or More	8.3%	88	16.1%	166	76.5%	805	1,059
Unknown	s	<11	15.1%	19	61.4%	79	<b>Total s</b>
White	12.1%	1,498	11.1%	1,453	76.8%	4,157	7,108
Total	<b>Total s</b>	<b>Total s</b>	<b>Total s</b>	<b>Total s</b>	77.8%	12,118	<b>Total s</b>

**Note 1:** s = suppressed due to cell size < 11

**Note 2:** Total s = Total suppressed: row and column totals are suppressed when any cell in that row or column is suppressed to prevent the ability to calculate those cell numbers.

**Note 3:** Please note that the groups in this set are significantly different from one another. Given the small sample size of one of these groups relative to the rest, we recommend taking caution when interpreting results.

## For more information, see the following FAQs:

**Q. At what sample size should I suppress my data? Does FERPA have guidelines?**

**A.** While FERPA does not recommend a particular cell size under which data should be suppressed, we recommend suppression of both n-counts and associated data (e.g., percentages, averages) reported if the cell size is 10 or fewer individuals. Under the Every Student Succeeds Act ([ESSA](#)), each state must determine a set minimum number of students to present data on a given subgroup for federal accountability and reporting purposes. In California, this cell size for reporting assessment data is **more than 10**. While ESSA is focused on elementary and secondary education, our recommendation is based on this decision. However, in instances where the cell size is slightly larger than 10, but is substantially different in size from other groups, we recommend including a footnote reminding the reader to be cautious when interpreting results (See Recommendation 3 above).

**Note:** When suppressing data in a given table cell, you must also suppress the totals from any row or column, including a suppressed value (see Recommendation 2).

**Q. I'd like to present data on a cell size of 10 or fewer. How can I be sure my data have been de-identified enough to make it truly anonymous?**

**A.** While there are no hard and fast rules, we encourage researchers to adopt FERPA standard, which **states:**

*The FERPA standard for de-identification assesses whether a “reasonable person in the school community who does not have personal knowledge of the relevant circumstances” could identify*

*individual students based on reasonably available information, including other public information released by an agency, such as a report presenting detailed data in tables with small size cells (34 CFR §99.3 and §99.31(b)(1)). The “reasonable person” standard should be used by State and local educational agencies and institutions to determine whether statistical information or records have been sufficiently redacted prior to release such that a “reasonable person” (i.e., a hypothetical, rational, prudent, average individual) in the school community should not be able to identify a student because of some well-publicized event, communications, or other similar factors.*

— The U.S. Department of Education’s Privacy Technical Assistance Center (PTAC)

Please note that while FERPA guidelines go on to **state**:

*School officials, including teachers, administrators, coaches, and volunteers, are not considered in making the reasonable person determination since they are presumed to have inside knowledge of the relevant circumstances and of the identity of the students*

— The U.S. Department of Education’s Privacy Technical Assistance Center (PTAC)

As it relates to sharing data with entities (e.g., a teacher, administrator, etc.), we suggest one should also consider the inside knowledge they have and adjust accordingly to preserve anonymity.

**Note:** In our highly virtual world of late, it is important to consider everything shared online (even if only ‘live’ via a webinar) as public information given the technology available for capturing and sharing information (e.g., screenshots, unauthorized recording, photos, etc.).

**Q. My group size is >10 but is still substantially smaller than all the other groups in the set. Should I still suppress to ensure the reader does not infer more than they should about the results of that group?**

**A.** No - Suppressing data due to sample size differences could ultimately result in unnecessarily subjective decisions. Instead, we suggest including the following footnote:

Please note that the groups in this set are substantially different from one another. Given the small sample size of one of these groups relative to the rest, we recommend taking caution when interpreting results.

**Q. Are there any alternatives to suppression that allow the reader to get a general sense of the results, and ensure all students are included in equity decisions, while still protecting student confidentiality?**

**A.** Yes. However, the method you select should be based on the goal you have for your reader. Always err on the side of being thoughtful over thorough. If data are manipulated in any way, make sure that it is clear in a footnote. Possible alternative methods include:

- Using ranges
- Rounding
- Using top and bottom coding (e.g., <5%, >95%)
- Combining two or more smaller groups into one category (e.g., ‘other’)<sup>6</sup>

For more information on alternative techniques, [see this guidance from the Institute of Education Sciences \(IES\)](#)<sup>7</sup>, as well as [this information from the Data Quality Campaign](#).<sup>8</sup>

<sup>6</sup> This method helps ensure all students are included in equity decisions/students in traditionally small groups (e.g., Native American students) do not wind up always being overlooked. However, the actionability of the results can be limiting.

<sup>7</sup> The primary research arm of the US Department of Education (DOE), comprised of four “National Centers” devoted to supporting and disseminating scientific research related to education.

<sup>8</sup> A national, collaborative effort to encourage and support state policymakers to improve the collection, availability, and use of high-quality education data, and implement state longitudinal data systems to improve student achievement.

**Q. How do you apply data suppression when open-ended survey questions are involved?**

**A.** When sharing de-identified student-level data sets with another group:

- Decouple open-ended comments from quantitative data to ensure one could not identify the individual using a combination of data points (e.g., demographics) and information provided in the comments
- Redact elements in open-ended comments that could potentially reveal the identity of the respondent or the subject of the comment (e.g., names, titles, positions)

**Q. Do you have any recommended reading to learn more on the topic?**

**A.** In general, we strongly recommend that all IRPE professionals take the time to become familiar with the following resources:

- [\*FERPA and the Protection of Pupil Rights Amendment \(PPRA\) overview from United States Department of Education \(USDOE\)\*](#)
- [\*FERPA FAQs from USDOE\*](#)
- [\*FERPA Resources for Postsecondary School Officials\*](#)
- [\*Student Privacy Compass\*](#)