



---

## Masking Guidelines & Techniques

Masking is a general term used to describe methods that limit or hide original values in a data set. Data suppression, recoding, blurring, perturbation, and selective reporting are all forms of masking (NCES, 2010b; NCES, 2011; Privacy Technical Assistance Center [PTAC], 2012). In the case of the UH ERC, masking refers to the purposeful exclusion or removal of information prior to public release to protect individuals under FERPA.

A common situation where masking is needed is within small cells (see pp. 4-5). According to the U.S. Department of Education (2014), FERPA mentions but does not specifically designate a small cell count standard. Rather, states must define minimums within their respective State Accountability Plans approved by the federal government (U.S. Department of Education, 2014) and required by the Elementary and Secondary Education Act. The state of Texas has defined the minimum reporting requirement as fewer than five students (“Texas State Accountability Plan”, 2010, p. 43; also see Texas Education Agency, 2014a). Further, the state of Texas requires the consideration of cells with a count fewer than five in its memorandum of understanding (MOU) with ERCs in Texas:

“All research results must not disclose personally identifiable information. Data must exclude any data cell or subgroup that *may permit identification* [emphasis added]. Small data cells will be considered any cell containing between one and four individuals inclusive. Information may not be disclosed where small data cells can be determined through subtraction or other simple mathematical manipulations or subsequent cross-tabulation of the same data with other variables. Institutions may use any of the common methods for masking including: a) masking the small cell and the next larger cell on the row and column so the size of the small cell cannot be determined; b) masking the small cell and displaying the total for both the row and column as a range of at least ten; or c) any methodology approved by the Texas Higher Education Coordinating Board and the Advisory Board.”

The UH ERC has elected to fulfill the requirements of the MOU through option “c” by outlining our masking guidelines. The UH ERC has subsequently sought approval from the ERC Advisory Board for the required masking guidelines below. Masking guidelines and techniques are based on the practices of the Texas Education Agency and the Texas Higher Education



---

Coordinating Board, and the guidance of the U.S. Department of Education’s Privacy Technical Assistance Center (PTAC).

Remember, the protection of individual information is nuanced. The best practice is for researcher(s) to ask outlined by FERPA:

- Can a reasonable person in a school community use the information presented, alone or in combination with other publically available information, to identify an individual?
- If the answer is maybe or yes, then you must proceed with masking the data. Data that needs to be masked varies across projects, but the Commonly Used Variables subsection (p.4) provides examples.

### Required Masking

The subsequent three masking guidelines provide the basis of the required masking for UH ERC. For exemplars with rationale, see Appendix I.

**Small cells.** Any cell representing fewer than five individuals presents a small cell reporting issue that must be addressed with masking. Researcher(s) may have subgroups over five individuals, but individual cells fail to reach the minimum of five with subgroups are further broken down by outcomes or other measures of interest. Whether reporting subgroups or categorizations of subgroups, any cell with fewer than five must be masked if it *may permit identification*.

### Small Cell Masking Guide

Condition	Solution		
	Numerator	Denominator	Percent
<b>If denominator is &lt;5 including 0</b>	Mask (*)	Mask (*)	Mask (*)
<b>If percent is 100% or rounds to 100%</b>	Mask (*)	Mask (*)	Top Code %
<b>If percent is 0% or rounds to 0%</b>	Mask (*)	Mask (*)	Bottom Code %



<b>If the difference between the numerator and the denominator is fewer than 3</b>	Mask (*)	Mask (*)	%
<b>If the numerator is &lt;5 including 0</b>	Mask (*)	Mask (*)	%

*Note:* Based on Texas Education Agency Performance-Based Monitoring (2014b), but modified to include top and bottom coding requirements.

If reporting a percent, and the percent either is or rounds to 0% or 100%, then you must top and bottom code. The guideline for top and bottom coding depends on the size of the group you are reporting, and it a form of blurring. The table below provides the conditions, solutions, and exemplars of top-and bottom-coding practices.

### Top and Bottom Coding Guide

Size Range	Solution	Top Code	Bottom Code
<b>N &lt; 10</b>	Mask (*)	Mask (*)	Mask (*)
<b>10 &lt; N &lt; 15</b>	Change percent by 10%	≥90%	≤10%
<b>15 &lt; N &gt; 20</b>	Change by 7%	≥93%	≤7%
<b>20 &lt; N &gt; 30</b>	Change by 4%	≥96%	≤4%
<b>30 &lt; N &gt; 50</b>	Change by 3%	≥97%	≤3%
<b>50 &lt; N &gt; 300</b>	Change by 2%	≥98%	≤2%
<b>N &gt; 300</b>	Change by 1 %	≥99%	≤1%

*Note:* Based on guidance from PTAC described in the NCES (2011) regarding group size.

For instance, if the test passing rate is 100% with N=275, then the researcher(s) would report the passing rate as >98%. While a passing rate of 100% with N=25, would be reported as



---

>96%. This masking technique generally masks the percent to the difference between +/- one individual's score.

### **Complementary Cell Suppression**

Researcher(s) should be mindful of complementary cells following the masking of small cells. The National Center for Education Statistics (2010b) warned that by combining suppressed information with information in complementary cells, the “reported information can then be used to recover the suppressed data through a series of calculations” (p.9). If a *reasonable person* can reverse calculate the mask cells as a product of total count and/or percentages reported for each subgroup or category, then the researcher(s) must take additional measures to protect small cells.

Complementary cell suppression, also known as second least subgroup suppression, is the technique of identifying the next smallest subgroup or categorization to the cell with fewer than five individuals and masking it. The use of complementary suppression protects against the recovery of the suppressed cell information. UH ERC requires complementary cell suppression of small cell information if there is a chance for that the masked cell information can be recovered.

UH ERC urges researchers(s) to consider: Through the use of proportions, counts, and simple mathematic calculations, can a reasonable person recover masked information? If the answer is maybe or yes, then complementary cell suppression must occur.

### **Reporting Performance Based Indicators or Outcomes**

Performance based indicators, like student assessment scores, need additional precautions. Due to the reporting format of the Texas Education Agency publically available information, percentages for performance based indicators must be whole numbers. Round to the whole number for performance based indicators is also recommended by PTAC (see NCES 2011).

### **Common Issues**

Frequently, the process of masking information for FERPA compliance includes steps



---

beyond suppression. The situations and suggestions provided below represent common issues that arise for researchers.

**Profile & Context.** During the writing process of a research product, providing context is necessary to justify the approach, explain the results, and address the discussion component of the study. Researcher(s) may perfectly mask data according to the UH ERC guidelines, but undermine their efforts in the writing process.

To avoid compromising the masking efforts of results, researcher(s) should consider the following:

- In-depth profiles of schools, districts, regions, or states can provide a road map for use of secondary data sources to unmask the reported data. Be mindful of the profiles given in the write-up.
- Descriptions of a population/sample, in narrative or table form, may unmask results information. Researcher(s) should compare results and ~~descriptives~~ to ensure the combination of the materials does not lead to unintended disclosure of individuals.
- Ensure that the information provided in an end product (see Glossary) comes only from reviewed research products.
- In some cases, detailed descriptions of masking techniques can comprise masking efforts.

**Secondary Publications.** The existence of secondary publications, available through Texas Education Agency, Texas Higher Education Coordinating Board, and other entities reporting data that are found in the UH ERC Repository, may require additional considerations. The impact of secondary publications depends principally on the purpose and targeted population of a given study. The researcher(s) output from this study may be specific enough that when combined with annual public reporting could allow a reasonable person to identify individuals. In such a situation, additional precautions for FERPA compliance must be taken. These techniques may include range reporting for counts or the use of counts from a related, but different time period (e.g., fall enrollment counts and spring assessment by rate). UH ERC must approve any such contextual reporting of counts.

**Format.** Another common issue is the format of the information. Consistency is key. Researcher(s) should determine a given format for reporting and stay with the same basic format



---

for their research products. The lack of consistency may lead to the ability for a reasonable person to apply reverse calculations or simple mathematical formulas to recover masked data. An example of this issue may be reporting across different related subgroups or researcher created classifications. The review of research products typically occurs in several stages. While the UH ERC makes every effort to track the approved research products for each project, the responsibility for FERPA compliance falls on the researcher(s). The more consistent the format over the various stages of review, the easier it will be to ensure FERPA compliance.

Another helpful suggestion is the inclusion of a short description of how and where you masked data for the reviewer. This can be done when submitting the formal request for review. For instance, a brief descriptor may read:

“Excel file (8thgrade\_GT\_STAAR): File contains counts of grade 8 students classified as gifted and talented by STAAR performance levels within a specific, unnamed, district. Two cells, A17 & A49, represented fewer than five individuals. Those cells were masked. Additionally, corresponding cells of B17 and B51 were masked according to guidelines for complementary cell suppression. No totals rounded to are were zero or 100%. All other percentages were rounded to the nearest whole number since STAAR is a performance based indicator.”

**Output.** The statistical outputs must also be reviewed for possible disclosure of PII. Namely, the following types of output have shown to be problematic: ~~crosstabs, population~~ marginal means, fixed effects, least squared ~~means, some~~ regression outputs (e.g., logistic regression)

### **Other Masking Techniques**

The UH ERC also wants to acknowledge that there are several other techniques used to avoid the disclosure of information. Moreover, there may be a unique situation in which the guidelines listed above fail to meet the needs of the researcher(s) and/or FERPA compliance. In such a rare situation, where the above mentioned guidelines are inadequate, UH ERC reserves the right to require or suggest researcher(s) to take additional precautions through the application of other masking techniques. These techniques may include the blurring of data (e.g., aggregation, rounding, truncation, top/bottom coding), additional suppression (e.g., collapsing across outcome subgroups), and perturbation (e.g., introduction of noise, data swapping, creation



---

of synthetic data). These additional masking techniques, if necessary in an exceptional case, will follow the guidelines produced by PTAC, Federal Commission on Statistical Methodology, and National Center for Education Statistics.

The UH ERC aims for a seamless review process that allows researcher(s) to conduct analyses while protecting the confidentiality of individuals within a dataset. In an exceptional situation, UH ERC will work with researcher(s) to find the best possible solution to meet both researcher and FERPA compliance.

### **Contact Information**

If you have any questions, concerns, or issues with any UH ERC policies or procedures, please contact the Director or Principal Investigators. Additional information and all forms can be found on the UH ERC website.

<p><b>UH ERC Director</b></p> <p>Catherine Horn, Ph.D. (713)743-5032 clhorn2@uh.edu</p>	<p><b>UH ERC Principal Investigator</b></p>
---	---