

(CNMI Public School System)

Guidelines for Research Products & Masking Techniques

2022

Table of Contents

Purpose & Rationale	2
Personally Identifiable Information (PII)	3
Commonly Used Variables	3
Protocol for Pre-Release Review of Research Products	4
Common Issues	4
Unclear Variables or Output	4
Graphics	5
Degree of Detail	5
Masking Guidelines & Techniques	5
Required Masking	6
Small Cells	7
Complementary Cell Suppression	8
Reporting Performance Based Indicators or Outcomes	8
Common Issues	8
Profile & Context	8
Secondary Publications	9
Format	9
Output	10
Other Masking Techniques	10
Glossary	10
References	12
Appendix	13
Small Cell Reporting	13
Complementary Cell Suppression	14
Collapsing Categories	15
Secondary Publications & Top and Bottom Coding	15

CNMI PSS

Guidelines for Research Products & Masking Techniques

Purpose & Rationale

The federal law known as the Family Educational Rights and Privacy Act of 1974 (20 U.S.C. § 1232g; 34 C.F.R. § 99), commonly referred to as FERPA, protects the release of and access to educational records. Educational records include those maintained by an educational institution, agency, or party acting on behalf of the agency or institution that are directly related to a student (34 C.F.R. § 99.3). The law applies to all educational records generated by schools that receive funds under the applicable programs of the United States Department of Education. FERPA further defines who shall be granted access, what information is considered personally identifiable, and criteria for what is considered a legitimate educational interest allowing for access (see 34 C.F.R. § 99.3; 99.30; 99.31). The data housed at The CNMI PSS, known as the P20/Workforce Data Repository (Repository), contains educational records and personal information about students, educators, and employees in CNMI PSS that is considered confidential and protected by FERPA regulations.¹

The CNMI PSS aims to balance compliance with FERPA with our mission. The CNMI PSS mission is to bridge the gap between theory and policy by providing a cooperative research environment for study by both scholars and policy makers. Within our goal, the CNMI PSS ERC works with researchers, practitioners, state and federal agencies, and other policymakers to help inform upon today's critical issues in education. Access to longitudinal, student level data for scientific inquiry and policymaking purposes, however, must come with procedures to ensure the protection of the individuals behind the data. The review of research product guidelines set forth in this document are one aspect of the overall procedures undertaken to fulfill the obligation to protect individuals under FERPA while simultaneously providing access and support to authorized researchers.

Review of Research Products

To maintain compliance with FERPA, the CNMI PSS Advisory Board requires designated ERC staff members to review all research products derived from the Repository. Research product is a broad term, but for the purposes of the CNMI PSS ERC it includes any written outcome or display resulting from the confidential data. This includes, but is not limited to, output from statistical software, tables, or graphs. Consider any electronic file or written artifact intended to, or which potentially could be, viewed by anyone not currently approved to access the Repository as a research product. All research products must be reviewed prior to release to the researcher(s).

The review process serves as a safeguard to prevent the inadvertent release of any personally identifiable information (PII). The review provides a secondary check of the

¹ Details pertaining to CNMI PSS ERC's alignment with FERPA along with information on the oversight of the Education Research Center Advisory Board can be found in the Policies & Procedures manual.

researcher(s) efforts to mask and appropriately report research results in a manner that protects PII. The following section covers the process and procedures related to protecting confidential CNMI PSS ERC data. Additionally, it includes a discussion on what qualifies as PII, commonly used variables, common issues that arise during the review process, and suggestions for simplifying the review process for researcher(s).

Personally Identifiable Information (PII)

Under FERPA, PII is comprised of both personal identifiers and indirect identifiers (National Center for Educational Statistics [NCES], 2010a; 34 C.F.R. § 99.3). Personal identifiers include information like the student's name or Social Security Number. The data in the CNMI PSS ERC Repository is devoid of Social Security Numbers and names, and a state generated identification number links the information. Indirect identifiers include any

“other information that, alone or in combination, is linked or linkable to a specific student that would allow for a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty” (NCES, 2010a, p.2). Examples of indirect identifiers include race/ethnicity, program specific enrollment, grade level, or course enrollment (NCES, 2010a). The CNMI PSS ERC review process aims to ensure that researcher(s) have followed the established masking guidelines (see Masking Guidelines and Techniques) to appropriately protect PII.

Commonly Used Variables

The CNMI PSS ERC serves as a repository of data spanning numerous datasets with countless variables. While not all the variables within the dataset are considered PII, in conjunction with other information as related to educational performance records, they can act as indirect PII that needs masked to protect the individual. Due to the vast number of variables, it is impossible to list all the possible combinations of variable that might qualify as PII. Moreover, requirements for masking research products may depend on the research design or the method of displaying results.

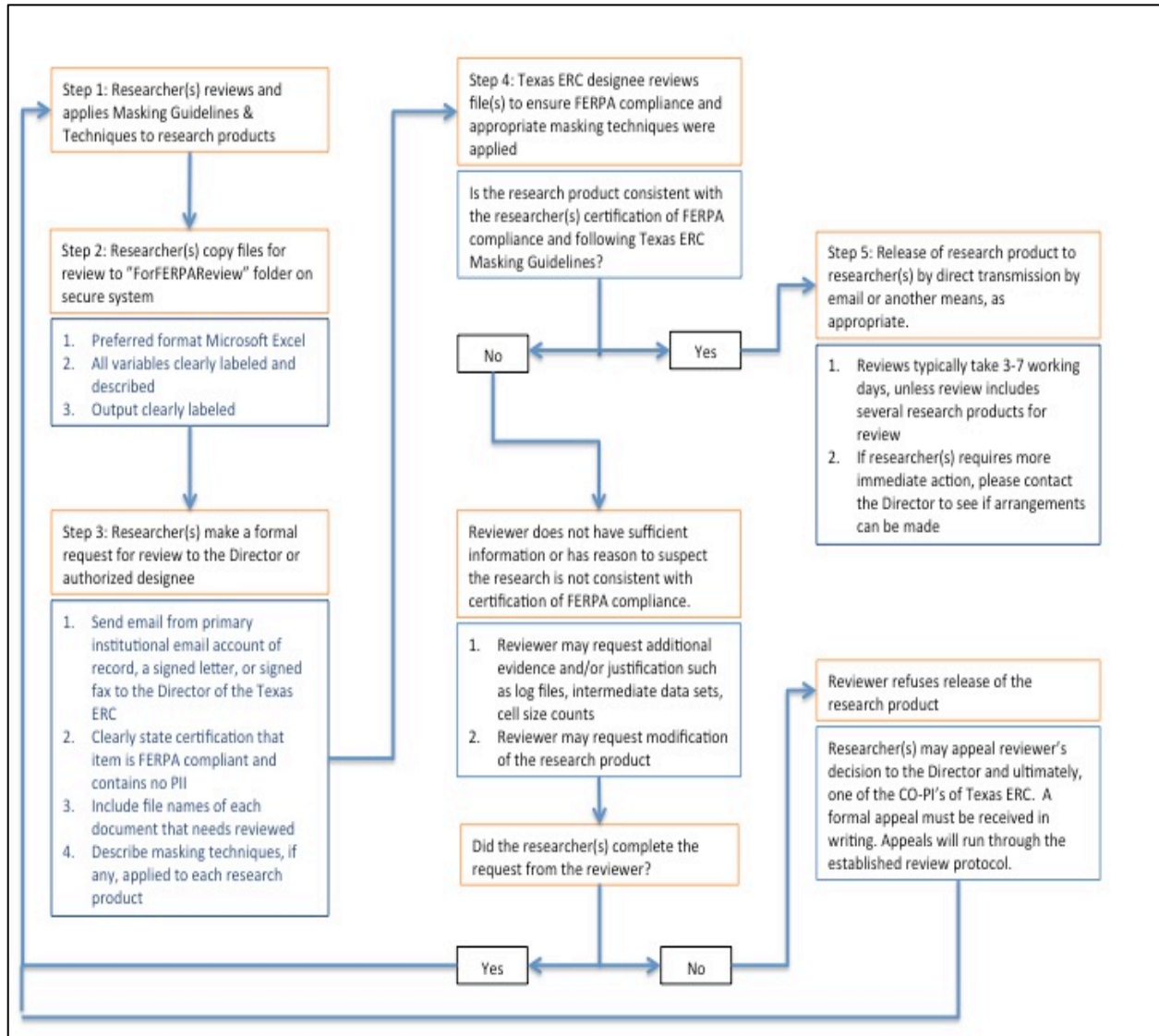
The CNMI PSS ERC does, however, see many commonly used variables that may produce results that fall under the protections of FERPA. This typically is caused by disaggregation (broken down, splitting up, or isolating) the data enough to create small groups (cells) that can reflect individual performance. Please note that these variables may not alone qualify as PII. It is when they are combined, specifically with performance indicators, that it might be possible to link the data to a specific student. Below is an initial, not all-inclusive, list of commonly used variables that *may* constitute PII:

- **Performance Based Indicators:** STAAR, TAKS, SAT/ACT scores, CNMI PSS Success Initiative (TSI), exit exams, cumulative pass rates, college readiness
- **Specialized or Sensitive Programs:** special education, gifted & talented, English Language Learner (ELL)/Limited English Proficiency (LEP), International Baccalaureate (IB), Developmental Education/Disability Services, Advanced Placement (AP) enrollment, college/dual credit courses
- **Student Demographics:** race/ethnicity, age, immigrant status, gender, at risk indicator, homeless status, foster care status
- **Completion:** dropout, graduation, time to completion, high school degree plan
- **Other:** attendance, school transfers, teacher value-added scores

Protocol for Pre-Release Review of Research Products

All research products derived from the Repository's must be reviewed prior to public release consistent with the following procedure:

Review of Research Products Flowchart



Note: See *Policies & Procedures: Approved Project Handbook* for more detail.

Common Issues

The research products vary greatly across the many ongoing research projects at the CNMI PSS ERC. In spite of this variability, CNMI PSS ERC staff members have identified common issues encountered during the review process.

Unclear Variables or Output

Considering the sheer number of variables and the creation of new variables during the research process (collapsing groups, new categorical variables, or indices), reviewers may

not be able to determine if the variables included require masking for FERPA compliance. This requires additional information from the researcher(s).

Researcher(s) should ensure:

- Variables/output are clearly labeled
- Each variable/output includes a definition
- Derived or created variables are identified

The easiest method for labeling and defining variables is to use Microsoft Excel. When exporting output from a statistical software package or creating new output, but sure to include the descriptors. Tables, graphics, and other displays may be done in the same manner.

Graphics

The statistical packages available at the CNMI PSS ERC provide many wonderful options for the visual display of information. The outputs of graphics, however, are often overlooked by researcher(s) when reviewing for PII. The same guidelines (see Masking Guidelines & Techniques) apply to visual displays such as graphs or charts. If the graph illustrates aggregated data about a subgroup with fewer than five persons, it must be appropriately masked.

Degree of Detail

The depth and richness of the dataset available in the CNMI PSS ERC database affords research opportunities from the individual to institutional level. This ability is invaluable for several lines of inquiry in education. These same fine-grained abilities of the dataset, however, may necessitate a greater degree of masking. The CNMI PSS ERC encourages researcher(s) to consider the following question when determining subgroup/categorical groupings, variable sections, reporting, and other methodological decisions:

- Considering your research purpose, what degree of detail is required to answer your research questions?

For instance, consider a research project focused on high school IB program enrollment and time to completion in higher education. Is it necessary to report the outcomes for all racial/ethnic subgroups, which will require more masking or would it be more beneficial to report the outcome for all IB, which would require less masking? The answer is up to the researcher(s), but the decision may impact the amount of time required to complete the review process.

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 CNMI PSS 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 requires 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). The state of CNMI PSS has defined the minimum reporting requirement as fewer than five students (“CNMI PSS State Accountability Plan”, 2010, p. 43; also see CNMI PSS Education Agency, 2014a). Further, the state of CNMI PSS requires the consideration of cells with a count fewer than five in its memorandum of understanding (MOU) with ERCs in the CNMI PSS:

“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 CNMI PSS Higher Education Coordinating Board and the Advisory Board.”

The CNMI PSS ERC has elected to fulfill the requirements of the MOU through option “c” by outlining our masking guidelines. The CNMI PSS ERC has subsequently sought approval from the Advisory Board for the required masking guidelines below. Masking guidelines and techniques are based on the practices of the CNMI PSS Education Agency and the CNMI PSS 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 (see p. 4). 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 CNMI PSS ERC. For exemplars with rationale, please see the Appendix.

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
If denominator is <5 including 0	Mask (*)	Mask (*)	Mask (*)
If percent is 100% or rounds to 100%	Mask (*)	Mask (*)	Top Code %
If percent is 0% or rounds to 0%	Mask (*)	Mask (*)	Bottom Code %
If the difference between the numerator and the denominator is fewer than 3	Mask (*)	Mask (*)	%
If the numerator is <5 including 0	Mask (*)	Mask (*)	%

Note: Based on CNMI PSS 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
N < 10	Mask (*)	Mask (*)	Mask (*)
10 < N < 15	Change percent by 10%	≥90%	≤10%
15 < N > 20	Change by 7%	≥93%	≤7%
20 < N > 30	Change by 4%	≥96%	≤4%
30 < N > 50	Change by 3%	≥97%	≤3%
50 < N > 300	Change by 2%	≥98%	≤2%
N > 300	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. CNMI PSS ERC requires complementary cell suppression of small cell information if there is a chance for that the masked cell information can be recovered.

CNMI PSS 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 CNMI PSS 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 CNMI PSS 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 CNMI PSS Education Agency, CNMI PSS Higher Education Coordinating Board, and other entities reporting data that are found in the CNMI PSS 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). CNMI PSS 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 CNMI PSS 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 CNMI PSS 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, CNMI PSS 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 CNMI PSS 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, CNMI PSS ERC will work with researcher(s) to find the best possible solution to meet both researcher and FERPA compliance.

Glossary

- **Complementary Suppression**– A masking technique used to further protect individuals represented in masked cells. Requires the suppression or masking of the second least subgroup or category to prevent recovery of suppressed information. Also know as Second Least Subgroup suppression.
- **Educational Record**– Any record that is directly related to a student and is maintained by an education agency, institution, or by a party acting for the agency or institution. FERPA further provides exclusions under this definition (see 34 C.F.R. § 99.3).
- **End Product**– Any written outcome or display developed from a reviewed and approved research product. This may include the following: a) abstracts, manuscripts, reviews for publication; (b) research proposals; (c) abstracts,

manuscripts, and reviews for submission for funding; (d) posters for display; (e) write-ups for a website; (f) dissertations; (g) term papers; (h) handouts; (i) presentations; (j) book chapters; and (k) policy briefs.

- **Masking**- Refers to the purposeful exclusion or removal of information prior to public release to protect individuals under FERPA.
- **Personally Identifiable Information (PII)**- Information that can be used to distinguish an individual directly or indirectly through linkages with other information. Many of the direct PII, such as student name and Social Security Number are removed, yet PII remains rich in the CNMI PSS ERC Repository. Information alone, or in combination with other reported information, can be linked back to a specific student.
- **Research Products**- Any written outcome or display derived from the Repository. Including but is not limited to, any research output (e.g., printouts of log files, statistical output, tables, or graphs) intended to, or potentially could be, viewed by anyone not currently approved to use the Repository.
- **Small Cell Reporting**- Referring to suppression of subgroups or categories represented in a given cell with fewer than five individuals. Several masking techniques are used to address small cell reporting (see Complementary Suppression).

References

Code of Federal Regulations, 34 C.F.R. § 99.

Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g.

National Center for Education Statistics. (2010a). *Data stewardship: Managing personally identifiable information in electronic student education records*. (Issue Brief 2, NCES 2011-602). U.S. Department of Education, Institute of Education Sciences. Retrieved from <http://nces.ed.gov/pubs2011/2011602.pdf>

National Center for Education Statistics. (2010b). *Statistical methods for protecting personally identifiable information in aggregate reporting*. (Issue Brief 3, NCES 2011-603). U.S. Department of Education, Institute of Education Sciences. Retrieved from http://ptac.ed.gov/sites/default/files/FAQs_disclosure_avoidance.pdf<http://nces.ed.gov/pubs2011/2011603.pdf>

Privacy Technical Assistance Center. (2012). *Frequently asked questions–Disclosure Avoidance*. U.S. Department of Education. Retrieved from http://ptac.ed.gov/sites/default/files/FAQs_disclosure_avoidance.pdf

Seastrom, M., Rodriguez, B., & Szuba, T. (2011, April). *SLDS technical brief #3: Statistical methods for protecting personally identifiable information in aggregate reporting*. Presented at the American Education Research Association, New Orleans, LA.

Texas Education Agency. (2014a). Accountability manual. Texas Education Agency. Retrieved from <http://ritter.tea.state.tx.us/perfreport/account/2014/manual/>

Retrieved from <http://ritter.tea.state.tx.us/perfreport/account/2014/manual/>

Texas Education Agency. (2014b). *Performance-based monitoring analysis system: Masking rules*. CNMI PSS Education Agency. Retrieved from [http://tea.texas.gov/Student_Testing_and_Accountability/Monitoring_and_Interventions/Performance-based_Monitoring_Analysis_System_\(PBMAS\)/PBMAS_Masking_Rules/](http://tea.texas.gov/Student_Testing_and_Accountability/Monitoring_and_Interventions/Performance-based_Monitoring_Analysis_System_(PBMAS)/PBMAS_Masking_Rules/)

Texas State Accountability Plan. (2010). U.S. Department of Education. Retrieved from <http://www2.ed.gov/admins/lead/account/stateplans03/txcsa.pdf>