SLDS Technical Brief

Guidance for Statewide Longitudinal Data Systems (SLDS)

Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting

Contents

SLDS Technical Briefs are intended to provide "best practices" for consideration by states developing Statewide Longitudinal Data Systems.

Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Introduction

Over the last decade, increased attention on education has led to an expansion in the amount of information on students and their schools and school districts reported to parents and the general public (20 U.S.C. § 6311). States now report student outcomes based on assessments of student achievement in specific subjects and grade levels for all students, as well as for subgroups defined by gender, race and ethnicity, English proficiency status, migrant status, disability status, and economic status. Typically, the data are reported as the percentage distribution of students in a subgroup across achievement levels. These reports are issued at the state, district, and school levels. Additional outcome measures, such as data on attendance, dropout rates, and graduation rates, are also reported frequently.

These reports offer the challenge of meeting the reporting requirements while also meeting legal requirements to protect each student's personally identifiable information (Family Educational Rights and Privacy Act [FERPA]) (20 U.S.C. § 1232g; 34 CFR Part 99). Recognizing this, the reporting requirements state that subgroup disaggregations of the data may not be published if the results would yield personally identifiable information about an individual student (or if the number of students in a category is insufficient to yield statistically reliable information). States are required to define a minimum number of students in a reporting group or subgroup required to publish results consistent with the protection of personally identifiable information (34 CFR § 200.7).

Individual states have adopted minimum group size reporting rules, with the minimum number of students ranging from 5 to 30 and a modal category of 10 (used by 39 states in the most recent results available on state websites in late winter of 2010). Each state has adopted additional practices to protect personally identifiable information about its students in reported results. These practices include various forms of suppression, top and bottom coding of values at the ends of a distribution, and limiting the amount of detail reported for the underlying counts. This Technical Brief includes a summary of key definitions, a brief discussion of background information, and a review and analysis of current practices to illustrate that some practices work better than others in protecting personally identifiable information reported from student education records.

The review led to the formulation of recommended reporting rules that are driven by the size of the reporting groups or subgroups. The reporting rules are intended to maximize the amount of detail that can be safely reported without allowing disclosures from student outcome measure categories that are based on small numbers of students. NCES welcomes input on these recommendations.

Definitions

Personally identifiable information includes the name and address of the student and the student's family; a personal identifier, such as the student's Social Security Number, student number, or biometric record; other indirect information, such as the student's date and place of birth and mother's maiden name; other information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of relevant circumstances, to identify a student with reasonable certainty; and information based on a targeted request.

Disclosure means to permit access to or the release, transfer, or other communication of personally identifiable information contained in education records by any means. To avoid disclosures in published tables, whenever possible, data about individual students should be combined with data from a sufficient number of other students to disguise the attributes of a single student. When this is not possible, data about small numbers of students should not be published.

Suppression refers to withholding information from publication. Some information is withheld from publication in a table to protect data based on small counts because the release of the information would likely lead to a disclosure. Other information is withheld from publication in a table to prevent the calculation of the data based on small counts from the published information; this is known as complementary suppression.

Recoding refers to reporting values as being within a specified range rather than as a specific value.

> **Top coding** refers to reporting values over a set value as greater than that value.

> **Bottom coding** refers to reporting values under a set value as less than that value.

Top coding and bottom coding are specific types of recoding. These procedures are used to protect data for individual students from disclosure.

Subgroups refer to students within a larger group who share specific characteristics, such as the subgroup of male students and the subgroup of female students within a school or within a grade in a school. Information from student records is often reported for subgroups of students by gender, race and ethnicity, English proficiency status, migrant status, disability status, and economic status.

Outcome measures refer to the student's educational experiences that are recorded in student's educational records. For example, student grades, courses completed, scores on standardized assessments, school attendance, graduation status, participation in extracurricular activities, and disciplinary actions are commonly reported measures of student outcomes.

Categories refer to groups of students that share specific experiences that comprise the range of possible outcomes for each educational measure. For example, the percent of students with passing as compared to failing grades, the percent of students who dropout as compared to completing high school, or the percent of students who scored at each of several achievement levels on a standardized state assessment.

Background

As the nation has focused its attention on education over the last decade, there has been a large increase in the amount of data reported to the general public on America's students and their schools and school districts (20 U.S.C. § 6311(h); 20 U.S.C. § 9607; U.S. Public Law 110-69; U.S. Public Law 111-5). Reporting requirements for public elementary and secondary institutions that receive federal funds include annual status and progress reports at the school, district, and state levels (20 U.S.C. § 6311(h)).[1] Among other requirements, these reports, identified as report cards, must include results from state assessments on the percent of students assessed, along with student achievement results across achievement levels in specific subjects and grade levels for all students and for reporting subgroups including gender, race/ethnicity, English proficiency status, migrant status, disability status, and economic status. The annual status and progress report cards also typically include data on attendance rates and report graduation rates for secondary schools. Dropout rates are also frequently reported at the district and school levels.

The current reporting requirements are typically met through state-, district-, and school-level reports that are published by each state's department of education. These reports offer the challenge of balancing the reporting requirements against legal requirements to protect each student's personally identifiable information (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99). To this end, the reporting requirements for Title I state that disaggregating the data for specific subgroups may not occur if the number of students in a reporting group or subgroup is insufficient to yield statistically reliable information or if the results would yield personally identifiable information about an individual student (20 U.S.C. § 6311(h); 34 CFR § 200.7).[2]

As part of the reporting requirements, each state is required to have an accountability plan that describes its system for monitoring adequate yearly progress with annual objectives for continuous and substantial improvement for all students and for each specified student subgroup. In addition to defining specific measures, each state's accountability plan is expected to include the state's definition of the minimum number of students in a subgroup required for reporting purposes and information as to how the State Accountability System protects the privacy of students when reporting results.

What does protecting student privacy mean in a reporting context? In order to protect a student's privacy, the student's personally identifiable information must be protected from public release. The broad, federal government-wide definition of personally identifiable information states "the term 'personally identifiable information' refers to information that can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc., alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc." (OMB Memorandum 07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information; Implementation Guidance for Title V of the E Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)*).

---

[1] The requirement specified in law is for an annual state report card and for annual district report cards that include information for the district and each school.

[2] The law states that reporting student assessment results disaggregated by economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency is not required if the number of students in a *category* is insufficient to yield statistically reliable information or the results would reveal personally identifiable information about an individual student (20 U.S.C. § 6311). However, the regulations use the term *subgroup* to refer to the disaggregated student data, and the regulations specify that a state may not report achievement results for a *subgroup* if the results would reveal personally identifiable information about an individual student (34 CFR § 200.7). This is further promulgated in the September 12, 2003 non-regulatory guidance on Report Cards Title I, Part A.

The FERPA definition of personally identifiable information (34 CFR § 99.3) follows the government-wide definition and includes the following:

Personally identifiable information includes, but is not limited to:

1. The student's name;

2. The name of the student's parent or other family members;

3. The address of the student or student's family;

4. A personal identifier, such as the student's Social Security Number, student number, or biometric record;[3]

5. Other indirect identifiers, such as the student's date of birth, place of birth, and mother's maiden name;

6. Other information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty;

7. Information requested by a person who the educational agency or institution reasonably believes knows the identity of the student to whom the education record relates.
(34 CFR § 99.3)

Protecting student privacy means publishing data only in a manner that does not reveal individual students' personally identifiable information, either directly or in combination with other available information. Another way of putting this is that the goal is to publish summary results that do not allow someone to learn information about a specific student.

States publish annual status and progress reports that are based on reports of outcome measures at the school, district, or state level. These reports aggregate, or combine, the results for individual students into summary statistics. These statistics include the number or percentage of students overall or in each of the reporting subgroups for specific outcome measures (e.g., the percentage of students in each racial and ethnic group who graduate from high school; the percentage of English language learners who score in each achievement level on a state assessment).

This report demonstrates how disclosures occur even in summary statistics. It describes various reporting practices and data protection techniques currently in use and illustrates how commonly used methods of data protection may fall short of their goal. The report then identifies "best practices" to avoid the unintended disclosure of personally identifiable information, including publishing the percentage distribution across categories of outcome measures with no underlying counts or totals; publishing a collapsed percentage distribution across categories of outcome measures with no underlying counts or totals; publishing counts but using complementary suppression at the subgroup level when a small subgroup is suppressed; limiting the amount of detail published for school background information; recoding the ends of percentage distributions; and recoding high and low rates. This information is used to develop recommendations for reporting rules that maximize the amount of information reported while protecting the privacy of each student's data.

Unintended Disclosure of Personally Identifiable Information

When personally identifiable information is revealed through information released to the public, it is called a disclosure.[4] When schools, districts, or states release information about educational progress, they typically release aggregated data—data for groups of

---

[3] FERPA 2008 regulations state that the term "biometric record, as used in the definition of personally identifiable information, means a record of one or more measurable biologic or behavioral characteristics that can be used for automated recognition of an individual. Examples include fingerprints; retina and iris patterns; voiceprints; DNA sequence; facial characteristics; and handwriting." (34 CFR § 99.3)

students—to prevent disclosure of information about an individual. Even with some methods of aggregation, unintended disclosure of personally identifiable information may occur. How could data reporting outcome measures for groups of students possibly reveal information on an individual student? The example that follows shows how information about individual students' achievement levels can be revealed, even in data reported for groups of students. Furthermore, it shows that the identity of groups of students can be revealed within combinations of achievement levels (e.g., Below Basic and Basic for students who scored below Proficient, or proficient and advanced for students who scored at or above Proficient).

Typically, each child's parents are given their child's score and achievement level on the state assessment as well as the report for their child's school. Table 1 provides the percentage distribution and number of students at each achievement level at the school level in grade 4 mathematics, for students overall and for several subgroups: White and Hispanic students, students with and without an individualized education plan, and students who are and are not English language learners. Any combination of these three subgroup variables that reveals the achievement level for a student or group of students with identifiable characteristics results in a disclosure.

Example 1: Unintended Disclosures

Consider a school report that includes results on the state assessment by grade and subject. No results are suppressed as a result of a small subgroup count, since each subgroup included more than the minimum reporting group size of 5. The report shows that there are 32 fourth-graders in this school and that they were all assessed in mathematics (table 1). Among these students, 12.5 percent, or 4 students, scored at the Below Basic achievement level; 31.3 percent, or 10 students, scored at the Basic level; 34.4 percent, or 11 students, scored at the Proficient level; and 21.9 percent, or 7 students, scored at the Advanced level. The data reported for the subgroups of students with and without an individualized education plan show that all fourth-graders with an individualized education plan scored below the Proficient level (4 students at the Below Basic level plus 3 at the Basic level). Assuming that other students in the class know who among their peers have individualized education plans, this is a disclosure because it reveals that each fourth-grader with an individualized

education plan failed to reach the Proficient level on the assessment.

Next, looking at the 10 Hispanic fourth-graders, the data show that 1 student in this subgroup scored at the Proficient level, while the other 9 students scored at either the Basic level (5 students) or the Below Basic level (4 students). Since parents receive their child's score and achievement level as well as a school report that shows the performance in mathematics by grade, the parents of the 1 Hispanic student who scored at the Proficient level know that the other 9 Hispanic students in the fourth grade each scored below the Proficient level in mathematics. This is a disclosure, because these parents now know that each of their child's ethnic peers failed to reach the Proficient level.[5]

The subgroup data in this table also show that each of the 4 fourth-graders who scored at the Below Basic level were Hispanic, received English language instruction, and had an individualized education plan. This is a considerable amount of information

about the characteristics of the 4 lowest performers. However, since there were Hispanic students who scored at the Below Basic, Basic, and Proficient achievement levels, students with individualized education plans who scored at both the Below Basic and Basic achievement levels, and students receiving English language instruction who scored at both the Below Basic and Basic achievement levels, the table only identifies the fact that there are four Hispanic fourth-graders with this set of three shared characteristics; it does not identify the 4 specific Hispanic students. Thus, the table considered alone does not result in a disclosure in this instance.

Suppose, however, that the students with individualized education plans receive observable special services (e.g., a tutor, extra time on tests, one-on-one test instruction) and that there are exactly 4 Hispanic students receiving these services; then it becomes apparent that these are the 4 Hispanic students who scored at the Below Basic achievement level.

---

[4] Under FERPA, disclosure means to permit access to or the release, transfer, or other communication of personally identifiable information contained in education records by any means, including oral, written, or electronic means, to any party except the party identified as the party that provided or created the record (34 CFR § 99.3).

[5] While this disclosure is based on the parents' personal knowledge of their child's score, the fact that each parent in the school receives his or her child's score raises this source of disclosure as a topic of concern (i.e., knowledge of one child's score revealing the performance of other students).

Table 1.   School-level grade 4 mathematics assessment results in a state with a minimum reporting group size of 5

		Percent assessed	Tested	Below Basic	Basic	Proficient	Advanced
Total	%	100	100	12.5	31.3	34.4	21.9
	N	†	32	4	10	11	7
White	%	100	100	0.0	22.7	45.5	31.8
	N	†	22	0	5	10	7
Hispanic	%	100	100	40.0	50.0	10.0	0.0
	N	†	10	4	5	1	0
Individualized education plan	%	100	100	57.1	42.9	0.0	0.0
	N	†	7	4	3	0	0
No individualized education plan	%	100	100	0.0	28.0	44.0	28.0
	N	†	25	0	7	11	7
English language learner	%	100	100	40.0	50.0	10.0	0.0
	N	†	10	4	5	1	0
Not English language learner	%	100	100	0.0	22.7	45.5	31.8
	N	†	22	0	5	10	7

† Not applicable.
NOTE: Details may not sum to totals because of rounding.

Recall that the reporting requirements acknowledge the risk associated with small numbers by indicating that results should only be published if the results would not reveal personally identifiable information about an individual student. The instructions for the state accountability plan also acknowledge this risk with the requirement for each state to establish a minimum subgroup size for reporting and with the requirement for each state to describe how the State Accountability System protects the privacy of students when reporting results.

Current Disclosure Prevention Practices that Retain Some Disclosure Risk

Typically, a state establishes the required minimum number of students in a subgroup for privacy protection and then does not report the results for outcome measures for any subgroup with less than this established minimum number. The groups not reported are identified as having been suppressed to protect student privacy. A review in late winter of 2010 of the most recent reported assessment results for each state and the District of Columbia found that 39 states use a minimum reporting group size of 10 students. Another 7 states set the minimum reporting group size at 5, and 5 states set the minimum higher, with values ranging from 15 to 30.

While subgroup suppression is a good start, it may not be enough to prevent disclosure of personally identifiable information. The descriptions of current practices include such potentially problematic methods as 1) suppressing data for small subgroups but not for small categories of outcome measures for reported subgroups; 2) suppressing data for small subgroups but reporting counts across the categories of the outcome measure for the overall group and the reported subgroups; 3) suppressing data for small subgroups but reporting the overall total count; and 4) suppressing data for small subgroups but reporting ranges for the overall totals and the reported subgroup totals.

Suppressing Data for Subgroups but not for Reporting Categories

The practice of suppressing data for small subgroups is a start. However, when subgroup results are reported for the categories of an outcome measure, there can also be a small number of students in one or more of the categories within the larger subgroups. Reporting results for small numbers of students within a category or within a subgroup can present a risk to student privacy because it increases the risk of unintentionally releasing information that identifies individual students. The minimum for categories within subgroups can be set lower than the size of the subgroup minimum, but there should be a minimum size specified for individual categories to guard against unintentional disclosures. This minimum, which is sometimes referred to as the threshold rule, defines those categories in a table that are defined as sensitive because the number of students is less than the specified number. Some data collection agencies set this number at 5, while others set it as 3. (Federal Committee of Statistical Methodology, Working Paper 22). Sensitive categories are illustrated in the following example.

Example 2: Suppression of Small Subgroups but not Small Categories

In this example, when a minimum reporting size of 10 is applied to the data from table 1, the assessment results for the 7 students with individualized education plans are presumed to be protected from disclosure because the results are suppressed (see table 2). Thus, the result in example 1 showing that all students with an individualized education plan failed to reach the Proficient level of the state assessment is presumed to be protected from disclosure. However, when the assessment results of the 10 Hispanic students and the 10 English language learners are reported across the four achievement levels, the number of students at each achievement level falls below the established minimum reporting size. In both subgroups, there are 4 students in the Below Basic achievement group, 5 students in the Basic achievement group, and 1 student in the Proficient achievement group; nevertheless, the results are reported since the minimum size rule is applied at the subgroup reporting level. As described in example 1, reporting that only one Hispanic child scored at or above the Proficient level discloses information about that child and about the achievement level of the other students in the subgroup. Anyone who is able to identify the Hispanic child with a high score then knows that the other Hispanic children in the same grade failed to reach the proficient achievement level.

Table 2.   School-level grade 4 mathematics assessment results in a state with a minimum reporting group size of 10

		Percent assessed	Tested	Below Basic	Basic	Proficient	Advanced
Total	%	100	100	12.5	31.3	34.4	21.9
	N	†	32	4	10	11	7
White	%	100	100	0.0	22.7	45.5	31.8
	N	†	22	0	5	10	7
Hispanic	%	100	100	40.0	50.0	10.0	0.0
	N	†	10	4	5	1	0
Individualized education plan	%	100	100	*	*	*	*
	N	†	7	*	*	*	*
No individualized education plan	%	100	100	0.0	28.0	44.0	28.0
	N	†	25	0	7	11	7
English language learner	%	100	100	40.0	50.0	10.0	0.0
	N	†	10	4	5	1	0
Not English language learner	%	100	100	0.0	22.7	45.5	31.8
	N	†	22	0	5	10	7

† Not applicable.
* Not reported to protect subgroups with fewer than 10 students.
NOTE: Details may not sum to totals because of rounding.

Suppressing Data for Subgroups but Reporting Too Much Detail in Underlying Counts

Suppressing data for small subgroups is a first step. However, when data are suppressed to protect student privacy, care must also be taken to avoid publishing information that can be used to retrieve or recover the suppressed information. The next three examples illustrate disclosure problems that can occur in reporting student outcome measures.

The released data in each example table are displayed with a white background. The shaded portions of the example tables represent data that were suppressed. The data entries in the shaded portions of the table were recovered from the released data.

Counts for overall group and reported subgroups

In 38 states, the data are suppressed for subgroups that fall below the minimum reporting group size; however, the number of students and the percentage distributions across the categories of the outcome measure are reported for the overall group and the remaining reporting subgroups. The reported information can then be used to recover the suppressed data through a series of calculations. This can be done using the following steps:

1. Convert the percentages across the outcome categories for the overall group to proportions.

2. Multiply the proportions by the number of students in the overall group to yield the number of students in each category of the outcome measure in the overall group.

3. Identify a suppressed subgroup and the related reported subgroup(s).

4. Repeat steps 1 and 2 for the related reported subgroup(s) to yield the number of students in each category of the outcome measure in the reported subgroup.

5. Subtract the number of students in each category of the outcome measure for the reported subgroup from the overall count for that outcome category to yield the number of students in each category of the outcome measure for the suppressed subgroup.

6. If there are more than 2 subgroups for one disaggregation (e.g., race/ethnicity), compute the counts across the categories of the outcome measure for each reported subgroup, sum subgroup counts for the reported subgroups across each outcome category, and then subtract from the overall number for that category of the outcome measure to yield the number of students in each category of the outcome measure for the suppressed subgroup(s).

All students are in one of two subgroups when student outcome measures are reported by gender, economic status, English proficiency status, migrant status, or disability status. When the data for one of the two subgroups are suppressed and the data for the other subgroup and the total are published, the suppressed data can be fully recovered. When student outcome measures are reported for race and ethnicity, subgroup data are frequently suppressed for more than one subgroup. However, the difference between the counts computed for the outcome categories of students overall and the summation across the outcome categories for the reported subgroups can be used to recover data for the suppressed subgroup(s). This recovery may yield identifying information about the students in the reporting subgroup(s) with suppressed data.

The recovery of suppressed results does not always pose a serious threat to students' personally identifiable information, but in some instances it does—the risk of identifying an individual student is a function of the distribution of students across the recovered categories.

Example 3: Suppressing Outcomes but Reporting Counts for Subgroups

The reported data in table 3 show that among 82 students who were assessed in third-grade reading, 7.3 percent (6 students) scored at the Below Basic achievement level, 42.7 percent (35 students) scored at the Basic level, 37.8 percent (31 students) scored at the Proficient level, and 12.2 percent (10 students) scored at the Advanced level. Seventy-five of the 82 students did not have an individualized education plan, and the reported data show that 8.0 percent (6 students) in this reporting subgroup scored at the Below Basic level, 42.7 percent (32 students) scored at the Basic level, 36.0 percent (27 students) scored at the Proficient level, and 13.3 percent (10 students) scored at the Advanced level.

Although the data were suppressed for students with an individualized education plan, the recovered data show that 7 of the 82 students assessed in third-grade reading were in this suppressed reporting subgroup. Further, a comparison of the overall assessment results with those for the 75 students without an individualized education plan shows that 3 of the 7 students with an individualized education plan scored at the Basic level and 4 scored at the Proficient level. These data do not provide the information needed to identify which students with an individualized education plan scored at the Proficient level and which did not. Thus, this table does not disclose an individual student's performance; however it does reveal the fact that no student with an individualized education plan scored at the Advanced level or at the Below Basic level.

In contrast, the recovered data for 8 low-income students show that 3 of these students scored at the Below Basic achievement level and 5 scored at the Basic achievement level. Thus, all students identified as low-income scored below the Proficient achievement level. If an individual student is known to be from a low-income family, the information in this table discloses that student's score as below Proficient.

The recovered data for 8 students receiving English language instruction show that 3 scored at the Below Basic achievement level, 4 scored at the Basic achievement level, and 1 scored at the Proficient level. Since parents receive their child's score along with the school report, the parents of the child who scored at the Proficient level could use the information in the published table for their child's grade to learn that each of their child's peers who received English language instruction failed to score at the Proficient achievement level.

Table 3.   School-level grade 3 reading assessment results for a state with a minimum reporting size of 10

		Tested	Below Basic	Basic	Proficient	Advanced
Total	%	100	7.3	42.7	37.8	12.2
	N	82	6	35	31	10
Individualized education plan	%	100	0.0	42.9	57.1	0.0
	N	7	0	3	4	0
No individualized education plan	%	100	8.0	42.7	36.0	13.3
	N	75	6	32	27	10
English language learner	%	100	37.5	50.0	12.5	0.0
	N	8	3	4	1	0
Not English language learner	%	100	4.1	41.9	40.5	13.5
	N	74	3	31	30	10
Low income	%	100	37.5	62.5	0.0	0.0
	N	8	3	5	0	0
Not low income	%	100	4.1	40.5	41.9	13.5
	N	74	3	30	31	10

NOTE: Details may not sum to totals because of rounding.

Counts for the overall group

Some states report the percentage distribution across achievement levels for the overall population in a grade and subject along with the percentage distributions for each subgroup, but only publish the number of students tested overall for that grade and subject. This seems like it would provide more protection to students' personally identifiable information, since the number of students in each subgroup is not published. However, in many cases—especially at the school or district level for the data reported by grade and subject—there is only one unique mathematical solution that could yield the reported subgroup percentage distributions for the reported number of students overall.

Example 4: Suppressing Outcomes but Reporting Counts for Groups

In this school, 46 students were assessed in third-grade reading (table 4), and this number is known. Note that the shaded cells in the table display the data that were recovered from the reported information. Multiplying the proportions from the percentage distribution times the number in the overall group (46) shows that the 6.5 percent who scored at the Below Basic level represents 3 students (i.e., 0.65 × 46=3). The data reported by gender show that the 3 students who scored at the Below Basic level are all males. Thus, by dividing 8.3 by 3, the data show that each male student represents 2.77 percent of the number of males in the subgroup. Dividing each of the remaining percentages by 2.77 shows that there are 10 males who scored at the Basic level, 20 who scored at the Proficient level, and 3 who scored at the Advanced level.

Next, the number of males at each achievement level is subtracted from the number of students at that achievement level to recover the suppressed data for females. These calculations show that there are no females at the Below Basic level, no females at the Basic level, 7 females at the Proficient level, and 3 females at the Advanced level. The recovered data do not reveal which females scored at each of these two levels. However, when the focus of the reporting or interpretation of the data shifts to performance at or above versus below the Proficient level, the data for students scoring at the Below Basic and Basic level are combined to show the percent of students who scored below the Proficient level and the percent of students who score at the Proficient and Advanced levels are combined to show the percent of students who scored at the Proficient level. In this example, the recovered data show that all of the third-grade females in this school scored at the Proficient level or above in reading. This then discloses information about the reading achievement level of each of the third-grade females in this school.

Table 4. School-level grade 3 reading assessment results for a state with a minimum reporting size of 10

		Tested	Below Basic	Basic	Proficient	Advanced
Total	%	100	6.5	21.7	58.7	13.0
	N	46	3	10	27	6
Male	%	100	8.3	27.8	55.6	8.3
	N	36	3	10	20	3
Female	%	100	0.0	0.0	70.0	30.0
	N	10	0	0	7	3

NOTE: Details may not sum to totals because of rounding.

Counts for the overall group and subgroups reported as ranges

Another reporting approach recognizes the problem with reporting exact population counts for students assessed and, instead, reports the counts in ranges (i.e., as a categorical variable). With this approach, the percentage distribution is reported for each grade and subject overall and for each of the reporting subgroups that do not require suppression; then, instead of reporting the exact number of students in each group or subgroup, a range that includes the exact number is all that is reported for the count (e.g., instead of reporting 33 students, the number is reported as 30–39). As with the last approach, this would seem to provide more protection to students' personally identifiable information, since the exact number of students is not published. However, the range of possible values for the number of students can be used to identify the number of students that, when applied to the proportion of students at each achievement level, yields estimates that are the closest to whole numbers. Once these counts are established for the overall group and for a reported subgroup, the suppressed counts for a related subgroup can be recovered.

Example 5: Suppressing Outcomes but Reporting Ranges for Counts

The number of third-graders assessed in reading was reported as 40–49 (table 5). The percentage distribution of third-graders overall, across the achievement levels, was reported with 2 decimal places. The percentage distribution across the achievement levels was reported for the 30–39 students who did not have an individualized education plan, but the achievement results were suppressed for the 6–9 students who had one. First, the proportions from the distribution across the achievement levels were applied to each of the 10 numbers in the 40 to 49 range. The number that resulted in estimates that were closest to whole numbers is 41. This showed that, overall, 2 students scored at the Below Basic level, 5 scored at the Basic level, 15 scored at the Proficient level, and 19 scored at the Advanced level. Next, this set of steps was repeated for the 10 numbers in the 30–39 range, using the proportions from the percentage distribution across the achievement levels for students who did not have an individualized education plan. This showed that there were 34 students in this group, with none at the Below Basic level, none at the Basic level, 15 at the Proficient level, and 19 at the Advanced level.

Finally, the counts for students who did not have an individualized education plan were subtracted from the overall counts to recover the suppressed number for the students with an individualized education plan—there were 7 students in this group. Within this group, 2 scored at the Below Basic level, 5 scored at the Basic level, none scored at the Proficient level, and none scored at the Advanced level. These counts can then be used to compute the suppressed percentage distribution. The recovered data show that each of the 7 third-graders with individualized education plans scored below the Proficient level in reading. This is a disclosure of the reading achievement-level information for these 7 students

Table 5. School-level grade 3 reading assessment results for a state with a minimum reporting size of 10 and counts reported as ranges

		Percent assessed	Number tested	Below Basic	Basic	Proficient	Advanced
Total	%	100	†	4.88	12.20	36.59	46.34
	N	40–49	41	2	5	15	19
Individualized education plan	%	100	†	28.57	71.43	0.00	0.00
	N	6–9	7	2	5	0	0
No individualized education plan	%	100	†	0.00	0.00	44.12	55.88
	N	30–39	34	0	0	15	19

† Not applicable.
NOTE: Details may not sum to totals because of rounding.

Best Practices: Practices that Mitigate Disclosure Risk

The review of each state's online reporting of assessment results for schools uncovered three approaches that can help in protecting against the release of information needed to recover personally identifiable information. The first such approach involves not reporting any of the enrollment data that were used to compute the percentage distributions across the achievement-level results. The second approach starts with the first approach (i.e., the underlying enrollment counts are not reported) and collapses across outcome categories to further limit the amount of detail published. This increases the number of students included in each reported outcome category. The third approach involves suppressing subgroups other than the subgroups with less than the minimum reporting size in order to prevent the recovery of the suppressed results for the small subgroups.

Additional practices that support public reporting while protecting student privacy were identified and are discussed in this section. The first involves the reporting of background data on enrollment by grade and enrollment by student characteristics for a school or district. The second involves protecting data at the ends of the distribution, or at the low and high values for a rate, to avoid reporting that a small number of students (or nearly all students) have a specific outcome.

Each of these practices taken alone does not necessarily address each of the potential sources of disclosure, but they do reflect practices that, when taken in combination, may lead to improved protection of personally identifiable information about individual students in published tables.

No Counts Published

Eight states were identified that publish student assessment results by grade and subject for the overall student population and for the reportable subgroups (i.e., those subgroups that do not require suppression) only as a percentage distribution across the achievement levels. In these states, the school reports do not include counts of the number of students assessed overall or of the number of students assessed in each of the reporting subgroups. However, since too much precision in the percentages can limit the possible options for the underlying counts, limiting the

percentages reported to whole numbers increases the number of possible options for the underlying counts. This helps protect the suppressed data for small groups. It also helps protect the counts for small categories within outcome measures for the reported subgroups. The following example of school-level third-grade reading results shows that while the relative relationships across achievement levels within and across subgroups are evident, the absence of the counts used to compute the percentage distributions prevents the recovery of the suppressed data.

Table 6 shows assessment results only as percentage distributions reported as whole numbers. This, coupled with the fact that no counts are reported, protects the suppressed data from disclosure (table 6). The table shows that 13 percent of the students scored at the Below Basic level, 44 percent scored at the Basic level, 27 percent scored at the Proficient level, and 16 percent scored at the Advanced level. Relatively more male than female students and more low-socioeconomic status than non-low-socioeconomic status students performed at the Below Basic level. The data are suppressed for the English language learner subgroup because there are fewer than 10 students in the subgroup.

Table 6.   Percentage distribution of school-level grade 3 reading assessment results in a state with a minimum reporting size of 10 and no counts

	Below Basic	Basic	Proficient	Advanced
Total	13	44	27	16
Male	17	47	23	13
Female	9	42	30	18
Low SES	28	39	22	11
Not low SES	7	47	29	18
English language learner	*	*	*	*
Not English language learner	6	44	31	19

* Not reported to protect subgroups with fewer than 10 students.
NOTE: Details may not sum to totals because of rounding. SES = Socioeconomic status.

Collapsing Across Outcome Categories

Seven states limited their reporting of achievement results to two categories—those at or above the level established by the state for successful performance and those who did not score in the successful range. Collapsing across outcome categories is useful when there are a small number of students in one or more of the outcome categories. This approach, combined with the decision to not report the underlying counts, is another way of increasing the protection of student privacy in reported summary tables.

Example 7: Best Practices: Collapsing across Outcome Categories

Collapsing across outcome categories and displaying the assessment results only as a percentage distribution protects the underlying counts from disclosure. Collapsing the data used in the previous example, 57 percent of the students scored at or below the Basic level, and 43 percent scored at or above the Proficient level (table 7). Relatively more male then female students (64 percent versus 51 percent) and low socioeconomic status than not low socioeconomic status students (67 percent versus 53 percent) scored at the Below Basic level. The data are suppressed for the English language learner subgroup because there are less than 10 students in the subgroup.

Table 7. Percentage distribution of school level, grade 3 reading assessment results collapsed in a state with a minimum reporting size of 10 and no counts

	Basic or below	Proficient or above
Total	57	43
Male	64	36
Female	51	48
Low SES	67	33
Not low SES	53	47
English language learner	*	*
Not English language learner	50	50

* Not reported to protect subgroups with fewer than 10 students.
NOTE: Details may not sum to totals because of rounding. SES = Socioeconomic status.

Counts Published with Additional Suppression

One state provides counts for the overall number of students assessed in a specific grade and subject and for students in reportable subgroups. However, instead of suppressing only the subgroups that do not meet the minimum reporting size, subgroups related to the suppressed group are also suppressed. This is referred to as "complementary suppression." That is, a subgroup with less than 10 students is suppressed, and one (or more) of the other subgroups that combine with the small subgroup to account for a larger share of the students in the overall group is also suppressed. The following example of school-level third-grade reading results provides an illustration of this approach.

Example 8: Best Practices: Schools Counts Published with Additional Suppression

This example includes two schools. The school-level report is designed to display results by gender, race and ethnicity, low-income status, and individualized education plan status. School 1, with 30 students, had a number of reporting subgroups with fewer than 10 students. Suppressing the assessment results for the small subgroups and suppressing the outcome measure for a related category (i.e., complementary suppression of additional rows of the table) protects the reported data at the school level, but leads to the loss of information. As shown in table 8, data were suppressed for the 27 White students because there were fewer than 10 students in each of the other racial and ethnic subgroups (i.e., 2 Native American students and 1 Black student). Data were suppressed for the 21 low income students because there were fewer than 10 students who were not low income. Data were also suppressed for the 21 students without an individualized education plan, because only 9 students had individualized education plans. By comparison, assessment data were reported for the 30 third-grade students overall, and for the 12 male and 18 female students because the minimum reporting threshold of 10 students was exceeded in each case.

Table 8. School 1: Number tested and percentage distribution of grade 3 reading assessment results with a minimum reporting size of 10 and complementary row suppression

	Number tested	Below Basic	Basic	Proficient	Advanced
Total	30	16.7	56.7	20.0	6.7
Male	12	25.0	58.3	16.7	0.0
Female	18	11.1	55.6	22.2	11.1
White	27	*	*	*	*
Native American	2	*	*	*	*
Black	1	*	*	*	*
Low income	21	*	*	*	*
Not low income	9	*	*	*	*
Individualized education plan	9	*	*	*	*
No individualized education plan	21	*	*	*	*

* Not reported to protect subgroups with fewer than 10 students.

School 2, with 45 students, had 10 or more students in each reporting group. As a result, no data were suppressed and the third-grade reading assessment results were reported for each of the reporting variables—gender, race and ethnicity, low income status, and individualized education plan status (table 9).

Table 9. School 2: Number tested and percentage distribution of grade 3 reading assessment results with a minimum reporting size of 10 and complementary row suppression

	Number tested	Below Basic	Basic	Proficient	Advanced
Total	45	2.2	22.2	62.2	13.3
Male	18	5.6	27.8	55.6	11.1
Female	27	0.0	18.5	66.7	14.8
White	20	0.0	10.0	65.0	25.0
Native American	10	10.0	40.0	50.0	0.0
Black	15	0.0	26.7	66.7	6.7
Low income	14	7.1	21.4	64.3	7.1
Not low income	31	0.0	22.6	61.3	16.1
Individualized education plan	11	9.1	72.7	18.2	0.0
No individualized education plan	34	0.0	5.9	76.5	17.6

These two schools are the only schools in a district that include the third grade. When the data for the two schools were combined at the district level, there were 10 or more students in each reporting group. The resulting data are displayed in the next example.

Example 9: Best Practices: District Counts Published with Additional Suppression

Since there were more than 10 students in each reporting subgroup at the district level, the district table based on the schools in example 8 (tables 8 and 9) was produced with full details reported for each reporting group. Table 10 displays these results.

Table 10. Number tested and percentage distribution of district-level grade 3 reading assessment results with a minimum reporting size of 10 and complementary row suppression

	Number tested	Below Basic	Basic	Proficient	Advanced
Total	75	8.0	36.0	45.3	10.6
Male	30	13.4	40.0	40.0	6.7
Female	45	4.4	33.3	48.9	13.3
White	47	6.4	38.3	40.4	14.9
Native American	12	16.7	41.7	41.7	0.0
Black	16	6.3	25.9	62.5	6.3
Low income	35	17.1	54.3	25.7	2.8
Not low income	40	0.0	20.0	62.5	17.5
Individualized education plan	20	30.0	55.0	15.0	0.0
No individualized education plan	55	0.0	29.1	56.4	14.5

But with all of the details published for school 2 and for the district, the percentage distribution across the achievement levels in each row can be converted to proportions. The proportions can then be applied to the number of students in the reporting subgroup to compute the number of students at each achievement level in each reporting group. Once this is done at the district level and for school 2, all of the suppressed data for school 1 can be recovered. For example, 38.3 percent of the 47 White third graders in the district scored at the Basic achievement level. Multiplying 0.383 times 47 shows that 18 White third graders in the district scored at the Basic achievement level. The results for White third graders in school 2 show that 10 percent of the 20 students in this subgroup scored at the Basic achievement level. Multiplying 0.10 times 20 shows that 2 White third graders in School 2 scored at the Basic achievement level. Subtracting the 2 students from School 2 from the 18 students in the district reveals the fact that there were 16 White third graders in School 1 who scored at the Basic achievement level. These 16 students comprise 59.3 percent of the 27 White third graders in school 1. These procedures were repeated to recover each of the percentages that were suppressed for school 1 in table 8. The recovered results for school 1 are shown in the shaded cells in table 11 which show that the 2 Native American third graders scored at or below Basic, the 1 Black third grader scored below Basic, and 23.8 percent of the 21 low income students scored below Basic and the other 76.2 percent scored at the Basic level. When the results for students who scored at the below Basic and Basic levels are combined to show the percent who scored below proficient, the data show disclosures of the fact that all students who were Native American, Black, or low income scored below the Proficient level. Furthermore, the parents of the 1 third grade student in school 1 with an individualized education plan who scored at the Proficient achievement level (i.e., 11.1 percent of 9 students is 1 student) know that the other third graders with individualized education plans each failed to reach the Proficient achievement level.

Table 11. School 1: Number tested and percentage distribution of grade 3 reading assessment results with suppressed percents recovered

	Number tested	Below Basic	Basic	Proficient	Advanced
Total	30	16.7	56.7	20.0	6.7
Male	12	25.0	58.3	16.7	0.0
Female	18	11.1	55.6	22.2	11.1
White	27	11.1	59.3	22.2	7.4
Native American	2	50.0	50.0	0.0	0.0
Black	1	100.0	0.0	0.0	0.0
Low income	21	23.8	76.2	0.0	0.0
Not low income	9	0.0	11.1	66.7	22.2
Individualized education plan	9	55.6	33.3	11.1	0.0
No individualized education plan	21	0.0	66.7	23.8	9.5

* Not reported to protect subgroups with fewer than 10 students.

This example illustrates the fact that it is not enough to simply suppress results at the school level, since comparisons of data published for other schools and the district can be used to recover suppressed results within a school. To avoid the recovery of suppressed school level results, the results for other schools in the district and the results for the district must also be taken into account. If the results for a specific subgroup are suppressed in at least two schools, the suppressed results for each school cannot be recovered from the results reported for other schools and the district. However, when the results are suppressed for a specific subgroup in only one school, to protect the suppressed results from recovery, the results for that subgroup must be suppressed for either another school in the district or for the district.

To protect results that are suppressed at the district level, the same precautions must be taken across district and state results. To protect suppressed results from recovery, if the results are suppressed for a specific subgroup in one district, the results

for that subgroup must be suppressed for a second district in the state.

It is important to note that this problem is not limited to applications that use complementary suppression across related subgroups. The same comparisons between district results and the results reported for other schools in the district or between state results and the results reported for other districts in the state can be applied when the results are suppressed for a single subgroup (i.e., without complementary subgroup suppression).

Care must be taken to ensure that the suppressed results for a subgroup in a single school or single district cannot be recovered using reported data for other schools in the district or other districts in the state. This can be achieved by ensuring that the results for a suppressed subgroup are suppressed in two schools. Alternatively, in districts with only one school for a grade, the results for the suppressed subgroup must also be suppressed at the district level. Similarly, the results for a suppressed subgroup must be suppressed for two districts in a state.

Reporting School-, District-, or State-Level Background Information

In reports of outcome measures, some school-, district-, or state-level reports display background information on the distribution of students in a school, district, or state in two separate summary tables. One summary table reports the total number of students enrolled and the percentage of students enrolled by grade. The second summary table reports the total number of students enrolled and the percentage of students in each of the reporting subgroups (e.g., gender, race and ethnicity, English proficiency status, migrant status, disability status, and economic status). Thus, rather than providing the exact number or percentage of students in each grade in each reporting subgroup, the report gives a portrait of the school, district, or state. However, if the number of students reported for an individual grade is the same as the number of students enrolled on the assessment date, that number, along with the report of the percentage of the students who participated in the assessment, can

be used with the percentage distribution across the achievement levels to recover the underlying numbers of students who scored at each achievement level.

Three things can be done to counter this problem. First, use background enrollment counts for a day other than that of the assessment administration and clearly label the date of the background enrollment counts and the date of the assessment in public reports to establish the fact that they are different. Second, report the percentage distribution for the background data and for the results reported across the achievement levels only in whole numbers. This decreases the precision of the reported percentages, which lowers the chance of an accurate recovery of the numbers of students in both reported and suppressed results. Third, report the percentage of students assessed as a whole number.

Example 10: Best Practices: Reporting Background Information

Table 12 provides an example of school-level data for enrollment by grade for an elementary school with grades K–6. The shaded cells are not included in the reported table, but are included here to illustrate the added protection from reporting the percentage distribution without any decimal places. For example, 4 of the 7 grades are reported as being 14

percent of the school's enrollment; the underlying data show that the more precise percentages are 13.9, 14.5, 13.6, and 14.2. The state assessment in this state is administered in March of each school year; reporting enrollment data from 5 months earlier in the school year is likely to result in some differences from the enrollment data at the time of the assessment.

Table 13 displays school-level enrollment data reported by student characteristics for the same elementary school. Again, the patterned cells are not included in the reported table. Taken together, these tables provide a profile of the school without providing the level of detail needed to recover the underlying counts for the outcome measures reported for the school.

Table 12. Elementary school enrollment, by grade

	Number	Unrounded percent	Percent
Total	359	100.0	100
Kindergarten	50	13.9	14
Grade 1	52	14.5	14
Grade 2	54	15.0	15
Grade 3	49	13.6	14
Grade 4	48	13.4	13
Grade 5	51	14.2	14
Grade 6	55	15.3	15

Table 13. Elementary school enrollment, by selected characteristics

	Number	Unrounded percent	Percent
Total	359	†	†
Male	185	51.5	52
Female	174	48.5	48
White	221	61.6	62
Black	70	19.5	19
Hispanic	59	16.4	16
Asian	*	*	*
Native American	*	*	*
Low income	100	27.9	28
Not low income	259	72.1	72
Individualized education plan	59	16.4	16
No individualized education plan	300	83.6	84
English language learner	40	11.1	11
Not English language learner	319	88.9	89

† Not applicable.
* Not reported to protect subgroups with fewer than 10 students.

Recoding the Ends of the Distribution

Another protection implemented by a number of states involves bottom or top coding the results at the tails of the percentage distribution, or for high and low rates. This is typically done by coding all percentages above 95 percent as greater than 95 percent and coding all percentages below 5 percent as less than 5 percent. This is done to avoid reporting the fact that all, or nearly all, of the students in a reporting subgroup share the same achievement level or the same outcome or that very few or none of the students have a particular outcome.

Ideally, this approach is intended to protect categories with 0 to 2 fewer than all students in a reporting category or, conversely, categories with 0 to 2 students. However, with reporting subgroups of 10 to 19 students, all of the percentages of 10 percent or less are based on only 1 student (e.g., 1 of 19 students is 5 percent and 1 of 10 students is 10 percent, while 2 of 19 is 11 percent and 2 of 10 is 20 percent). As a result, with reporting subgroups of 10 to 19 students, even reporting a category as 10 percent or less is no different than reporting that there is at most only 1 student in the category.

The extent of recoding required to protect small categories is related to the size of the subgroup, with a larger recoded range required for smaller subgroups. At a minimum, results should not be published for outcomes based on the experiences of 1 student. The goal is to ensure that each recoded percent could include at least 2 students. Additional protection is provided by including counts of students in the range of recoded percentages where the recoded percent could include at least 3 students (i.e., the threshold rule of 3). For example, in reporting outcome measures for subgroups of 10 to 20, recoding the ends of the distribution to 20 percent or less and 80 percent or more would result in recoding all percentages for categories based on 0 to 2 students (i.e., 20 percent of 10 is 2).[6] In addition, categories of 3 students would be included in the recoded category when there are 15 or more students in the subgroup (i.e., 3 out of 15 is 20 percent).

In reporting outcome measures for groups of 21 to 40, recoding the ends of the distribution to 10 percent or less and 90 percent or more would result in recoding all percentages based on categories of 0 to 2 students. In this recode, categories of 3 students would be included in the recoded category when there are 30 or more students in the subgroup (i.e., 3 out of 30 is 10 percent).

When there are 41 to 100 students, recoding the ends of the distribution to 5 percent or less and 95 percent or more ensures results based on 0 to 2 students when there are 41 students and 0 to 4 students when there are 100 students (above 59 students, this recode would include categories of 3 students). Similarly, for groups of 101 to 300 students, recoding the ends of the distribution to 2 percent or less and 98 percent or more ensures reporting results based on 0 to 2 students when there are 101 students and 0 to 6 students when there are 300 students (above 149 students this recode includes categories of 3 students) . Finally, for groups of more than 300 students, recoding the ends of the distribution to 1 percent or less and 99 percent or more ensures results based on 0 to 3 students at a minimum

Recoding the percentages at one end of a percentage distribution is not necessarily enough to protect the original contents of the recoded category, since the sum of the reported categories subtracted from 100 percent yields the percent that was recoded.

To protect the recoded categories, additional recoding is needed. For groups of 10 to 20 students, the results should be collapsed into two categories and percentages between 21 and 79 should be reported in 10 percentage point ranges. For groups of 21 to 40 students, the percentages in categories of an outcome measure should be recoded in 10 percentage point ranges. For groups of 41 to 200 students, the percentages in categories of an outcome measure should be recoded in 5 percentage point ranges. For groups of 201 or more students, reporting the percentages in categories of an outcome measure as whole numbers provides sufficient recoding (i.e. there are at least 2 counts that could yield each reported percent).

To further protect small categories, if one subgroup includes 200 or fewer students, any related subgroups (i.e., those that combine to sum to the total) with more than 200 students should be recoded using the ranges for 200 students.

---

[6] Reporting results based on fewer than 10 students while ensuring that there could be at least 2 students in a reported category requires more extensive top and bottom coding and would limit the number of reportable outcomes to a small enough set of possible outcomes that they would not be well protected. For example, with results based on 6 students, 2 students account for 33 percent, and recodes of 33 and 67 percent leave only 1 response option that could be reported. Similarly, with 7 students, the recodes would be 29 and 71 percent, leaving 2 response options for reporting; with 8 students, the recodes would be 25 and 75 percent, leaving 3 response options for reporting; and with 9 students, the recodes would be 22 and 78 percent, leaving only 5 response options for reporting.

Example 11: Best Practices: Recoding the Distribution

Table 14 in this example shows the number of students and the actual and recoded percentage distributions for the school-level third-grade reading assessment results for 32 students for this reporting option. The shaded cells are not publicly reported. Table 14 displays the data with reporting subgroups less than 10 suppressed and the categories of other subgroups recoded to protect small categories. For the overall results of the 32 students, each category is recoded into a 10 percentage point range to protect small categories in the subgroups in the table. Given that there are only 10 students in the Hispanic subgroup, the 0 in the Advanced category is combined with the 10 percent in the proficient category and recoded to less than or equal to ($\leq$) 20 percent at or above proficient, and the 50 percent at the Basic level is combined with the 40 percent at the Below Basic level and recoded to greater than or equal to 80 percent. The data for the 22 White students are recoded, with the 0 percent in the Below Basic category recoded to less than or equal to 10 percent and the other three categories recoded into 10 percentage point ranges. Since there are fewer than 10

students with individualized education plans, the data for this subgroup and the data for students who do not have individualized education plan are suppressed. The outcome measures for the 12 English language learners and the subgroup of 20 students who are not English language learners are reported for those students scoring at the proficient or above level and those performing at or Below the Basic level.

Table 15 follows the same format and shows the results for the district-level third-grade reading assessment results. With 320 students in the group, the results for the 3 students in the advanced category that account for 1 percent of the total are recoded to less than or equal to ($\leq$) 1 percent, and the other three categories are reported as percentages that are rounded to whole numbers. With 198 White students and 122 Hispanic students, the results for the 3 Advanced students in the White subgroup and for 0 Advanced students in the Hispanic subgroup are both recoded to less than or equal to ($\leq$) 2 percent, and the other three categories in each subgroup are recoded into 5 percentage point ranges. With 40

students with individualized education plans, the Advanced category for these students is recoded to less than or equal to ($\leq$) 10 percent, and the remaining categories are recoded into 10 percentage point ranges. The data for the 280 students in the related subgroup who do not have individualized education plans are recoded following the procedures that apply to 200 students, with the 1 percent at the Advanced level recoded to less than or equal to ($\leq$) 2 percent and the other three categories recoded into 5 percentage point ranges. Finally, because there are only 12 students who are English language learners, the Advanced category for these students is combined with the Proficient category and reported as 21 to 29 percent, and the Below Basic and Basic categories are combined and reported as 70 to 79 percent. The data for the 308 students in the related subgroup who are not English language learners are recoded, with the percent at the Advanced level reported as less than or equal to ($\leq$) 2 percent and the other three categories recoded into 5 percentage point ranges.

Table 14.   School-level grade 3 reading assessment results for a state with a minimum reporting size of 10

		Percent assessed	Tested	Below Basic	Basic	Proficient	Advanced	
Total	N	†	32	4	10	11	7	
	%	100	100	13	31	34	22	Actual
	%	100	100	11–19	30–39	30–39	20–29	Reported
White	N	†	22	0	5	10	7	
	%	100	100	0	23	45	32	Actual
	%	100	100	≤10	21–29	40–49	30–39	Reported
Hispanic	N	†	10	4	5	1	0	
	%	100	100	40	50	10	0	Actual
	%	100	100	†	≥80	≤20	†	Reported
Individualized education plan	N	†	7	4	3	0	0	
	%	100	*	*	*	*	*	<10
	%	100	*	*	*	*	*	<10
No individualized education plan	N	†	25	0	7	11	7	
	%	100	100	0	28	44	28	Actual
	%	100	*	*	*	*	*	Suppressed
English language learner	N	†	12	4	5	2	1	
	%	100	100	33	42	17	8	Actual
	%	100	100	†	70–79	21–29	†	Reported
Not English language learner	N	†	20	0	5	9	6	
	%	100	100	0	25	45	30	Actual
	%	100	100	†	21–29	70–79	†	Reported

† Not applicable.
* Not reported to protect subgroups with fewer than 10 students.
NOTE: Details may not sum to totals because of rounding and recoding.

Table 15. District level, Grade 3 reading assessment results for a state with a minimum reporting size of 10

		Percent assessed	Tested	Below Basic	Basic	Proficient	Advanced	
Total	N		320	40	167	110	3	
	%	100	†	13	52	34	1	Actual
	%	100	†	13	52	34	≤1	Reported
White	N		198	0	105	90	3	
	%	100	†	0	53	45	2	Actual
	%	100	†	≤2	50–54	45–49	≤2	Reported
Hispanic	N		122	40	62	20	0	
	%	100	†	33	51	16	0	Actual
	%	100	†	25–29	50–54	15–19	≤2	Reported
Individualized education plan	N		40	25	15	0	0	
	%	100	†	63	38	0	0	Actual
	%	100	†	60–69	30–39	≤10	≤10	Reported
No individualized education plan	N		280	15	152	110	3	
	%	100	†	5	54	39	1	Actual
	%	100	†	5–9	50–54	35–39	≤2	Reported
English language learner	N		12	4	5	2	1	
	%	100	†	33	42	17	8	Actual
	%	100	†	†	70–79	21–29	†	Reported
Not English language learner	N		308	36	162	108	2	
	%	100	†	12	53	35	1	Actual
	%	100	†	10–14	50–54	35–39	≤2	Reported

† Not applicable.
NOTE: Details may not sum to totals because of rounding and recoding.

Recommendations

This review and analysis of current reporting practices illustrates that some practices work better than others in protecting suppressed results and, thus, in protecting against disclosures of personally identifiable information about individual students. It is important to note that each of the practices requires some loss of information. The challenge rests in identifying practices that protect information about individual students while minimizing the negative impact on the utility of the publicly reported data. Drawing upon the review and analysis presented in this brief leads to recommended reporting rules to be used in producing reports of percentages and rates to describe student outcomes to the public. These rules are intended for use in the public release of new data.

Rules 1 through 4 and 6 and 7 are general reporting rules. Rule 5 is guided by the number of students in the reporting group or subgroups; the underlying principle is that the amount of detail that can be reported while protecting each student's privacy is related to the number of students in a reporting group or subgroup—that is, more detail can be reported for larger groups. Rule 5a applies to instances in which there are more than 300 students in each of a set of related reporting subgroups (e.g., in each race/ethnicity group, for students with and without an individualized education plan, for students receiving or not receiving instruction as an English language learner). Rule 5b applies to instances in which the smallest reporting subgroup within a set of related reporting subgroups has 201 to 300 students. Rule 5c applies to instances in which the smallest reporting subgroup within a set of related reporting subgroups has 101 to 200 students. Rule 5d applies when the smallest reporting subgroup in a set of related subgroups has 41 to 100 students. Rule 5e applies when the smallest reporting subgroup in a set of related subgroups has 21 to 40 students. Rule5f applies when the smallest reporting subgroup in a set of related subgroups has 10 to 20 students.

Reporting Rules

1. Minimize the amount of enrollment details reported in the profile of the school, district, or state in reports of outcome measure results. If possible, use enrollment data for a different date than that of the reported outcome measures and label the different dates (e.g., report enrollment data for a date different from the assessment date, such as fall enrollment for a spring assessment). In so doing, tell the readers that the data on student enrollment by grade and by selected student characteristics are included to provide context for the results presented but should not be assumed to exactly match the student composition at the time the outcome was measured.

    a. Report the percentage distribution of students by grade at the school, district, or state level in a standalone table without any of the outcome measures or reporting subgroup details.

    b. Report the percentage distribution of students by reporting subgroup at the school, district, or state level in a standalone table without any of the outcome measures or enrollment by grade details.

    c. Do not report the details of the enrollment data within each reporting subgroup by individual grades.

4. Use a minimum of 10 students for the reporting subgroup size limitation.

    a. Suppress results for all reporting groups with 0 to 9 students.

    b. Suppress results for reporting subgroups with 0 to 9 students and suppress each of the related reporting subgroups regardless of the number of students in the subgroup (i.e., suppress the other subgroup(s) of the set of subgroups that sum to the overall group). In instances with 3 or more subgroups, the subgroups with 0 to 9 students can be combined with each other or with the smallest reportable subgroup to form an aggregated subgroup of 10 or more students to allow for the reporting of data for larger subgroups.

3. Use only whole numbers when reporting the percentage of students for each category of an outcome measure (e.g., the percentage assessed).

4. Do not report the underlying counts for the subgroup or group totals (i.e., the denominators of the percentages); also do not report the underlying counts of students in individual outcome categories (i.e., the numerators).

5. **To implement the next step in the data protection procedure in the remaining reporting groups and subgroups, the approach used is determined by the number of students in the smallest reporting subgroup among a set of related groups or subgroups (i.e., groups that in combination sum to the total). To protect student privacy:**

   a. **For reporting variables/outcome measures with more than 300 students and no related subgroup with fewer than 200 students, use the following approach:**

      i. Recode categories with values of 99 to 100 percent to greater than or equal to 99 percent (≥ 99 percent).

      ii. Recode categories with values of 0 to 1 percent to less than or equal to 1 percent (≤ 1 percent).

      iii. Otherwise, report the percentage of students in each category using whole numbers.

   b. **For reporting variables/outcome measures with 201 to 300 students and no related subgroup with fewer than 200 students, use the following approach:**

      i. Recode categories with values of 98 to 100 percent to greater than or equal to 98 percent (≥ 98 percent).

      ii. Recode categories with values of 0 to 2 percent to less than or equal to 2 percent (≤ 2 percent).

      iii. Otherwise, report the percentage of students in each category using whole numbers.

   c. **For reporting variables/outcome measures in which the number of students ranges from 101 to 200,** use the following option in this group and all related subgroups with more than 200 students:

      i. Recode categories with values of 98 to 100 percent to greater than or equal to 98 percent (≥ 98 percent).

      ii. Recode categories with values of 0 to 2 percent to less than or equal to 20 percent (≤ 2 percent).

      iii. Recode the percentage in each remaining category in all reporting groups or subgroups to intervals as follows (3–4, 5–9, 10–14, 15–19, . . ., 85–89, 90–94, 95–97).

   d. **For reporting variables/outcome measures in which the number of students in the smallest reporting group or subgroup ranges from 41 to 100,** use the following option in that group or subgroup and use option 5c for each related reporting group or subgroup with more than 100 students:

      i. Recode categories with values of 95 to 100 percent to greater than or equal to 95 percent (≥ 95 percent).

      ii. Recode categories with values of 0 to 5 percent to less than or equal to 5 percent (≤ 5 percent).

      iii. Recode the percentage in each remaining category in all reporting groups or subgroups to intervals as follows (6–9,10–14, 15–19, 20–24, . . ., 85–89, 90–94).

e. **For reporting variables/outcome measures in which the number of students in the smallest reporting group or subgroup ranges from 21 to 40,** use the following option for that group or subgroup, use option 5d for each related reporting group or subgroup with 41 to 100 students, and use option 5c for those with more than 100 students:

   i. Recode categories with values of 90 to 100 percent to greater than or equal to 90 percent (≥ 90 percent).

   ii. Recode categories with values of 0 to 10 percent to less than or equal to 10 percent (≤ 10 percent).

   iii. Recode the percentage in each remaining category in all reporting groups or subgroups to intervals as follows (11–19, 20–29, . . ., 80–89).

f. **For reporting variables with 10 to 20 students in the smallest subgroup,** use the following option for that group or subgroup, use option 5e for each related group or subgroup with 21 to 40 students, use option 5d for those with 41 to 100 students, and use option 5c for those with more than 100 students:

   i. Collapse all outcome measures to only two categories, using the same collapsing rules across all subgroups for each outcome measure (e.g., assessment results collapsed to below the proficient level and at or above the proficient level by sex, racial and ethnic groups, disability status, etc.).

   ii. Recode categories with values of 0 to 20 percent to less than or equal to 20 percent (≤ 20 percent), and recode the other category to greater than 80 percent (> 80 percent).

   iii. If both collapsed categories have percents of 21 to 79 percent, recode the percentage in each collapsed category to intervals as follows (21–29, 30–39, . . ., 70–79).

6. For each outcome measure reported at the district level, if results for a group or subgroup have been collapsed, recoded, or suppressed in only one school in the district, apply the same collapsing, recoding, or suppression rule for that group or subgroup in a second school or at the district level (i.e., for any specific measure and group or subgroup, there must be either no school-level data suppressed for a specific subgroup or the data for that subgroup must be suppressed for at least 2 schools or for one school and the district).

7. For each outcome measure reported at the state level, if results for a group or subgroup have been collapsed, recoded, or suppressed in only one district in the state, apply the same collapsing, recoding, or suppression rule for that group or subgroup in a second district (i.e., for any specific measure and group or subgroup, there must be either no district-level data suppressed for a specific subgroup or the data for that subgroup must be recoded or suppressed for at least 2 districts).

Summary

This Brief discusses the potential for the disclosure of personally identifiable information in summary school-, district-, and state-level reports from education records using current reporting practices. Building on current best practices, the Brief outlines reporting recommendations. Primarily, the goal of these reporting recommendations is to maximize the reporting of student outcomes while protecting students' personally identifiable information.

While it would be easier to have only one set of reporting recommendations, the reporting rules are intended to maximize the amount of detail that can be safely reported without allowing the disclosure of student outcome measure categories based on small numbers of students. A secondary goal of these recommendations is to maximize uniformity in reporting practices across states in order to facilitate cross-state comparisons.

The recommendation to provide data on enrollment by grade and enrollment by student characteristics that are not identical to those for the day the outcome is measured is intended to prevent the statistical manipulation of the data to recover protected student information. However, this may not always be possible, and in some instances, these data may not change over the course of a school year. Thus, the reporting rules that are linked to the number of students included in a subgroup are intended to add additional protections by ensuring that, if the subgroup size is known, each reported category could include at least two students. Further, if the subgroup size is not known, each reported category could include at least three students.

There are multiple approaches to statistical data protection. The recommendations here were selected with the goal of maximizing the amount of information that can be released while protecting personally identifiable student information through a relatively straightforward set of rules that can be easily implemented. For those readers wanting to read further on the topic of statistical data protection, please see Duncan et. al. (1993) *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics;* Willenborg and de Waal (2001) *Statistical Disclosure Control in Practice;* Federal Committee on Statistical Methodology Working Paper 22, *Report on Statistical Disclosure Limitation Methodology;* and the American Statistical Association, Committee on Privacy and Confidentiality website, *Key Terms/Definitions in Privacy and Confidentiality.*

NCES welcomes input on these recommendations.

References

American Statistical Association, Committee on Privacy and Confidentiality. *Key Terms/ Definitions in Privacy and Confidentiality.* Alexandria, VA: Retrieved from http://www. amstat.org/committees/pc/keyterms.html on 6/17/2010.

Code of Federal Regulations, Title 34—Education, Part 200. *Improving the Academic Achievement of the Disadvantaged, Section 200.7, Disaggregation of Data,* (34CFR200.7). Washington, DC: GPO Access CFR. Retrieved from http://edocket.access.gpo.gov/cfr_2010/ julqtr/34cfr200.7.htm on 10/10/2010.

Code of Federal Regulations, Title 34—Education, Part 99. *Family Educational and Privacy Rights,* (34CFR99). Washington, DC: GPO Access e-CFR. Retrieved from http://ecfr.gpoaccess.gov/ cgi/t/text/text-idx?c=ecfr&sid=44d350c26fb9c ba4a156bf805f297c9e&tpl=/ecfrbrowse/ Title34/34cfr99_main_02.tpl.

Duncan, George T., Jabine, Thomas B. and de Wolf, Virginia A., Editors. (1993). *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics.* Panel on Confidentiality and Data Access, National Research Council. Washington, DC: National Academy Press.

Federal Register, Office of Management and Budget, *Implementation Guidance for Title V of the E-Government Act, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA),* Washington DC: Vol. 72, No. 115 / Friday, June 15, 2007. Retrieved from http://www.whitehouse.gov/sites/default/files/ omb/assets/omb/fedreg/2007/061507_cipsea_ guidance.pdf on 9/9/2010.

Office of Management and Budget, Federal Committee on Statistical Methodology, (2005). Statistical Policy Working Paper 22, *Report on Statistical Disclosure Limitation Methodology.*

Retrieved from http://www.fcsm.gov/ workingpapers/spwp22.html on 9/9/2010.

Office of Management and Budget, OMB Memorandum M-03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002.* Retrieved from http://www.whitehouse.gov/omb/memoranda_ m03-22/ on 9/9/2010.

Office of Management and Budget, OMB Memorandum M-07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information.* Retrieved from http:// www.whitehouse.gov/sites/default/files/omb/ assets/omb/memoranda/fy2007/m07-16.pdf on 9/9/2010.

U.S. Code, Title 20—Education, Chapter 31— General Provisions Concerning Education, Subchapter III—General Requirements and Conditions Concerning Operation and Administration of Education Programs: General Authority of Secretary, Part 4—Records, Privacy, Limitation on Withholding Federal funds, Section 1232g. *Family Educational and Privacy Rights,* (20USC1232g). Washington, DC: GPO Access. Retrieved from http://frwebgate4.access. gpo.gov/cgi-bin/TEXTgate.cgi?WAISdocID=799 486197532+0+1+0&WAISaction=retrieve.

U.S. Code, Title 20—Education, Chapter 70—Strengthening and Improvement of Elementary and Secondary Schools, Subchapter I—Improving the Academic Achievement of the Disadvantaged, Part A—Improving Basic Programs Operated by Local Educational Agencies, Subpart 1—Basic Program Requirements, Section 6311. *State Plans,* (20USC6311). Washington, DC: GPO Access. Retrieved from http://frwebgate2.access.gpo. gov/cgi-bin/TEXTgate.cgi?WAISdocID=bULwJ H/21/1/0&WAISaction=retrieve.

U.S. Code, Title 20—Education, Chapter 76— Education Research, Statistics, Evaluation, Information, and Dissemination, Subchapter II—Educational Technical Assistance, Section 9607. *Grant Program for Statewide, Longitudinal Data Systems,* (20USC9607). Washington, DC: GPO Access. Retrieved from http://frwebgate3.access.gpo.gov/cgi-bin/ TEXTgate.cgi?WAISdocID=FKr6BA/0/1/0&WA ISaction=retrieve on 9/9/2010.

Public Law 110-69, America Competes Act, Title VI—Education, Section 6401. Washington, DC: GPO Access. Retrieved from http://frwebgate. access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_ cong_public_laws&docid=f:publ069.110 on 9/10/2010.

Public Law, 111-05, American Recovery and Reinvestment Act, Title VIII—Education, Institute of Education Sciences. Washington, DC: GPO Access. Retrieved from http:// frwebgate.access.gpo.gov/cgi-bin/ getdoc.cgi?dbname=111_cong_public_ laws&docid=f:publ005.111 on 9/10/2010.

Willenborg, L., and De Waal, T. (2001). *Elements of Statistical Disclosure Control,* Vol. 155, Lecture Notes in Statistics, New York, NY: Springer.