Privacy Violations in Accountability Data Released to the Public by State Educational Agencies

Krish Muralidhar
Gatton Research Professor, Gatton College of Business & Economics, University of Kentucky, Lexington KY 40506
Rathindra Sarathy
Ardmore Professor, Spears College of Business, Oklahoma State University, Stillwater OK 74078

Abstract

As a part of the No Child Left Behind Act of 2001, every state is required to release information to the public regarding the performance of schools in their state. In accordance with this requirement, education agencies at practically every state now release information either at the district or school level. When releasing data, the education agencies must also ensure that the released data does not violate the requirements of the Family Educational Rights and Privacy Act (FERPA) of 1974. Except in special circumstances, FERPA specifically requires that individual performance information regarding students cannot be released without the written permission. In this study, we show that the current methodologies employed by many state educational agencies across the country do not satisfy the FERPA privacy requirements. Using accountability data from several states, we illustrate the violations of privacy that occur when data is released at the school or district level and show that individual performance information for individual students and/or small (primarily ethnic, gender, or other disadvantaged) subgroups are easily computed using the data that is released to the public. We describe the efforts of the office of Assessment and Accountability at the Kentucky Department of Education to identify these privacy violations. It is likely that education agencies continue to release detailed performance information even if they are not required by law. The analysis provided in this paper will enable education agencies to provide useful data to the public without violating the privacy of individuals or small subgroups.

Introduction

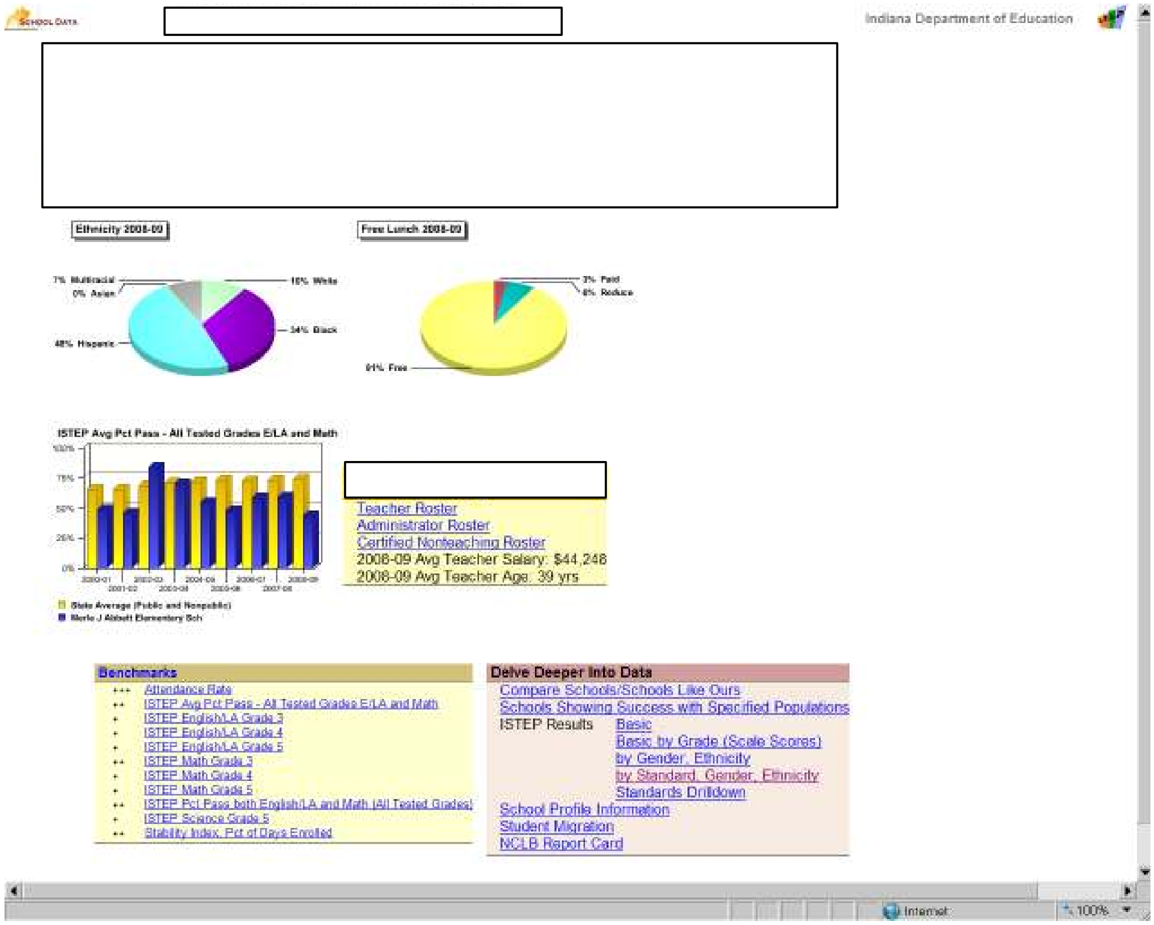
As a part of the No Child Left Behind Act (NCLB) of 2001, every state is required to release information to the public regarding the performance of schools in their state. In accordance with this requirement, education agencies at practically every state now release information either at the district or school level. While many state agencies have released performance information even before NCLB, the specific requirements in NCLB provided a more detailed structure to the reports that were released. Many states also have their own requirements for performance data that must be released to the public. Some of these requirements predate NCLB.

In accordance with the NCLB and other requirements, many state educational agencies release aggregate performance data. The level of detail in the reports and the unit level at which this information is released vary by state. One of the most detailed reports is provided by the Indiana Department of Education on their web site. Referred to as Indiana Accountability Systems for Academic Progress, this web site provides extensive information on student performance.

Of particular interest is the information provided to the public at the school level with the following information:

> The **School Data** section provides informational, demographic, and achievement data about Indiana Schools. You can disaggregate the data using multiple variables and graphically display the results. This section also allows for comparison to other similar-schools and links to possible strategies for improvement.

An example of the data provided by the Indiana Department of Education on their web site is provided below for a particular school. In this and in all other examples, we have intentionally hidden all indentifying information.



[redacted]

[redacted]

[redacted]

Teacher Roster
Administrator Roster
Certified Nonteaching Roster
2008-09 Avg Teacher Salary: $44,248
2008-09 Avg Teacher Age: 39 yrs

Benchmarks

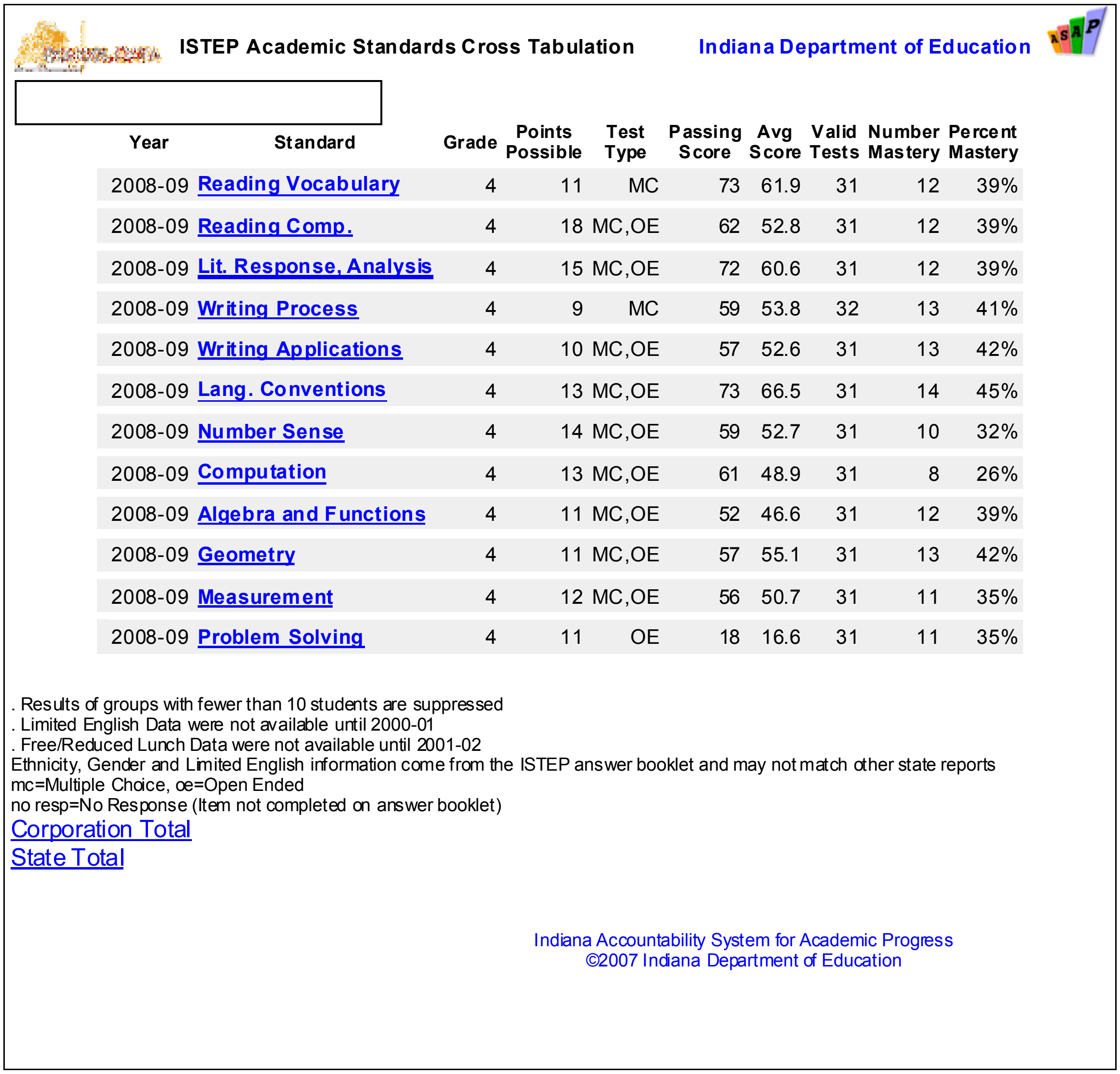
+++ Attendance Rate
++ ISTEP Avg Pct Pass - All Tested Grades E/LA and Math
+ ISTEP English/LA Grade 3
+ ISTEP English/LA Grade 4
+ ISTEP English/LA Grade 5
++ ISTEP Math Grade 3
+ ISTEP Math Grade 4
+ ISTEP Math Grade 5
++ ISTEP Pct Pass both English/LA and Math (All Tested Grades)
+ ISTEP Science Grade 5
++ Stability Index, Pct of Days Enrolled

Delve Deeper Into Data

Compare Schools/Schools Like Ours
Schools Showing Success with Specified Populations
ISTEP Results Basic
Basic by Grade (Scale Scores)
by Gender, Ethnicity
by Standard, Gender, Ethnicity
Standards Drilldown
School Profile Information
Student Migration
NCLB Report Card

As you can see from the “benchmarks”, the report provides extensive information regarding the performance of students at this particular school by grade. In addition, the web site also provides the ability to “delve deeper into data” by being able to view the results for subsets of individuals. Students can be divided into subsets of Gender, Ethnicity, whether they receive free lunch or not, Special Education, and Limited English proficiency. The web site also offers us the ability to identify a student by all combinations of the above.

In terms of the details of the actual scores, the Indiana Department of Education not only provides a count of the number of students who passed or failed a subject (such as Math or Reading), but also provides detailed breakdowns in terms of individual standards within each test. For example, the math test is broken down into Number Sense, Computation, Algebra and Functions, Geometry, Measurement, Data Analysis and Prob., and Problem Solving. The average score in each of these standards is also provided.



ISTEP Academic Standards Cross Tabulation

Indiana Department of Education

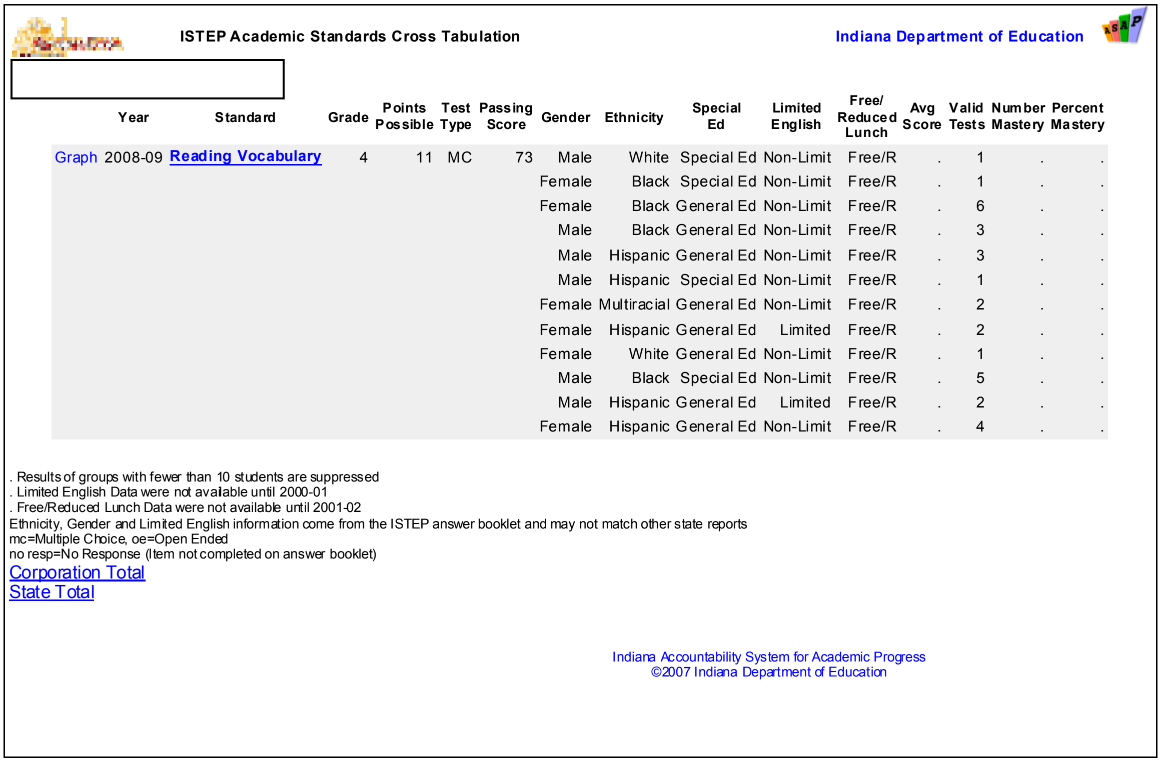
Year	Standard	Grade	Points Possible	Test Type	Passing Score	Avg Score	Valid Tests	Number Mastery	Percent Mastery
2008-09	Reading Vocabulary	4	11	MC	73	61.9	31	12	39%
2008-09	Reading Comp.	4	18	MC,OE	62	52.8	31	12	39%
2008-09	Lit. Response, Analysis	4	15	MC,OE	72	60.6	31	12	39%
2008-09	Writing Process	4	9	MC	59	53.8	32	13	41%
2008-09	Writing Applications	4	10	MC,OE	57	52.6	31	13	42%
2008-09	Lang. Conventions	4	13	MC,OE	73	66.5	31	14	45%
2008-09	Number Sense	4	14	MC,OE	59	52.7	31	10	32%
2008-09	Computation	4	13	MC,OE	61	48.9	31	8	26%
2008-09	Algebra and Functions	4	11	MC,OE	52	46.6	31	12	39%
2008-09	Geometry	4	11	MC,OE	57	55.1	31	13	42%
2008-09	Measurement	4	12	MC,OE	56	50.7	31	11	35%
2008-09	Problem Solving	4	11	OE	18	16.6	31	11	35%

. Results of groups with fewer than 10 students are suppressed
. Limited English Data were not available until 2000-01
. Free/Reduced Lunch Data were not available until 2001-02
Ethnicity, Gender and Limited English information come from the ISTEP answer booklet and may not match other state reports
mc=Multiple Choice, oe=Open Ended
no resp=No Response (Item not completed on answer booklet)

Corporation Total
State Total

Indiana Accountability System for Academic Progress
©2007 Indiana Department of Education

On top of all this, the Indiana Department of Education also provides the ability to break down the student population by Gender, Ethnicity, Free Lunch, Special Education, and Limited English Proficiency. Hence, the 31 students in grade 4 above can be further broken down into subgroups formed *by any combination* of the categories above. The result is that even large classes can be clustered into small subgroups. For the same data above, the following table provides the students broken down in smaller subgroups for a single performance category (Reading Vocabulary).



ISTEP Academic Standards Cross Tabulation

Indiana Department of Education

	Year	Standard	Grade	Points Possible	Test Type	Passing Score	Gender	Ethnicity	Special Ed	Limited English	Free/ Reduced Lunch	Avg Score	Valid Tests	Number Mastery	Percent Mastery
Graph	2008-09	Reading Vocabulary	4	11	MC	73	Male	White	Special Ed	Non-Limit	Free/R	.	1	.	.
							Female	Black	Special Ed	Non-Limit	Free/R	.	1	.	.
							Female	Black	General Ed	Non-Limit	Free/R	.	6	.	.
							Male	Black	General Ed	Non-Limit	Free/R	.	3	.	.
							Male	Hispanic	General Ed	Non-Limit	Free/R	.	3	.	.
							Male	Hispanic	Special Ed	Non-Limit	Free/R	.	1	.	.
							Female	Multiracial	General Ed	Non-Limit	Free/R	.	2	.	.
							Female	Hispanic	General Ed	Limited	Free/R	.	2	.	.
							Female	White	General Ed	Non-Limit	Free/R	.	1	.	.
							Male	Black	Special Ed	Non-Limit	Free/R	.	5	.	.
							Male	Hispanic	General Ed	Limited	Free/R	.	2	.	.
							Female	Hispanic	General Ed	Non-Limit	Free/R	.	4	.	.

. Results of groups with fewer than 10 students are suppressed
. Limited English Data were not available until 2000-01
. Free/Reduced Lunch Data were not available until 2001-02
Ethnicity, Gender and Limited English information come from the ISTEP answer booklet and may not match other state reports
mc=Multiple Choice, oe=Open Ended
no resp=No Response (Item not completed on answer booklet)
Corporation Total
State Total

Indiana Accountability System for Academic Progress
©2007 Indiana Department of Education

Note that a large group of students (31 in total) now result in subgroups *all of which are less than 10*. By using these breakdowns, we are also able to identify many students individually. For example, there is only one White Male student who is in a Special Education class. Every student is now classified into a small subgroup allowing many of them to be "individually identified."

The Indiana Department of Education is to be commended on their effort to provide maximum information to the public. The information provided is indeed very useful to the public in evaluating the performance of Indiana schools. This very ability to "drill down" the details of the performance of subgroups of students also raises the danger that the information that is released could potentially lead to disclosure of performance information regarding individuals or small subgroups. Unfortunately, in the their desire to provide detailed performance information, the Indiana Department of Education seems to have *ignored the fact that they are also required to prevent the disclosure of individually identifiable student performance data*.

The Family Educational Rights and Privacy Act (FERPA) explicitly requires all agencies releasing performance data to ensure that the privacy of individuals or small subgroups is not violated by the release of such information. In this study we seek to illustrate that current procedures implemented to ensure the privacy requirements may not satisfy the standards set by FERPA. Using publicly available data, we show that the actual performance of individuals or small subgroups can be easily inferred from the data.

Disclosure of Performance Identifiable Information

In this section, we attempt to define what constitutes disclosure of personally identifiable information. Not being lawyers, we do not claim to provide a legal definition although we believe that our definition is likely to be more liberal; it is likely that the legal definition would be far more conservative. Most of these definitions have been developed based on the information provided at the United States Department of

Education web site (www.ed.gov) and two recent Federal Registers (hereafter FR1[1] and FR2[2]) relating to FERPA.

The first definition deals with the application of FERPA and can be found on the Department of Education website:

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

The privacy requirement based on FERPA is defined in the FR1 (page 15576) as follows:

> Statute: 20 U.S.C. 1232g(b)(1) and (b)(2) provides that an educational agency or institution subject to FERPA may not have a policy or practice of releasing, permitting the release of, or providing access to personally identifiable information from education records without prior written consent.

In this context, FR1 (page 15576) defines disclosure as follows:

> The regulations in § 99.3 define the term disclosure to mean permitting access to or the release, transfer, or other communication of personally identifiable information from education records to any party by any means. The regulations do not address issues relating to the return of records to the party that provided or created them.

Personally identifiable information is defined as follows (FR1, page 15583):

> Personally identifiable information is defined in § 99.3 to mean information that can be used to identify a student, including direct identifiers, such as the student's name, SSN, and biometric records, alone or combined with other personal or identifying information that is linked or linkable to a specific individual, including indirect identifiers such as the name of the student's parent or other family member, the student's or family's address, and the student's date and place of birth and mother's maiden name, that would allow a reasonable person in the school or its community, who does not have personal knowledge of the relevant circumstance, to identify the student with reasonable certainty. The Department does not hold educational agencies and institutions responsible for knowing the status of all non-educational records about students (e.g., law enforcement or hospital records). However, the Department encourages educational agencies and institutions to be sensitive to publicly available data on students and to the cumulative effect of disclosures of student data.

In FR2 (page 74833), the term "school or its community" was modified as "school community." It is interesting to note that by this definition, educational agencies are required to be sensitive to data that is available from others sources and to *cumulative effect* of disclosures. In addition, FR2 (page 74833) also provides the following definition for indirect identifiers:

> In order to provide maximum flexibility to educational agencies and institutions, we did not attempt to define or list all other ''indirect identifiers''. We believe that the examples listed in paragraph (3) of the definition of personally identifiable information—date of birth, place of birth, and mother's maiden name – indicate clearly the kind of information that could identify a student. Race and ethnicity, for example, could also be indirect

[1] Federal Register, Vol. 73, No. 57, Department of Education, 34 CFR Part 99, Family Educational Rights and Privacy, Proposed Rule, Monday, March 24, 2008.

[2] Federal Register, Vol. 73, No. 237, Department of Education, 34 CFR Part 99, Family Educational Rights and Privacy, Final Rule, Tuesday, December 9, 2008.

> identifiers. It is not possible, however, to list all the possible indirect identifiers and ways in which information might indirectly identify a student.

The definition of a de-identified record is provided in FR1 (page 15583) as follows:

> The proposed regulations would amend § 99.31(b) to provide objective standards under which educational agencies and institutions may release, without consent, education records, or information from education records, that has been de-identified through the removal of all personally identifiable information.

The release of aggregated data of the nature that we are discussing in this paper would certainly fall under this category since all identifiers are removed. However, just because all identifiers have been removed does not mean that the record cannot be re-identified as belonging to a particular student as the following statement in FR1 (page 15583) observes:

> The Department recognizes that avoiding the risk of disclosure of identity or individual attributes in statistical information cannot be completely eliminated, at least not without negating the utility of the information, and is always a matter of analyzing and balancing risk so that the risk of disclosure is very low. The reasonable certainty standard in the proposed definition of personally identifiable information requires such a balancing test.

In other words, these regulations make it clear that when data is released, the agency releasing the data must make some efforts to consider if the release of such data would lead to disclosure of personally identifiable information. If the agency determines that releasing the data would result in the disclosure of personally identifiable information, then the agency must take necessary precautions in order to prevent such disclosure. FR2 (page 74835) directly addresses this issue:

> In response to requests for guidance on what specific steps and methods should be used to de-identify information (and as noted in the preamble to the NPRM, 73 FR 15584), it is not possible to prescribe or identify a single method to minimize the risk of disclosing personally identifiable information in redacted records or statistical information that will apply in every circumstance, including determining whether defining a minimum cell size is an appropriate means to protect the confidentiality of aggregated data and, if so, election of an appropriate number.

On the same page, the issue of releasing data for NCLB requirements, FR2 (page 74835) states:

> With regard to issues with NCLB reporting in particular, determining the minimum cell size to ensure statistical reliability of information is a completely different analysis than that used to determine the appropriate minimum cell size to ensure confidentiality.

The discussion in FR2 refers the readers to Statistical Policy Working Paper #22 titled “Report on Statistical Disclosure Methodology” which can be found at http://www.fcsm.gov/working-papers/spwp22.html for further guidance on this issue.

The discussion following these definitions also provides a simple description of a situation that would constitute disclosure of personally identifiable information. Assume for the sake of argument that in a school with 100 Female Hispanic students, one female Hispanic student failed to graduate. FR2 (page 74835) provides the following discussion:

> Simply knowing that one out of 100 Hispanic females failed to graduate does not identify which of the Hispanic females it might be. But suppose this female is an English language learner who is also enrolled in special education classes. The school also publishes tables on participation in special education classes by race, ethnicity, and grade, and tables that include the graduation status of Hispanic females disaggregated in one table by English language proficiency status, and by participation in special education

> classes in another. Suppose that these three tabulations each show separately that there is one 12th grade Hispanic female enrolled in special education classes, that the one Hispanic female who did not graduate was enrolled in special education classes, and that the one Hispanic female who did not graduate was an English language learner. With this information, the discerning observer knows that the one Hispanic female who failed to graduate is an English language learner and that she was the only 12th grade Hispanic student enrolled in special education classes. Any number of people in the school would be able to identify the Hispanic female who did not graduate with these three pieces of information.

In this case, the regulations actually provide a nice example of what would constitute disclosure. While the above example dealt with a single student, the discussion goes on to say that if 3 students can be identified in the same manner, it would constitute disclosure as well.

Thus, a close reading of the proposed and final rules clearly indicate that, in releasing any data relating to students, educational agencies must be careful to evaluate whether releasing such data could lead to disclosure of personally identifiable information. If the agency believes that such disclosure could occur, then they should take necessary precautions to prevent such disclosure. However, as we show in the following section, the aggregate performance data released by many state agencies would fail this disclosure test.

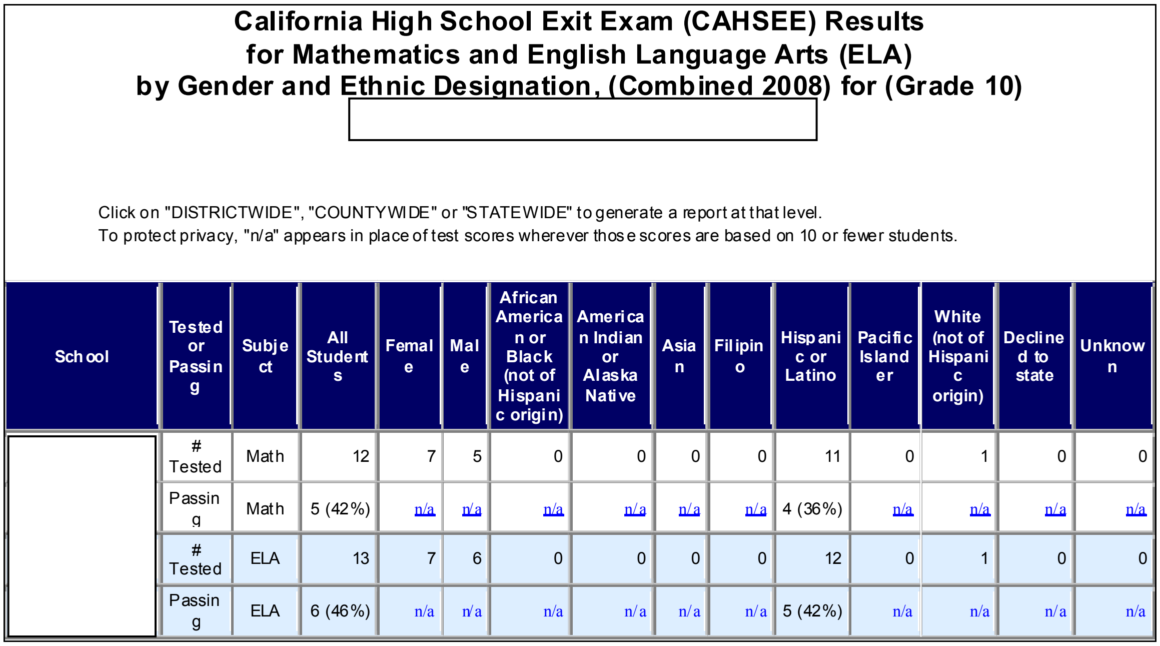
Example of Disclosure of Personally Identifiable Information

In this section we show that in many cases, the data being currently released to the public results in disclosure of personally identifiable information directly contradicting the requirements of FERPA. It is important to note that we *could not identify a single instance* where the released data leads to the *direct disclosure* of personally identifiable information. All state agencies have used some sort of *minimum cell size* requirement whereby when the size of a particular group or subgroup is less than the minimum cell size, the information for this group or subgroup is suppressed. For instance, if there is only one Asian student in a particular school, the information for this student is always suppressed.

The problem is that minimum cell size restrictions prevent the *direct disclosure* of personally identifiable data, but they do not prevent the *indirect disclosure* of personally identifiable information. According to FR2, agencies which release data *must* consider both direct and indirect disclosure when releasing the data. The example of the Female Hispanic student is a clear illustration that indirect disclosure must be prevented. After all, we did not *directly* identify that the Hispanic female did not graduate, but *indirectly* based on other sources of information that was released by the agency. In our illustrations, the disclosure is much more easily identified than the case of the Hispanic female. In all these cases, we show that disclosure of personally identifiable information *occurs using only the data that is available in that particular report without using any other source of information*.

Examples from California

As an illustration, consider the following report published by the California Department of Education providing school level data for the California High School Exit Exam that is available on the web. As indicated earlier, we have intentionally suppressed the identification information regarding the school. The California Department Education notes that “To protect privacy, "n/a" appears in place of test scores wherever those scores are based on 10 or fewer students.” The objective in suppressing this information is to prevent disclosure of confidential information regarding individual students or small subgroups (defined by ethnicity in this case) to be identified. Yet, looking at this illustration, it is easy to see that suppressing this information does not prevent disclosure.

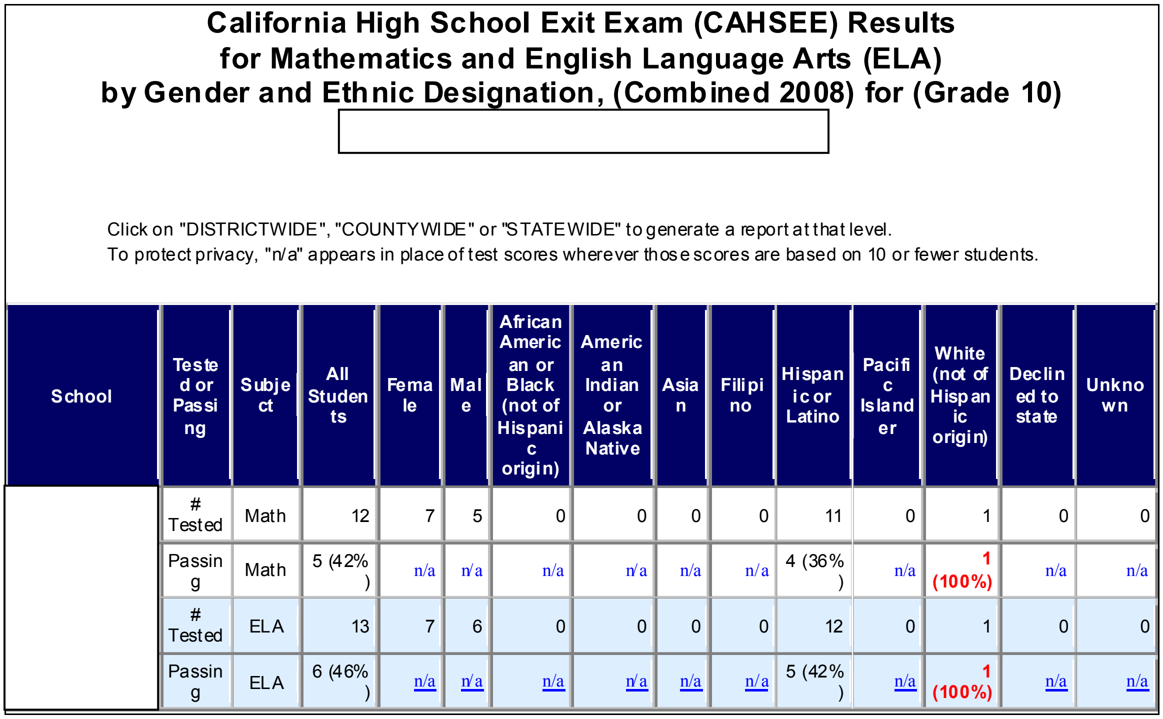


California High School Exit Exam (CAHSEE) Results
for Mathematics and English Language Arts (ELA)
by Gender and Ethnic Designation, (Combined 2008) for (Grade 10)

Click on "DISTRICTWIDE", "COUNTYWIDE" or "STATEWIDE" to generate a report at that level.
To protect privacy, "n/a" appears in place of test scores wherever those scores are based on 10 or fewer students.

School	Tested or Passing	Subject	All Students	Female	Male	African American or Black (not of Hispanic origin)	American Indian or Alaska Native	Asian	Filipino	Hispanic or Latino	Pacific Islander	White (not of Hispanic origin)	Declined to state	Unknown
	# Tested	Math	12	7	5	0	0	0	0	11	0	1	0	0
	Passing	Math	5 (42%)	n/a	n/a	n/a	n/a	n/a	n/a	4 (36%)	n/a	n/a	n/a	n/a
	# Tested	ELA	13	7	6	0	0	0	0	12	0	1	0	0
	Passing	ELA	6 (46%)	n/a	n/a	n/a	n/a	n/a	n/a	5 (42%)	n/a	n/a	n/a	n/a

From the summary information, we know that there are 12 students in this class at this high school. The ethnic breakdown of these students is 11 Hispanic or Latino students and 1 White student. Of the 12 students in the school, a total of 5 students passed the Math test. Of the 11 Hispanic students, 4 passed the Math test. From this information, *it is easy to infer that the only White student also passed the Math test*. Similarly, using the information for ELA, *we can also infer that the only White student also passed the ELA test*. For all practical purposes, it is as if the information was not suppressed in the first place and was released in the following manner.



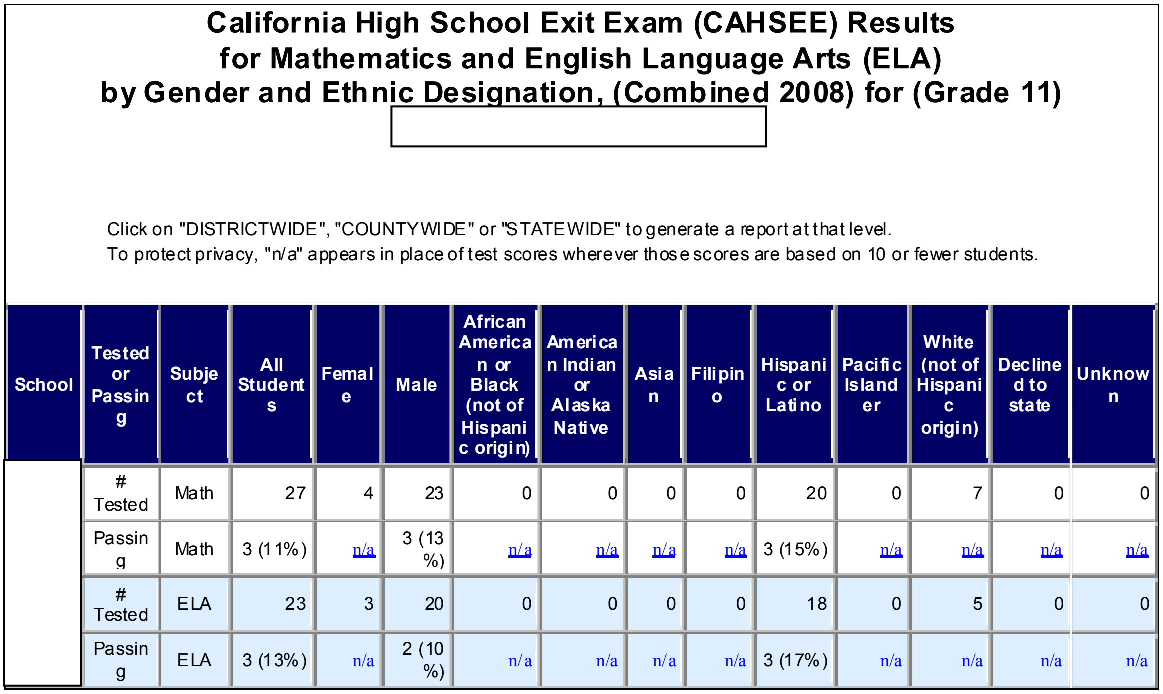
California High School Exit Exam (CAHSEE) Results
for Mathematics and English Language Arts (ELA)
by Gender and Ethnic Designation, (Combined 2008) for (Grade 10)

Click on "DISTRICTWIDE", "COUNTYWIDE" or "STATEWIDE" to generate a report at that level.
To protect privacy, "n/a" appears in place of test scores wherever those scores are based on 10 or fewer students.

School	Tested or Passing	Subject	All Students	Female	Male	African American or Black (not of Hispanic origin)	American Indian or Alaska Native	Asian	Filipino	Hispanic or Latino	Pacific Island er	White (not of Hispanic origin)	Declined to state	Unknown
	# Tested	Math	12	7	5	0	0	0	0	11	0	1	0	0
	Passing	Math	5 (42%)	n/a	n/a	n/a	n/a	n/a	n/a	4 (36%)	n/a	**1 (100%)**	n/a	n/a
	# Tested	ELA	13	7	6	0	0	0	0	12	0	1	0	0
	Passing	ELA	6 (46%)	n/a	n/a	n/a	n/a	n/a	n/a	5 (42%)	n/a	**1 (100%)**	n/a	n/a

The above example clearly illustrates that even though the information regarding this single White student is suppressed, we can easily compute the actual performance of this student. In addition, in most cases, anyone in the school community can easily identify this student by his/her ethnicity. It is precisely because this student can be personally identified by his/her ethnicity that the information is suppressed in the first place. Yet, using the information available in this single report, we are able to infer the performance of an individually identifiable student. This constitutes a direct violation of FERPA requirements.

Some might argue that, since the student passed both tests, this does not constitute disclosure. Unfortunately, this is not the case. FERPA does not make a difference between good or bad results; it requires that personally identifiable information should not be disclosed. The above report clearly violates this requirement. In addition, for every case where the disclosure occurs for a “good” result, we can find disclosure for a “bad” result as shown in the report below.



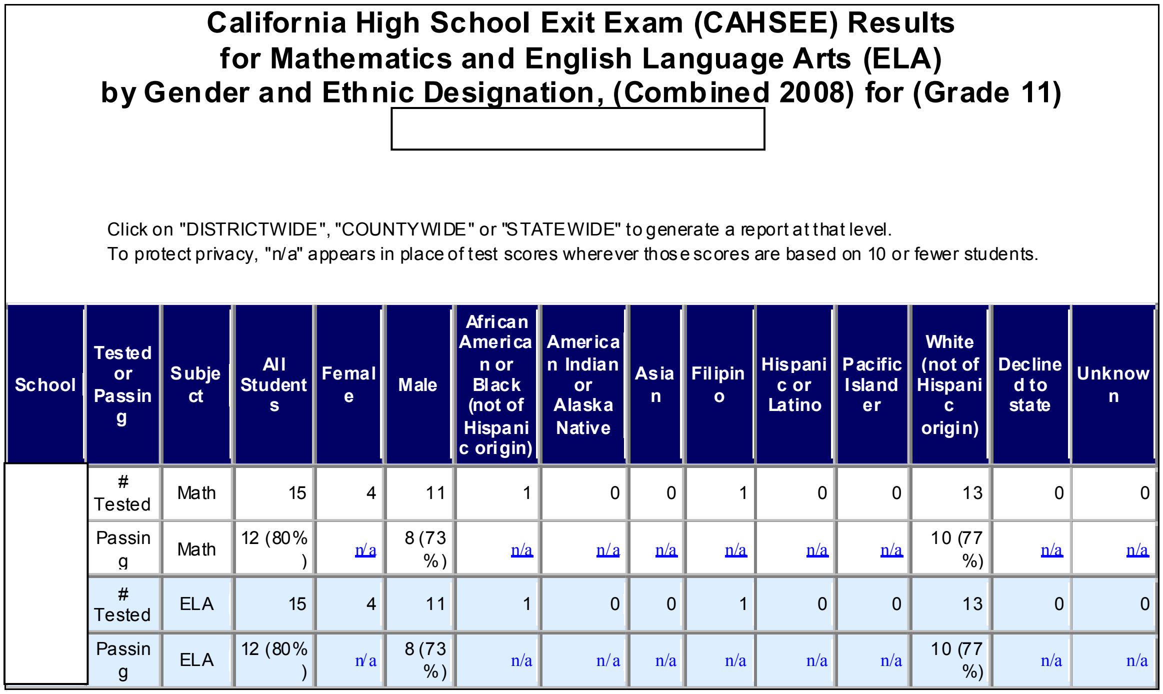
California High School Exit Exam (CAHSEE) Results
for Mathematics and English Language Arts (ELA)
by Gender and Ethnic Designation, (Combined 2008) for (Grade 11)

Click on "DISTRICTWIDE", "COUNTYWIDE" or "STATEWIDE" to generate a report at that level.
To protect privacy, "n/a" appears in place of test scores wherever those scores are based on 10 or fewer students.

School	Tested or Passing	Subject	All Students	Female	Male	African American or Black (not of Hispanic origin)	American Indian or Alaska Native	Asian	Filipino	Hispanic or Latino	Pacific Islander	White (not of Hispanic origin)	Declined to state	Unknown
	# Tested	Math	27	4	23	0	0	0	0	20	0	7	0	0
	Passing	Math	3 (11%)	n/a	3 (13%)	n/a	n/a	n/a	n/a	3 (15%)	n/a	n/a	n/a	n/a
	# Tested	ELA	23	3	20	0	0	0	0	18	0	5	0	0
	Passing	ELA	3 (13%)	n/a	2 (10%)	n/a	n/a	n/a	n/a	3 (17%)	n/a	n/a	n/a	n/a

From the above information, we can see that a total of 3 students passed the two tests and further that all 3 students who passed the test were Hispanic. *This leads to the inevitable conclusion that all White students failed both tests*. That this information is suppressed has no impact on our ability to infer the performance of the White students.

Consider the following report also from the State of California. In this case, we focus on gender rather than ethnicity. From the report, we can see that in the 11th grade there are a total of 15 students (4 female and 11 male) who took the Math test. From the report, we also know that a total of 12 students passed the Math test. Of the 12 students who passed the test, the report also indicates that 8 are male students. It follows directly from this information that 4 of the students who passed the Math test must be female. In other words, we know that every female student passed the Math test. Using similar logic, we can also conclude that female student also passed the ELA test. This provides another example of the futility of simply suppressing one piece of information. It simply does not work.

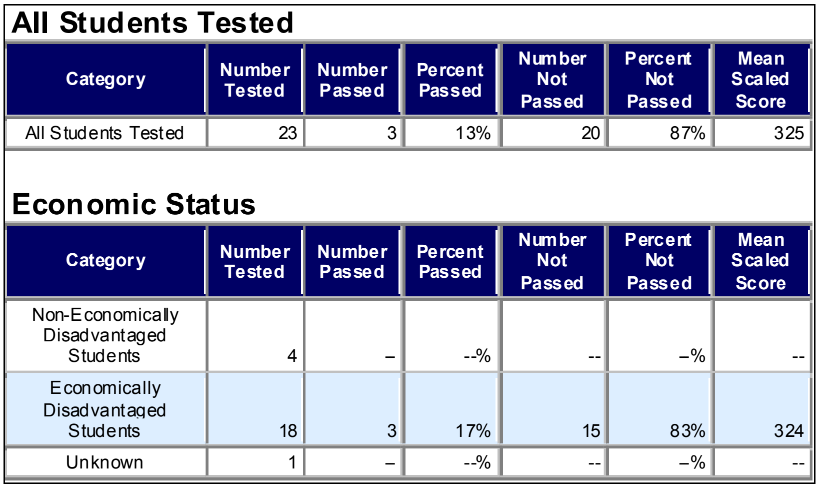


California High School Exit Exam (CAHSEE) Results for Mathematics and English Language Arts (ELA) by Gender and Ethnic Designation, (Combined 2008) for (Grade 11)

Click on "DISTRICTWIDE", "COUNTYWIDE" or "STATEWIDE" to generate a report at that level.
To protect privacy, "n/a" appears in place of test scores wherever those scores are based on 10 or fewer students.

School	Tested or Passing	Subject	All Students	Female	Male	African American or Black (not of Hispanic origin)	American Indian or Alaska Native	Asian	Filipino	Hispanic or Latino	Pacific Islander	White (not of Hispanic origin)	Declined to state	Unknown
	# Tested	Math	15	4	11	1	0	0	1	0	0	13	0	0
	Passing	Math	12 (80%)	n/a	8 (73 %)	n/a	n/a	n/a	n/a	n/a	n/a	10 (77 %)	n/a	n/a
	# Tested	ELA	15	4	11	1	0	0	1	0	0	13	0	0
	Passing	ELA	12 (80%)	n/a	8 (73 %)	n/a	n/a	n/a	n/a	n/a	n/a	10 (77 %)	n/a	n/a

In addition to gender and ethnicity breakdown, the State of California also provides information by various other factors. As an illustration consider the following example based on Economic Status.



All Students Tested

Category	Number Tested	Number Passed	Percent Passed	Number Not Passed	Percent Not Passed	Mean Scaled Score
All Students Tested	23	3	13%	20	87%	325

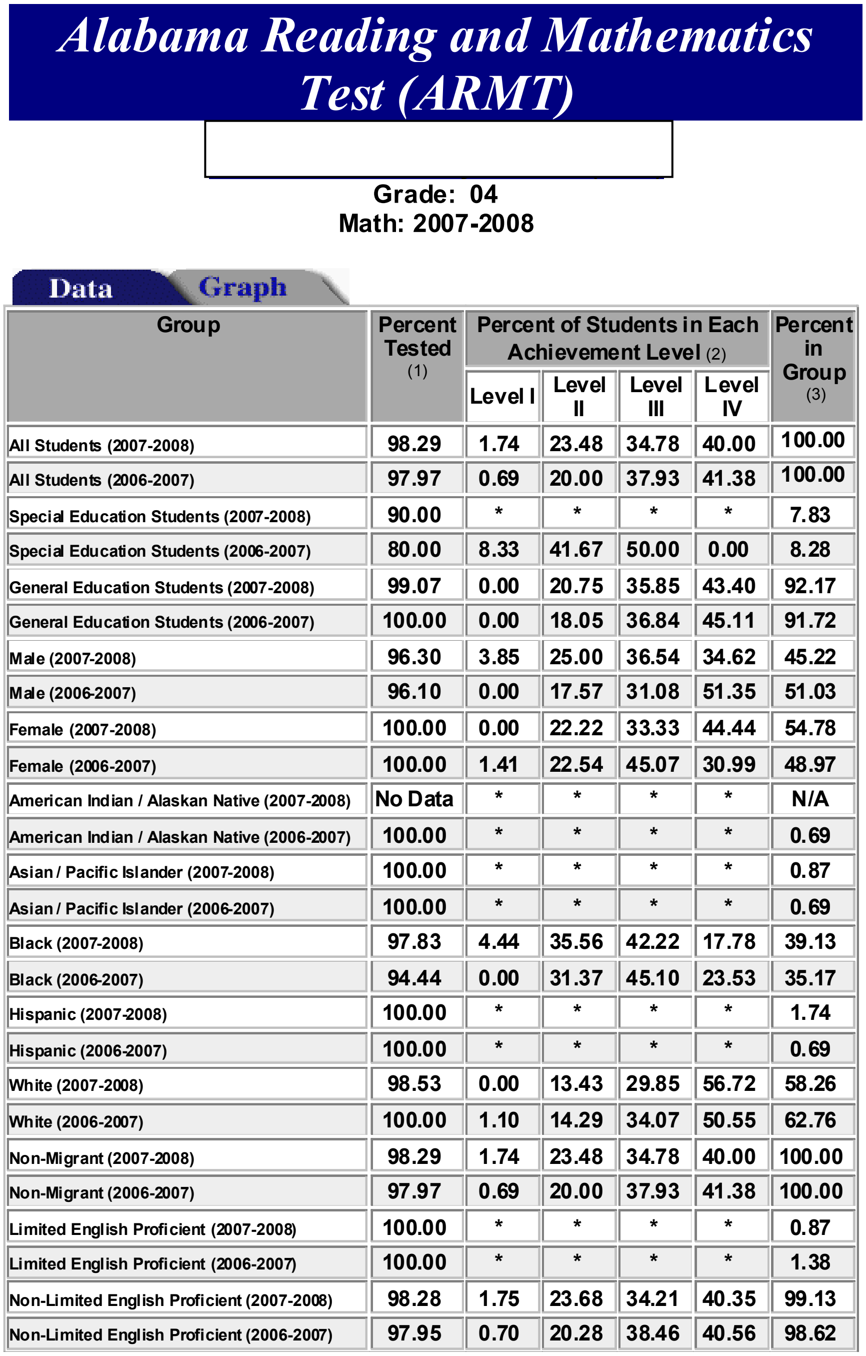
Economic Status

Category	Number Tested	Number Passed	Percent Passed	Number Not Passed	Percent Not Passed	Mean Scaled Score
Non-Economically Disadvantaged Students	4	–	--%	--	–%	--
Economically Disadvantaged Students	18	3	17%	15	83%	324
Unknown	1	–	--%	--	–%	--

Since only 3 students passed the test and all 3 of them are economically disadvantaged students, we can conclude that all four of the non-economically disadvantaged students failed the ELA test.

In conclusion, in order to comply with FERPA regulations which dictate that individually identifiable information should not be disclosed, the State of California suppresses information regarding individual students or small subgroups. Our illustrations above clearly show that suppression alone is not adequate. Even if the data is suppressed, the other information provided in the data allows us to compute individually identifiable performance information. Unfortunately, disclosure of personally identifiable information for individuals and small subgroups is the norm rather than the exception. In addition, this situation is not limited to the information released by the California, but in many other states as well as illustrate below.

Examples from Alabama
The examples provided below from the State of Alabama are results from the Alabama Reading and Mathematics test. A sample report from the web site is provided below.



Alabama Reading and Mathematics Test (ARMT)

Grade: 04
Math: 2007-2008

Group	Percent Tested (1)	Percent of Students in Each Achievement Level (2)				Percent in Group (3)
		Level I	Level II	Level III	Level IV	
All Students (2007-2008)	98.29	1.74	23.48	34.78	40.00	100.00
All Students (2006-2007)	97.97	0.69	20.00	37.93	41.38	100.00
Special Education Students (2007-2008)	90.00	*	*	*	*	7.83
Special Education Students (2006-2007)	80.00	8.33	41.67	50.00	0.00	8.28
General Education Students (2007-2008)	99.07	0.00	20.75	35.85	43.40	92.17
General Education Students (2006-2007)	100.00	0.00	18.05	36.84	45.11	91.72
Male (2007-2008)	96.30	3.85	25.00	36.54	34.62	45.22
Male (2006-2007)	96.10	0.00	17.57	31.08	51.35	51.03
Female (2007-2008)	100.00	0.00	22.22	33.33	44.44	54.78
Female (2006-2007)	100.00	1.41	22.54	45.07	30.99	48.97
American Indian / Alaskan Native (2007-2008)	No Data	*	*	*	*	N/A
American Indian / Alaskan Native (2006-2007)	100.00	*	*	*	*	0.69
Asian / Pacific Islander (2007-2008)	100.00	*	*	*	*	0.87
Asian / Pacific Islander (2006-2007)	100.00	*	*	*	*	0.69
Black (2007-2008)	97.83	4.44	35.56	42.22	17.78	39.13
Black (2006-2007)	94.44	0.00	31.37	45.10	23.53	35.17
Hispanic (2007-2008)	100.00	*	*	*	*	1.74
Hispanic (2006-2007)	100.00	*	*	*	*	0.69
White (2007-2008)	98.53	0.00	13.43	29.85	56.72	58.26
White (2006-2007)	100.00	1.10	14.29	34.07	50.55	62.76
Non-Migrant (2007-2008)	98.29	1.74	23.48	34.78	40.00	100.00
Non-Migrant (2006-2007)	97.97	0.69	20.00	37.93	41.38	100.00
Limited English Proficient (2007-2008)	100.00	*	*	*	*	0.87
Limited English Proficient (2006-2007)	100.00	*	*	*	*	1.38
Non-Limited English Proficient (2007-2008)	98.28	1.75	23.68	34.21	40.35	99.13
Non-Limited English Proficient (2006-2007)	97.95	0.70	20.28	38.46	40.56	98.62

This data is available starting 2001-2002. The examples provided here are from test results for the year 2007-2008. Alabama provides both a printable version of the data as well as a downloadable version that can be saved as a text file. One key difference between the printable version and the downloadable version is that the printable version does not have the total number of students tested while the downloadable version has the total number of students tested. The following represents one simple case where the Mathematics test results for a particular grade at a particular school are provided for “All students” (115 students), “Non-Limited English Proficient” (114 students), and “Limited English Proficient” (1 student).

Alabama has a minimum cell size of 10 and hence for the single student with limited English proficiency, the true result is replaced by an * with the notation that "* Indicates less than ten students of a particular group tested."

Category	Number of Students Tested	Percentage of Students Tested	% Level 1	% Level 2	% Level 3	% Level 4	Percent in Group
All Students	115	98.29	1.74	23.48	34.78	40	100
Non-Limited English Proficient	114	98.28	1.75	23.68	34.21	40.35	99.13
Limited English Proficient	1	100	*	*	*	*	0.87

Using simple arithmetic, using the percentages that are provided, we can easily compute the *number of students* in the each of the Levels for "All students" and "Non-Limited English Proficient" as shown below. For example, from the table above, we know that 1.74% of the 115 students tested at Level 1. From this we know that a total of 2 students tested at Level 1 (1.74% of 115). We can repeat these computations for "All Students" and "Non-Limited English Proficiency" students. The results are provided below.

Category	Number of Students Tested	Percentage of Students Tested	# Level 1	# Level 2	# Level 3	# Level 4	Percent in Group
All Students	115	98.29	2	27	40	46	100
Non-Limited English Proficient	114	98.28	2	27	39	46	99.13
Limited English Proficient	1	100	**0**	**0**	**1**	**0**	0.87

Once we perform this computation, the performance of the only "Limited English Proficient" student becomes immediately obvious (as being Level 3). Note that this situation is very similar to the Hispanic female student situation illustrated in FR2. Members of the school community would be able to identify a student whose has Limited English Proficiency. Hence, this constitutes a disclosure of personally identifiable performance information of a single student. As before, we can provide the same illustration for the situation where the student has performed poorly. However, as we observed earlier, FERPA does not differentiate between good and bad performance when disclosure is concerned.

As another illustration, consider the following example, also from the Alabama Reading and Mathematics Test results. In this example, there is one Asian/Pacific Islander and one Hispanic student in this class. Again, since the minimum cell size is 10, the information for the two students is suppressed and replaced with an *.

Category	Number of Students Tested	Percentage of Students Tested	% Level 1	% Level 2	% Level 3	% Level 4	Percent in Group
All Students	116	100	0	25.86	57.76	16.38	100
Asian / Pacific Islander	1	100	*	*	*	*	0.86
Black	45	100	0	40	55.56	4.44	38.79
Hispanic	1	100	*	*	*	*	0.86
White	69	100	0	17.39	57.97	24.64	59.48

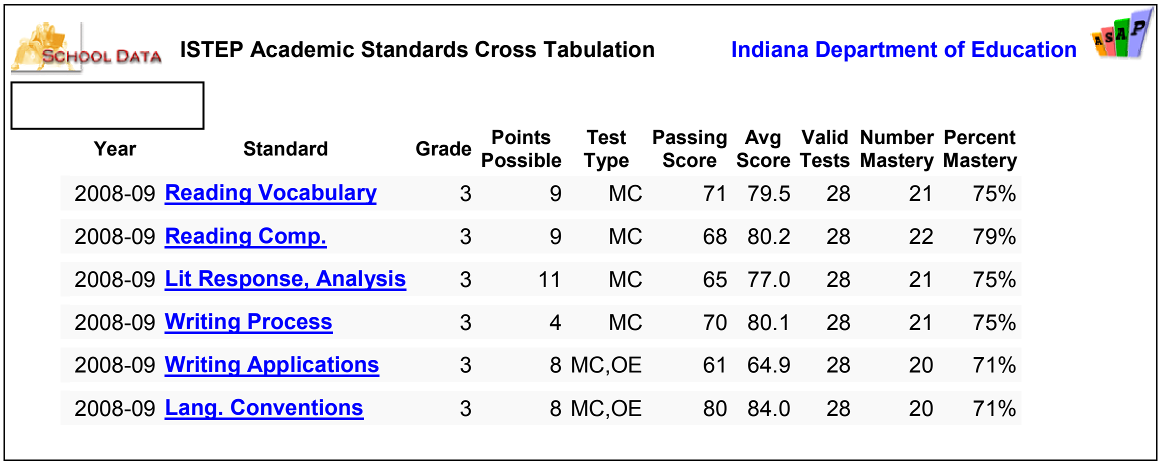
As before, using the percentages in each Level and the total number of students tested, we can easily replace the percentage values with the actual number of students as shown below. Once computed, it is easy to figure out that both the Asian and the Hispanic student scored at Level 3. Members of the school community would be able to easily identify these students just by their ethnicity. Thus, the released data *results in complete disclosure of personally identifiable performance data for these two students*. This is a clear violation of FERPA requirements.

Category	Number of Students Tested	Percentage of Students Tested	# Level 1	# Level 2	# Level 3	# Level 4	Percent in Group
All Students	116	100	0	30	67	19	100
Asian / Pacific Islander	1	100	**0**	**0**	**1**	**0**	0.86
Black	45	100	0	18	25	2	38.79
Hispanic	1	100	**0**	**0**	**1**	**0**	0.86
White	69	100	0	12	40	17	59.48

For the Alabama State report, we have identified disclosures relating to practically every ethnic subgroup, both genders, and practically every subgroup for which results are reported. Disclosure occurs in most reports at every grade level and is the norm rather than the exception.

Examples from Indiana

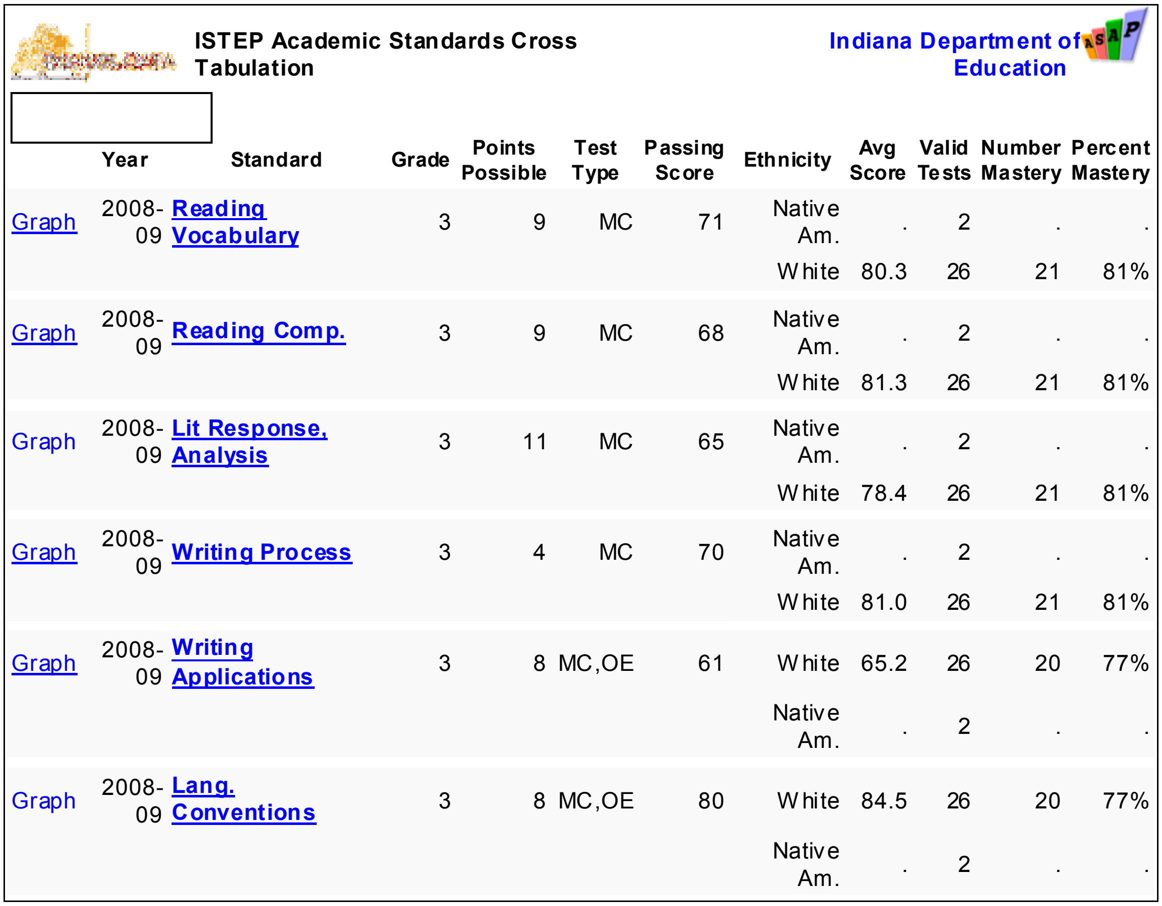
Indiana Department of Education provides extensive individual grade level analysis of the data. One of the key aspects of the reports provided by Indiana is the fact that they are very detailed, even breaking down reading scores into its individual components. While the effort to provide detailed analysis must be appreciated, such detailed data also results in disclosure at the same detailed level. As an illustration consider the results from a particular school for students in the 3rd grade shown below.



SCHOOL DATA ISTEP Academic Standards Cross Tabulation Indiana Department of Education ASAP

Year	Standard	Grade	Points Possible	Test Type	Passing Score	Avg Score	Valid Tests	Number Mastery	Percent Mastery
2008-09	Reading Vocabulary	3	9	MC	71	79.5	28	21	75%
2008-09	Reading Comp.	3	9	MC	68	80.2	28	22	79%
2008-09	Lit Response, Analysis	3	11	MC	65	77.0	28	21	75%
2008-09	Writing Process	3	4	MC	70	80.1	28	21	75%
2008-09	Writing Applications	3	8	MC,OE	61	64.9	28	20	71%
2008-09	Lang. Conventions	3	8	MC,OE	80	84.0	28	20	71%

For the same school and same grade level, we can also get a report that provides the results broken down by ethnicity which is provided below.



ISTEP Academic Standards Cross Tabulation Indiana Department of Education ASAP

	Year	Standard	Grade	Points Possible	Test Type	Passing Score	Ethnicity	Avg Score	Valid Tests	Number Mastery	Percent Mastery
Graph	2008-09	Reading Vocabulary	3	9	MC	71	Native Am.	.	2	.	.
							White	80.3	26	21	81%
Graph	2008-09	Reading Comp.	3	9	MC	68	Native Am.	.	2	.	.
							White	81.3	26	21	81%
Graph	2008-09	Lit Response, Analysis	3	11	MC	65	Native Am.	.	2	.	.
							White	78.4	26	21	81%
Graph	2008-09	Writing Process	3	4	MC	70	Native Am.	.	2	.	.
							White	81.0	26	21	81%
Graph	2008-09	Writing Applications	3	8	MC,OE	61	White	65.2	26	20	77%
							Native Am.	.	2	.	.
Graph	2008-09	Lang. Conventions	3	8	MC,OE	80	White	84.5	26	20	77%
							Native Am.	.	2	.	.

From report detailing ethnic breakdown, we note that there are 26 white students and 2 Native American students. As with all other states, Indiana also suppresses all results when the number of students in any particular subgroup is less than 10. The objective of course is to prevent disclosure of performance information for these two students who can be identified by their ethnicity. However, as with previous cases, it is easy to see that suppression alone is inadequate.

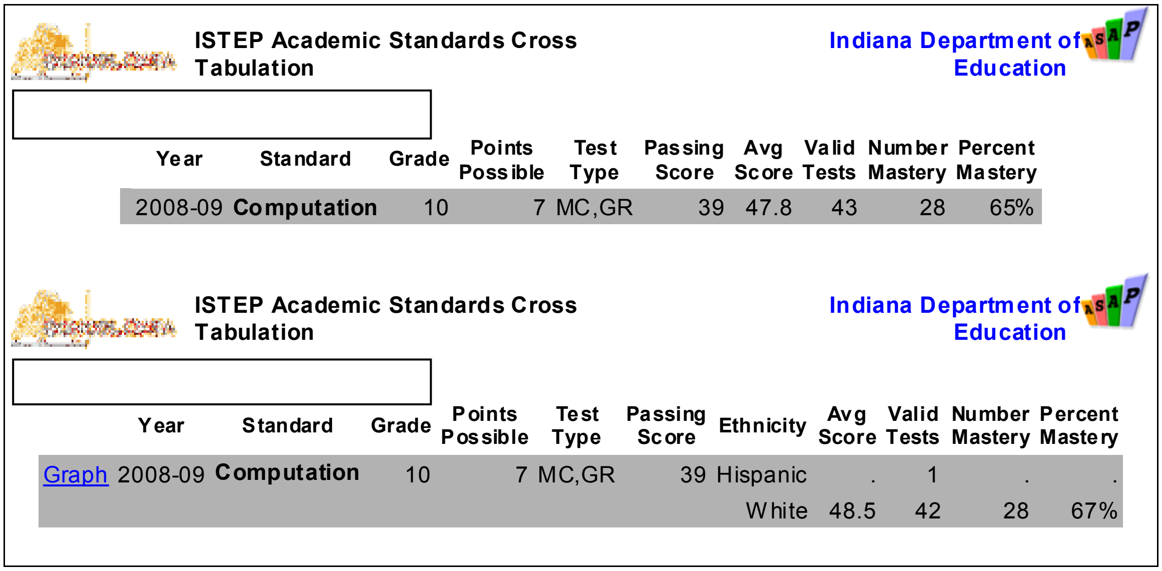
From the above report, we know that of 28 valid tests, a total of 21 exhibited "mastery" in Reading Vocabulary. From the report below, we also know that there are a total of 28 valid tests (same as the total above) of which 26 valid tests are those of White students while 2 belong to Native American students. We also know from the table below that, 21 of the 26 White students exhibited "mastery" in Reading Vocabulary. Even though the information regarding the 2 Native American students is suppressed, we can *infer that both the Native American students did not exhibit mastery in Reading Vocabulary*.

By repeating this process, we can completely recreate the performance of the 2 Native American students. The results in this particular case are provided below. We know that one of the Native American students passed Reading Comp. and more importantly, that the Native American students failed (to exhibit mastery) in all other tests.

Standard	Total Number of Students	Number of White Students	Number of Native American Students	Total Number of Students Exhibiting Mastery	Number of White Students Exhibiting Mastery	Number of Native American Students Exhibiting Mastery	Percent of Native American Students Exhibiting Mastery
Reading Vocabulary	28	26	2	21	21	**0**	**0%**
Reading Comp.	28	26	2	22	21	**1**	**50%**
Lit. Response, Analysis	28	26	2	21	21	**0**	**0%**
Writing Process	28	26	2	21	21	**0**	**0%**
Writing Applications	28	26	2	20	20	**0**	**0%**
Language Conventions	28	26	2	20	20	**0**	**0%**

The disturbing part of this disclosure is the fact that we can develop very detailed personally identifiable performance information for individual students or small subgroups. This illustrates the tradeoff in releasing detailed performance information; on the one hand they provide the public with additional information, but also result in disclosure that is very detailed.

In addition to providing information on whether students exhibited mastery or not, the Indiana Department of Education also provides information on the average score for the students. This allows us to estimate the score of individual or small subgroups of students, thereby disclosing even more additional information. For the purpose of illustration, consider the following report from Indiana for all students for Math Computation and the same information broken down by ethnicity.



ISTEP Academic Standards Cross Tabulation

Indiana Department of Education

Year	Standard	Grade	Points Possible	Test Type	Passing Score	Avg Score	Valid Tests	Number Mastery	Percent Mastery
2008-09	**Computation**	10	7	MC,GR	39	47.8	43	28	65%

ISTEP Academic Standards Cross Tabulation

Indiana Department of Education

	Year	Standard	Grade	Points Possible	Test Type	Passing Score	Ethnicity	Avg Score	Valid Tests	Number Mastery	Percent Mastery
Graph	2008-09	**Computation**	10	7	MC,GR	39	Hispanic	.	1	.	.
							White	48.5	42	28	67%

From the two reports above, we know that of the 43 students in the 10th grade, 28 passed the test. In addition, we know that 28 of the 42 White students passed the test. Based on this information, we know, with certainty that the Hispanic student failed (to exhibit number mastery) of the test.

In addition to the above information, we can also attempt to compute the *actual score* of the Hispanic student in Math Computation using the following simple arithmetic procedure. We know that the average score for all 43 students was 47.8 and from this we can compute that the total score for all students as (43 × 47.8 =) 2055.4. Similarly, using the information that the average score of the 42 White students was 48.5, we can compute the total score for all students as (42 × 48.5 =) 2037. An estimate of the score of the Hispanic student is 18.4 (2055.4 – 2037.0).

It should be noted that because the scores are rounded off to one decimal place, the above value is not an exact estimate. With some simple derivations, we can easily show that the *actual score* for the Hispanic student is in the range 18.4 ± 4.25 (14.15 to 22.65). Thus, although we cannot compute the exact score because of rounding, we can get what is a relatively precise estimate of the true score of this student. In our opinion, this represents a further illustration of disclosure that could possibly occur. Using the information in the Indiana reports, we were able to compute the scores for this Hispanic student in all Math categories as shown below.

Standard	Passing Score	Range of scores		Pass or Fail
		Lower	Upper	
Number Sense	42	48.25	56.75	Pass
Computation	39	14.15	22.65	Fail
Algebra and Functions	30	49.55	58.05	Pass
Geometry	32	50.35	58.85	Pass
Measurement	40	63.35	71.85	Pass
Data Analysis and Prob.	64	48.45	56.95	Fail
Problem Solving	23	46.85	55.35	Pass

In our opinion, this represents a clear violation of the rights of the Hispanic student under FERPA. Not only are we able to determine whether this student passed of failed the test, we can even estimate the true value accurately. Some will argue that since we cannot be certain what the specific score is, it does not constitute disclosure. However, a careful reading of the literature on statistical disclosure limitation indicates that the ability to estimate the grade for a particular student with level of precision does constitute disclosure.

Statistical Working Policy Paper 22[3] titled "Report on Statistical Disclosure Limitation Methodology" defines disclosure as follows:

> Disclosure relates to inappropriate attribution of information to a data subject, whether an individual or an organization. Disclosure occurs when a data subject is identified from a released file (identity disclosure), sensitive information about a data subject is revealed through the released file (attribute disclosure), or the released data make it possible to determine the value of some characteristic of an individual more accurately than otherwise would have been possible (inferential disclosure).

This definition makes it very clear that even if we are not able to compute the exact score of the Hispanic student, the information provided in the Indiana reports certainly make it *possible to determine the math scores* of the Hispanic student *more accurately than otherwise would have been possible*. Hence, by this definition, Indiana reports result in disclosure of scores of individuals and small subgroups who can be easily personally identified by the school community. It is possible that the Indiana Department of Education has somehow altered the values that are presented in these reports. If so, then they should indicate that modifications have been made to the reports. Failure to do so is misleading the public regarding the information presented in these reports.

Among the different states, Indiana perhaps provides the greatest detail in the information released to the public. While this is to be commended, we do not believe that the same can be said of their approach to protecting personally identifiable information. We found disclosures in practically every report, in every school, and in every district.

Summary

We have chosen three different states to illustrate that data that is made available to the public results in disclosure of personally identifiable information in all three cases. As indicated in the examples, unfortunately, disclosure is the rule rather than the exception in these cases. This problem is not limited to these three states, but can be found in the reports released by most states.

We chose these three reports since they provide three different types of presentation:

(1) In the case of Alabama, only a single report is provided with no additional ability to breakdown the data into further smaller subgroups,
(2) In the case of California, we do not have the ability to subdivide the data into smaller subgroups, but we do have the ability to request reports for specific dates on which the tests were administered, and
(3) In the case of Indiana, we have the ability to subdivide the data into any combination of subgroups formed by different categories such as Gender, Ethnicity, Limited English Proficiency, etc.

The interesting aspect is that providing the ability to breakdown the data further results in a greater probability of disclosure. While all three reports result in disclosure, it is far easier to identify the disclosures that occur in the reports provided by Alabama while it is extremely difficult to identify all the potential disclosures that occur in the reports provided by Indiana. In fact, in the case of Indiana, in some school and grade combinations, every student falls in a small subgroup (of size less than 10) resulting in a situation where *disclosure of personally identifiable information could occur for every student*.

Disclosure Audit to Prevent Disclosure of Personally Identifiable Information

The only way to prevent disclosure of personally identifiable information is to conduct a disclosure audit. Such an audit would require that every potential disclosure of personally identifiable information be identified. Once the different types of disclosures have been identified, then every report that is released to

[3] STATISTICAL POLICY WORKING PAPER 22 (Second version, 2005), "Report on Statistical Disclosure Limitation Methodology," Federal Committee on Statistical Methodology, http://www.fcsm.gov/working-papers/SPWP22_rev.pdf.

the public must be evaluated to assess whether these reports result in disclosure of personally identifiable information. Obviously, it is not an easy task to evaluate every report. In addition, we cannot identify disclosure simply based on the number of students since in many cases creating subgroups results in a smaller number of students than the threshold for small subgroups. Such an audit must also identify whether disclosure occurs just in count data (as is the case for Alabama) or for both count data and numerical score data (as is the case with Indiana).

A preliminary disclosure audit was performed by the Department of Education in the Commonwealth of Kentucky to evaluate the extent to which the data released to the public, Kentucky Performance Reports (KPR), results in disclosure of personally identifiable information. In Kentucky, a KPR is released for every subject in every grade in every school in the Commonwealth. The disclosure audit was limited to KPR from one subject area since it was determined that the presence of disclosure in any one subject area is likely to be present in other subject areas as well. A total of 3998 KPR's were evaluated from 1186 different schools. There were 14861 KPR's reflecting multiple subjects from each school/grade combination. The following is an analysis of the disclosures that occurred by grade.

Grade	Total Number of KPR's	KPR's without Disclosure Issues		KPR's with Disclosure Issues		Average Number of Disclosure Issues per KPR	Total number of Disclosure Issues for the Grade
		Number of KPR's	Percent of KPR's	Number of KPR's	Percent of KPR's		
3	738	18	2.44%	720	97.56%	2.63	1941
4	739	13	1.76%	726	98.24%	2.73	2017
5	734	16	2.18%	718	97.82%	2.77	2036
6	421	19	4.51%	402	95.49%	2.55	1075
7	330	17	5.15%	313	94.85%	2.32	765
8	328	9	2.74%	319	97.26%	2.36	774
10	236	15	6.36%	221	93.64%	2.12	501
11	236	11	4.66%	225	95.34%	2.03	479
12	236	9	3.81%	227	96.19%	2.36	558
Total	3998	127	3.18%	3871	96.82%	2.54	10146

The above table indicates that *disclosure risk is a serious problem in the data that is currently being released to the public with approximately 97% of all KPR's having at least one instance of disclosure*. Only a small number of KPR's (127) had no disclosure. The problem is consistent across all grade levels, although the percentage is slightly smaller for grades 10, 11, and 12 but only by a small margin. Even in the best case (grade 10), approximately 94% of the KPR's had at least one disclosure.

The above table also shows that, on average, there were approximately 2.5 disclosures associated with each KPR. The number of disclosures ranged from 0 to as many as 7. The average number of disclosures is remarkably consistent across all grade levels ranging from 2.03 to 2.63. This implies that, *if we randomly selected a KPR from among all KPR's, there are likely to 2 to 3 instances of disclosure in the KPR*. There were a total of 10146 instances of disclosure from the 3998 KPR's that were investigated.

In order to investigate if there any patterns at the school level, we repeated the analysis at the school level. There were a total of 1186 schools in the Commonwealth at three levels (Elementary, Middle, and High). Of the 1186 schools, *KPR's from 1182 (99.66%) schools showed at least one instance of disclosure*.

School Level	Number of Schools	Schools without Disclosures		Schools with Disclosures		Average Number of KPR's with Disclosures	Total number of Disclosures for the Level
		Number of Schools	Percent of Schools	Number of Schools	Percent of Schools		
Elementary	654	1	0.15%	653	99.85%	3.17	5546
Middle	296	2	0.68%	294	99.32%	3.68	2686
High	236	1	0.42%	235	99.58%	3.53	1914
Total	1186	4	0.34%	1182	99.66%	3.37	10146

The table shows that more than 50% of the disclosures occur at the Elementary School level. This is not surprising for two reasons. There are more elementary schools (654) than there are middle (296) or high schools (236). In addition, the average number of students at the Elementary School level (65.28) is typically much lower than that at the Middle school level (145.15) or High school level (185.03). Hence, subgroups of 9 or fewer individuals are much more likely at the Elementary school level. However, when we adjust for the number of schools at each level, it seems as if the occurrence of disclosure across the different level schools is very similar. We see that each level we approximately the same number of KPR's that result in disclosure. In fact there are, on average, fewer KPR's with disclosure per school at the Elementary level than at the other two levels. The average number of disclosures per school is also very close with 8.48 disclosures at the Elementary, 9.07 disclosures at the Middle, and 8.11 at the High school level. Table below provides the most frequent reason for the disclosure.

Reason for Disclosure	Level			Total	
	Elementary	Middle	High		
Gender	183	53	19	255	
Ethnicity	1466	540	457	2463	
Title I	29	7	30	66	
Migrant	198	124	44	366	
Limited English Proficiency	644	276	168	1088	
Extended School Services	673	198	138	1009	
Gifted and Talented Program	924	212	78	1214	
Lunch Program	528	121	35	684	
Vocational/2 credit	0	0	84	84	
Disabled	1301	307	184	1792	2917
Disabled with Accommodations	32	12	19	63	
Disabled without Accommodations	419	361	282	1062	
Total	6397	2211	1538	10146	

As noted before, the subgroup where disclosure occurs most frequently is for students with disabilities, followed by ethnic subgroups. Vocational programs are offered only at the High school level and hence there are no disclosures associated with this subgroup at the Elementary and Middle school level. As discussed earlier, it is the subgroups that can be easily identified (disabled students and ethnic subgroups) that most frequently result in disclosure.

In summary, the disclosure audit conducted for the Kentucky Department of Education shows that the disclosure of personally identifiable confidential information occurs in over 96% of the reports that is made available to the public. Our analysis of data available from other states indicates that in many cases these reports are susceptible to the same level of disclosure. In some states, such as Indiana where subgroups can be created by request, disclosure is likely to be higher.

The Kentucky Department of Education is currently investigating methods by which the disclosures identified in the audit can be eliminated. The procedures that are being investigated include complementary suppression, perturbation, and other procedures that have been suggested in Statistical Working Policy Paper 22. We have not had the opportunity to investigate the extent to which the new policies and procedures have reduced disclosure. However, we believe that by being aware of the issue and adopting appropriate policies and procedures for reducing disclosure, the Kentucky Department of Education is likely to significantly reduce the disclosure of personally identifiable information.

Conclusions

All states are required to release performance data from schools as a part of the NCLB statutes. The level of detail in the data that is released to the public varies by state. Most states provide very detailed data which allows parents and the general public to assess and compare the performance of a particular school or district to others. We commend the states in their effort to provide this detailed information.

When releasing such data, the agencies releasing the data are also required by FERPA to ensure that no personally identifiable information is released. Our analysis of the data seems to indicate that states have not paid attention to this requirement. In most states, the only disclosure prevention technique that seems to have been adopted is to suppress the data when the number of students in a subgroup is less than a minimum number. This minimum cell size also varies by states (10 in most cases and 5 in some). However, as our analysis above indicates, such suppression alone is inadequate to prevent disclosure of personally identifiable information. *We have shown that even if the data is suppressed, anyone with basic mathematical skills can easily compute the results for the suppressed group using the other information that is available in the report*. In states where the user is allowed to further subdivide categories on request (such as Indiana), the risk of disclosure is even higher. In addition, for those states that provide numerical scores in addition to count data, it is even possible to precisely estimate the score for subgroups.

Our analysis in this paper is at the individual school level. However, disclosure of personally identifiable information may also occur at the district level or other higher levels of aggregation. Hence, it would be necessary to carefully evaluate the data released even at this level. *The only way to prevent disclosure of personally identifiable information is to conduct a comprehensive disclosure audit of every report that can potentially be generated using the information provided by the state agencies*. Where such audit identifies disclosure of personally identifiable information, state agencies must adopt appropriate procedures that are available to prevent such disclosure.

Finally, the analysis in this paper raises the important issue of the tradeoff between the public's needs to know and preserving the privacy of individuals or subgroups. While we completely agree that providing information to the public is an important objective, we do not believe that it should be done at the cost of the privacy of individuals or small subgroups. We believe that in situations where releasing the data could potentially result in disclosure, state agencies need to err on the side of the individual's privacy rather than the public's need to know. In addition, we believe that some states have carried this desire to provide the public with data to a level where disclosure is almost guaranteed. This is the case with the state of Indiana. By allowing the users to specify the subgroups, Indiana has almost guaranteed that disclosure will occur. While we understand and agree with most of the information provided in the reports, one does have to wonder whether it is really necessary for the public to be able to request results regarding White Male students who receives Special Education and Free Lunch? We believe that by being aware of disclosure issues and adopting the appropriate policies and procedures, state agencies can release information this is useful to the public but also prevents disclosure of personally identifiable data regarding individuals or small subgroups.