Technical Brief

Statistical Methods for Protecting Personally Identifiable Information in the Disclosure of Graduation Rates of First-Time, Full-Time Degree- or Certificate-Seeking Undergraduate Students by 2-Year Degree-Granting Institutions of Higher Education

Contents

Introduction

The Higher Education Opportunity Act of 2008 (HEOA) requires that Title IV degree-granting institutions[1] disclose annually the graduation rates of first-time, full-time degree- or certificate-seeking undergraduate students,[2] disaggregated by gender, each major racial/ethnic subgroup, and receipt or non-receipt of a federal Pell grant or subsidized Stafford loan.[3] The law requires that this information be made available, through appropriate publications, mail, or electronic media, to current and prospective students.[4] All 4-year degree-granting institutions are expected to implement this mandated disclosure requirement immediately upon HEOA enactment, but a provision of the law requires this disclosure requirement to apply to 2-year degree-granting institutions beginning in the academic year 2011–12 (HEOA § 488(a)(3)).

The public disclosure of graduation rate data, while important, potentially risks disclosure of personally identifiable information. The challenge of meeting disclosure requirements lies in releasing as much information as required by the law, while also meeting legal requirements to protect each student's privacy under the Family Educational Rights and Privacy Act (FERPA) (34 CFR § 99.3). Recognizing this, HEOA states that disaggregated graduation rates are to be disclosed only "if the number of students in subgroups is sufficient to yield statistically reliable information and reporting will not reveal personally identifiable information about an individual student. If such number is not sufficient for such purposes, then the institution shall note that the institution enrolled too few of such students to so disclose or report with confidence and confidentiality" (HEOA § 488(a)(3)).

The purpose of this brief is to provide technical guidance to Title IV 2-year degree-granting institutions in meeting the statutory disclosure requirement related to graduation rates while minimizing the risk of revealing the graduation status of individual students.[5]

---

[1] Title IV degree-granting institutions are postsecondary institutions that are eligible for Title IV federal financial aid programs and that grant an associate's or higher degree. Eligibility for participation in Title IV financial aid programs requires that an institution offer a program of at least 300 clock hours in length, have accreditation recognized by the U.S. Department of Education, have been in business for at least 2 years, and have signed a participation agreement with the Department.

[2] First-time, full-time degree- or certificate-seeking undergraduate students are students who have no prior experience attending any postsecondary institution at the undergraduate level, are enrolled for 12 or more semester/quarter credits or 24 or more contact hours a week each term, and are enrolled in courses for credit and recognized by the institution as seeking a degree, certificate, or other formal award. High school students who are enrolled in postsecondary courses for credit are not considered degree- or certificate-seeking.

[3] The original statutory language in HEOA uses recipients of "subsidized Federal Family Education Loan (FFEL)." Beginning July 1, 2010, however, subsidized FFEL was replaced by subsidized Stafford loans. More information about subsidized FFEL and Stafford loans can be found at http://studentaid.ed.gov/PORTALSWebApp/students/english/studentloans.jsp.

[4] More specifically, the law requires that the graduation rate data be available by July 1 each year for the most recent cohort that has had 150 percent of normal time for completion by August 31 of the prior year, and that if the information is requested by a prospective student, it must be made available prior to the student's enrolling or entering into any financial obligation with the institution (NPEC 2009).

[5] Although this technical brief targets 2-year postsecondary institutions, disclosure practices suggested in the brief can be applied more generally to all types of postsecondary institutions.

Adapted from *Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting* (Seastrom 2010b),[6] this brief includes a summary of key definitions, a brief discussion of background information, a review of current disclosure practices used by institutions, and a discussion of some practices for balancing disclosure and confidentiality. The brief concludes with a set of recommended rules that postsecondary institutions can apply to disclosure of graduation rate data required by the HEOA. It is important to emphasize that disclosure rules described in this brief aim only to meet the requirements of HEOA for disclosure of graduation rates. These rules may not apply to other disclosure or reporting requirements.

Definitions of Terms

This box provides the definitions of some commonly used terms related to the disclosure and protection of personally identifiable information. These definitions may not be universally accepted and are intended to provide understanding of the terms used in this brief. These definitions come directly from *Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting* (Seastrom 2010b), and *Information Required to Be Disclosed Under the Higher Education Act of 1965: Suggestions for Dissemination* (NPEC 2009).

Personally identifiable information includes the name and address of the student and his or her 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 in such a way as to allow a reasonable person in the school community without personal knowledge of relevant circumstances to identify a student with reasonable certainty; and information based on a targeted request (34 CFR § 99.3; Seastrom 2010b).

Disclosure means to permit access to or 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 or the party that provided or created the record. To avoid disclosures, 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 (34 CFR § 99.3; Seastrom 2010b).

Disclosure requirement is information that an institution is required to disseminate or make available to other parties, such as students or their families, through appropriate publications, mail, or electronic media. Disclosure requirement differs from "reporting requirement." The latter is information that must be submitted to the U.S. Department of Education or other agencies. Disclosure and reporting requirements, however, sometimes overlap. For certain information, such as graduation rates, both requirements may apply, requiring postsecondary institutions both to make the information publicly available and to report the data to the U.S. Department of Education (NPEC 2009).

Suppression refers to withholding information from publication. Some information is withheld from publication in a table to protect data based on small counts, because releasing the information likely would lead to a disclosure. Other information is withheld from publication in a table to prevent calculations of the data based on small counts that could lead to identifying specific students; this is known as "complementary suppression" (Seastrom 2010b).

---

[6] Released by the National Center for Education Statistics at the U.S. Department of Education in 2010, this technical brief provides guidance to states in meeting reporting requirements for public elementary and secondary institutions that receive federal funds and also meeting the legal requirements to protect each student's personally identifiable information as required by FERPA.

Background

The federal government's involvement in collecting and reporting postsecondary institutions' graduation rates started with the enactment of the Student Right to Know Act (SRK) in 1990 (Cook and Pullaro 2010). This Act requires postsecondary institutions receiving Title IV funds to submit an annual report to the Department of Education containing information on enrollment and graduation rates disaggregated by gender, race/ethnicity, receipt of athletic scholarships, and type of sport. To help colleges comply with SRK, the National Center for Education Statistics (NCES) of the Department of Education created a new survey component, the Graduation Rate Survey (GRS), in the Integrated Postsecondary Education Data System (IPEDS)[7] in 1997 to collect annual data on graduation rates from all Title IV degree-granting institutions.[8] To push for more institutional accountability and better consumer information on graduation rates, the 2008 HEOA called for wider disclosure of institutional graduation rates and mandated that each institution annually disclose graduation rates at 150 percent of normal time for first-time, full-time degree- or certificate-seeking undergraduate students by gender, race/ethnicity, and receipt of Pell grants or subsidized Stafford loans (see the box below for an example of computation of graduation rates at 150 percent of normal time). An institution may choose to disclose this information through its website or distribute a paper copy to current or prospective students upon request.

Computation of Graduation Rates in the Graduation Rate Survey (GRS)

The Graduation Rate Survey (GRS) of IPEDS collects data on graduation rates annually from all Title IV degree-granting institutions. The survey focuses only on first-time, full-time degree- or certificate-seeking undergraduate students and calculates graduation rates for individual institutions (i.e., the rates of students who start and finish at the same institution). Specifically, graduation rates in the GRS are calculated as the total number of first-time, full-time degree- or certificate-seeking undergraduates in a cohort who attained a degree, certificate, or other formal award (from the institution where they entered) within 100 percent, 150 percent, and 200 percent of normal completion time divided by the total number of first-time, full-time degree- or certificate-seeking undergraduates in the cohort. Normal completion time is the amount of time necessary for a student to complete all requirements for a degree or certificate, according to the postsecondary institution's policy. This is typically 4 years for a bachelor's degree and 2 years for an associate's degree. The percentages of 100, 150, and 200 of normal completion time correspond, respectively, to 4, 6, and 8 years for a bachelor's degree and 2, 3, and 4 years for an associate's degree. More information about graduation rates in the GRS is available at http://nces.ed.gov/ipeds/glossary/index.asp?searchtype=term&keyword=graduation+rate.

HEOA permits institutions to exclude from graduation rate calculations those students who leave school to serve in the Armed Forces, on official church missions, or with a federal foreign aid service, or those who die or become totally and permanently disabled. If an institution has 20 percent or more of such students, however, the institution may include these students in the graduation rate calculations, but allow for the time period these students were not enrolled due to their service by adding that time period to the time used in the calculations (NPEC 2009).

There are many ways of calculating graduation rates (Cook and Pullaro 2010). Though the GRS method generates standardized graduation rates that can be compared across institutions, institutions are not required to compute graduation rates this way when complying with HEOA graduation rate disclosure requirements.

---

[7] Created in 1986, the Integrated Postsecondary Education Data System (IPEDS) is a set of interrelated surveys conducted annually by NCES. Institutions participating in federal student aid programs (Title IV) are required by the Higher Education Act of 1965 to report data on enrollments, degree completions, graduation rates, faculty and staff, finances, institutional prices, and student financial aid. IPEDS is the primary vehicle for reporting this information to the U.S. Department of Education. Data collected through IPEDS are available to the public through the IPEDS Data Center (http://nces.ed.gov/ipeds/datacenter/).

[8] Beginning in 2007, GRS no longer required institutions to report graduation rates for scholarship athletes (http://nces.ed.gov/ipeds/glossary/?charindex=G).

To help postsecondary institutions identify and meet their obligation to disclose information as required by HEOA, the Department of Education commissioned a working group of the National Postsecondary Education Cooperative (NPEC) to prepare a report providing technical guidance on how institutions can make the HEOA-required disclosure information more accessible and understandable to consumers (NPEC 2009). Specifically, the group recommended that institutions: (1) focus both on compliance and communication; (2) develop a single web page on the institution's website to provide hyperlinks to the disclosed information; (3) minimize searching by adopting a "3-click" approach; (4) use consumer-friendly labels and language and avoid technical jargon; and (5) use a common set of content titles.

Although full compliance with HEOA graduation rate disclosure for 2-year institutions is not required until academic year 2011–12, many 2-year institutions already have started to publish graduation rates online. A review of 30 community college websites revealed wide variation in methods used and content disclosed.[9] Some colleges posted graduation rates directly on their website, others published them as reports, and a few provided links to the community college system or institutional research office that maintains these statistics in a more centralized way. While a few colleges published only overall graduation rates for first-time, full-time degree- or certificate-seeking students, others published the disaggregated rates by gender and race/ethnicity. Some provided rates only as percentages, others published detailed unsuppressed counts, and a few supplied both percentages and counts. No community college in the sample of 30 used suppression or other methods to ensure nondisclosure of individually identifiable information.

---

[9] Authors performed a web search using "Student Right-to-Know 2-year community college" as key words. Thirty colleges from the first four pages of hits were examined. These colleges were located in 20 different states. All but one college posted graduation rates on their websites.

Unintended Disclosure of Personally Identifiable Information

There are three types of disclosure—*authorized, unauthorized,* and *inadvertent* (Seastrom 2010a). Authorized disclosure allows specific users (e.g., school officials) to access personally identifiable information in student records without the written consent of the parent or eligible student. An unauthorized disclosure occurs when personally identifiable information from a student's record is made available to a third party who has no legal authority to access the information. An inadvertent disclosure occurs when information about an individual is unintentionally revealed through information released to the public. For example, some data in a table might allow users to identify an individual student or reveal sensitive or confidential information about a student. This brief focuses only on preventing *inadvertent disclosure*.

When colleges release information about students' outcomes, they typically release aggregated data for groups of students rather than information about individual students. Even with aggregation, however, unintended disclosure of personal information may occur. Several examples below illustrate how some disclosure practices used by community colleges reveal the graduation status of individual students in reporting aggregated data. Note that all data in the examples below are hypothetical to prevent identification of such colleges.

Example 1. Community College A posted on its website the counts of first-time, full-time degree- or certificate-seeking students in fall 2007 and students in this cohort who graduated within 150 percent of normal program completion time (i.e., completed a 2-year program within 3 years) (table 1). Data are reported for both male and female students and for five major racial/ethnic groups. The table shows that there was one American Indian/Alaska Native student enrolled as a first-time, full-time degree- or certificate-seeking student in 2007, and that this student did not graduate within 150 percent of normal time. This information results in disclosure because anyone who knows this individual student would then learn his or her graduation status from the table.

The table further shows that among 14 Black students enrolled as first-time, full-time degree- or certificate-seeking students in Community College A in 2007, one graduated within 3 years. This also is a disclosure because this graduate then knows that the remaining 13 Black students were not awarded a degree or certificate by Community College A within 3 years.

Table 1. **Number of first-time, full-time degree- or certificate-seeking students at Community College A in 2007 and number of these students who graduated from this college within 150 percent of normal time, by gender and race/ethnicity**

	Number of students enrolled in 2007	Number of students who graduated within 150 percent of normal time
Total	125	53
Gender		
Male	37	17
Female	88	36
Race/ethnicity		
White	70	40
Black	14	1
Hispanic	35	9
Asian/Pacific Islander	5	3
American Indian/Alaska Native	1	0

Example 2. Community College B posted on its website the enrollment counts of first-time, full-time degree- or certificate-seeking students in 2007 and their graduation rates within 150 percent of normal program completion time (table 2). Though this table omitted the counts of students who graduated, the information on graduation rates can be recovered fully by multiplying the proportion of students who graduated by the number of students enrolled. The recovered data show that only one of three Asian/Pacific Islander male students who were enrolled as first-time, full-time degree- or certificate-seeking students in Community College B in 2007 graduated from this college within 3 years. Anyone able to identify this graduate would then know that none of the other first-time, full-time degree- or certificate-seeking Asian/Pacific Islander male students in the 2007 cohort graduated from this college within 3 years. The table also discloses that none of the American Indian/Alaska Native male students had graduated from Community College B within 150 percent of normal time.

Sometimes colleges publish student enrollment information and graduation rates in separate tables, often on different web pages or locations in a report (tables 3a and 3b). Anyone able to locate these tables, however, can reconstruct outcome results that potentially may lead to disclosures of personal or sensitive information about individual students.

Table 2. Number of first-time, full-time degree- or certificate-seeking students at Community College B in 2007 and percentage of these students who graduated from this college within 150 percent of normal time, by gender and race/ethnicity

	Number of students enrolled in 2007	Percent of students who graduated within 150 percent of normal time	Number of students who graduated within 150 percent of normal time
All students			
Male	461	41.6	192
Female	539	49.9	269
White			
Male	368	45.4	167
Female	421	52.3	220
Black			
Male	35	11.4	4
Female	47	27.7	13
Hispanic			
Male	53	37.7	20
Female	64	50.0	32
Asian/Pacific Islander			
Male	3	33.3	1
Female	7	57.1	4
American Indian/Alaska Native			
Male	2	0.0	0
Female	0	†	†

†Not applicable.
NOTES: The shaded column displays data that are not publicly reported but can be derived from reported information in the table.

Table 3a. Demographic snapshot of students enrolled at Community College C in 2006

	All students	First-time, full-time degree- or certificate-seeking students
Total	1,040	348
Gender		
Male	456	128
Female	584	220
Race/ethnicity		
White	779	316
Black	224	23
Hispanic	18	6
Asian/Pacific Islander	5	2
American Indian/Alaska Native	3	1
Unknown	11	0

Table 3b. Percentage of 2006 first-time, full-time degree- or certificate-seeking students at Community College C who graduated from this college within 150 percent of normal time, by gender and race/ethnicity

	Percent of students who graduated within 150 percent of normal time	Number of students who graduated within 150 percent of normal time
Total	25.0	87
Gender		
Male	12.5	16
Female	32.3	71
Race/ethnicity		
White	25.9	82
Black	8.7	2
Hispanic	16.7	1
Asian/Pacific Islander	100.0	2
American Indian/Alaska Native	0.0	0
Unknown	†	†

†Not applicable.

NOTE: Tables 3a and 3b are posted on different web pages. The shaded column displays data that are not publicly reported but can be derived from reported information in tables 3a and 3b.

Practices That Mitigate Disclosure Risk

Adapted from *Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting* (Seastrom 2010b), this section discusses several practices that can help prevent the unintentional disclosure of personal information. The first practice involves not publishing any of the student counts. The second practice calls for suppressing small groups, as well as at least one more group related to the suppressed small groups to prevent the recovery of the suppressed results for the small groups. The third practice involves recoding results into a range to avoid disclosing that a small number of students have or do not have a specific outcome (e.g., graduated from college).

Each of these practices results in some loss of information. These practices are developed on the principle of maximizing the amount of information that can be safely released while protecting individual student privacy. Each practice used alone may not entirely avoid potential disclosure, but used in combination, they may provide improved protection for personal information when postsecondary institutions disclose information about graduation rates. Finally, it is important to note that the practices proposed below are suggestive, and institutions are not required to adopt them, although they are responsible for adhering to the confidentiality requirements of FERPA when disclosing graduation rates.

Not Publishing Counts

While HEOA does not require disclosing enrollment and graduation counts, many postsecondary institutions post these data on their websites. Publishing detailed counts for small numbers of students within a subgroup can jeopardize student privacy because it increases the risk of unintentionally identifying individual students. To mitigate this risk, institutions can publish graduation rates only as percentages without publishing underlying counts of students in the cohort and subgroups. Using this strategy, the example in table 4 includes only graduation rates and excludes detailed counts of students who graduated.

To provide additional safeguards, institutions can also consider rounding graduation rates into whole numbers. This is a useful strategy, particularly when disclosing graduate rates for small groups. Because small groups can only support a relatively small number of exact graduation rates (e.g., an enrollment of 10 students can only support 11 possible graduation rates), reporting graduation rates with decimal place precision may disclose information. Rounding the rates to whole numbers would decrease the precision in the rates and introduce additional protection (Federal Committee on Statistical Methodology 2005).

Table 4. Percentage of 2006 first-time, full-time degree- or certificate-seeking students at Community College D who graduated from this college within 150 percent of normal time, by gender, race/ethnicity, and financial aid

	Before removing counts		After removing counts and rounding percents
	Number of students who graduated within 150 percent of normal time	Percent of students who graduated within 150 percent of normal time	Percent of students who graduated within 150 percent of normal time
Total	148	40.3	40
Gender			
Male	68	52.3	52
Female	80	33.8	34
Race/ethnicity			
White	99	52.9	53
Black	15	19.7	20
Hispanic	12	22.2	22
Asian/Pacific Islander	20	48.8	49
American Indian/Alaska Native	2	12.4	12
Financial aid			
Pell Grant recipients	40	40.8	41
Subsidized Stafford Loan recipients	9	21.4	21
Received neither Pell nor subsidized Stafford Loan	99	43.6	44

NOTE: The shaded columns display data that are not publicly reported.

Suppressing Results for Small Groups

Sometimes postsecondary institutions may want to publish enrollment counts of a cohort as contextual information when releasing graduation rate data.[10] In such cases, colleges can suppress graduation rate data for small groups, since publishing results for such groups can increase the risk of unintentionally releasing individually identifiable information (Federal Committee on Statistical Methodology 2005). To do this, institutions should first establish a minimum number of students in a group for privacy protection (e.g., 10 students) and then suppress results for outcome measures for any group of a size less than this established minimum number.

Sometimes suppression of a small subgroup alone may not be sufficient because information for a suppressed group could be recovered by simple calculations such as subtraction from the marginal total. For example, one can derive the number of female students by subtracting the number of male students from the total number of students. To prevent such data recovery, institutions should apply suppression to at least one other subgroup related to the suppressed small subgroup. This method is commonly known as complementary suppression. That is, a subgroup of a size less than the established minimum number for disclosure 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 (e.g., suppressing data for Hispanics due to the small size and complementarily suppressing data for a related subgroup, such as Blacks, to prevent data recovery).

The example in table 5 illustrates this approach. After a minimum reporting size of 10 is applied, the data for American Indian/Alaska Native students are suppressed. Suppressing data only for this subgroup,

Table 5.   Number of first-time, full-time degree- or certificate-seeking students at Community College E in 2006 and percentage of these students who graduated from this college within 150 percent of normal time, by gender, race/ethnicity, and financial aid

	Before suppression		After suppression of a small subgroup		After suppression of a small subgroup, complementary suppression of a related subgroup, and rounding percents	
	Number of students enrolled in 2006	Percent of students who graduated within 150 percent of normal time	Number of students enrolled in 2006	Percent of students who graduated within 150 percent of normal time	Number of students enrolled in 2006	Percent of students who graduated within 150 percent of normal time
Total	367	40.0	367	40.0	367	40
Gender						
Male	130	51.3	130	51.3	130	51
Female	237	33.8	237	33.8	237	34
Race/ethnicity						
White	191	51.9	191	51.9	191	52
Black	76	19.7	76	19.7	76	20
Hispanic	54	22.2	54	22.2	54	22
Asian/Pacific Islander	41	48.8	41	48.8	([1])	([1])
American Indian/Alaska Native	5	20.0	([1])	([1])	([1])	([1])
Financial aid						
Pell Grant recipients	98	40.8	98	40.8	98	41
Subsidized Stafford Loan recipients	42	21.4	42	21.4	42	21
Received neither Pell nor subsidized Stafford Loan	227	43.3	227	43.3	227	43

[1]Suppressed to protect student privacy.

NOTE: The shaded columns display data that are not publicly reported.

---

[10] Detailed student enrollment data for individual institutions also can be easily retrieved from the IPEDS database.

however, is inadequate to conceal the fact that one American Indian/Alaska Native student graduated because this information can be recovered by using the following calculations: (1) multiply the total number of enrolled students (367) by the overall graduation rate (40 percent) to obtain the total number of graduates (147); (2) use the same method to obtain the number of graduates among Whites (99), Blacks (15), Hispanics (12), and Asians/Pacific Islanders (20); and (3) subtract from the total number of graduates the numbers of graduates in other racial/ethnic subgroups to obtain the number of American Indian/Alaska Native graduates (147-99-15-12-20=1). To prevent recovery of the suppressed data, complementary suppression of a related subgroup (e.g., Asian/Pacific Islander students in this example) is necessary. With this additional suppression, it is impossible for anyone to recover the suppressed result for American Indian/Alaska Native students.

Recoding Into Ranges

Recoding low or high graduation rates into ranges is another approach to protecting data from disclosure in cases where institutions want to publish enrollment counts as contextual information for graduation rate data. For example, instead of showing a graduation rate of 3 percent, an institution can bottom code[11] this rate to "< 5 percent." Conversely, an institution can top code[12] a graduation rate of 98 percent to "> 95 percent." These types of recoding typically are done to avoid disclosure of the fact that very few or none of the students in a group or nearly all of the students in a group have a particular outcome (e.g., graduated from college).

The extent of recoding required to protect small groups is related to the size of the group, with a larger recoded range required for smaller groups. For the purpose of protecting personally identifiable information, results based on the experience of one student should not be disclosed. The minimum goal is to ensure that each recoded range includes at least two students (Seastrom 2010b). For example, with a group of 10 to 19 students, any graduation rates that are 10 percent or lower are based on one student in the group (e.g., one graduate out of 19 students results in a 5 percent of graduation rate and one graduate out of 10 students results in a 10 percent of graduation rate). In other words, even reporting graduation rates in a range of "≤ 10 percent" would reveal that there is at most one graduate in the group of 10 to 19 students. Thus, a larger recoded range is required to avoid such disclosure. In this case, with counts of 10 to 19 students, an institution can recode any graduation rates of 20 percent or below into the range of "≤20 per-

cent" to ensure that the rates based on 0 to 1 students are protected.

Conversely, if the institution publishes a graduation rate of 95 percent for a subgroup of 20 students, it would reveal that 19 out of 20 students graduated and one student failed to graduate. This would be a disclosure. Thus, in this case, an institution can recode any graduation rates of 80 percent or more into the range of "≥ 80 percent" to protect disclosure of the small number of students who did not graduate.

Adapted from *Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting* (Seastrom 2010b), the following recoding ranges are recommended when institutions disclose graduation rates for groups of various sizes.[13] The recoding ensures that each recoded range includes at least two students.

- For groups of 10 to 20 students, recode graduation rates that are 20 percent or less into the range of "≤ 20 percent" to protect rates that are based on 0 to 2 graduates; and recode graduation rates that are 80 percent or more into the range of "≥ 80 percent" to protect rates where small numbers of students do not graduate.

- For groups of 21 to 40 students, recode graduation rates that are 10 percent or less into the range of "≤ 10 percent" to protect rates that are based on 0 to 2 graduates; and recode graduation rates that are 90 percent or more into the range of "≤ 90 percent" to protect rates where small numbers of students do not graduate.

- For groups of 41 to 100 students, recode graduation rates that are 5 percent or less into the range of "≤ 5 percent" to protect rates that are based on 0 to 2 graduates; and recode graduation rates that are 95 percent or more into the range of "≥ 95 percent" to protect rates where small numbers of students do not graduate.

- For groups of 101 to 300 students, recode graduation rates that are 2 percent or less into the range of "≤ 2 percent" to protect rates that are based on 0 to 2 graduates; and recode graduation rates that are 98 percent or more into the range of "≥ 98 percent" to protect rates where small numbers of students do not graduate.

- For groups of more than 300 students, recode graduation rates that are 1 percent or less into the range of "≤ 1 percent" to protect rates that are based on 0 to 2 graduates; and recode graduation rates that are 99 percent or more into the range of "≥ 99 percent" to protect rates where small numbers of students do not graduate.

---

[11] Bottom coding refers to reporting values under a set value as less than that value (Seastrom 2010b).
[12] Top coding refers to reporting values over a set value as greater than that value (Seastrom 2010b).

---

[13] For groups of fewer than 10 students, the general approach, as suggested above, is to suppress results.

The example in table 6 illustrates data disclosure protection using this approach. Since there are fewer than 10 American Indian/Alaska Native students, the graduation rate is suppressed. To prevent recovery of the suppressed data, the graduate rate for Asian/Pacific Islander students is also suppressed (i.e., complementary suppression). Given that 2 out of 58 Hispanic students graduated, the graduation rate (3 percent) is recoded to "≤ 5 percent." Furthermore, since 21 out of 22 students who received subsidized Stafford loans graduated, the graduation rate (96 percent) is recoded to "≥ 90 percent."

Table 6. Number of first-time, full-time degree- or certificate-seeking students at Community College F in 2007 and percentage of these students who graduated from this college within 150 percent of normal time, by gender, race/ethnicity, and financial aid

	Before suppresssion and recoding			After suppression of a small subgroup, complementary suppression of a related subgroup, recoding, and rounding percents	
	Number of students enrolled in 2007	Number of students who graduated within 150 percent of normal time	Percent of students who graduated within 150 percent of normal time	Number of students enrolled in 2007	Percent of students who graduated within 150 percent of normal time
Total	336	50	14.9	336	15
Gender					
Male	130	16	12.3	130	12
Female	206	34	16.5	206	17
Race/ethnicity					
White	186	36	19.4	186	19
Black	63	10	15.9	63	16
Hispanic	58	2	3.4	58	≤5
Asian/Pacific Islander	22	1	4.5	([1])	([1])
American Indian/Alaska Native	7	1	14.3	([1])	([1])
Financial aid					
Pell Grant recipients	98	6	6.1	98	6
Subsidized Stafford Loan recipients	22	21	95.5	22	≥90
Received neither Pell nor subsidized Stafford Loan	216	23	10.6	216	11

[1]Suppressed to protect student privacy.

NOTE: The shaded columns display data that are not publicly reported.

Disclosure Rules

Drawing upon the disclosure prevention practices discussed above, this brief recommends several disclosure rules that postsecondary institutions can use to release graduation rates to the public. These rules were developed on the principle of maximizing the amount of information to be disclosed in compliance with HEOA while protecting student privacy. These rules were also developed to provide a relatively straightforward set of disclosure rules that can be implemented easily. Another goal of these rules, echoing the suggestion by NPEC (2009) on how to disclose HEOA-required information, is to maximize uniformity in disclosure practices across institutions to facilitate cross-institution comparisons.

Rules 1 and 2 are general rules. Rule 3 is guided by the number of students in the underlying group.

Rule 1. *Do not publish detailed student counts when disclosing graduation rates*. Round graduation rates into whole numbers to decrease the risk of disclosure.

Rule 2. *Use a minimum of 10 students for the reporting group size limitation*. Suppress results for all subgroups with fewer than 10 students and suppress at least one subgroup (regardless of size) related to the suppressed small subgroup.

Rule 3. *Recode small and large graduation rates based on the group size*. The recoding range depends on the group size; smaller groups require larger ranges (see the recoding rules above).

Summary

Starting in academic year 2011–12, all 2-year degree-granting institutions participating in Title IV federal student aid programs are required to disclose the graduation rates of first-time, full-time degree- or certificate-seeking undergraduate students disaggregated by gender, by each major racial and ethnic subgroup, and by receipt of federal Pell grants or subsidized Stafford loans. A review of data disclosure practices currently used by many 2-year institutions indicates that a majority do not implement protective measures when releasing data on graduation rates either on their website or in downloadable reports. This brief demonstrates how the unintended disclosure of personal information can occur even in summary statistics.

Building on recommendations for privacy protection in reporting student outcome measures at the elementary and secondary levels (Seastrom 2010b), this brief identifies several practices to avoid unintended disclosure of personally identifiable information in the context of postsecondary education. These practices include establishing minimum cell sizes; publishing graduation rates with no detailed underlying counts; rounding graduation rates into whole numbers; suppressing small groups; and recoding rates based on fewer than two graduates into ranges. These practices are used as the basis for three recommended disclosure rules that are relatively straightforward and easily implemented.

There are multiple approaches to statistical data protection. For readers interested in learning more about the topic of statistical data protection, please see Duncan, Jabine, and de Wolf (1993); Willenborg and De Waal (2001); and Federal Committee on Statistical Methodology (2005).

References

Code of Federal Regulations, Title 34—Education, Part 99. *Family Educational and Privacy Rights,* (34CFR99). Washington, DC: Government Printing Office. Retrieved April 12, 2011, from http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=44d350c26fb9cba4a156bf805f297c9e&tpl=/ecfrbrowse/Title34/34cfr99_main_02.tpl.

Cook, B., and Pullaro, N. (2010). *College Graduation Rates: Behind the Numbers*. Washington, DC: American Council on Education. Retrieved April 14, 2011, from http://www.acenet.edu/AM/Template.cfm?Section=Search&section=reports2&template=/CM/ContentDisplay.cfm&ContentFileID=8515.

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

Federal Committee on Statistical Methodology. (2005). *Report on Statistical Disclosure Limitation Methodology*. Statistical Policy Working Paper 22 (Second version, 2005). Washington, DC: Statistical and Science Policy, Office of Information and Regulatory Affairs, Office of Management and Budget, Executive Office of the President. Retrieved May 5, 2011, from http://www.fcsm.gov/working-papers/SPWP22_rev.pdf.

Higher Education Opportunity Act of 2008, P.L. 110-315, 122 Stat. 3078 (2008). Retrieved April 12, 2011, from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_public_laws&docid=f:publ315.110.pdf.

National Postsecondary Education Cooperative (NPEC). (2009). *Information Required to Be Disclosed Under the Higher Education Act of 1965: Suggestions for Dissemination (Updated)* (NPEC 2010-831v2). Prepared by Carol Fuller and Carlo Salerno for Coffey Consulting. Washington, DC.

Seastrom, M. (2010a). *Basic Concepts and Definitions for Privacy and Confidentiality in Student Education Records* (NCES 2011-601). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Seastrom, M. (2010b). *Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting* (NCES 2011-603). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

The Student Right to Know Act, P.L. 101-542, 34 C.F.R. 668 (1990).

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