Reader's Guide

The Condition of Education is available in two forms: this print volume for 2005 and a web version on the NCES website (http://nces.ed.gov/programs/coe). The web version includes special analyses, essays, and indicators from this and earlier print volumes of *The Condition of Education*. (See page xxiv for a list of all the indicators that appear on *The Condition of Education* website.)

Each section of the print volume of *The Condition of Education* begins with a summary of the general topic areas covered by the indicators in this volume and on *The Condition of Education* website. All indicators contain a discussion, a single graph or table on the main indicator page, and one or more supplemental tables. All use the most recent national data available from the National Center for Education Statistics (NCES) or other sources serving the purposes of the indicator. The "eye" icon at the bottom of the page and to the side of the graph or table directs readers to supplemental notes, supplemental tables, or another source for more information.

When the source is an NCES publication, such as *The Digest of Education Statistics 2003* (NCES 2005–025), that publication can be viewed at the NCES website (http://nces.ed.gov/pubsearch).

The supplemental tables (appendix 1) provide more detailed breakouts for an indicator, such as household income, students' race/ethnicity, or parents' education. Supplemental notes (appendix 2) provide information on the sources of data used, describe how analyses were conducted, or provide explanations of categories used in an indicator. Tables of standard errors (see below) are also included for applicable indicators. A glossary of terms and a comprehensive bibliography of items cited in *The Condition of Education* appear at the end of the volume.

DATA SOURCES AND ESTIMATES

The data in this report were obtained from many different sources, including state education agencies, local schools, and colleges and universities using surveys and compilations of administrative records. Users of *The Condition of Education* should be cautious when comparing data from different sources. Differences in procedures, timing, question phrasing, interviewer training, and so forth can all affect the comparability of results.

Data reported in this volume are primarily from two types of sources. Some indicators report data from entire populations, such as *indicator 36* (public elementary and secondary expenditures per student by district poverty). With these kinds of data, information is collected from every member of the population surveyed. This "universe" could be all colleges and universities or every school district in the country. Other indicators report data from a statistical sample of the entire population. When a sample is used, the statistical uncertainty introduced from having data from only a portion of the entire population must be considered in reporting estimates and making comparisons.

In contrast, when data from an entire population are available, estimates of the size of the total population or a subpopulation are made simply by counting, or summing, the units in the population or subpopulation. In the case of subpopulations, the size is usually reported as a percentage of the total population. In addition, estimates of the average (or mean) values of some characteristic of the population or subpopulation may be reported. The mean is obtained by summing the values for all members of the subpopulation and dividing the sum by the size of the subpopulation. An example is the annual mean salaries of professors at 4-year colleges and universities.

Another population measure sometimes used is the median. The median is the value of a population characteristic above which 50 percent of the population is estimated to fall. An example is the median annual earnings of full-time, full-year wage and salary workers (see *indicator 16*).

Although estimates derived from universe surveys are not affected by sampling, they are affected by a wide range of potential data collection errors such as coverage errors, response errors, coding errors, and data entry errors. These errors in datasets with the entire population may be larger than the error due to collecting data on a sample of the population. Estimates of the size of these errors are typically not available.

A universe survey is usually expensive and time consuming, so researchers often collect data from a small sample of the population of interest. Through (stratified) random sampling and other methods, researchers seek to ensure that this sample accurately represents the larger population to which they wish to generalize. As an illustration, the Early Child Longitudinal Study–Birth Cohort, upon which *indicator 35* is based, surveyed a representative sample of over 10,500 families of babies born in 2001 across the country. Based on this sample, conclusions can be drawn about all babies, such as their race/ethnicity, the education of their parents, parent-child interactions, and their early childhood mental and motor skills.

Estimating the size of the total population or subpopulations from a data source based on a sample of the entire population requires consideration of several factors before the estimates become meaningful. However conscientious an organization may be in collecting data from a sample of a population, there will always be some margin of error in estimating the size of the actual total population or subpopulation because the data are available from only a portion of the total population. Consequently, data from samples can provide only an estimate of the true

or actual value. The margin of error or the range of the estimate depends on several factors, such as the amount of variation in the responses, the size and representativeness of the sample, and the size of the subgroup for which the estimate is computed.[1] The magnitude of this margin of error is measured by what statisticians call the "standard error" of an estimate.

Most indicators in *The Condition of Education* summarize data from sample surveys conducted by NCES or the Bureau of the Census with support from NCES. Brief explanations of the major NCES surveys can be found in *supplemental notes 3* and *4* of this volume. More detailed explanations can be obtained at the website noted above, under "Survey and Program Areas." Information about the Current Population Survey, another frequent source of survey data used in *The Condition of Education*, can be obtained at http://www.bls.census.gov/cps/cpsmain.htm (and also in *supplemental note 2*).

STANDARD ERRORS

When data from samples are reported, as is the case with most of the indicators in *The Condition of Education*, the standard error is calculated for each estimate provided in order to determine the "margin of error" for these estimates. The standard errors for all the estimated means, medians, or percentages reported in the graphs and text tables of *The Condition of Education* can be found in appendix 3, Standard Error Tables. The corresponding standard errors for the supplemental tables can be viewed at the NCES website at http://nces.ed.gov/programs/coe.

The standard errors of the estimates for different subpopulations in an indicator can vary considerably. As an illustration, *indicator 11* reports on the mathematics scores of 4th- and 8th-graders in the United States and 44 other countries in 2003. The average score of 8th-graders in the United States was 504, compared with an average score of 505 in Australia (see

Reader's Guide

Continued

supplemental table 11-1). In contrast to the similarity of these scores, their standard errors were 3.3 and 4.6, respectively (see table S11-1 in http://nces.ed.gov/programs/coe/2005/section2/table.asp?tableID=222).

The percentage or mean score with the smaller standard error provides a more reliable estimate of the true value than does the percentage or mean score with a higher standard error. Standard errors tend to diminish in size as the size of the sample (or subsample) increases. Consequently, for the same kinds of data, such as kindergarten entry rates in the Early Childhood Longitudinal Study, Kindergarten Class of 1998, or scores on the National Assessment of Educational Progress, standard errors will almost always be larger for Blacks and Hispanics than for Whites, who represent a larger proportion of the population.

DATA ANALYSIS AND INTERPRETATION

Due to standard errors, caution is warranted when drawing conclusions about the size of one population estimate in comparison to another or whether a time series of population estimates is increasing, decreasing, or staying about the same. Although one estimate of the population size may be larger than another, a statistical test may find that there is no measurable difference between the two estimates because there may appear to be a large standard error associated with one or both of the estimates.

Whether differences in means or percentages are statistically significant can be determined using the standard errors of the estimates. When differences are statistically significant, the probability that the difference occurred by chance is small; for example, it might be about 5 times out of 100. Some details about the method primarily used in *The Condition of Education* for determining whether the difference between two means is statistically significant are presented in the introduction to appendix 3, Standard Error Tables.

For all indicators in *The Condition of Education* based on samples, differences between means or percentages (including increases or decreases) are stated only when they are statistically significant. To determine whether differences reported are statistically significant, two-tailed t tests, at the .05 level, are typically used. The t test formula for determining statistical significance is adjusted when the samples being compared are dependent. When the difference between means or percentages is not statistically significant, tests of equivalence will often be run. An equivalence test determines the probability (generally at the .15 level) that the means or percentages are statistically equivalent; that is, with the margin of error that the two estimates are not substantively different. When the difference is found to be equivalent, language such as x and y "were similar" or "about the same" will be used.

When the variables to be tested are postulated to form a trend, the relationship may be tested using linear regression, logistic regression, or ANOVA trend analysis instead of a series of t tests. These other methods of analysis test for specific relationships (e.g., linear, quadratic, or cubic) among variables.

Discussion of several indicators illustrates the consequences of these considerations. *Indicator 17* shows a larger percentage of Hispanic persons ages 25–34 (6 percent) were unemployed than White persons (4 percent) in 2004 (see supplemental table 17-1). Although the difference of the unrounded estimates is relatively small (1 percentage point), so are the standard errors associated with each estimate (0.20 and 0.18 for Hispanics and Whites, respectively) (see table S17-1), and the difference is statistically significant and supports the statement. In contrast, *indicator 30* discusses the incidence of school violence against students ages 12–18. The data in supplemental table 30-2 indicate there were 27 violent crimes committed at school against males per 1,000 students in 2002, compared with 21 violent crimes committed at

school against females per 1,000 students. This difference of 6 percentage points is larger than in the previous example, but the standard errors are also larger (2.8 and 2.5, respectively) (see table S30-2). The difference is not statistically significant, and therefore, the data do not support a conclusion that males are more likely than females to be victims of violent crime at school. The introduction to appendix 3 explains in some detail how the statistical significance of the difference between two estimates is determined.

VARIATION IN POPULATIONS

In considering the estimated means in the tables and figures shown in this volume and on the website, it is important to keep in mind that there may be considerable variation among the members of a population in the characteristic or variable represented by the population mean. For example, the estimated average mathematics literacy score of 15-year-olds in the United States in 2003 was 483 (see supplemental table 13-1). In reality, many students scored above 483 points, and many scored below 483 points. Likewise, not all faculty salaries, benefits, and total compensation at postsecondary institutions were the same at each type of institution in 2002–03 (see *indicator 32*).

Because of this variation, there may be considerable overlap among the members of two populations that are being compared. Although the difference in the estimated means of the two populations may be statistically significant, many members of the population with the lower estimated mean may be above the estimated mean of the other population and vice versa. For example, some percentage of young adults with a high school diploma or GED have higher earnings than young adults with a bachelor's degree or higher (see *indicator 16*). The extent of such overlap is not generally considered in the indicators in this volume.

Estimates of the extent of variation in such population characteristics can be computed

from the NCES survey datasets or are available in published reports. For example, estimates of the variation in students' assessment scores can be found using the NAEP Data Tool at http://nces.ed.gov/nationsreportcard/naepdata/ or in the appendices to most NAEP reports.

ROUNDING AND OTHER CONSIDERATIONS

Although values reported in the supplemental tables are generally rounded to one decimal place (e.g., 76.5 percent), values reported in each indicator are rounded to whole numbers (with any value of 0.5 or above rounded to the next highest whole number). Due to rounding, cumulative percentages may sometimes equal 99 or 101 percent, rather than 100.

In accordance with the recently revised *NCES Statistical Standards*, many tables in this volume use a series of symbols to alert the reader to special statistical notes. These symbols, and their meaning, are as follows:

—   Not available.
     Data were not collected or not reported.

†   Not applicable.
     Category does not exist.

Rounds to zero.
     The estimate rounds to zero.

!   Interpret data with caution.
     Estimates are unstable (because standard errors are large compared with the estimate).

‡   Reporting standards not met.
     Did not meet reporting standards.

* $p < 0.05$ Significance level.[2]

NOTES

[1] If there are five racial/ethnic groups in a sample of 1,500, the researcher would have less confidence in the results for each group individually than in the results for the entire sample because there are fewer people in the subgroup than in the population.

[2] The chance that the difference found between two estimates when no real difference exists is less than 5 out of 100.