DOCUMENT RESUME

ED 272 543 TM 860 459

AUTHOR David, Jane L.
TITLE Improving the Quality and Utility of NCES Data.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE 21 Jun 85
NOTE 8p.; In: Invited Papers: Elementary/Secondary Education Data Redesign Project, October 1985; see TM 860 450.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Collection; *Data Interpretation; *Educational Research; Elementary Secondary Education; *Reliability; *Research Methodology; Surveys; Technological Advancement
IDENTIFIERS *National Center for Education Statistics

ABSTRACT

Three goals must be met in order for the National Center for Education Statistics (NCES) to improve the quality and utility of its data collection: (1) the choice of what to collect must be driven by the questions of interest to decisionmakers and the public; (2) procedures must insure validity and reliability of the data; and (3) the data must be reported in ways that facilitate use by the intended audience. The reliability of the data reported is the biggest challenge facing NCES. Careful choices about what to collect and clear reporting cannot compensate for inaccuracy. Reporting data should include the following points: (1) description of similarities and discrepancies from different sources; (2) indication of how various sources were used, when applicable; (3) precise descriptions of what the numbers represent and interpretations of how their known weakness influence the results; (4) knowledgeable interpretation of data; (5) methods of reporting which minimize inappropriate comparisons; (6) full and accurate citation of data sources; and (7) inclusion of a glossary to help readers. NCES should investigate the use of minicomputers for different kinds of assessment instruments and should begin to design a computerized data collection system utilizing telecommunications capability. Comments on "Indicators of Education Status and Trends" are listed with page references. (JAZ)

IMPROVING THE QUALITY AND UTILITY OF NCES DATA

Jane L. David
Consultant
3144 David Avenue
Palo Alto, CA 94303

21 June, 1985

Introduction

The National Center for Education Statistics (NCES) seeks to redesign its data collection efforts to improve the quality and utility of the data for decisionmakers and the general public. For this to occur, three goals must be met. First, the choice of what data to collect must be driven by the questions of interest to decisionmakers and the public. Second, procedures must be in place that insure the data are valid and reliable. Third, the data must be reported in ways that facilitate use by the intended audiences.

For NCES, the second goal is the most critical. Careful choices about what data to collect and clear reporting cannot compensate for inaccurate data. Below I first comment on data accuracy and then discuss reporting, citing examples from Indicators of Education Status and Trends (January 1985). I then list specific comments on Indicators of Education Status and Trends with page references.

Data Accuracy

The biggest challenge facing NCES is that of insuring the validity and reliability of the data they report. If the data continue to be as inaccurate in the future as they have in the past, all other issues are moot. The more levels of aggregation the data pass through, the more sources for error. Because NCES must rely on second, third, and fourth hand data, it is essential to put into place a set of procedures designed to check the validity and reliability of the data.

Given the need to rely on data from other sources (particularly state administrative data which are notoriously inaccurate), NCES must, at the least, develop a system that permits crosschecking the data with other sources for the same information. Judging from the description of current data sources, it appears that there are multiple sources of data for certain types of information (e.g., data on staffing and teacher characteristics reported by the states are also collected in NCES's Public School Survey). To the extent that multiple data sources already exist, NCES should make comparisons across data sources and report on both the extent to which discrepancies are found and plausible explanations for the discrepancies. NCES should exploit all opportunities to corroborate data sources over which they have little direct control, such as state administrative data and Census data.

Data for which multiple sources do not currently exist should be collected through alternative means designed explicitly as a crosscheck. The Fast Response Survey System could readily be adapted to this end. Of course, this presumes that care is

taken to ask precise questions. Unfortunately, my only experience with NCES Fast Response Survey data suggests that these data are often inaccurate as well. On the one occasion in which I used these data (1979), my own telephone surveys corroborated by field work produced figures quite different from those of NCES (e.g., NCES reported 66 schools in California participating in the Schoolwide Projects Provision of what was then ESEA Title I; I located 107 such schools).

Perhaps NCES could also establish samples of schools within states (if such a sample does not exist) as an extension of the Fast Response Survey System. NCES could also build items into contracted longitudinal studies for purposes of corroborating other data sources. In addition, NCES should be aware of other national data collection efforts (particularly federally funded studies in education, labor, and health and annual surveys such as Gallup, Louis, etc.) and develop agreements for sharing data with the funding or data collection agency.

Reporting

Given the nature of the data to be collected, and the reliance on indirect sources and multiple levels of aggregation, there will always be issues of validity and reliability. However, NCES can take steps (a) to maximize validity and reliability and (b) to inform readers of the weaknesses that remain.

In reporting the data, it is absolutely essential to have indicators of the validity and reliability of the data. As a user of data, I am always suspicious of any type of survey data or compilation across levels of government. However, when I know how the question was asked, I can draw my own conclusions about the bias of the responses. When I know the sample size and a standard deviation or a confidence interval, I can draw my own conclusions about its credibility and utility. Through television and other media, even lay audiences are accustomed to confidence intervals and other indicators of measurement error.

Based on a close reading of the Indicators of Education Status and Trends, I urge NCES to consider the following recommendations for reporting the data in addition to reporting how questions were asked and estimates of reliability.

1) Following the previous recommendations for corroborating data, reported data should include a brief description of the similarities and discrepancies from different sources and an analysis of what accounts for the differences. To the extent that the discrepancies influence interpretation of the data, the text should alert the reader to the limits on interpretation.

2) When multiple sources are cited, as is currently the case for many of the tables and charts in Indicators, indicate how the multiple sources were used. Multiple sources always suggests to me that data were merged without regard to comparability. When multiple sources are cited, note which data came from which source and, if two sources were merged for one estimate, describe how and why this was done.

3) Precise descriptions of what the numbers represent and interpretations of how their known weaknesses influence the results is crucial. Some inaccuracy is inevitable in such a mammoth system of data collection. No one expects otherwise. The utility of the data rests on attaining an acceptable level of quality and alerting the reader to its remaining weaknesses. Unfortunately, the text in Indicators accomplishes neither of these goals. Sloppy table titles and category labels are inexcusable as is vague and ungrammatical prose. (See examples below.) Readers must understand the limits of the data.

4) Knowledgable interpretation of data is useful to readers; uninformed or sloppy interpretation is dangerous. NCES should provide interpretation but not without a system that corroborates the interpretation(s). One approach is to create panels of outside reviews in different areas (e.g., one for student performance, another for human resources) who would comment on a draft of the presentation. Agreement is not essential; presenting conflicting interpretations is also extremely useful to readers.

5) Collect and report the data in ways that minimize inappropriate comparisons. For example, comparing states by comparing trends over time within states decreases the problems due to different measures and definitions. NCES does this well in several places. In addition, the text around tables and charts should draw the readers' attention to the appropriate comparisons.

6) NCES should cite data sources more accurately and fully. When "NCES estimates" are cited as the data source, my eyebrows go up. All sources should include dates.

7) Use the glossary to help readers. Dictionary definitions of enrollment, attendance and biology, for example, are not helpful. The Glossary should reflect the way questions were asked and indicate differences in definitions across states or data sources.

Comments on "Indicators"

The concept of a report on indicators of education status and trends is excellent. This kind of annual report has the potential to provide an invaluable picture of our educational

system over time. At the least, it provides a backdrop against which to interpret other data and educational issues at all levels--national, state and local. At the most, the data can inform the public and policymakers about expected trends in such critical areas as teachers supply and student enrollment.

Recognizing that the January 1985 Indicators of Education Status and Trends is a first attempt which seeks reactions from potential users, I offer the following specific reactions by report section and page.

Outcomes

p. 3 I find the NAEP data by assessment area within subject particularly illuminating. I realize presenting subscales greatly increases the quantity of the data. This is an instance in which some clear text around the tables would be useful. The text could report conclusions from an inspection of more detailed tables and reference other publications. (A minor point: describing shifts in performance "over the past decade or so" is misleading, especially for science which includes no data from the past eight years.)

p. 6 I don't know what to conclude from these data. How were these topics measured? Did the items correspond to a particular curriculum? Did they emphasize computation or problem solving? Are the numbers medians of 18 means? Do the results look different if the comparison is between the United States and the five countries with the same proportion enrolled in math?

p. 8 Because this is an area in which many states are changing their requirements, it would be useful to see shifts over time and a breakdown by state. Are these only comprehensive high schools?

p. 10 This is an instance in which how the question was asked and to whom is critical. Can GEDs be reported separately? (It isn't clear whether they are included on p. 10 or not).

Resources

p. 24 Basing pupil/teacher ratios on all instructional staff results in a gross underestimate of class size. Readers will draw from this table an image that severely distorts reality. These data must be presented for teachers with regular classroom assignments or not at all. There is no indication of what the sample is for the class size data; are they based on regular teachers only?

Why is the research literature mentioned here and not elsewhere? If research is mentioned, there should be complete citations. I don't think it belongs here; an adequate summary is difficult to do in one sentence.

"In 1971 and subsequent years, the data by level are estimated." The footnote should say how and why the estimates were made. Citing "unpublished data and estimates" only invites suspicion.

p. 26-27 The headings and labels for the table on p. 26 are a good illustration of the need to use precise language. The title is extremely misleading; this is not a table about the distribution of academic ability in the teaching force. Throughout the present report, typically the charts on the right are far easier to understand than the tables on the left. This is partly because graphs are often easier to grasp, but more due to the clarity of the titles and headings. The title and labels of the graph are much more precise and hence clearer than those on the left. (E.g., Percent Scoring in Highest Fifth is much clearer than Highest Rank with a footnote saying the sample was broken into five ranks).

The content of these two pages also raises questions. First, it is unconscionable to refer to these data as measures of the "Quality of the Teaching Force." No single test score can capture teacher quality. Moreover, given the small percent of teachers who take the SAT (and no indication of this in the table), I conclude that the numbers are extremely misleading. I have no problem with the idea of looking at teacher performance on academic measures, but I need to be convinced that these data speak to that issue. Perhaps this is an instance in which the National Longitudinal Study is not the best source of data. (The samples seem quite small.)

p. 28 For data about teacher supply and demand to be useful, they need to be reported separately for elementary and secondary school and by subject area. It would also be useful to see these figures by region and the same breakdowns for teachers "teaching out of subject"--that is, teaching in fields for which they are not certified. The table doesn't state that the entries are in thousands.

Context

p. 37 Do data on teachers' perceptions of problems exist over time? The wording of the second bullet suggests that the Metropolitan Life/Harris Survey was conducted in other years as well. It would be useful to see trends in these data. Were the questions posed to the public and to teachers in the same way? If not, since the tables invite comparisons, it would be useful to know how they differ.

p. 38 Certainly school environment indicators are of interest. However, the key factors associated with effective schools are not amenable to the kind of measurement that could be aggregated across schools and districts and states. Most of us don't even know how to measure them on site. The kinds of perceptions to be reported in the 1984 follow-up for High School and Beyond may prove useful. I suspect that the greatest utility will lie in the items that have been asked over time. As with trends within state, shifts over time are easier to interpret than absolute levels of factors like "environment conducive to student achievement." This suggests a critical need for ongoing longitudinal studies of this type.

p. 40 How many students actually receive these various types of services?

p. 42 Where are the data cited in the text on pages 42 and 43? To estimate school enrollment trends, are census data available on numbers of babies/children ages 0 - 3?

p. 44 Referring to state required Carnegie units as "State Governance" seems odd.

Additional data that I would find useful include:

-median age of teachers by state
-data on preschool attendance
-measures of student mobility/turnover
-information on number of hours worked (and types of jobs) for high school students by state and minority status

Looking to the Future

The proliferation of microcomputers in district offices and schools, combined with growing sophistication about and access to telecommunications, has far reaching implications for future data collection. Now is not the time to implement such a system; neither access to the technology nor user sophistication is sufficiently widespread. But now is the time to begin to design a computerized data collection system utilizing the telecommunications capability that most districts and schools will have within a decade or less. Such a system will require considerable planning and testing; waiting until the technology is completely in place will put NCES a decade behind. If General Motors can design a system that automatically translates an individual customer's order into instructions for what parts to manufacture and into a custom made car, NCES ought to be able to gather basic descriptive information about our schools through similar applications of technology.

The second application of technology that NCES should now be investigating is the use of microcomputers for different kinds of assessment instruments. The limits of paper and pencil tests are well known. Designing new measures that go beyond simple multiple choice questions should be underway.