DOCUMENT RESUME

ED 365 734 TM 020 979

AUTHOR Franklin, Bobby J.; And Others
TITLE Are the Data Clean? Data Verification Procedures: Louisiana and the Nation.
PUB DATE Nov 93
NOTE 16p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 10-12, 1993).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Attendance; *Data Collection; Dropout Research; Educational Quality; Elementary Secondary Education; *Profiles; *Research Methodology; Research Reports; *School Districts; School Statistics; *State Departments of Education; State Programs; State Surveys; Suspension
IDENTIFIERS Aggregation (Data); *Louisiana; *Verification

ABSTRACT

This paper documents the multistep procedure used by the Louisiana Department of Education, Bureau of School Accountability to correct and verify the data that drive the state's statewide school performance indicator program, the "Progress Profiles." The paper also reports the findings of a nationwide survey in which state departments of education describe their data-verification procedures. To correct problems identified with the first "Profiles," new data collection and aggregation procedures were developed, beginning with educating staff about the importance of quality data. A second initiative was to involve the department staff and the owners and users of data in the verification and correction of raw and aggregated data. Profile coordinators were used to coordinate data verification and correction efforts, and reports were prepared with raw data for each local education agency (LEA). Each LEA reviewed raw data reports for correctness. The 50 states were surveyed about the procedures they follow in verifying data, and 42 states replied with descriptions of their procedures. In general, edit checks are performed on collected data, and data are returned to LEAs if discrepancies are found. In Louisiana, the verification process appears to have improved attendance and dropout data, but has not had the same impact on suspension and expulsion data. The upcoming Student Information System should further improve data quality. Three tables are included. (Contains 4 references.) (SLD)

Are the Data Clean? Data Verification Procedures: Louisiana and the Nation

Bobby J. Franklin,
Linda J. Crone,
Catherine Glascock,
&
Susan Kochan

Louisiana Department of Education
Office of Research & Development
Bureau of School Accountability

A paper presented at the annual meeting of the Mid-South Educational Research Association in New Orleans, Louisiana, November 10-12, 1993.

Are the Data Clean? Data Verification Procedures: Louisiana and the Nation

Should data be accurate? Most would say the need for accurate, quality data is a given. Without it policymakers are less likely to make sound decisions and research findings are misleading. Although data collection is a basic part of research text books, there is little research on the issue of ensuring the collection of quality data. This issue is especially important for State Departments of Education (SDE) who are major collectors and disseminators of education data.

In the decade since the publication of a *Nation at Risk*, virtually every state in the nation has implemented some system of educational performance indicators aimed at monitoring the condition of education. These accountability movements are a consequence of "the public's desire to know the results of education for all America's students" (Beller-Simms, Brauen & Danielson, 1993, p. 15). Though such systems are intended to support educational improvement by enabling policymakers to make informed decisions, access to inaccurate or unreliable information is more dangerous than no information at all. Hence, an integral part of the administration of any education indicator system must include a comprehensive data verification process which ensures data accuracy.

This paper documents the multi-step procedure used by the Louisiana Department of Education (Department), Bureau of School Accountability, to

correct and verify those data that drive Louisiana's statewide school performance indicator program, the *Progress Profiles.* In addition to documenting Louisiana's verification process, this paper will report the findings of a nationwide survey in which State Departments of Education describe their data verification procedures. This report also provides statistical evidence that shows how data quality can impact data analysis. Implications and recommendations also are discussed.

The Need

The Department produced its first accountability report, *Louisiana Progress Profiles*, in December 1991. As with those states that preceded us, errors within the data were made public. For example, one school reported that 85% of the student body had been suspended during the school year. This caused the *Profiles* to be labeled as the "error-plagued report cards" (Myers, 1991, March 19) and because of the political situation we were forced to make corrections to the database and reprint the documents. At the time of the December release, the dropout data collection had not been completed, therefore, the decision was made to include dropout information on the *Profiles* for the second printing. This also proved disastrous because of data errors and the news media once again labeled the report as "mistake-riddled" (Myers, 1991, August 18). Again, we were required to cleanup the data and reprint the *Profiles.*

The Reason

Why was this occurring? There were several factors that contributed to this dilemma. The first involved the data collection procedure itself. The data were collected in an aggregate form at either the school or grade level. This information was typically recorded on paper at the school site and sent to the central office where clerical staff keypunched the information directly into the Department computer. Collecting data in an aggregated form (coupled with a process that requires two levels of data entry) lends itself to, not only numerous errors, but errors that are sometimes difficult to correct. The second source of error was a result of Local Education Agencies (LEAs) viewing the Department as a *blackhole* for data. The data went in but nothing ever came out. So why worry about data quality, especially if it does not affect funding? Other errors were the result of inconsistencies inherent in Department databases. For example, school identification codes often varied among databases and teacher identification numbers were sometimes inconsistent between the certification and annual school report databases. Still, other errors evolved from miscommunication and multiple interpretations of terminologies among Department and LEA staff such as numerous definitions for attendance, dropouts, and suspensions.

The Solution

In order to correct and eliminate the types of problems identified with the first *Profiles*, new procedures were developed featuring several new checks and

balances to the data collection and aggregation process. The first step involved educating the Department staff, LEA superintendents and their staff, and school principals on the importance of quality data and the urgent need for implementing data verification and correction procedures. The negative press provided some of the impetus for developing a better working relationship among Department staff and between the Department and LEAs. This process was accomplished by improving communication lines between Department staff, through in-service workshops on data collection, and presentations at state data conferences.

The second initiative was to require the active involvement of the Department Data Responsibility Center (DRC) staff (owners and users of the data) to assist in the verification and correction of all raw and aggregated data to be used for *Progress Profile* reports. The DRCs had to work closely with the school districts to verify and correct data. The school districts are the original owners and producers of the data and their active involvement in the verification and correction procedure is essential.

Thirdly, it appeared necessary to use LEA liaisons (Profile Coordinators) for coordinating all data verification and correction efforts. In order to implement a comprehensive and yet efficient data verification and correction procedure, the Department dealt with LEA-designated liaisons who in turn dealt with their respective data contact staff. In theory this appeared to be a workable solution. However, working through a middle person added another

layer of complexity and slowed the process. This was an additional burden on the Coordinators and it placed them in a supervisory capacity with no authority over other personnel. This step was abandoned after the first year and the Department dealt directly with the LEA staff responsible for submitting the data.

The fourth strategy included sending reports containing both non-aggregated raw data (data as submitted to the Department by LEAs), aggregated *Profile* data (data as will be presented in the *Progress Profiles*), and tolerance limits to each LEA. The LEAs review the raw data reports to determine if the data received by the Department are correct. The aggregated reports and the tolerance limit reports are used to highlight extreme cases that bear closer scrutiny. The tolerance limit reports showed schools with data that exceeded one standard deviation from the mean. In some situations only extreme cases were notified for further review. For example, high schools with zero dropouts, schools with 100% attendance, or schools that had suspended more that one third of the student population would be asked to reexamine their data. Schools that showed large changes in their data from the previous year are also contacted. Approximately four weeks are given to the LEAs to review the data and make the necessary corrections. The person in the LEA office who is responsible for the data and the LEA superintendent are required to complete and sign a verification form indicating the data are correct. Once this has taken place then the data are released for analysis and reporting.

The initial verification procedure was a three step process:

(1) The data were returned in a printed format for review and the LEAs were allowed to make online corrections.

(2) The data were returned a second time for those who made changes during step one for a second review. If additional changes were necessary, the LEA requested these additional changes and the DRC within the Department made the changes.

(3) The data were returned a third time for those who requested changes during step two. Once this third step had been completed then the data are deemed ready for processing.

This three step process has been reduced to a single step due to an increased awareness of the importance of data reporting by the LEAs. In addition, the LEAs are becoming more automated with their data collection and reporting which has accelerated the process.

Definitions

The term "data verification" appears to have somewhat different meanings depending on who is discussing the issue. This term is sometimes used interchangeably with edit checks. In this paper an edit refers to the process of using algorithms to identify computational errors which may occur within a database or between databases. Discrepancies are identified as errors and solutions are sought to correct the situation. Verification may include edit checks but is not limited to them. Verification goes beyond edits to include

those errors the edits will not detect. For example, failure to report a teacher, reporting the wrong social security number, or reporting the incorrect number of students in a class. Often edits are designed to identify situations that exceed policy standards, but do not catch accidental underreporting. Having someone close to the data source review the data improves the chance of correcting errors not detected through edits.

The States

The fifty states were surveyed as to the procedures they follow in verifying their data. A letter was mailed to each Chief State School Officer providing them with a brief description of the process used by Louisiana. Each state was asked to provide a brief description of the process they use. Forty-two of the 50 states responded. Of these 42, thirty-five indicated they have some type of verification process in place. All the reporting states use edit checks. Whereas, some states appear to place more emphasis on data tied to funding others more closely scrutinize their non-fiscal data. Data tied to funding were reported to be edited and/or audited by at least 24 states. Non-fiscal data were verified or edited by at least 28 states.

Generally the procedure appears to be (a) edit checks are performed once the data are collected, (b) if discrepancies are detected, the LEAs are notified by phone or in writing, and (c) the state works with the LEA to resolve the discrepancy. Many states are moving toward an electronic data transfer system, therefore, a large number of those responding indicated the use of

online systems, tape transfer, or floppy disks to make data corrections. Several states, like Louisiana, return the data to the LEAs in a paper format for review. One state, however, reported that because their data is transmitted on paper there was no need to return it for verification.

Statistical Differences

Descriptive analysis, as well as, correlation and regression statistics were used to examine data collected from the LEAs before and after verification. If the verification process is having any impact on the data, then some differences should be identifiable when the results of the two separate analyses are compared. Specifically of interest is the predictive validity of the data before and after verification. In other words, will the verified process variables better predict test scores (outcome variable) than will the unverified data.

Two sets of data files were maintained for four indicators reported on the *Profiles*, attendance, suspensions, expulsions, and dropouts. The first file contained data before verification and the second file contained data after verification. The after-file data were used to produce the *Profiles*. Some membership data were also examined because of their inclusion in calculating the percentages for the above indicators.

Using the "Proc Compare" command in *SAS*, school totals for the indicators listed in Table 1 were compared to identify changes that occurred as a result of corrections made during the verification process. Please note that

only changes to the school total are presented in Table 1. Corrections that did not result in a change in the total value for the indicator are not detectable with this method. The greatest number of changes occurred among those indicators dealing with discipline and fewer changes occurred among those related to student membership and attendance.

Generally the changes resulted in a increase in the mean value (Table 2) of the indicators examined with the exception of in-system gains which showed a decrease. Only the dropout data show a significant difference between the before and after means.

Pearson correlations between test scores and the variables listed in Table 3 were conducted with the before and after data. Very similar results were obtained for suspension and expulsion data regardless of the database used. For attendance and dropouts, however, higher correlations occurred with the database after verification. For these same four indicators, the after data also produced a higher explained variance (R^2=.44) than did the before data (R^2=.31).

Conclusions/Implications

For attendance and dropout data, the verification process appears to have positively impacted the quality of these data as evidenced by changes in their correlation and predictive ability with test scores. Likewise, these same effects were not observed for suspension and expulsion data rendering these findings less conclusive than those with attendance and dropouts. However,

the effectiveness of the verification process is based on the assumption that schools not making changes have correct data. What impact simply refusing to make corrections has on the data remains unanswered.

The level and method of data collection presently used possesses inherent problems because of collecting data in an aggregated form and the lack of consistent procedures and definitions. The upcoming Student Information System, centralizing all department data collection efforts, and the development of common indicator definitions may help reduce gross errors in the data and is an area of further study.

Decisions related to education which are impacted by these and other educational data can be seriously hampered with faulty data. Although no one would argue with this statement, there doesn't appear to be a groundswell of concern about reporting accurate information. As these data become more visible, attention will be directed toward their accuracy. Likewise, State Departments of Education cannot expect schools to be concerned about data quality unless good use is made of the data collected.

Table 1

The Number and Percent of Schools with Different Before and After Verification Totals for Certain School Indicators

Indicators	Number of Schools with different Before and After Totals	Percent Different
Aggregate Days of Attendance[1]	50	3.48
Aggregate Days of Membership[1]	51	3.55
Registration[2]	49	3.41
In-system gains[2]	46	3.20
Out-system gains[2]	17	1.81
Number Suspended	121	8.42
Number Expelled	88	6.12
Number Dropouts	121	8.42

[1]Used to calculate percent attendance.
[2]Used in calculating percent suspensions, expulsions, and dropouts.
N=1436

Table 2

School Indicator Means Before and After Verification

Indicators	Means Before Verification	Means After Verification	Difference
Aggregate Days of Attendance	89,278.0	90,310.0	1,032.00
Aggregate Days of Membership	95,478.0	96,373.0	895.00
Percent Attendance	93.8	94.0	0.20
Registration	521.0	531.0	10.00
In-system gains	57.1	51.5	-5.60
Out-system gains	38.9	39.1	0.20
Cumulative Enrollment	617.0	621.0	4.00
Number of Suspensions	53.3	56.8	3.50
Percent Suspensions	7.7	8.1	0.40
Number Expulsions	2.5	2.8	0.30
Percent Expulsions	0.3	.3	0.00
Number Dropouts	11.0	17.4	6.40
Percent Dropouts	1.8	2.4	0.60

Please note that these are population means.

Table 3

Pearson Correlations Comparing Data Before and After Verification for School Indicators

Indicators	Before r	After r
Percent Attendance	.34	.48
Percent Suspensions	-.22	-.23
Percent Expulsions	-.21	-.21
Percent Dropouts	-.19	-.39

References

Beller-Simms, N., Brauen, M. & Danielson, L. (1993). *Proceedings Seventh Annual Conference On the Management of Federal/State Data Systems.* Bethesda, Maryland: U. S. Department of Education, Office of Special Education & Westat.

Myers, Doug. (1991, March 19). Error-plagued school reports to be replaced. Morning Advocate, 1A, 6A.

Myers, Doug. (1991, August 18). Cody wants "report cards' accurate. Morning Advocate, 1B, 4B.

Stigler, S. M. (1992, November). A historical view of statistical concepts in phychology and educational research. *American Journal of Education, 101, 60-70.*