

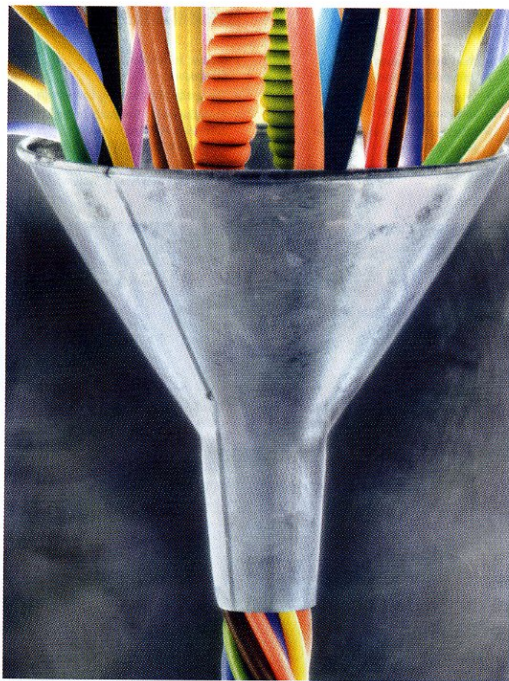
Data Warehousing: Beyond Disaggregation

Recent developments in data analysis can help educators meet new reporting requirements and provide a more precise tool for improving education.

Lawrence M. Rudner and Carol Boston

Educators have long disaggregated test scores by race, sex, and socioeconomic class. The new No Child Left Behind legislation requires breaking down results further to consider disabilities and English-language status. The legislation also demands an annual analysis of trends to demonstrate yearly progress and prove that program participation improves achievement. Many states and school districts are scrambling to ensure that their data collection and reports comply with these new requirements.

Rather than work toward mere compliance, educators should seize this opportunity to retool and systematize their data collection, reporting, and analysis. Properly constructed, a data collection and analysis system can go beyond simple disaggregation to provide information more fully responsive to local, state, and federal data needs. Such a system allows many school personnel, not just in-house statisticians and programmers, to generate reports and analyses that supply information, provide accountability, explore relationships among different kinds of data, and inform decision making. A data warehouse is a system specifically structured for such query and analysis (Kimball & Ross, 2002).



Uses of Education Data

Once a system is in place, educators can use the collected data for traditional purposes, including monitoring compliance with federal and state law and standards, responding to federal reporting requirements, determining state funding allocations, tracking school performance, assigning sanctions and awards, responding to data requests from state legislatures and local boards of education, producing annual statewide summary publications, and designing school improvement plans.

In addition to these traditional purposes, data warehouses can transform mountains of data into useful information and help policymakers identify and plan responses to key trends. Like businesses that use information to maximize their

competitive stance, educators can use a system with high-quality data to improve and maximize learning (Kimball & Ross, 2002). Data warehouses can present the entire picture of education—including school buildings, teachers, students, support staff, attendance data, achievement data, and various programs—and allow the study of its functioning over time. These analyses can predict events and help guide action to anticipate them. For example, does a district have a serious

pending teacher shortage? A data warehouse can help examine trends in retirement rates, new hires, certification types, teacher age, and student enrollment. Once policymakers obtain a clear understanding of these issues, they can undertake appropriate short- and long-term planning.

Data warehouses support two types of data—cross-sectional and longitudinal—to create a more complex and nuanced picture of school performance than does the mere reporting of test score changes from year to year.

Cross-Sectional Data

Much of the public reporting of data in education has consisted of repeated cross-sectional views. Each year, the reports provide information for different groups within one time period—for example, reading scores in 3rd grade. Over time, repeated cross-sectional views present the evolving picture of the overall education system. Without sophisticated analysis and reporting, however, many mistakenly assume that the cross-sectional approach reflects the overall effectiveness of the education system from year to year.

Consider, for example, the cross-sectional 3rd grade reading achievement data for one county in Maryland (see fig. 1). The data reflect the overall achievement levels in the county. At first glance, it appears that the quality of education in the county improved from 1997 to 1998, then dropped, slowly at first, and then rather precipitously, from 1998 to 2001.

During that time, however, corresponding variations occurred in county enrollment and class size; the percentage of students receiving free and reduced-price meals; the percentage of students in special education or in English as a Second Language classes; the percentage of new teachers and teachers with provisional certification; and the average number of years of teaching experience. What effect might these factors have on test scores? The statistical techniques of regression analysis allow us to assess the relationship between the dependent variable—the

test scores—and the combined effect of several independent variables. The multiple correlation of the pass rate with only two other factors—free and reduced-price lunches and special education enrollment—is 0.887, which means that 78.6 percent of the variation in test scores correlate with just those two demographic factors. The demographic shifts were so large and so powerful that they overshadowed any changes caused by differences in school effectiveness.

Longitudinal Data

The education data warehouse permits analyses of groups of students as they progress through grades. Such tracking yields more meaningful data about education gains than those obtained merely by comparing test results of this year's 3rd graders with test results from last year's 3rd graders.

Longitudinal analyses are crucial for meaningful evaluation of program success. A case in point is Tennessee's Project STAR (Student Teacher Achievement Ratio), experimental research involving the random assignment of 11,000 K-3 students and teachers in 79 elementary schools into one of three groups: small classes of about 15 students, regular classes of about 25, and regular classes of approximately 25 with a full-time aide. Students assigned to the smaller classes posted significantly higher scores on the Stanford Achievement Test and a project-developed basic skills test (Mosteller, 1995).

In the second phase of the study, researchers tracked the progress of these students from 5th to 9th grade and discovered that the positive academic outcomes carried over for students who were in small classes for three or four years in the early grades (Finn, Gerber, Achilles, & Boyd-Zaharias, 2001). Further longitudinal studies have linked the small-class experience with higher graduation rates and more honors diplomas (Boyd-Zaharias & Pate-Bain, 2000). Inspired by the success of the STAR class-size experiment, California and at least 18 other states and large districts legislated or volunteered to reduce class size in the primary grades—an expensive education reform.

The caveat that results don't always generalize is applicable here. To evaluate its experience, the Los Angeles Unified School District conducted several studies examining reading, mathematics, and language arts skills in the smaller classrooms and reported mixed, but promising, results (Fidler, 1999). Examining the data more carefully, however, Hamilton (2002) reported that several schools showed dramatic increases when the analysis used repeated cross-sectional data but no increase when it followed groups of students during the course of several years. Hamilton argued that California's data collection program should examine a wide range of longitudinal data, which would enhance the accuracy of school-level achievement data and help disentangle the impact of schools and

F I G U R E 1

Percent of 3rd Graders Whose Test Performance Is at Least Satisfactory, by Year

Year	1997	1998	1999	2000	2001
Percent Satisfactory or Better	51.8	54.0	51.4	50.2	44.8

A quick view of these cross-sectional data would suggest that school effectiveness has declined. Combining these data with corresponding school and demographic data reveals that the data reflect dramatic changes in the school environment.

Extraction from Maryland Student Performance Assessment Program data by Lawrence M. Rudner.

teachers from the effects of factors not under their control. This example underscores the need for local data warehousing.

Beyond Simple Disaggregation

A well-organized, easily accessible data warehouse allows a wide range of important analyses using both cross-sectional and longitudinal data. Such analyses can shed light on one of our most vexing challenges: What programs and policies are effective in addressing the achievement gap? Simple data disaggregation might suggest that demographics play a determining role in education outcomes, but data warehousing can help tease out what factors make a difference for at-risk students who are successful. Quite simply, the data warehouse supports more detailed breakdowns that can help answer such questions as

- Within what kinds of settings do initially high-achieving students maintain their high levels of achievement for several years?

- Do certain programs work better for students with different skills?

- What specific mathematics learning outcomes mastered in 3rd grade best predict overall mathematics achievement in grades 5 and 8?

- What is the relative impact of teacher experience, class size, and mentoring on student learning?

- What mathematics curriculum is most effective in closing the achievement gap and increasing the participation and performance of underrepresented groups in algebra and other higher mathematics courses?

One interesting application of data warehousing is the Tennessee Value-Added Assessment System (Sanders, 1998). In addition to linking student records over time, Tennessee links students' records to their teachers' characteristics in an effort to determine the effects of teachers on student achievement. Because the major test publishers' data records contain the names of both students and teachers, linking the student and teacher databases is simple and makes possible a wide range of

Data warehouses can transform growing mountains of data into useful information and help policymakers identify and plan responses to key trends.

Resources on Data for Educators

Audio

Seize the Data! Maximizing the Role of Data in School Improvement Planning. An ASCD 2000 Teaching and Learning Conference presentation covers data mining, data analysis, data communication, and the use of data for making decisions. Available: <http://shop.ascd.org/ProductDisplay.cfm?ProductID=200318>

Multimedia

Disaggregation without Aggravation. Southwest Educational Development Laboratory's multimedia training package shows how to disaggregate student data and use that information to improve instruction. Available: www.seidl.org/pubs/catalog/items/teaching06.html

The Internet

The American Association of School Administrators provides links to resources that districts can use for data-driven decision making, including *Using Data to Improve Schools*, an easy-to-read guide for using data for school improvement. Available: www.aasa.org/cas

The North West Regional Laboratory's *Evaluating Whole-School Reform Efforts* is an online book about analyzing education data. Available: www.nwrac.org/whole-school/index.html

interesting analyses. For example, which teacher characteristics are most highly associated with achievement gains for different groups of students? What support systems for beginning teachers yield the highest gains for disadvantaged students?

Prerequisites

Until now, educators have met traditional reporting requirements by using a patchwork system of single-use tools to generate specific reports. Although useful, such a system depends on the expertise of statisticians and programmers and is not conducive to exploring data or informing policy. Fortunately, the field of data analysis has moved forward. Today, with a well-designed system, educators can obtain information with a few keystrokes from a desktop computer.

Installing easy-to-use data analysis tools to query and filter database records can produce a wide range of reports. SAS Institute, Cognos, Oracle, IBM, and others have been developing such tools for the business community for years. Their online analytical processing tools allow sophisticated, multidimensional data analyses.

Efforts to build a functional education data warehouse should emphasize the overall system planning and the collection of first-rate data. McIntire (2002) recommends a four-stage process:

- Conduct an information inventory;
- Standardize the management of data;
- Analyze the data; and
- Make changes and define new strategies.

Ensuring that the data are of high quality is crucial. Database records should be complete, with valid and appropriate entries for each data element. Too often data are missing because of typographical errors: incorrect values (*B* to indicate gender, for example); unrealistic values (a salary of \$670,000); or wrong data types (a *Y* instead of the numeral 3).

A well-designed data entry and validation system will ensure that correct data are available for analysis and decision

making. Assigning clear responsibilities for individual data elements is important, as well as using data entry and validation tools as back-ups. For these tools to be successful, the data collection system must be easy for staff to use, facilitate efforts to provide good data, use a range of data verification techniques, and provide flexibility for dealing with changes in requirements and data entry.

The Benefits

To be of use for improving education, data must be of high quality, accessible, and in a format that the requester can use. Developing a data warehouse to support the development of reports, analyses, and decision making may require a modest investment of resources. Once in place, the investment should rapidly pay for itself by simplifying the effort needed to generate many required reports. The

real payoff will come, however, when we use data to make decisions about improving education. We can then tie outcome measures more precisely to school efforts rather than amorphously to school efforts plus shifts in regional demographics. Using data this way will enable us to improve programs that need help and develop those that work well. ■

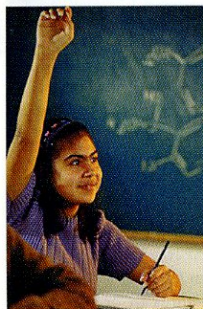
References

- Boyd-Zaharias, J., & Pate-Bain, H. (2000). *The continuing impact of elementary small classes*. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, Louisiana.
- Fidler, P. (1999). The impact of class-size reduction on student achievement. Los Angeles Unified School District. Planning, Assessment and Research Division Publication No. 109.
- Finn, J. D., Gerber, S. B., Achilles, C. M., & Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103(2), 145-183.
- Hamilton, L. (2002, May). California student data system: In need of improvement. *EdSource* [Online journal]. Available: www.edsource.org/pdf/EDFctDataSystem_Final.pdf
- Kimball, R., & Ross, M. (2002). *The data warehouse toolkit* (2nd ed.). New York: John Wiley & Sons.
- McIntire, T. (2002). The administrator's guide to data-driven decision making. *Technology and Learning*, 22(11), 18-33.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 5(2), 113-127.
- Sanders, W. L. (1998, December). Value-added assessment. *School Administrator*, 55, 101-113.

Lawrence M. Rudner (lrudner@ericae.net) is Director and Carol Boston (boston@ericae.net) is Assistant Director of the ERIC Clearinghouse on Assessment and Evaluation, housed in the Department of Measurement, Statistics, and Evaluation at the University of Maryland, 1129 Shriver Laboratory, College Park, MD 20742.



Do You Know What Your Students Know?



Educators Agree....

With *The Curriculum Mapper*™ it's easy to work together to create a coherent, data-driven curriculum. Input, edit and share curriculum maps from any computer with an Internet connection. Attach any documents (lesson plans, rubrics, assessments) directly into your maps! Call us 800-318-4555 for information on how to get your school or district started down the mapping path.

Interested in attending a **FREE** workshop on Curriculum Mapping? *The Curriculum Mapper*™ will be hosting the workshop Analyzing and Aligning your Curriculum Using *The Curriculum Mapper*™ at the ASCD convention Sunday March 9, 2003 in San Francisco. Call us toll free 800-318-4555 for workshop details.

Can't make the workshop? Call us for a code to try it yourself.

www.curriculummapper.com

"I recommend The Curriculum Mapper™ as a straightforward and thoughtful tool for assisting schools and districts in carrying out the mapping process"

Heidi Hayes Jacobs, Ed.d.: author of "Mapping the Big Picture: Integrating Curriculum and Assessment K-12" (1997 ASCD)