For *Educational Leadership*
draft 1 10/06 11:30 am

Data Warehousing: The Next Step Beyond Data Disaggregation

Lawrence Rudner and Carol Boston
ERIC Clearinghouse on Assessment and Evaluation
University of Maryland, College Park.

This is an era of unprecedented demand on the public schools. Schools are called upon to provide broader, more in-depth, and a higher level of instruction to a wider range of student abilities. Schools today serve a diverse population that never has been so varied in terms of its language backgrounds, economic status, mobility, and physical and cognitive strengths. At the same time, local and state financial resources are disappearing, the public call for accountability is increasing, and federal data collection requirements are growing at a seemingly exponential rate.

This is a complex environment for the state and local education agencies. Inform decisions and policy need quality data. Yet the support staff of LEA's are among the first to be hit when budgets are reduced. Data quality suffers and deadlines become hard to meet. As a result, relevant data is either not consulted because it is not trusted or decisions are compromised because they are informed by the available, but questionable, data. This is not an era to compromise data collection and analysis

Properly constructed, a data collection and analysis system can go well beyond simple disaggregation to provide information responsive to LEA, SEA and federal data needs. Such a system will allow a wide range of personnel, no just in-house statistician and programmers, to generate ad-hoc as well as routine reports and analyses to help inform decision-making, explore relationships, provide accountability, and supply needed information. With the passage of the 1,168 page No Child Left Behind Act last July, increased accountability and analysis is mandated at the federal level. Accountability and reporting systems must cover all public schools, students, and subgroups of students. These reports must include trends over time to demonstrate annual yearly progress. The additional mandate for "scientifically-based" educational programs requires showing that program participation improves achievement. With increased reporting and analysis at the state level, one can only anticipate a demand for increased reporting and analysis at the LEA, and probably the school, level

Uses

Once a system is in place, collected data can be used for traditional purposes, including

- monitoring compliance with federal and state law and standards
- designing of school improvement plans
- determining school performance and sanctions and awards
- preparation of the ever increasing federal reporting requirements
- responding to State legislative and Board of Education data requests
- production of annual Statewide summary publication, and

- determination of State funding allocations.

In addition to these traditional purposes, a system containing high quality data can be used for a wide range of analyses, by a wide range of individuals, to inform a wide range of decisions and policies. In the corporate world, *business information* has been key to strategic positioning in order to maximize one's competitive stance (Kimball, 1996). The corollary in education is to use the warehoused data to inform policies that will lead to continuous quality improvement and maximize learning. In addition to the traditional uses outlined above, online analytical processing (OLAP) tools can be used to inform state and local decisions based on sophisticated, multi-dimensional data analysis.

Data warehouses exist to help transform the growing mountains of data info useful information and to help managers identify key trends. It helps present the entire picture of education - a picture composed of some districts, teachers, students, buildings, support staff, attendance data, achievement data, and various programs, and over time. It can be used to foresee predictable events and help guide action in anticipation of those events. For example, do we have a serious pending teacher shortage? One could query a quality data warehouse to model teacher demand by examining trends in retirement rates, new hires, certification types, teacher age, and student enrollment. With a good systematic understanding of these events, the issue is placed in perspective and the appropriate reactions can be planned.

Cross-sectional data

Much of the public reporting of data in education has been repeated cross-sectional views. Each year, information is reported for different groups using a common time frame. This approach holds districts accountable for all students in their schools, as it should be. When repeated cross-sectional views are examined over time, they present the evolving picture of overall education. However, this is often confused and taken to reflect the effectiveness of the educational system.

Consider the following cross-sectional grade three reading achievement data for one county in Maryland (actual data).

Table 1: Percent of third graders whose test performance is at least satisfactory, by year.

year	1997	1998	1999	2000	2001
Percent Satisfactory	51.8	54.0	51.4	50.2	44.8

The data reflects the overall achievement levels in the county. The naive view would be that the quality of education in the county improved from 1997 to 1998 and then dropped slowly and then rather precipitously from 1998 to 2001. However, during that time, there were corresponding variations in county enrollment, the percent of students on free and reduced lunch, the percent of students in English as a Second Language programs, the percent of special

education students, the percent of new teachers, the percent of teachers with provisional certification, class size and the average teaching experience. The multiple correlation of pass rate with FARMS and Special education enrollment is .887. In other words some 78.6% of the variation in test scores could be attributed to those two demographic factors alone. The demographic shifts were so large and so powerful that they overshadowed any nuances that may exist due to differences in school effectiveness. Here the cross-sectional view of the data was aided by examining other cross-sectional data.

Longitudinal Data

With consistent data keys, the educational data warehouse will also be able to permit analyses of cohorts. Such an analyses are critical for a meaningful evaluation of program success. Inspired by success of the STAR program in Tennessee (Mosteller, 1995), California and at least 18 other states legislated or volunteered to have reductions in primary grade class size - a very expensive educational reform. To evaluate their experience, Los Angeles Unified School District conducted several studies examining reading, mathematics and language arts skills and felt the results were mixed but promising (Fidler, 1999). Hamilton (2002) reports that when examining LAUSD data, several schools showed dramatic increases when repeated cross sectional data was used and no increase when cohorts were followed.

Hamilton raises the point arguing that the California School Information Services program (a data collection program mandated by the California state legislature) should make a wide range of linked (keyed) longitudinal data a central feature of an improved state data collection system. In addition to meaningful program evaluations based on the students served by a program, Hamilton cites other advantages of maintaining longitudinal data: enhanced accuracy of school level achievement data by highlighting the impact of student mobility, and help disentangle the impact of schools and teachers from the effects of factors not under their control.

Beyond simple disaggregation

By having a well organized, easily accessible data warehouse, a wide range of important analyses using both cross-sectional and longitudinal data can be conducted. As mentioned earlier, one would be able to disentangle the impact of schools and teachers from demographic factors not under the schools control and provide more accurate evaluations of programs and policies. Education has had a long history of disaggregating data. Achievement levels by various subpopulations (race, ESOL, income) are routinely reported. The data warehouse will enable more detailed breakdowns. Do initially high achieving students maintain their high levels? Do certain programs work better for students with different skill matrices? Which of the learning objectives are most critical for later learning? What is the relative importance of teacher experience, class size, and special training (e.g. mentoring programs)?

One interesting application is the Tennessee Value-Added Assessment System (Sanders, 1998). In addition to linking student records over time, Tennessee links students records to their teacher database. They then attempt to model the effects of teachers on student achievement. Since the data tapes from the major test publishers contain student and teacher names, linking the student and teacher databases should be relatively simple. This opens the possibility of a wide range of

interesting analyses, e.g. What are the characteristics of teachers most highly associated with achievement gains for different groups of students? What support system for beginning teachers yield the highest gains for disadvantaged students?

Prerequisites

The clear prerequisites are that the data must be of high quality, accessible, and in a format that can be used by the requestor. In education, as in business, much of the emphasis to date has been on collecting data and meeting traditional reporting requirements. In many cases, a patchwork system of non-unified, single-use tools are used to generate specific reports. Statisticians and programmers are required to extract information. While utilitarian and functional, such a system is not conducive to data exploration or informing policy. The data analytic world has progressed well beyond this legacy approach. Today, with a well designed system, information can be obtained with a few key stokes, rather than hours or days of professional time.

Much of this introduction has emphasized the long term capabilities of a well coordinated data warehouse. It is fairly easy to accomplish the back-end - installing easy-to-use data analysis tools to query and filter database records in order to produce a wide range of very specific pre-packaged as well as ad-hoc reports and analyzes. SAS, Cognos, Oracle, IBM and others have been developing such tools for the business community for years.

Efforts to build a quality education data warehouse should emphasize the front-end - 1) system planning and 2) the collection of high-quality data. McIntire (2002) outlines the four steps that he successfully employed in establishing a data warehouse when he was the technology director for a suburban New York City School District.

- conduct an information inventory
- standardize data management
- analyze the data
- make changes and define new strategies

Critical in this effort, is the collection of high-quality data. Data should be complete with valid and appropriate entries for each data element. Too often data is missing, out-of-bound values are entered (e.g. *B* is used to indicate gender), unrealistic values are encountered (e.g. a salary of $670,000), and wrong data-types are found (e.g. *Y* is entered when a *1* should have been used). A well-designed data entry and validation system will go a long way toward ensuring that high-quality data are available for analysis and decision making. Responsibility for individual data elements should be clearly defined. Data entry and validation tools should be employed. For these tools to be successful, the data collection system must:

- be easy for staff to use,
- aide, not hinder, efforts to provide quality data
- use a range of data verification techniques
- provide flexibility because requirements, data fields, and valid data entries change.

Developing a data warehouse to support report generation, analyses, and decision making may require a modest investment of resources. Once in place, however, the investment should rapidly pay for itself just by simplifying the effort needed to generate the multitude of required and expected reports. The real pay-off will come when decisions are based on data. Outcome measures can be attributed to school efforts rather than school efforts plus shifts in regional demographics. Programs and activities that work well can be recognized and expanded. Areas and groups in need of improvement can be properly identified and assisted.

Select Resources

Audio & Multimedia

Seize the Data! Maximizing the Role of Data in School Improvement Planning, The Association for Supervision and Curriculum Development - An audiotape covering data mining, data analysis, data communication, and data use for decision making. http://shop.ascd.org/ProductDisplay.cfm?ProductID=200318.

Disaggregation without Aggravation, Southwest Educational Development Laboratory - A multimedia training package designed to help educators learn to disaggregate student data and use that information to improve instruction. http://www.sedl.org/pubs/catalog/items/teaching06.html

On-line reading

Data Driven Decision Making, the American Association of School Administrators - A website designed to be helpful to districts interested in data driven decision making and to provide a list of resources for those districts http://www.aasa.org/cas/

Evaluating Whole-School Reform Efforts, the North West Regional Laboratory - An on-line book on educational data analysis. http://www.nwrac.org/whole-school/index.html

Using Data to Improve Schools, the American Association of School Administrators - An on-line easy-to-read guide to using data to drive school improvement. http://www.aasa.org/cas/UsingDataToImproveSchools.pdf

References

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

Hamilton, L. (2002). California Student Data System. *EdSource*, May.
Kimball, R (1996). *The Data Warehouse Toolkit*. John Wiley and Sons, Inc.

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, 113-27.

Sanders, W.L. (1998). Value-Added Assessment. *School Administrator*, December.